Universiteit Utrecht

Personalized Workflow Optimization for University Staff Empowerment: Introducing WorkflowAId

Universiteit Utrecht

author:	Luuk Janssen
1st supervisor:	Prof. dr. ir. H.A. Reijers
2nd supervisor:	Dr. I.M. Beerepoot

Master Thesis
Business Informatics
Luuk Janssen
July 9, 2024

ABSTRACT

Understanding personal characteristics and patterns that impact productivity on a personal level and using this information to support individuals has been recognized as being essential in improving personal productivity. However, existing studies often lack a versatile, personalized approach. This research aims to contribute to this problem by exploring the impact of work-related factors on perceived productivity among university staff. The main goal of this research is designing a workflow support system that identifies personal work-related factors and aids the user in being more productive whilst working. These work-related factors include time-related features like day of the week and sequential patterns of work activities. A combination of neural networks, pattern mining and data analytics is used to investigate the impact of work-related factors on productivity. Data was collected from four university staff members over a total period of around six months, and consists of log data regarding work-related activity tracked by Active Window Tracking software. This data is combined with survey data consisting of daily perceived productivity scores to gain insights on work-related factors and their impact on productivity.

The results show the significant impact of time-related features, work activities, and sequential patterns of work activities on perceived productivity. Additionally, the individual differences in work-related features that impact perceived productivity are shown, supporting the need for a personalized solution. However, the results of designing and evaluating the personalized workflow support system show the challenges of designing such a personalized solution that will be adopted by the user.

This research addressed gaps in the existing literature by focusing on work-related factors that impact perceived productivity over an extended period of time and can serve as a foundation for future personalized solutions that support individuals in being more productive. Additionally, it emphasises the need for future research in applying our approach in different contexts with a broader set of work-related factors, more participants and a more detailed measure of productivity. Finally, implementing and assessing the user experience of workflow support systems in real-world scenarios will be crucial for developing effective tools for improving personal productivity.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor, Prof. dr. ir. H.A. Reijers. His insightful guidance, support, and expertise have been instrumental throughout the course of my research. I have learnt a lot from the talks we have throughout this process and the useful feedback that was given on my work.

I would also like to thank Dr. I.M. Beerepoot for her constructive feedback and sparking my interest in studying this topic. Besides my supervisors, I would also like to take a moment to thank all the other authors whose work laid the foundation for my research [5]. This study truly initiated the thought process for my entire thesis, inspiring me to delve deeper into this fascinating subject.

Finally, I would like to thank everyone around me who showed unwavering support throughout the bumps along the way and helped me push through. Your encouragement and understanding have been invaluable.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

In the twentieth century, people had already begun to notice a clear pattern: the use of computers in the workplace was increasing at such a high rate that it would become the most prevalent working tool in the future. From 1984 to 1993, the number of people using computers at work rose from one-quarter to nearly one-half, indicating an annual increase of 2.4 percent in the workforce per year [3]. Nowadays, computers are of great value in any workplace and almost everyone uses them at some point during their workday. It has not only been found that computerization improves productivity on the work floor [4], but the data generated during computer use has also been shown to provide us with unique insights into the work processes [5].

This study is built upon preliminary research of Beerepoot et al. [5] and uses data from university staff at Utrecht University, tracked during working hours, to better understand their work characteristics and its relation to productivity. The software used to provide these insights is called Tockler, which utilizes the Active Window Tracking (AWT) technique to log variables related to work processes, including applications used and timestamps [14]. In this research, this data is combined with work activity labels and daily surveys measuring university staff members perceived productivity. The goal is to design a workflow support system that shows the unique work patterns and characteristics of university staff members, with the aim of helping them achieve maximum productivity in a personalized manner.

In the next section the problem that this research originates from will be addressed, after which the research questions associated with this problem will be stated. Finally, we will give an explanation on what neural networks and workflow support systems exactly are. In Chapter 2, a thorough review of the relevant literature in this field will be presented. After this, in Chapter 3, the methods used to conduct this research will be explained. In Chapter 4, we will present the results of our study, where after they will be discussed in Chapter 5. Finally, we will provide a conclusion on our research in Chapter 6

1.1 PROBLEM STATEMENT

A lot of research has been done in the field of analyzing productivity of employees, where various factors have been found to have significant impact on productivity, both positively and negatively. For example, the study of Mark et al. [16], which showed that the more time employees spent using email during the day, the less productive they felt at the end of the day. Or the study of Mark et al. [15], which showed that participants assessed their productivity significantly higher whilst using blocking software to prevent unwanted distractions during work. Next to this, interruptions, task switching, context switching, time management, user input, task size, breaks, and physical environment all have been found to significantly impact the perceived productivity of employees [6, 9, 11, 17, 16, 18]. Most of these studies share a common focus on identifying factors that significantly impact productivity at the collective or group level, rather than the personal level. Although this is the case for most studies, there are some cases that study the effect of personalized tooling on productivity. One such tool is called TimeAware [13]. This tool uses an ambient widget combined with an information dashboard to show users their productivity and application use, displayed in a positive and negative frame. In this tool productivity is measured on the basis of a self-assigned productivity score for all applications used. The outcome of this study showed that participants increased their productivity only, and significantly, in the negative framing situation.

The self-assessment of a productivity score to an application does raise a question: How accurate is this assessment? Consider Facebook, which is often seen as distracting during work and, therefore, unproductive for many employees. However, there may be situations where an employee is temporarily stuck on a task, and a distraction could actually improve overall productivity. In such cases, taking a step back and looking at the entire workflow might be necessary to gain a broader understanding of the individual's workflow in different situations and its impact on productivity. This point is also raised by studies of Kaur et al. and Kim et al. [11, 13], both of which suggest that future research should focus on delivering more personalized strategies for individuals, including when and how they can optimize their productivity whilst taking into account their whole workflow.

This research aims to address the problem of taking into account the whole workflow when assessing its productivity by examining the individual's workflow whilst not only considering specific events during the process, but also examining the broader work activities, patterns of work activities and their relation with other characteristics involved. The goal is to design a workflow support system that helps university staff members become more productive by providing them with unique, personalized insights into their workflow.

Key concepts: Work activities and patterns

A work activity refers to a distinct task performed by the university staff member during their work. Examples include conducting research, assessing student papers, and giving lectures. We can investigate the computer behavior regarding these work activities, and ultimately study the computer behavior while performing these work activities and its relation to productivity. Since a substantial amount of work is now conducted on computers, studying the effect of these activities on productivity is crucial. By analyzing these effects, we can identify work activities that contribute positively or negatively to productivity on a personal level.

As mentioned earlier, we believe that it is important to take into account the whole workflow whilst assessing its impact on productivity. We aim to do this by studying the impact of the sequence of work activities on productivity. In our case, a work pattern is a sequential sequence of work activities, representing the order and context in which different tasks are carried out throughout the day. Analyzing these work patterns can help us understand the cumulative and contextual effects of tasks sequences on productivity. For instance, while a specific work activity might independently be associated with high productivity, the context of sequential patterns it is performed in may reveal that there are certain combinations that result in a higher productivity than others. This highlights the importance of not only the individual work activity, but also the context in which they are performed, providing deeper insights into optimizing the workflow for enhanced productivity on a personal level.

1.2 RESEARCH GOAL AND QUESTIONS

In this section, we will pose the main research goal and the accompanying research questions that form the foundation of our research. This goal and the following questions are designed to shed light on the problem of discovering work patterns and characteristics within workflows that significantly impact productivity. Our aim is to provide users with unique insights into their work processes by visualizing the relationships between work activities and work-related characteristics with productivity. The goal of providing these personalized insights through a workflow support system that attempts to help users improve on their productivity during work. This brings us to the main research goal:

RG: *To design a workflow support system that utilizes a combination of a personalized neural network, data analytics, and pattern mining which effectively integrate:*

- *low-level variables, such as application usage.*
- *high-level work activities.*

With the goal of improving productivity of university staff members by:

- *identifying personal work patterns and characteristics that significantly improve productivity.*
- *aiding university staff members by optimizing and visualizing their workflow with the use of a workflow support system.*

Low-level variables and high-level work activities
Before we move onto the research questions that will help us achieve our main research goal, a clear distinction between different variables used in our research needs to be made. The Tockler software used in our research creates a log of the workers' activities, where every line consists of information about the application used, a title of the action performed in this application, the starting time, the ending time and the date. These variables are considered low-level, as they represent the data at the finest level of granularity, providing the most detailed information about the application use. The higher-level work activities are present in the log files of two participants and are also represented in each line of the log data, but mostly span across many lines of the log. These work activities are a result of manually labeling the log data and they represent a sequence of actions performed within on ore more applications and are therefore considered as high-level.

To achieve our main research goal we constructed the following research questions:

RQ1: *Can we identify the work activity performed by a participant at any given moment in time by studying the low-level variables and patterns within the log?*

This first research question aims to discover low-level variables and patterns within the log that are highly correlated with certain work activities. Successfully answering this research question will allow us to detect these high-level work activities in log data and enrich the unlabeled data with it, resulting in data with low-level variables and high-level work activities for each individual. This enrichment of our data will allow us to study a wider time span of data, resulting in a workflow support system that is based on a larger amount of personal work patterns and characteristics, and therefore would be more versatile.

RQ2: *Can we identify workflow patterns by examining different work activities performed throughout the day, and do these patterns significantly impact personal productivity?*

This research question aims to investigate if there are sequential patterns of work activities that have a significant impact on an individual's productivity. As mentioned earlier, we believe that it is essential to investigate the context in which work activities are performed, which we do by investigating the sequence of work activities and its impact on productivity. Discovering such patterns can give university staff members valuable insights into their personal workflow, empowering them to make informed changes or keep working in a certain manner that has a positive impact on productivity.

RQ3: *Do specific features, such as time-related features, application switches and the nature of the work activity, significantly impact personal productivity?*

While we explore the sequences of work activities, as discussed in our introduction, other factors also have been found to impact productivity significantly. Therefore we believe that it is necessary that these features are also studied. For instance, a work activity may occur frequently on days with high productivity, but its impact could vary by day or certain periods during the day. Hence, we will analyze the impact of these time-related features on productivity. Additionally, application use and switching between them can impact productivity, as discussed in Chapter 1. Thus, our analysis will also focus on switching between applications whilst performing work activities. Finally, the nature of each work activity might impact productivity. For example, we might find that reviewing papers has an overall negative impact on productivity for an individual. Whilst focusing on the chain of work activities, we believe that accounting for these kind of findings in our analysis is important and will lead to useful insights.

1.3 ADDED VALUE

1.3.1 Practical contributions

In contrast to existing studies that focus on factors that impact productivity on a collective level, our research targets drivers and barriers of productivity on a personal level. This personalized approach provides university staff members with tailored insights into their workflows, aiming to enhance their personal productivity.

Next to this, our workflow support system could serve as an ethical alternative to surveillance software which lets management keep tabs on its employees [10]. Unlike systems that monitor employees and measure productivity by automated metrics like completion time, our approach relies on a self-assessed productivity score. This approach puts back the control into the hands of the employees by making it a system that studies optimal workflow patterns and characteristics from their perspective. The implementation of such a support system can ultimately result in feeling more productive, not measured by efficiency metrics but by your own subjective experience.

1.3.2 Academic contributions

Our research expands on the current literature by investigating a spectrum of work activities and their impact on productivity, providing a different perspective than current research which focuses on specific factors, like email use. Additionally, the effect of work patterns on productivity will be studied, which is an area of research that is underexplored in the current literature. Finally, developing individualized neural networks represents a field of research which is still quite young, with limited prior exploration. The combination of these personalized neural networks with a personalized workflow support system can advance the understanding of solutions which aim to improve personal productivity.

1.4 BACKGROUND

This section provided essential background information on two important topics within our research, ensuring clarity. First of all, we will explain what a neural network exactly is and its connection to our research. Secondly, we will give background information on workflow support systems, our definition of it and the connection to the neural network.

1.4.1 Neural Network

A neural network is a structure of nodes that learns from each other through interaction. These networks are typically organized into input, hidden and output layers, which can be seen in Figure 1.

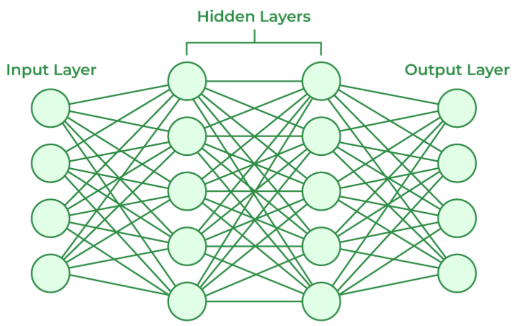


Figure 1: Architecture of a Neural Network. [1].

The neural network uses weighted connections between the different nodes to process data whilst minimizing errors during training. This way of adaptive learning allows the network to discover interesting patterns and features, even if the data is complex in nature [1].

Given that, in our study, we are dealing with time-series data and want to discover patterns and complex non-linear relationships, a neural network seems to be an appropriate solution. In the context of our study, the input layer processes the log data in a chronological manner, enabling the nodes to learn patterns over time. Each of the nodes in the input layer represents a certain feature of this log data, such as application usage. The hidden layer(s) extract features and dependencies related to the work activity performed, adjusting the node weights according to the feature importance. The final output layer contains the work activity discovered based on individual patterns and features within the data. This prediction will be used to analyze the interesting features and patterns that significantly impact the self-assessed productivity. These insights will ultimately provide the user with feedback to improve upon this with the use of our workflow support system.

1.4.2 Workflow Support System

The term "workflow support system" can be interpreted in a lot of ways, so we define it in this section in the context of our research. Our goal is to aid university staff members by visualizing their workflow using a personalized system. Because of this, our workflow support system will contain visualizations that originate from the resulting data produced by the personalized neural network. This approach aims to showcase how different factors and sequences of work activities impact productivity. It is important to note that we aim to develop a system design which is used as a guideline, and not as a set of rules that users must unquestionably follow; it is designed to provide assistance and support rather than enforce strict rules.

2 LITERATURE REVIEW

In this chapter, a review is presented of the literature that is relevant to our research. The literature discussed, results from following a Literature Research Protocol. This protocol is explained in more detail in Chapter 3.2. This literature review consists of content in two main topics, namely productivity and personal informatics.

2.1 PRODUCTIVITY

The main goal of our research is improving the productivity of university staff members by aiding them with a workflow support system. To achieve this, a deeper understanding of productivity is needed, which will be discussed in the next sections. In section 2.1.1, the different kind of ways productivity is measured in the literature will be covered. In section 2.1.2, the internal and external factors that impact on workers' productivity during work will be discussed. After this, the effect of work patterns on productivity will be covered. Finally, we will discuss the differences individuals experience with regards to their productivity in section 2.1.4.

2.1.1 Productivity measures

The measurement of productivity in the literature is found to be measured from objective, subjective and combined perspectives. Objective measures focus on hard numbers, such as classifying applications as productive or unproductive or associating productivity to higher grades [13, 25]. The subjective perspective focuses on the feeling of being productive, being assessed in some kind of manner by the individual themselves [9, 11, 12, 14, 16, 17]. Finally, we encountered one study that tries to combine both approaches, balancing feelings and hard numbers [3].

The first objective measure of productivity is from Kim et al. [13], who developed TimeAware, a feedback tool using positive and negative framing to enhance productivity. They found that only workers in the negative framing condition improved their productivity significantly. Each application in their study was classified into productive, neutral, or distractive. This classification could be manually edited, but this was barely used because as it was seen as a burden by the worker. The resulting productivity score is a weighted average of all the applications used during the day. A problem we see here is that giving each application a set level challenges the applicability of such a classification, as users may perceive the productivity of applications differently.

Whittaker et al. [25] deployed the meTime application, which shows users' top five applications used in the last 30 minutes to improve productivity by raising awareness about application use. In their study, the impact of the tool on student grades was investigated during a two day period, associating better grades with being more productive. The study showed that raising awareness resulted in higher grades among the students. However, as discussed in this study, there are limitations in using class grades as indicators of productivity, emphasizing the need for future research to consider broader measures and long-term effects [25]. Our study will contribute by examining variables and patterns affecting a self-assessed productivity score over nearly 6 months.

In contrast to objective measures, many studies adopt a subjective approach to measure productivity, aligning with our perspective. For example, Epstein et al. [9] explored how interruptions during breaks impact knowledge workers' self-assessed productivity negatively. Kim et al. [12] identified six themes that individuals use to assess productivity, showing a wide range of activities considered as productive, from desk work to personal tasks like getting a haircut. Another study of Kaur et al. [11] attempts to model opportune moments for transitions and breaks and investigates the effect of this on self-assessed productivity. Participants felt that the Flowzone tool made them feel more productive at work. One participant stated the following: "it can be a digital assistant looking after you and your well-being, what more could you want?! [11]", expressing the importance of tools that offer support by taking into account your feelings.

Finally, the literature also explores a combined approach to measuring productivity, integrating both objective and subjective perspectives into complex models for assessing an individual's productivity. Balakayeva et al. [3] developed DERES, a 360-degree employee rating system that assesses individuals based on self-assessment and peer-assessment to increase overall enterprise productivity, whilst also taking into account personal development. An increased productivity within teams was shown using the DERES employee rating system.

Our study relies on a subjective measure of productivity, namely a self-assessment of how productive the person felt at the end of each day. While combining objective and subjective approaches is interesting, it often tends to lean more towards achieving enterprise productivity [3]. We also believe that, in our case, a subjective approach can tailor to individuals in a more personalized way than an objective approach because of the diverse and varied nature of tasks assigned to university staff members, resulting in a better solution for aiding individuals in feeling more productive.

2.1.2 Factors impacting productivity

Internal factors

Now that we have discussed the different measures of productivity, a deeper understanding of the factors that have an impact on productivity is needed. If our goal is to create a system that aids workers in achieving better productivity, an understanding of where to improve upon is necessary. Internal factors, initiated by individuals themselves, significantly impact productivity. Epstein et al. [9] studied how work breaks affect self-assessed productivity, highlighting that participants who took more digital breaks, like checking social media, felt less productive. Studies by Mark et al. [16, 17] on email use found that the more time an individual spends on email during the day, the less productive the individual felt at the end of the day. Next to this, their study also showed that the tactic individuals use to handle their emails impact their productivity assessment, emphasizing the impact of personal behavior on productivity.

These studies highlight the importance of understanding and providing insights into personal behaviors and its relation to productivity. Our goal is to contribute to this by designing a personalized workflow support system that analyzes the impact of internal factors, like application use, on productivity.

External factors

Next to internal factors, literature has also shown the impact of external factors on the productivity of individuals. These are factors that are not initiated by the individual themselves, but rather from an external source. Kim et al. [13] demonstrated that negative framing of productivity tools can improve productivity by showing time spent performing 'unproductive activities'. Studies by Mark et al. [14, 15] explored how distractions impact productivity, showing that cutting off distractions effectively can enhance focus but may increase stress. These results point out the importance of systems that know when to cut off distractions and when not. Epstein et al. [9] highlighted the negative effect of interrupted breaks on productivity. Finally, study of Balakayeva et al. [3] showed that peer-assessment systems like DERES can improve team productivity by sparking motivation, though it may also lead to higher stress.

Our study tries to mitigate the possible increase in stress by designing a workflow support system which is meant to improve your feeling of productivity, making it a tool for individuals themselves, and not for their employers. We do realise that there are external factors, like offline interruptions, that could impact productivity in our study which will not be measured, as this is out of our scope.

2.1.3 Patterns impacting productivity

To improve university staff member productivity, our research aims to identify personal patterns that significantly impact productivity, recognizing the individual nature of work resulting in personal patterns. We will elaborate on the research that has been done regarding this topic.

Barata et al. [4] introduced AppInsight, with the goal of allowing users to view their overall application usage patterns over time. Participants were positive about the ability to study their application use patterns. For instance the monthly usage view, which provided useful insights into the individuals' overall patterns of activity. This excitement about studying patterns raises the question of how to identify which patterns to follow or avoid for improved productivity. Our study addresses this by using a workflow support system to discover and display productive and unproductive patterns based on user data, enabling the user ability to make informed decisions on avoiding or pursuing certain patterns.

Freed et al. [10] developed RADAR, a personal assistant tool that reduces email overload by learning productive task sequences. The downside to this tool is that it only manages tasks on the basis of email content of individuals. However, most work activities do not involve email.

Cranefield et al. [5] present a case study with the goal of understanding the role of personalized work-based analytics with respect to productivity improvement and work-life balance. In this paper, they talk about Microsoft MyAnalytics, which is a tool that uses machine learning to coach individuals in being more productive. This tool rates productivity based on four key work patterns: Focus, Wellbeing, Collaboration, and Network. Focusing on a set of patterns may overlook personal patterns that impact productivity. For instance, the 'Focus' pattern, which is associated with higher productivity, scores based on time not being engaged in email or meetings. But what if an individual is actually focused whilst being in meetings all day? Our study explores a broad range of work activity patterns, rather than predefined ones, to better understand the impact of an unspecified set of patterns on productivity.

As mentioned earlier, Kaur et al. [11] developed Flowzone to model opportune moments for transitions and breaks, finding that current actions are influenced by past actions, indicating the presence of interesting time series patterns. Our study will investigate these patterns and their effects on productivity, providing support through our workflow support system based on these insights.

Lastly, Epstein et al. [9] found a positive relationship between productivity before a break and feeling ready to work afterward, suggesting a productive work-break-work pattern. The period of work before the break might not impact productivity significantly on its own, but taking a break at the right time causes the participant to feel ready to get back to work after. This indicates the importance of taking into account the whole workflow and studying patterns and their impact on productivity.

2.1.4 Individual differences

The final important effect impacting productivity are the differences individuals experience whilst performing work, which we will discuss in this section. First of all, the study of Kim et al. [12] found that knowledge workers' perceptions of productivity varied based on task significance and emotional state. This study shows the different factors individuals feel are important when assessing their perceived productivity.

The personalized models that form the basis of the Flowzone tool, developed by Kaur et al. [11] performed really well regarding goodness-of-fit and low error values on suggesting opportune moments for transitions and breaks. They also found that cluster models, which group participants based on job roles, performed similarly or better than personalized models. This suggests significant differences in work patterns between individuals. While cluster models are beyond our research scope, we aim to identify productive and unproductive patterns and their influencing factors.

Finally, Cranefield et al. [5] and Mark et al. [15] emphasized the need for personalized solutions. Participants in Cranefield's study found general AI solutions like Microsoft MyAnalytics too general and not applicable to them. For instance, academics in this study categorized their work into teaching, research, and administration, however the MMA tool assigned all their time to 'focus' or 'collaboration,' leading to a mismatch between the tool and participants. The study of Mark et al. [15] showed individual differences in stress and productivity when distractions were blocked.

These studies highlight individual differences in experienced productivity and its causes, underscoring the need for personalized solutionss. In our study, we will try to offer this solution in the form of a personalized workflow support system, which analyses personal data and its relation to productivity to provide personalized support.

2.2 PERSONAL INFORMATICS

In this study, we aim to identify productive patterns and aid university staff by optimizing and visualizing their workflow using a support system. Achieving this goal requires personal data, so we will explore how such data is collected, processed, and used in support tools. Personal informatics is the class of tools that uses this personal data for self-monitoring and self-reflection, which we will discuss in the next sections. Section 2.2.1 covers productivity tools, followed by an exploration of Machine Learning in personal informatics in section 2.2.2. Section 2.2.3 addresses design implications and guidelines. Finally, section 2.2.4 discusses the ethical considerations associated with the use of sensitive personal data by these tools.

2.2.1 Support systems

The first category of productivity tools we find are blocking tools. Mark et al. [14, 15] studied the impact of the Freedom software, which blocks non-essential websites like Facebook and Instagram, on productivity. Both studies showed increased perceived productivity when utilizing this kind of blocking technique. In another study by Tseng et al. [24], a system called UpTime is introduced, which helps workers transition from breaks by blocking distractions. While UpTime did not significantly impact perceived productivity, participants did report significantly lower stress levels.

The second type of productivity tools we find utilize visualizations of work behavior to improve productivity [2, 4, 8, 9, 13, 21, 25]. One of these tools is called Activity River, presented by Aseniero et al. [2], for which the main interface can be seen in Figure 2.

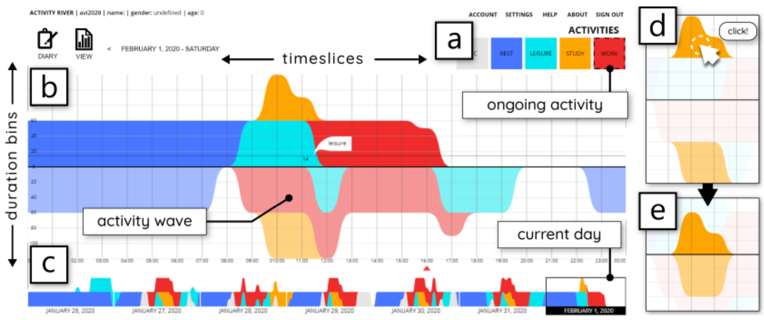


Figure 2: The main interface of Activity River, introduced by Aseniero et al. [2].

This tool provides an overview of time spent on activities versus planned time, proving to help participants stay on track. Another tool which uses visualizations is called AppInsight, presented by Barata et al. [4]. The main interface of this tool can be seen in Figure 3.

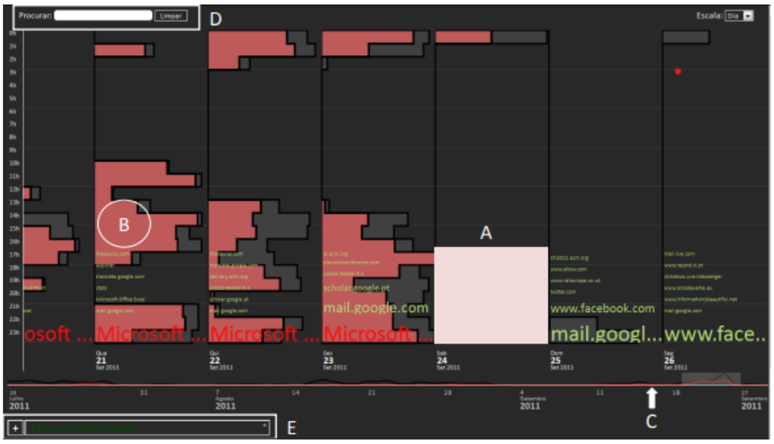


Figure 3: The main interface of AppInsight, introduced by Barata et al. [4].

The tool provides an overview of users' most used applications, helping them understand their usage patterns over time. Most users found it useful for monitoring and evaluating productivity.

The third type of tool we come across in the literature goes beyond visualizing user activities by using personal data to create predictions and guide the user in being more productive [5, 7, 10, 22]. Cranefield et al. [5] studied Microsoft MyAnalytics, which uses machine learning to improve productivity through visualizations and behavior suggestions. However, participants found it too general and not personalized. Another study of Saha and Iqbal [22] examined the effect of automatically scheduling time for focused work using the Focus Time feature on Outlook. Participants desired software to help them focus. While the study did not measure the effect on productivity, it found increased well-being and reduced anger, frustration, tiredness, and stress for participants.

These support system tools share a common potential: the potential of improving personal productivity through various means. However, they face challenges such as being interpreted as annoying and lacking in insights, which can undermine their effectiveness in improving productivity [24]. Tools relying solely on visualizations require users to independently redesign their work behavior, rather than providing direct guidance for productivity improvement. Like the users in the study of Dugan et al. [8], who wanted their visualizations to be actionable and show time management problems that matter to them, thus calling for a personalized solution.

In our study, we aim to develop a personalized workflow support system integrating visualizations and actionable suggestions for improving perceived productivity. Our approach emphasizes non-interference, ensuring minimal disruption to users during work.

2.2.2 Machine learning

In this section, we will discuss studies integrating machine learning with personal informatics to analyze productivity and patterns. The study of Di Lascio et al. [6] utilized data from devices like smartwatches and laptops to automatically classify work and break periods. Their study tested various machine learning methods, with an

XGBoost model achieving the highest F1-scores: 94% for work activities and 69% for breaks. RescueTime, the logging software used, relies on self initiated user logging, potentially missing small breaks and affecting break classification accuracy. We are interested to see how different neural networks will perform in automatically recognizing work activities compared to, for instance, a XGBoost method, by studying a combination of variables and patterns in our data.

A personalized machine learning approach for task management, called RADAR, was proposed by Freed et al. [10]. RADAR's 'Attention Manager' (AM) learns productive task sequences and alerts users when they fail to progress on important or urgent tasks. The machine learning approach significantly outperformed a non-learning approach, improving user performance and reducing information load. Post-test surveys indicated positive effects, such as increased confidence, immersion, and productivity. However, RADAR only tracks email-related tasks. Our system addresses this limitation by tracking computer activity across a range of work activities, resulting in a comprehensive personalized approach to guide users throughout their workday. Inspired by Freed et al.'s study, we will compare the performance of our personalized neural networks with baseline methods and gather user feedback to assess the potential productivity benefits of our workflow support system.

Finally, we found no literature on using machine learning to study personal work patterns and their impact on perceived productivity. As mentioned in the study by Cranefield et al. [5]: "As the ongoing sensemaking of AI continues, we call for more research to enhance human-AI collaboration for the regulation of productivity and wellbeing at work". We hope to contribute to this by offering a personalized solution which studies the effect of personal work patterns on perceived productivity.

2.2.3 System design

As our primary research goal involves developing a workflow support system, understanding how to design such a system is crucial. Studies have shown that these systems are often used for only about one minute per day [13] and that user feedback is essential for quality designs. This section discusses relevant feedback and design suggestions from the literature.

Several studies highlighted that users wanted visualizations of their work behavior in support systems [2, 5, 8, 21, 24, 25, 19]. Dugan et al. [8] found that frequent access to visualizations of time-based metrics positively impacted their time management. Another study of Aseniero et al. [2], reported that visually showing plans helped participants stay on track with their tasks.

Users also desired actionable visualizations. Dugan et al. [8] found that participants wanted clear, actionable insights from their data in order to make it easier to take action and improve time spending. One participant noted, "when I see these numbers: now what do I do with these numbers?" indicating that visualizations might not be enough on its own to improve workers' time management, as they might need actual advice from the system on what to do to improve their time management.

Personalization was another key aspect when investigating feedback on support system. Tseng et al. [24] reported that users wanted personalized lists of blocked websites based on their work behavior. Our study aims to create a personalized system that provides advice based on individual work behavior.

However, we found that users wanted to interact with visualizations within the system, they did not want the system to interact with them. Study of Whittaker et al. [25] found that most participants rejected pop-up messages, finding them annoying. The problem with some systems is that they rely on the input from these pop-up messages to learn about the worker. Cranefield et al. [5] reported that many participants found it stressful to provide quality inputs for AI-based support system.

Meyer et al. [19] offered design recommendations based on an analysis on the use of the WorkAnalytics tool, which is a self-monitoring tool made to increase self-awareness about productivity at work. These included interactive visualizations, natural language components, self-reports, activity reflections, and actionable insights. These recommendations were based on user preferences and results showing increased productivity and awareness.

Studies on design suggestions for future support system design echoed similar topics as we mentioned earlier [9, 11, 14, 17]. First of all, participants in Epstein et al. and Mark et al. [9, 14] suggested that visualizations containing information about their workflow would be useful for better time management. Mark et al. [14] also found that users wanted actionable insights, like break recommendations.

In conclusion, despite some design implications, there is strong user interest in systems that improve productivity. Mark et al. [17] suggested that tools protecting users from email overload could enhance productivity and health. Kaur et al. [11] reported positive feedback on systems recognizing optimal moments for breaks and transitions, with one participant expressing, "it can be a digital assistant looking after you and your well-being, what more could you want?!" We aim to design a system that integrates visualizations, actionable feedback, personalization, and minimal user input reliance to help workers improve productivity.

2.2.4 Ethical considerations

This section discusses the ethical considerations crucial when developing systems that guide workers in being more productive whilst using sensitive personal data. The first concern is the potential increase in stress these systems might cause. Studies have shown that support systems can sometimes lead to higher stress levels. For example, blocking software used in Mark et al.'s study [15] caused individuals with less self-control to work longer without breaks, increasing stress. A participant in another study by Mark et al. [14] noted, "It lets me focus on work more but then I'm not able to step back as easily. I didn't take breaks leaving the office and just ended up staying there." This points out the same concern that these kind of systems can overload the user, which can cause higher stress. Nyman et al. [20] suggest that shifting from using data to support dialogue to using data to improve work patterns transfers the burden from the company to the individual, potentially increasing stress.

Privacy is the second major ethical concern. Many studies highlighted concerns about the relationship between workers and management systems can bring with them [23, 11, 22, 25, 19]. For instance, participants in Meyer et al.'s study [19] feared that sharing workplace data with management could increase pressure and put their job in danger. To address this, they stored data locally and allowed participants to disable and alter data flows. Similarly, Saha and Iqbal's study [22] reported a participant's concern about feeling monitored: "I stopped going into 'inappropriate' sites because it was like I was being watched." This concern was also voiced in Kaur et al.'s study [11], with a participant stating, "feeling like you're being watched all the time would just be bad." These concerns highlight the need to protect worker privacy and ensure management cannot misuse the data.

Potential issues in the worker-management relationship were also discussed. Nyman et al. [20] noted that knowledge workers might not fully understand how these tools work, leading to oversimplified views of their work behavior. If management does have knowledge about these tools that the worker does not have, their relationship can be put into danger. Thompson and Molnar [23] also pointed out that the rationale for adopting monitoring software can differ from its application, shifting from increasing productivity to simply monitoring workers.

The literature also discusses how to build these tools in an ethical manner. Whittaker et al. [25] emphasized balancing privacy concerns and questioning the need and efficacy of such tools in the workplace. We believe that focusing on aiding individuals to improve their perceived productivity can positively impact workers, giving these types of tools huge potential in aiding the worker in a positive manner. However, ethical considerations must be taken into account during development of such tools.

2.3 CONCLUSION

In this chapter, we explored literature in the field of productivity and personal informatics. Research shows productivity is measured in various ways, often missing the individual's sense of productivity [3, 13, 25]. Factors affecting productivity include individual behavior, patterns, and external interruptions [5, 9, 10, 11, 12, 16, 17], highlighting the need for personalized solutions [5, 9, 15]. We aim to offer a personalized workflow support system based on individual data. Tools like blocking software and tools that utilize visualizations to aid the user have shown positive impacts on productivity [2, 4, 14, 15]. There are some tools that have shown the potential of machine learning used in these support tools, which tend to lack a high variety of work activities [6, 10]. We aim to fill this gap by using machine learning to predict work activities, where after a personalized workflow support system utilizes this data to discover personal patterns, consisting of a high variety of work activities, that have a significant impact on productivity. Design decisions significantly affect tool usage [2, 8, 9, 14, 19], and studying design feedback has provided valuable insights for developing our workflow support system in a user-friendly manner.

While there is enthusiasm for support tools, studies highlight ethical concerns like stress, privacy issues, and potential threats to work-management relationships [11, 14, 15, 19, 20, 22, 23, 25]. We aim to address these by offering a personalized solution that provides guidance, not authority, and is designed for individual use.

3 RESEARCH METHOD

In this chapter, the research method used in our study will be explained. Section 3.1 explains the overall structure, primarily based on the design science methodology by [24], which proposes a design cycle of three iterative tasks. The first task, discussed in section 3.1.1, focuses on investigating the problem. The second task, outlined in section 3.1.2, emphasizes designing the treatment for our research. The final task, detailed in section 3.1.3, centers on validating the treatment. Finally, the method we used for investigating the current state of the literature will be explained in section 3.2.

3.1 DESIGN SCIENCE METHODOLOGY

The research method constructed for this study was based on the design science methodology introduced by Wieringa [24]. This methodology encompasses the design and investigation of artifacts in context, where iterating over solving design problems and answering knowledge questions is the main focus. A solution can be seen as a design, for which its effectiveness is measured by investigating how well it meets the stakeholder goals. In our study, the solution is a personalized workflow support system design which aids stakeholders in being as productive as possible during their work. This research falls within the design science domain of Information Systems and Software Engineering. As mentioned earlier, the structure of this research followed the three tasks in the design cycle, displayed in Figure 4.

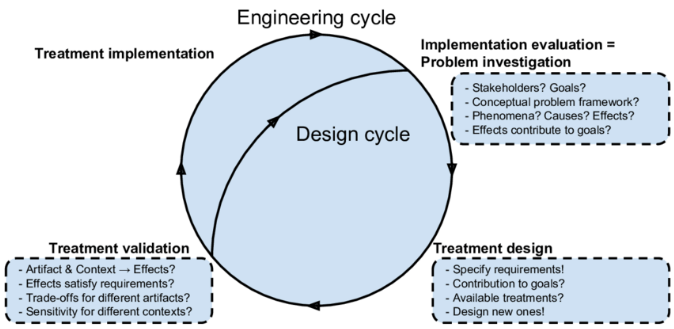


Figure 4: Engineering cycle introduced by [24].

3.1.1 Problem Investigation

The starting phase of this research project was performed in the 'Problem investigation' phase. In this phase, the preparation for the design of the treatment is done by investigating the problem at hand to gain a better understanding of it. To be more specific, in this phase existing solutions were analyzed and stakeholders and their goals were identified.

In this research, the stakeholders were University of Utrecht staff members, whose workflow was tracked using AWT software. Next to this, a daily survey gauged their perceived productivity, resulting in a dataset with descriptive information on their workflow and productivity over a time period of nearly five months. However, users could not leverage this data for personal insights to improve productivity. The more general problem here was a lack of solutions that provide personalized assistance in improving work productivity. The desired outcome in this study was for the university staff members to get unique personal insights into their workflow with the goal of improving their productivity. The existing solutions were studied and analyzed by performing a thorough literature review following a strict literature research protocol, explained in section 3.2.

3.1.2 Treatment Design

When the problem was understood well, the focus was shifted towards designing a solution for this problem. To achieve this, the requirements of the stakeholders and the context of the problem needed to be taken into account. These findings could have resulted in multiple design approaches, for which the one that met the requirements of the stakeholders and solved the problem in the best manner was chosen.

The second phase of this research project was divided into multiple tasks. A process diagram of these different tasks can be seen in Figure 5. First, the data produced by the AWT software needed to be collected, where after pre-processing needed to be done for enabling the creation of a neural network regarding work activity labels. This personalized neural network labeled previously unlabeled data, resulting in labeled datasets for each participant spanning over the productivity measurement period. These labeled datasets were combined with daily productivity scores to produce datasets containing work activities and productivity scores. Before we could showcase interesting relationships and patterns regarding work activities, the data had to be pre-processed again. After this, sequential pattern mining and data analytics were used to discover personal patterns and relationships with productivity. This process produced our final artifact: a personalized workflow support system design for each participant. Further details on each stage of our treatment design will be provided later in this section.

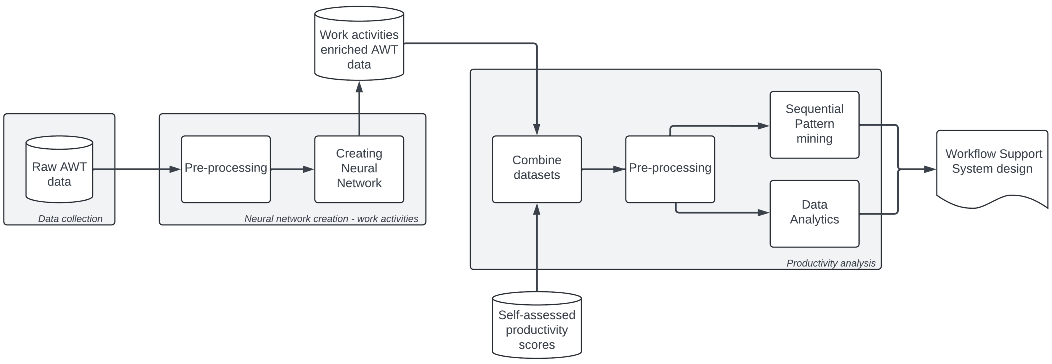


Figure 5: Process diagram of Treatment design.

Data Collection

Building a neural network that could analyze workflows required personal data. As mentioned in section 1, this data was collected by a software application called Tockler. This AWT software is free and easy to use. Once installed, Tockler automatically tracks user activity, including the application being used and a description of the activity, along with start and end times [14]. Tockler uses this data to create an overview within the application showing a timeline and a summary, containing overviews of when you are online and what kind of applications you are using. In Figure 6 an example of a part of the log data created by Tockler can be seen. In this example, the user is working in the 'Windows Shell Experience Host' for a couple of minutes, after which he switches to the 'Windows Explorer' application.

App	Type	Title	Begin	End
Windows Shell Experience Host	AppTrackItem	Quick settings	2023-10-15 23:55:21	2023-10-15 23:55:24
Windows Shell Experience Host	AppTrackItem	Quick settings	2023-10-15 23:55:22	2023-10-15 23:56:52
Windows Shell Experience Host	AppTrackItem	Quick settings	2023-10-15 23:57:04	2023-10-15 23:57:10
Windows Explorer	AppTrackItem	NO_TITLE	2023-10-15 23:57:10	2023-10-15 23:59:59

Figure 6: Start of a log created by Tockler.

In this research, Tockler data from four university staff members was collected. Preliminary research by Beerepoot et al. [5] provided a subset of this data with manually assigned work activity labels. Additionally, survey data on productivity scores was collected. This resulted in datasets from two different time periods. To give a good overview of the data that was collected, the next section will elaborate on the different sources separately, as well as the time frames for which data collection took place.

Data with work activity labels

As mentioned in section 1.2, studying the impact of higher-level work activities on productivity was a key part of this research. Thanks to preliminary research by Beerepoot et al. [5], two participants manually labeled their data over a four-week period. In Figure 7, some example lines of the log data resulting from this can be seen.

App	Type	Title	Begin	End	Date	Time	WP flow activity	Case
Google Chrome	AppTrackI	Untitled - Google Chrome	3/16/23 8:50	3/16/23 8:50	donderdag 16 maart 2023	31-12-1899 08:50:24	Communicating about events	BPM conference 2023
Google Chrome	AppTrackI	about:blank - Google Chrome	3/16/23 8:50	3/16/23 8:50	donderdag 16 maart 2023	31-12-1899 08:50:27	Communicating about events	BPM conference 2023
Google Chrome	AppTrackI	https://www.google.nl/search?hl=nl&q=anywhere+on+earth+time	3/16/23 8:50	3/16/23 8:50	donderdag 16 maart 2023	31-12-1899 08:50:33	Communicating about events	BPM conference 2023
Google Chrome	AppTrackI	anywhere on earth time - Google Zoeken - Google Chrome	3/16/23 8:50	3/16/23 8:50	donderdag 16 maart 2023	31-12-1899 08:50:36	Communicating about events	BPM conference 2023
Google Chrome	AppTrackI	Anywhere on Earth: 19:50 - Google Chrome	3/16/23 8:50	3/16/23 8:50	donderdag 16 maart 2023	31-12-1899 08:50:39	Communicating about events	BPM conference 2023

Figure 7: Start of a log created by Tockler including assigned work activity label.

This enriched data includes variables automatically tracked by Tockler, as well as the work activity and the case the candidate was working on. In this case, the participant was using 'Google Chrome' for 'Communicating about events' related to the 'BPM conference 2023'. 'WP flow activity' indicates the specific work activity. The work activities were selected from the University Job Classification system used by Dutch universities, providing a standardized subset of activities for participants to choose from [22]. Data collection for this period spanned from March 6 to April 2, 2023.

Data with productivity scores

To study the impact of personal workflows on productivity, a productivity measurement was necessary. As mentioned in section 1, a daily survey was conducted to gauge how university staff members perceived their productivity, along with factors like energy, enthusiasm, and immersion. This survey, constructed by Beerepoot et al. [5], was sent via WhatsApp to four participants at the end of each working day. The format of the survey can be seen in Appendix A. Scores were measured on a Likert scale: 1 to 5 for productivity and 1 to 7 for other metrics, with the minimum representing the lowest score and the maximum the highest. Data collection through this survey lasted from August 13, 2023, to January 31, 2024. Tockler data from these four participants was connected to the productivity data. The format of this data is the same as in Figure 6, and this data was enriched with work activity labels with the use of our personal neural networks.

Table 1 provides an overview of all datasets and their collection time frames. Participants gave consent by signing the form in Appendix B, ensuring they understood the research goals and their rights regarding participation in our study.

Source	Time frame (day/month/year)	Candidates
Data with work activity labels 3.1.2	06/03/2023 - 02/04/2023	2
Data with productivity scores 3.1.2	16/08/2023 - 31/01/2024	4

Table 1: Time frames and participants counts of different data sources.

Data pre-processing - Work Activities
Before creating personalized neural networks capable of assigning work activity labels, the *Data with work activity labels* underwent essential pre-processing steps. First of all, the dataset contained some missing values. To address this, the Python data manipulation tool pandas was used [15]. This library allowed us to easily import the dataset into a pandas DataFrame, on which we could perform methods to delete rows with missing values.

The datasets used in this study contained sensitive information regarding students and colleagues, including their names appearing in logs. To protect privacy, the data underwent anonymization using the Python programming language and the Natural Language Processing (NLP) library spaCy [11]. Before names of people could be identified, the data needed to be tokenized. Tokenization is the process of parsing text into tokens, which are words in our case [23]. Subsequently, a Named Entity Recognition (NER) model was used to label each token. Tokens identified as 'PERSON' were anonymized by replacing them with the placeholder string 'PERSON', and common identifiers such as the individual's email address were also replaced. The anonymized entities from each participant's dataset are summarized in Table 2. This table shows the number of entities anonymized in relation to the total number present in the respective columns of the datasets.

Dataset	Column(s)	Total entities	Anonymized entities
p1: Data with work activity labels	Title & Case	51.353	4576
p1: Data with productivity scores	Title	358.346	27.867
p2: Data with work activity labels	Title & Case	30.860	5169
p2: Data with productivity scores	Title	180.599	18.781
p3: Data with productivity scores	Title	145.461	28.360
p4: Data with productivity scores	Title	197.911	44.749

Table 2: Anonymized entities for all datasets with respect to the total entities.

The full anonymization process is detailed in the Python script '`anonymization.py`', available in our GitHub repository where all study-related code is stored [13].

Neural Network creation - Work Activities

Following data pre-processing, personal neural networks were developed to assign 'WP flow activity' labels to unlabeled data using data from participants one and two, who labeled their data with work activities. Given the time series nature of our data, it was crucial to consider both the data before and after an instance to make accurate predictions, as each line represents a specific moment in time, and activities span over periods of time. To capture these long-term dependencies, we utilized a deep learning model, specifically a neural network [7]. As only participants one and two had labeled their data with work activity labels, we created three neural networks: one for participant one, one for participant two, and one for participants three and four, using the data from participants one and two. This task is presented as a single process in Figure 5, although it was constructed of multiple sub-tasks. The process diagram of this modeling process can be seen in Figure 8.

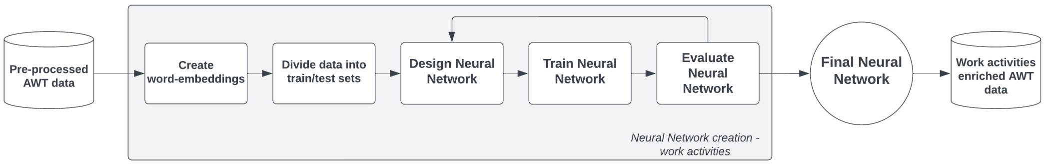


Figure 8: Process diagram of Data modeling regarding Work Activities.

First, we created word embeddings to feed data into the neural networks, using an embedding layer to generate vectors for each word while considering their similarities. For instance the words 'Google' and 'Chrome' had similar vectors as they appeared next to each other frequently in our dataset.

These types of word embeddings were created for the 'Title' feature, as this feature consisted of a description containing semantic value. Additionally, the 'Title' feature contained a wide range of values for both participants, making word embeddings appropriate for capturing the semantic meaning and variability, as shown in Table 3. To create word embeddings, we used the BERT language model, which stands for Bidirectional Encoder Representations from Transformers. This model considers the meaning of words within a sentence by jointly conditioning on both right and left context in all layers. The Python code for transforming the 'Title' feature into a BERT embedding is called `'nn_work_practices.py'`. It loops through the feature values, loads the pre-trained BERT models, calls a function to replace web links with the placeholder 'LINK', replaces the feature content with BERT embeddings using a thread pool, and returns the dataframe with features transformed into BERT embeddings.

Participant	Instances	Work activity count	Case count	Title count	App count
p1	10064	28	35	1871	34
p2	8717	37	38	1289	33
p1+p2	18781	55	71	3127	58

Table 3: Unique value counts of feature values from datasets of participant one and two.

As we can see, the features 'Work activity', 'Case', and 'App' are more categorical in nature as they contain a relatively small set of unique values, and were therefore transformed into one-hot encodings. These encodings are binary arrays, with the array length equal to the number of unique values in the dataset for that feature. For example, the 'App' feature in participant one's dataset resulted in a vector of length 34.

The final feature transformation involved time features. The 'Begin' and 'End' features, which recorded the start and end times of events, were transformed into 'Hour', 'DayofWeek', and 'Month' features. These were represented as one-hot encodings to capture the different times a participant performed a work activity. Additionally, a 'DurationSeconds' feature was created to represent the event's duration, scaled between 0 and 1 to ensure balanced learning across all features. The final features after transformation can be seen in Table 4

Feature	Before transformation	Transformation	After transformation	Final feature
WP flow activity	string	One-hot encoding	[1, 0, ..., 0]	Work activity
App	string	One-hot encoding	[1, 0, ..., 0]	App
Case	string	One-hot encoding	[1, 0, ..., 0]	Case
Title	string	BERT embedding	[0.053462, ..., 0.2254345]	Title
Begin	datetime	One-hot encoding	[1, 0, ..., 0]	Hour
Begin	datetime	One-hot encoding	[1, 0, ..., 0]	DayOfWeek
Begin	datetime	One-hot encoding	[1, 0, ..., 0]	Month
End - Begin	datetime	MinMaxScaler	0.5, range 0-1	DurationSeconds

Table 4: Transformation results of all features.

Data split

Secondly, we split the data into training and testing sets to ensure that the neural networks performed well on unseen data. For the neural networks of participants one and two, we used an 80/20 split, with 80% of the data for training and 20% for validation. For participants three and four, we concatenated the splits from participants one and two. This splitting method is illustrated in Figure 9, and the Python script for this is called `'nn_work_practices.py'`.

First, the data for participants one and two was split separately. The resulting splits were used to train their respective neural networks. For participants three and four, we combined these splits to create a versatile neural network. The data, consisting of time series where order and relation are important, was grouped by day. The splits were performed at the cut-off point closest to 80% to ensure both training and testing sets contained complete days of data. After this, the training and testing sets were divided into X (features) and y (target variable) sets. A 10% hold-out validation set, extracted from the training set, was used during the training process. Figure 9 illustrates the training, testing, and validation split applied to the datasets.

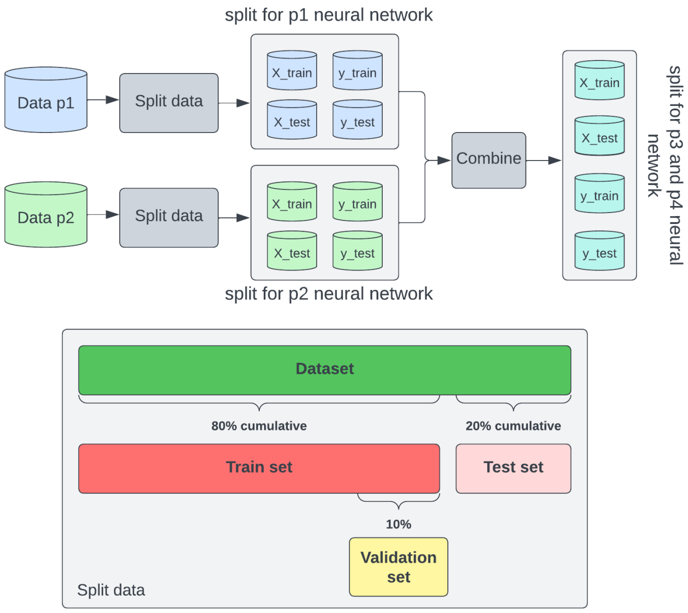


Figure 9: Training, testing and validation split applied to the datasets.

Sequences

In our study, where we aimed to predict work activities based on complex feature combinations, the lengths of work activities varied significantly, with an average length of approximately 22 lines but extending up to 600 lines for some instances. Therefore, instead of feeding the neural network data per line, we wanted to experiment with different sequence lengths.

The neural network aimed to accurately predict work activities from data sequences, which typically averaged 20 lines in length and varied. To accommodate this variability, sequences of varying lengths were used during training. Including overlap between these sequences helped prevent loss of critical information at sequence boundaries.

For example, consider incorporating the 'App' feature as a predictor for participant one. Initially, the shape of this feature in the training set was (8719, 34), indicating 8719 data lines with a 34-dimensional one-hot encoded vector. After generating sequences of length 20 with an overlap of 1, the shape of the feature became (8700, 20, 34). This resulted in 8700 sequences, each consisting of 20 consecutive data lines with an overlap of one, while preserving the original vector length of 34 for each line. The last 19 lines of data are eliminated, as we can not make an extra sequence with those. A visualization of generating such sequences from the training set of participant one can be seen in Figure 10.

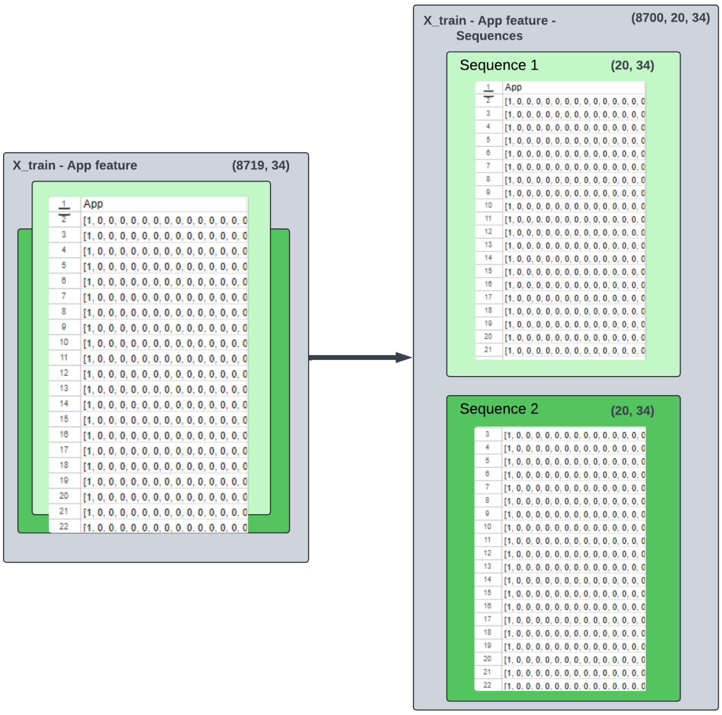


Figure 10: Example of sequence generation for p1 with a sequence length of 20 and an overlap of 1.

Neural network design

The datasets were now ready to use as input for our neural networks. After this, we designed the neural networks by defining the model type, number of layers, neurons, and the type of connection between the layers of our models. In section 1.4.1, we discussed the neural network's ability to learn over time, identifying complex patterns and non-linear relationships within the data. This motivated us to use a neural network for our task. Given the time series nature of our data, which captures users' activities over time, we chose a Bidirectional Long Short-Term Memory network (Bi-LSTM).

The Bi-LSTM architecture was selected for its ability to capture dependencies in both past and future contexts. Unlike other architectures, Bi-LSTM processes input forwards and backwards, capturing dependencies across entire sequences [21].

We initially started with a limited set of features due to the large size of our datasets, which made adding features time-consuming during training. After training and comparing several neural networks, additional features were incorporated into the best performing model to assess potential improvements. We started with 'Title' and 'App' features because we believed that combining information about the application a person is using with a textual description of their activity could provide valuable insights into their work activities.

Following these decisions, the neural network architecture became clear. Each network included two input layers, one for each feature. Additionally, two Bidirectional

LSTM layers were included, corresponding to each feature, with their outputs concatenated. The networks concluded with a Dense output layer, which had the ability to classify instances based on the different work activity feature values. A visualization of this architecture can be found in Figure 11

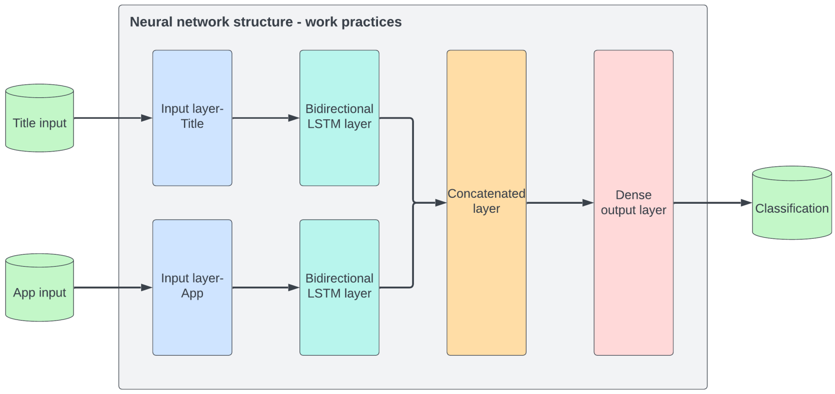


Figure 11: Structure of neural network for identifying work practices.

Hyperparameter tuning

After this, we could train the models on our training sets. During this task, decisions had to be made regarding parameters that needed to be defined during the neural network creation. These were the following:

- **Sequence length**: the line length of the chunks of data the neural network is trained with.
- **Overlap**: the overlap in similar lines between the chunks of data.
- **Epochs**: the amount of times the chunks are passed back and forward through the neural network, where one epoch consists of one back and forward pass.
- **Batch Size**: the amount of chunks that are passed to the neural network in each training iteration.
- **Units**: the number of neurons in each layer.
- **Optimizer**: the algorithm used during training of the neural network to minimize the loss function.

To identify the optimal set parameters, we conducted a grid search across 640 different combinations for our neural networks. We evaluated each combination based on accuracy, loss, and f1-score, prioritizing the neural network with the highest accuracy, and considering other metrics if accuracy scores were comparable. Accuracy captures the overall correctness of the model by measuring the correctly predicted labels in relation to the total number of predictions within the testing set, loss evaluates the performance during training by comparing the predicted values with the actual values and f1-score battles possible class imbalance by considering both precision and recall, which take into account both false positives and false negatives. These results will be presented in section 4.1. The final code for performing this grid search is called 'nn_work_practices.py'.

Parameter	Value(s) p1	Value(s) p2	Value(s) p3 and p4
Sequence length	5, 10, 25, 50, 100	10, 25, 50, 100	10, 25, 50, 100
Overlap	1, 10, 20, 50	1, 2, 5	1, 2, 5
Epochs	10, 25	10	10
Batch Size	10, 25	10, 25	10, 25
Units	64	64	32, 64
Optimizer	sgd, adam, rmsprop, adagrad, adadelta, adamax, nadam, ftrl	adam, rmsprop, adamax, nadam	adam, rmsprop, adamax, nadam

Table 5: Parameter values used during the grid search of participants.

The final artifact from this process was a personal neural network that was used to assign work activity labels to unlabeled data.

Combine datasets - Productivity
With the neural network artifact created, we could now enrich all unlabeled data with work activity labels. This resulted in the ability to transform data in the format shown in Figure 6, to data in the format shown in Figure 7, meaning that we could now predict the work activity the individual was working on based on their raw AWT data. After this, we combined the self-assessed productivity scores with the enriched AWT data, creating a dataset that linked productivity scores with the corresponding work activities for each participant.

The Python script used for this process is called 'merge_productivity_scores.py'. Productivity scores were added to data recorded before participants filled out their surveys. For instance, if a survey was completed at 8pm, all events ending before 8pm were labeled with the productivity score. For some days, multiple scores were found, which was mostly due to participants filling in the questionnaire of multiple days in the past at one moment in time. We recorded these cases for each participant and checked this data by hand. For the days that were not completed, the rows in the labeled data were dropped. The resulting days of data with a related productivity score per participant after this pre-processing step can be found in Table 6.

Participant	Data before merge	Data after merge
p1	160 days	142 days
p2	137 days	72 days
p3	112 days	48 days
p4	138 days	108 days

Table 6: Data instances before and after merging productivity scores with labeled data.

Pre-processing - Productivity
As noted in section 1.2, we were interested in sequences of work activities and their related features. The combined datasets contained numerous data points, with instances ranging from seconds to minutes. To focus on work activity sequences, the

data was consolidated so each line represents a sequential working period of one work activity. This involved creating new features like application switches and the duration of each work activity. The Python script '`grouped_per_activity_per_day.py`' facilitated this process, including counting application switches and recording time-related features. An example of this data consolidation can be seen in Figure 12.

Figure 12: Data pre-processing steps for our Data Analysis.

Data Analytics and Sequential Pattern mining - Productivity

After pre-processing, we started building the main artifact of this research: the workflow support system design. The main objective of this artifact was to provide university staff members with personalized aid during their work to improve productivity. As mentioned in section 1.2, our goal was to investigate both low- and high-level in workflows. Given the possibility of both horizontal and vertical relationships in our data, a combination of data analytics and pattern mining was used to investigate these relationships.

First of all, we analyzed low-level variables impacting productivity, such as time-related features, application switches, and the nature of the work activity. Were there for instance certain work activities that occurred more often on productive days? Or were there certain days that were more productive in general compared to others? These are the type of questions we tried to answer with the use of data analysis, including statistical tests.

Data Analytics - feature engineering

Due to the variation in work activities, time-related features, and application switches, we created levels for our variables based on the distribution within each participant's data. The Python script '`create_levels.py`' was used for this, resulting in the following features:

- **Application switches level**: To account for the distribution heavily leaning towards zero, indicating that participants mostly used one application per work activity, we created two levels: no switches and switches. Thus, this level indicated if an individual used one or more applications whilst performing a work activity.

- **Work activity duration level**: Using the 25th and 75th percentiles as cut-off points, we created three levels for each participant: short, medium, and long. Due to recording inconsistencies, we dropped rows with zero duration for participant two. The resulting levels are detailed in Appendix I.

- **Day Part**: To investigate the impact of different times of the day on productivity, we created a 'Day Part' feature with levels: morning, afternoon, evening, and night. Each level required a minimum of eight hours of data to be considered. The result of this can be seen in Appendix I.

The result of these feature engineering steps can be seen in the example rows, shown in Figure 12. Because we were dealing with a daily productivity score and we wanted to study different factors within these days, we applied a time weighted average with our statistical analysis. For instance, when investigating the impact of different work activities on productivity, a participant 'Adapting a publication' for four hours on a day with a productivity score of five contributes more heavily to the average productivity score of 'Adapting a publication' compared to a day where the participant was 'Adapting a publication' for one hour with a productivity score of two.

Data Analytics - statistical techniques

To analyze different relations between features and productivity, we applied several statistical techniques:

- **Linear Regression**: Used to assess the linear relationship between the different days of the week and productivity.

- **Kruskal-Wallis test**: Non-parametric test used to compare productivity across different feature groups, such as with part of the day and work activity.

- **Mann-Whitney U test**: Non-parametric test used to compare productivity across different feature groups when groups had less than five instances, such as with work activity.

- **One-way ANOVA**: Parametric test used to compared productivity across different feature groups when assumptions were met.

- **Post-hoc tests**: Used to further analyze which specific relations were significant between feature groups.

Before applying the statistical techniques, we tested the assumptions that had to be met before utilizing them. These assumptions were residual independence, linearity, homoscedasticity, normality of residuals and multicollinearity. To confirm these assumptions, we utilized various techniques such as residual plots, histograms, QQ plots, Variance inflation factor (VIF) calculations and parametric tests. By following this systematic approach of testing assumptions before applying statistical techniques, we ensured the validity of our findings.

Pattern mining

Secondly, we investigated sequential patterns of work activities. We extracted patterns of various lengths and created a new dataset where each line represented a sequential pattern of work activities with related features like duration, application switches, productivity score, start time, and end time. The resulting dataset represented the sequential patterns of work activities performed for each participant. This allowed us to investigate relationships between patterns, their features and their impact on productivity.

First of all, before any type of analysis could be done, the patterns had to be mined from the datasets. To achieve this, we decided to write our own Python function called '`create_patterns.py`', which creates sequential patterns of work activities. The script initiates a loop for each participant, iterating through the final dataset created during the Data Analysis phase (Figure 12). In this loop, sequential sequences of work activities of length n are extracted, specifically patterns of lengths two through four. To make sure that each sequence is sequential, the 'Begin' and 'End' time of the neighboring items within the sequence were checked. If they matched, the sequence was valid.

During the creation of the patterns, multiple variables were created. Total Work Duration, which is the sum of the 'Work activity duration' of each element within the pattern. Total Switches, which is the sum of the 'Application switches' of each element within the pattern. Day Part, which is the part of the day in which the pattern was performed the longest based on the 'Begin' and 'End' time. Productivity, which is the productivity score of the day the pattern was performed. The final variable was Date, which is the date the pattern was performed.

Pattern mining - feature engineering

After the sequential patterns of work activities were identified, within the same Python script, the features of these patterns were also created. For this, we followed a similar approach as in the Data Analysis phase. We created levels for feature values because of the high variability. This resulted in the following features, which were determined separately for each participant:

- **Total switches level**: For each pattern length, the 25th and 75th percentile were calculated based on a list containing all total 'Application switches' of each pattern with that length. These were used as cut-off points and resulted in the levels low, medium, and high. The values of these levels can be seen in Appendix I.

- **Total work duration level**: For each pattern length, the 25th and 75th percentile were calculated based on a list containing all total 'Work activity duration (s)' of each pattern with that length. These were used as cut-off points and resulted in the levels short, medium, and long. The values of these levels can be seen in Appendix I.

- **Average Productivity**: The sequential patterns were grouped together based on the variables 'Pattern', 'Total work duration (level)', 'Total switches (level)' and 'Day Part'. Whilst doing this, each 'Productivity' score associated with the pattern was appended to a list. The average of this list was taken to determine the 'Average Productivity' score of that unique sequential pattern.

- **Productivity scores**: This is a list of productivity scores on which the average is base. This is recorded for statistical testing later.

- **Count**: A count representing the amount of times the grouped pattern occured throughout the dataset.

An example of the resulting data, regarding sequential patterns, can be seen in Figure 13.

1	Pattern	Total Work Duration (level)	Total Switches (level)	Day Part	Average Productivity	Count
2	('Drafting conference papers', 'Proposing events')	medium	low	morning	4.043478261	23

Figure 13: Example of an instance of the sequential pattern data.

The resulting amount of sequential patterns extracted from the datasets can be seen in Table 7.

Participant	Patterns (n=2)	Patterns (n=3)	Patterns (n=4)	Total patterns
p1	6.166	18.197	25.539	49.902
p2	1.821	2.642	2.492	6.955
p3	4.302	8.625	10.177	23.104
p4	2.902	5.804	7.125	15.831

Table 7: Amount of patterns extracted regarding different lengths for all participants.

Finally, the findings from data analysis and sequential pattern mining were integrated into the workflow support system design. Visualizing these findings provided users with unique insights into their work activity patterns and overall behavior in relation to productivity, which will be further explained in section 3.1.3.

3.1.3 Treatment Validation

In the final phase, 'Treatment validation,' we evaluated the artifact from our study to see if it contributed stakeholder goals and satisfied its requirements. If this was the case, the artifact would be ready to be implemented into the real world.

Workflow support system
As explained in section 3.1.2, we explored various relationships between features and productivity through data analysis and sequential pattern mining. However, the artifact needed to be of use for stakeholders. One goal of this research, as mentioned in section 1.2, was to aid the university staff members by optimizing and visualizing their workflow with the use of a workflow support system. Figure 14 shows an example of a visualization providing both visual insights and textual aid, implemented in our workflow support system design.

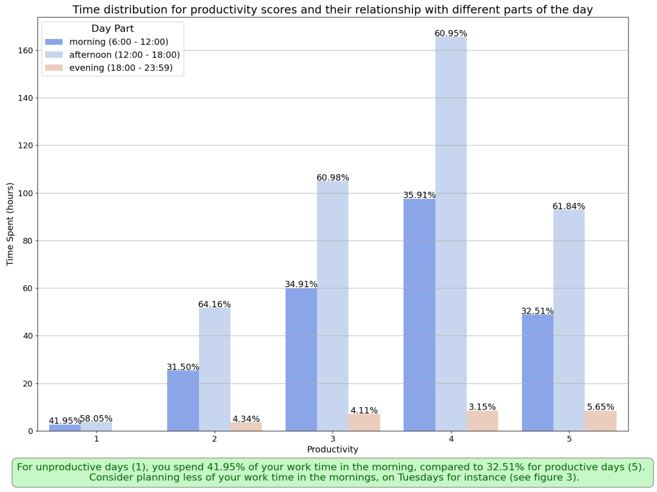


Figure 14: Example of a visualization with textual aid present in our final workflow support system design.

In this visualization, we can see information on how the individual spent time working during different parts of their day, and its relation to productivity. We will further elaborate on the different components of the dashboard in this section. The workflow support system design consists of five pages, each containing multiple visualizations, possibly accompanied by prompts which help to further support the user. For each participant, these designs contain the same visualizations, but giving different insights which are personalized to their data. We will explain the design in more detail in the upcoming sections. The complete dashboards designs for all participants can be found in Appendices L, M, N and O.

General metrics

On the first page of the dashboard, six visualizations are displayed, created by the Python script '`show_activities.py`'. These visualizations include the 'Time Weighted Average Productivity' showing overall productivity scores. They also show the distribution of productivity scores given, and productivity scores by day of the week and total work time. A scatter plot illustrates productivity scores and total time spent for different work activities. Additionally, two bar plots compare productivity scores with time spent in different parts of the day and the use of single versus multiple applications.

Best versus worst weekday

On the second page of the dashboard, visualizations compare the day with the lowest average productivity to the day with the highest average productivity. This detailed view expands on the third plot from the first page. The visualizations were created using the Python script '`show_worst_best_day.py`'. The first bar plot shows differences in percentages of total time spent on work activities between these days, helping identify the work activities that are performed more on productive days compared to unproductive days and vice versa. The second bar plot compares average work time across different parts of the day for the worst and best day. The third bar plot compares the use of one application versus multiple applications on the worst and best day. The final bar plot shows the average work duration on the most productive day versus the least productive day, accompanied by the average productivity score.

Work activity

On the third page of the dashboard, an in depth look into a single work activity is given, focusing on one of the work activities for which the individual performed the most time. The visualizations were created using the Python script 'show_activity_relations.py'. The first and second bar plots show the overall time-weighted average productivity and the distribution of productivity scores, but specific to the work activity. The third and fourth bar plot were also used on page one of the dashboard. They showcase the same principles, but now purely focusing on the time spent working on the specific work activity. Next to this, both plots give advise on how to improve productivity on the work activity. The fifth plot introduces results from sequential pattern mining, showing the top five most productive patterns containing the work activity. The final plot shows the top five least productive patterns containing the work activity.

Sequential pattern metrics

The fourth page of the dashboard is dedicated to the results of sequential pattern mining. This page attempts to give the user an overview of opportunities for improvement regarding time of the day, pattern length and application use whilst performing different sequential patterns of work activities. The visualizations on this page were created with the Python script called 'show_patterns.py'. The first bar plot shows the sequential pattern for which the biggest improvement can be made by performing the pattern at a different time of the day. The second bar plot shows the pattern where changing the duration could improve productivity. The third bar plot highlights the pattern where adjusting the number of applications used could improve productivity. The two bottom plots are similar to the ones on the third page of the dashboard. The difference here is that it shows the top five most and least productive sequential patterns of work activities over all the patterns extracted. For the pattern to be relevant the count had to be at least the mean count. Next to this the patterns were sorted on average productivity first, where after they were sorted on count.

Sequential pattern metrics - work activity insights

The final page of the dashboard also focuses on sequential pattern mining results, featuring three bar plots similar to those on page four: examining part of the day, pattern length, and application use. Next to this, the two plots showing the top five most and least productive patterns are also included. These visualizations, generated using the Python script 'show_patterns_details.py', allow users to specify any work activity or set of work activities. This can give the user a more in depth view on sequential patterns, whilst focusing on a certain work activity or work activities.

User Feedback
While measures and statistical tests validated our findings, stakeholder adoption and usability of the system were crucial. To validate this, we used the Method

Evaluation Model (MEM) by [18], which is based on the Technology Acceptance Model (TAM) and Methodological Pragmatism [8, 20]. TAM is widely accepted and used to investigate user acceptance of new systems, focusing on perceived usefulness and ease of use [12, 19]. Methodological Pragmatism, asserts that the validation of knowledge methods relies on their real-world success in emphasizing efficiency and effectiveness over proving correctness [20]. An overview of the model can be seen in Figure 15.

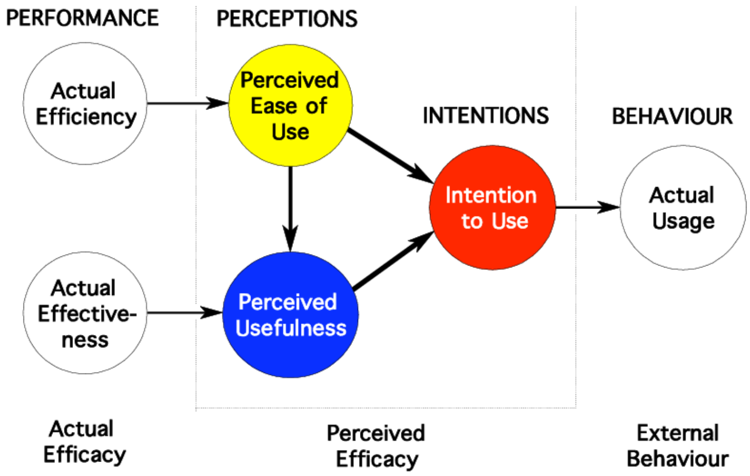


Figure 15: Method Evaluation Model (MEM) [18].

In our study, performance was measured by the effort required to apply our workflow support system design and how well it met its objectives. Stakeholder feedback on ease of use determined how useful the system would be in practice. The more useful the stakeholders felt the system was, the more they would intent to use it. This feedback influenced whether users would adopt the workflow support system to improve productivity if the system were to be deployed. A questionnaire, detailed in Appendix Q, was created to evaluate the system based on the different aspects of the MEM.

In the literature, valuable feedback on system design, detailed in Chapter 2.2.3, mostly came from both open and closed questions. While the TAM model questionnaire included only closed questions, we added open questions to gather deeper insights. For instance, we asked not just if the workflow support system would improve productivity but also why they thought it would or would not, and which visualization they found most useful. This approach provided more valuable insights into our system design.

During the three phases of Wieringa's engineering cycle [24], various methods were employed to address our research questions and achieve our main goal. Table 8 provides an overview of these methods used across the three iterative phases.

Phase	Literature review	Data analysis	User surveys
Problem investigation	X		
Treatment design		X	X
Treatment validation		X	X

Table 8: Research methods used for the different phases of the Engineering cycle.

3.2 LITERATURE RESEARCH PROTOCOL

A thorough and comprehensive literature review is an essential part of any research project. This provides researchers with knowledge about the current status of the research matter, the validity of the research and the position of the research within the field. The difficult part was finding our way through all the literature in the field, published by many online sources. Therefore, it was important to develop a strong framework to guide us through the vast literature and ensure that all relevant aspects of the field were covered. The full protocol we used can be found in Appendix P.

4 RESULTS

In this chapter, we present the results of our study. First, in section 4.1, we cover the results of evaluating the neural networks and labeling the data with work activities. Following this, in section 4.2, we will present the results of data analysis regarding different variables and productivity. Next, in section 4.3, we present the results analyzing the relations between sequential patterns and productivity. Finally, in section 4.4, we present the results of our workflow support system, which consists of analysis of the visualizations in our dashboards and feedback on these, gathered through a MEM questionnaire.

4.1 NEURAL NETWORK – WORK ACTIVITIES

This section presents the results of our neural networks developed to identify and assign work activity labels to unlabeled data. We first outline the performance metrics for each participant, including baseline accuracies and grid search outcomes. We then discuss the challenges in labeling due to the unique 'App' values and our solution involving sub-categorization. Finally, we provide the performance metrics of the re-trained neural networks and the results of labeling the datasets.

Neural network results - p1

To assess our neural network's performance, we established a baseline accuracy using the number of unique 'Work activity' labels, which was 28 (see Table 3). Random guessing would yield approximately 3.6% accuracy, which we set as our baseline accuracy. The best performing model resulting from the grid search achieved an accuracy of 84%, outperforming the baseline by a factor of approximately 23 (Figure 29). Additionally, this neural network also performs the best when looking at the f1-score. For the loss, the model does not get into the top ten, but still has a respectable value of 0.89. After attempting to improve the model by adding additional Bidirectional LSTM layers for each feature, no improvements were observed (Table 10). The final neural network structure for participant one is detailed in Table 9.

Neural network results - p2

For participant two, with 37 'Work activity' labels, the baseline accuracy was approximately 2.7% (Table 3). Initially, using parameters similar to participant one resulted in an accuracy of 49%. Despite outperforming the baseline by roughly 18 times, we tried to improve its performance. Thus, we conducted a grid search tailored to participant two, focusing on a subset of parameters and including overlaps of two and five. This approach yielded 96 parameter combinations. The top model achieved an accuracy of 59.6% and a f1-score of 0.61 (Figure 30). Regarding the loss, the model does not get into the top ten, but still achieves a respectable score of 3.0. After attempting to improve the model by adding additional Bidirectional LSTM layers for each feature, no improvements were observed (Table 10). The final neural network for participant two utilized parameters outlined in Table 9.

Neural network results - p3 & p4

For the final neural network regarding work practices, a baseline for participant three and four was created in the same manner as for the previous models. In Table 3, we can see that there are 55 unique 'Work activity' labels for the data of participant one and two combined, resulting in a baseline of 1.8%. A grid search was conducted with variations in units to optimize model efficiency due to the size of the dataset, resulting in 192 parameter combinations. The top model achieved an accuracy of 64%, significantly exceeding the baseline by approximately 35 times (Figure 31). Similar to previous participants, attempts to enhance the model with additional features did not improve performance (Table 10). The final neural network for participants three and four utilized parameters which can be seen in Table 9.

It is good to note that we did not experiment with adding the 'Month' feature, as these datasets only contain instances within the same month, which makes this feature useless. Next to this, the 'Case' feature is only present in the labeled datasets, and therefore not useful for these models.

Participant(s)	Baseline	Accuracy	Parameters	Features
p1	3.6%	84%	Sequence length = 100, Overlap = 1, Epochs = 10, Batch size = 10, Units = 64, Optimizer = adam	Title & App
p2	2.7%	60%	Sequence length = 100, Overlap = 1, Epochs = 10, Batch size = 25, Units = 64, Optimizer = adamax	Title & App
p3 + p4	1.8%	64%	Sequence length = 100, Overlap = 2, Epochs = 10, Batch size = 10, Units = 64, Optimizer = nadam	Title & App

Table 9: Parameters of best performing neural networks resulting from grid search for all participants.

Participant(s)	Metric	Title & App	+ Hour	+ DayOfWeek	+ DurationSeconds	Title & Sub category App
p1	Accuracy	0.84	0.78	0.61	0.82	0.83
p1	Loss	0.89	1.40	2.17	0.87	0.96
p1	F1-score	0.85	0.78	0.61	0.82	0.84
p2	Accuracy	0.60	0.42	0.54	0.52	0.55
p2	Loss	3.00	4.06	3.57	3.34	3.10
p2	F1-score	0.61	0.44	0.55	0.54	0.56
p3 & p4	Accuracy	0.65	0.54	0.55	0.60	0.61
p3 & p4	Loss	2.63	4.22	3.62	3.24	3.18
p3 & p4	F1-score	0.66	0.55	0.55	0.61	0.63

Table 10: Performance difference after adding features to neural networks.

Data labeling

The final step we needed to perform was to use the different neural networks from the participants to label the unlabeled datasets. When performing this labeling, we ran into a problem. The 'App' feature had more unique values in unlabeled datasets compared to the labeled datasets, on which the neural networks were trained. The input layers corresponding to this feature could only handle one-hot encodings from the 'App' values it was trained on and therefore, the neural networks could not process all the data. To resolve this, we categorized all 'App' values into 12 sub-categories, which can be seen in Appendix H. The Python script '`subcat_app.py` helped us with this, showing all unique 'App' values per participant, which we manually assigned to one of the 12 sub-categories. Subsequently, we re-trained the neural networks using 'Sub-category App' instead of 'App', resulting in a slight decrease in performance across all models, as indicated in Table 10. Despite this, all neural networks still outperformed our baseline by a minimum factor of roughly 20, confirming their suitability for labeling datasets.

The Python script '`label_data.py`' was used to label datasets. Before making predictions on 'Work activity', we preprocessed the data by transforming 'Title' into BERT embeddings and 'App category' into one-hot encodings. Data was then organized into sequences with overlaps, aligning with the optimal sequence length and overlap used in the best performing model. This approach resulted in each instance having up to 100 predictions across different sequences, where the most frequent label was chosen as the final prediction. Finally, the 'Title' and 'App category' were transformed back to the original strings, resulting in our final labeled datasets.

4.2 DATA ANALYTICS – PRODUCTIVITY

In this section, we will present the results of analyzing the impact of low and high-level variables, such as day of the week, part of the day, and work activity, on productivity. We investigated how these factors relate to productivity across all participants.

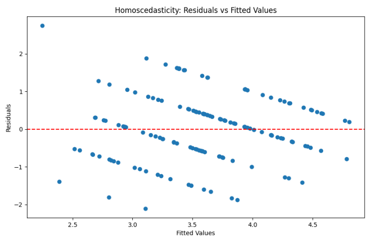
Day of the week

The first low-level variable that we were interested in analyzing was the actual day of the week. To be more precise, we wanted to find out if there were specific days during the week where an individual experienced a significantly different feeling of productivity compared to other days. To investigate this, we created a linear regression model for each participant. We will detail this process for participant one and then present the results for all participants. The Python script used during this process is called '`stats_day_of_week.py`'. First, we consolidated the data to create a subset with one row per unique date, including the productivity score, day of the week, and total work time. With this data we could construct the following linear regression formula:

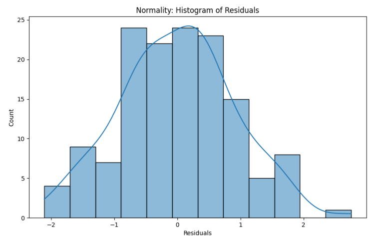
$$\text{'Productivity} \sim \text{C(day\_of\_week, Treatment("\{lowest\_productivity\_day\}"))} \times \text{duration\_m\_tot'} \tag{1}$$

With this model we were able to measure the effect of the dependent variables productivity against the independent variable day of the week. We treated the day of the week with the lowest time-weighted average as the reference day. Next to this, we included total work time in minutes for each day as an interaction variable to account for variations in work duration, ensuring that differences in productivity scores were not skewed by differing amounts of work time.

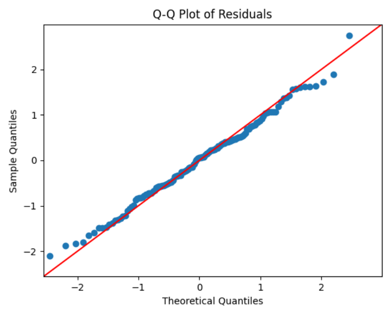
Before we analyzed the results of running our regression models, we needed to make sure the assumptions related to linear regression were met. The assumptions for independence, linearity, homoscedasticity and multicollinearity were met for all participants, which is shown for participant one in Figure 16.



(a) Residuals vs Fitted values for Linearity and Homoscedasticity assumption.



(b) Histogram of residuals for Normality assumption.



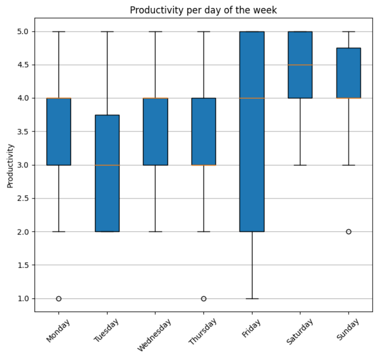
(c) QQ Plot of the residuals for Normality assumption.

Predictor	VIF
day_of_week_Monday	1.220956
day_of_week_Saturday	1.159332
day_of_week_Sunday	1.195269
day_of_week_Thursday	1.190270
day_of_week_Tuesday	1.176114
day_of_week_Wednesday	1.215683
duration_m_tot	1.167549

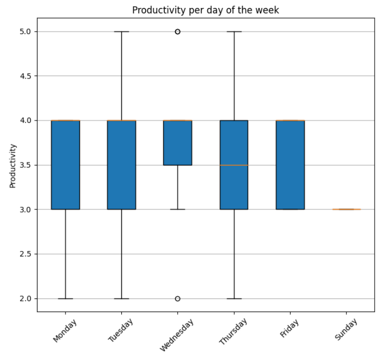
(d) VIF values of all independent variables for Multicollinearity assumption.

Figure 16: Assumption test for linear regression model of participant one.

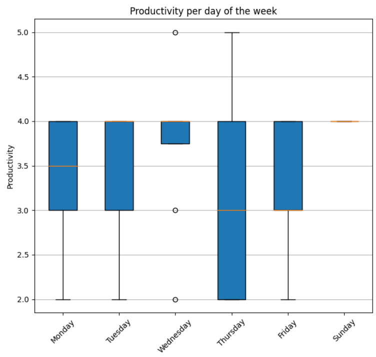
The resulting coefficients estimated by the regression models, including statistical significance and interactions, can be found in Appendix J. Each participant experienced a different day as being the least productive. Figure 17 shows box plots of productivity scores per day for all participants. Without considering work duration, some days were significantly different from the least productive day. For participant one, Saturday and Sunday was rated as being significantly more productive than Tuesday. For participant two, Saturday was rated as being significantly more productive than Sunday, despite the high amount of data for Saturday. Participants three and four had no days rated as being significantly more productive than their least productive day. When accounting for total work duration ('duration_m_tot'), no days were rated as being significantly more productive than the least productive day for any participant.



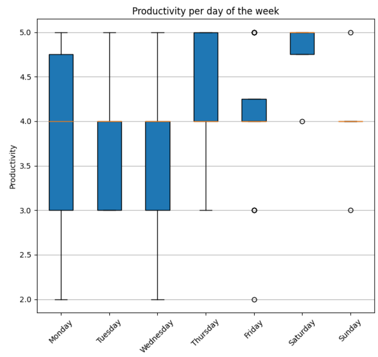
(a) participant one.



(b) participant two.



(c) participant three.



(d) participant four.

Figure 17: Boxplots of the spread of productivity scores for the different days of the week regarding all participants.

Part of the Day

The second low-level variable we were interested in was part of the day. We wanted to find out if there were certain parts of the day where the participant experienced a higher productivity. We do not see any value in comparing if a participant is more productive in the morning compared to the afternoon in general, as we believe that this information is not specific enough, therefore providing minimal actionable insight. For this reason we focused on each part of the day, being morning, afternoon and evening, and investigated if there was significance within these groups by comparing unproductive, neutral and productive days. This information enables individuals to identify unproductive days and adjust their schedule accordingly. For instance, if more time tends to be spent working in the afternoon on unproductive days, they can prioritize tasks during other parts of the day whenever possible.

We used the Python script '`stats_day_part.py`' to achieve this by re-categorizing the productivity ratings unproductive (1, 2), neutral (3) and productive (4, 5). After this, the data was grouped by day, part of the day, productivity category, and total work activity duration summed up. This resulted in one line per unique combination of day and day part, accompanied by the productivity category of that day and the total time spent in that day part during that day. For clarification, an example of two days can be seen here:

	Day	day_part	productivity_category	total_time_s	total_time_m
0	2023-08-16	afternoon	productive	1309	21.816667
1	2023-08-16	morning	productive	3063	51.050000
2	2023-08-17	afternoon	neutral	12740	212.333333
3	2023-08-17	morning	neutral	10559	175.983333

We compared the three groups (unproductive, neutral, and productive) using one-way ANOVA across the different day parts. This method is appropriate as it compares the means of multiple groups to see if at least one group is different from the others. The assumption of independence was satisfied in all instances, although we encountered non-normality and heteroscedasticity in some cases. Consequently, we used non-parametric tests, either a Kruskal-Wallis or Mann-Whitney U test, depending on the size of the groups involved. If the one-way ANOVA or Kruskal-Wallis test showed significance, we also performed a post-hoc test to investigate which specific relations were significant. Because of the varied options of tests and assumptions, a decision tree showing all groups for all participants can be seen in Figure 18.

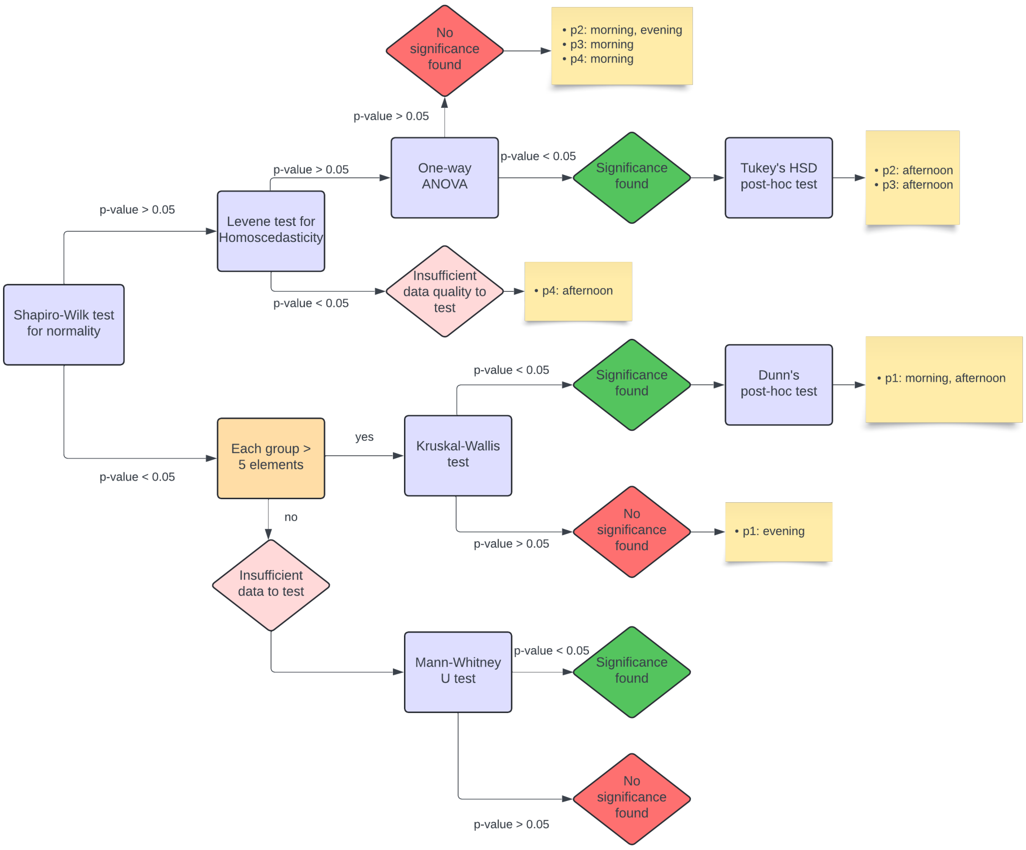
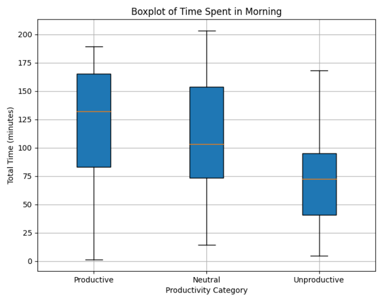


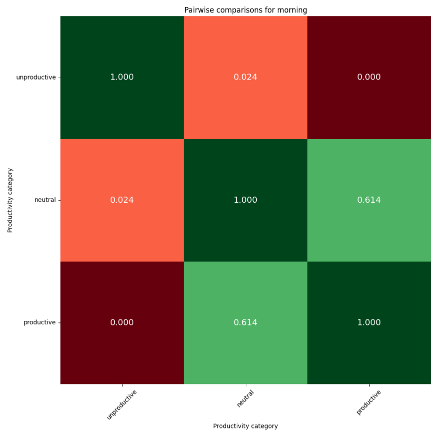
Figure 18: Decision tree for statistical testing of day part elements for all participants.

In the decision tree, we can see that we found significant differences for day parts regarding work time spent comparing productive, neutral, and unproductive days for participant one, two, and three. For this reason, we further analyzed these using post-hoc tests.

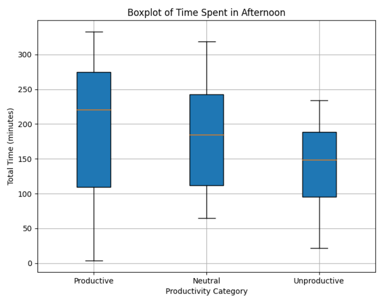
Significant results were found for participant one regarding the morning and afternoon. The box plots and the pairwise comparisons regarding these day parts can be seen in Figure 19.



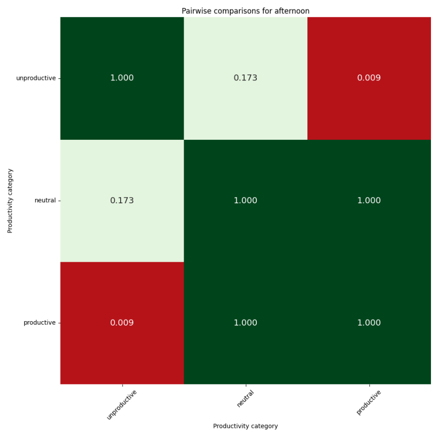
(a) Boxplot morning day part.



(b) Pairwise comparisons of productivity categories regarding morning day part.



(c) Boxplot afternoon day part.

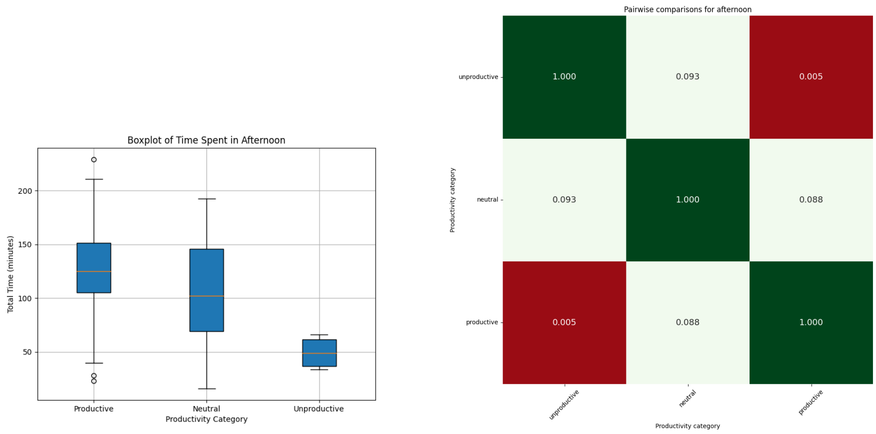


(d) Pairwise comparisons of productivity categories regarding afternoon day part.

Figure 19: Boxplots and pairwise comparisons for participant one regarding time spent in the morning and afternoon for different productivity categories.

For both day parts, we can see a similar pattern. On a productive day, the participant tends to spend more of their time working in the morning and afternoon compared to a neutral or unproductive day. When looking at the pairwise comparisons, we can state that participant one spends significantly more time in the morning on a productive and neutral day compared to an unproductive day ($p<0.05$). Next to this, we can also see that participant one spends significantly more time in the afternoon on a productive day compared to an unproductive day ($p<0.05$).

For participant two, we found significant results regarding the afternoon. The box plot and the pairwise comparison regarding this day part can be seen in Figure 20.



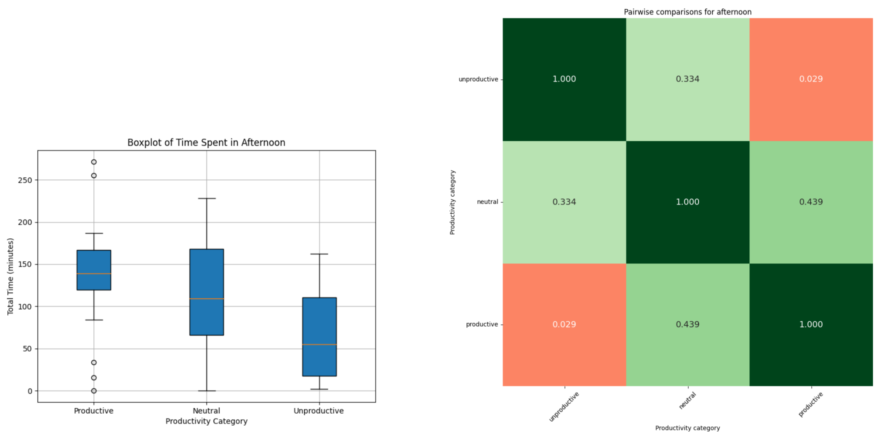
(a) Boxplot afternoon day part.

(b) Pairwise comparisons of productivity categories regarding afternoon day part.

Figure 20: Boxplots and pairwise comparisons for participant two regarding time spent in the afternoon for different productivity categories.

For the afternoon day part, we see that on a productive day, the participant tends to spend more of their working time in the afternoon, compared to a neutral or unproductive day. When inspecting the pairwise comparisons, we can confidently say that participant two spends significantly more time in the afternoon on a productive day, compared to an unproductive day ($p<0.05$).

Finally, we found significant results for participant three regarding the afternoon. The box plot and the pairwise comparison regarding this day part can be seen in Figure 21.



(a) Boxplot afternoon day part.

(b) Pairwise comparisons of productivity categories regarding afternoon day part.

Figure 21: Boxplots and pairwise comparisons for participant three regarding time spent in the afternoon for different productivity categories.

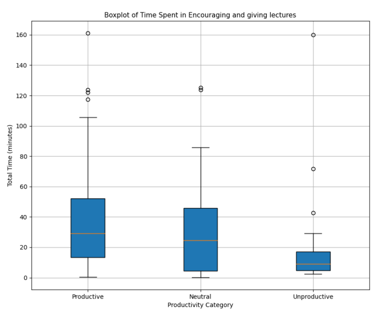
For the afternoon day part, the participant tends to spend more of their working time in the afternoon on a productive day, compared to a neutral or unproductive day. After inspecting the pairwise comparisons, we can state that participant three spends significantly more time in the afternoon on a productive day, compared to an unproductive day ($p<0.05$).

Work activities

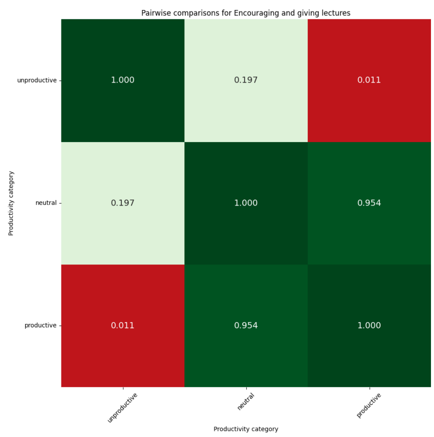
The high-level variable we investigated was 'Work activity,' spanning multiple log lines to provide a broader view of tasks performed over extended periods. Table 3 shows that each participant engaged in up to 55 distinct work activities, making a detailed model like we constructed for the 'Day of week' feature too complex. Therefore, we focused on the top five most frequent work activities per participant, determined by total time spent, using the Python script '`stats_wa.py`'. In Appendix K, the distribution of work activities and the time-weighted average productivity are shown.

As we did for our previous variables, we re-categorized the productivity ratings into unproductive, neutral, and productive. then investigated if significant differences in time spent on work activities existed between these categories for each participant. Using the decision tree logic from Figure 18, we found that none of the work activities met the normality assumption, so we used Kruskal-Wallis and Mann-Whitney U tests.

For participant one, the top five most performed work activities took up 50.6% of the total time spent on all work activities. However, these five activities only made up 18.5% of the different types of work activities participant one did, showcasing that work time was not equally spread over all work activities. Only 'Encouraging and giving lectures' showed a significant difference, being performed more on productive days than unproductive days, as indicated by a Kruskal-Wallis test ($p < 0.05$) and confirmed by post-hoc tests. Boxplots and pairwise comparisons with p-values are shown in Figure 22.



(a) Encouraging and giving lectures.

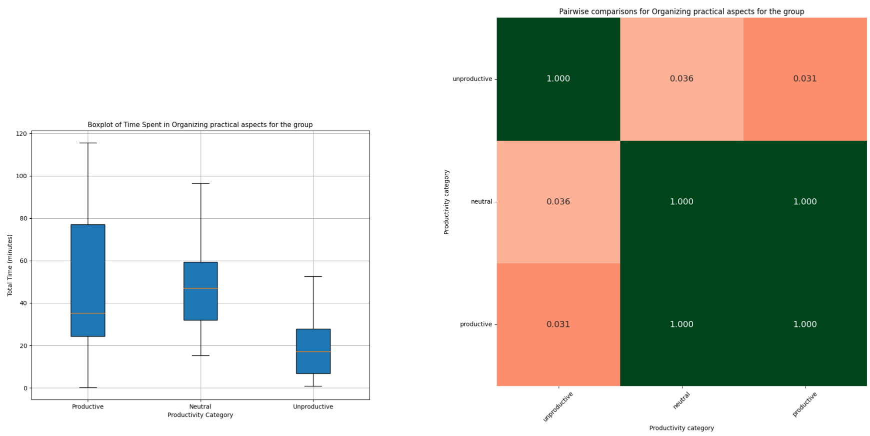


(b) Pairwise comparisons of productivity categories for 'Encouraging and giving lectures'.

Figure 22: Boxplots and pairwise comparisons of productivity categories for 'Encouraging and giving lectures'.

For participant two, the top five most performed work activities took up 78.5% of the total time spent on all work activities. However, these five activities only made up 16.7% of the different types of work activities participant two did. For all five work activities, we only found four instances in the unproductive category. Therefore, we could only statistically compare the productive and neutral category for these work activities, as a minimum of five groups is needed for statistically testing. Whilst doing this, we found no significant differences between productive and neutral days for all work activities.

For participant three, the top five most performed work activities took up 60% of the total time spent on all work activities. However, these five activities only made up 11.36% of the different types of work activities participant three did. Only 'Organizing practical aspects for the group' showed a significant difference, being performed more on productive and neutral days than unproductive days, as indicated by a Kruskal-Wallis test ($p < 0.05$) and confirmed by post-hoc tests. Boxplots and pairwise comparisons with p-values are shown in Figure 23.



(a) Organizing practical aspects for the group.

(b) Pairwise comparisons of productivity categories for 'Organizing practical aspects for the group'.

Figure 23: Boxplots and pairwise comparisons of productivity categories for 'Organizing practical aspects for the group'.

Finally, participant four spent 74.6% of his total time working on the top five work activities, which accounted for only 11.9% of all the unique work activities performed. Although we can see these interesting distributions for participant four, there were only three unproductive groups, which is too little for significance tests. For all work activities, we performed Mann-Whitney U tests to compare the time spent in productive and neutral categories. Only 'Assessing the students' assignments and submitting the assessment to the Board of Examiners' showed a significant difference, being performed more on productive days than unproductive days, as indicated by a Kruskal-Wallis test ($p < 0.05$) and confirmed by post-hoc tests. Boxplots and pairwise comparisons with p-values are shown in Figure 24.

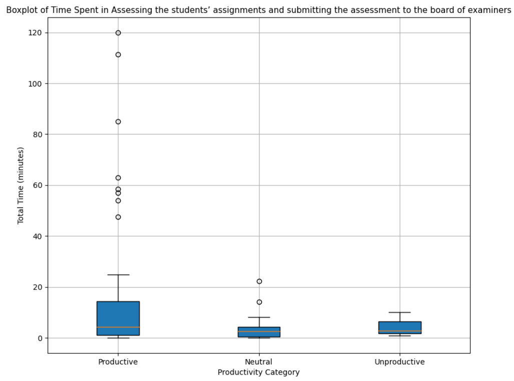


Figure 24: Assessing the students' assignments and submitting the assessment to the Board of Examiners.

4.3 SEQUENTIAL PATTERN MINING – PRODUCTIVITY

In this section, we will present our results from statistically analyzing the data resulting from mining sequential patterns of work activities. First of all, because of the variety of options for analyzing these patterns, we will explain the types of patterns we decided to statistically compare. After this, through statistical tests, we will investigate their impact on productivity.

As we can see in Table 7, there was a wide variety of sequential work activity patterns for each participant. We aimed to help users identify and improve on commonly used patterns by analyzing the top three most frequent patterns that include the initial work activity of the most common pattern for each participant. By looking at top three most common patterns containing a similar work activity, we could investigate the effect of the pattern on productivity. The data distributions were non-normal, and since patterns could occur on the same day, the groups were not independent. For these reasons, a non-parametric test, otherwise known as the Friedman test, was used. The patterns we compared for each participant differed in count, as some patterns occurred more than others. The Friedman test needs groups of the same length as input, and therefore we performed a series of Friedman tests whilst random sampling a set of productivity scores for which the length was equal to that of the lowest count value of the three patterns. We performed this a thousand times, where after the mean p-value of these thousand Friedman tests was used to determine if there were significant differences in productivity scores for the most common patterns. The Python code used for this is called '`stats_patterns.py`'.

For participant one, the three most common patterns containing the work activity 'Communicating about events' and their productivity score distributions are shown in Figure 25. The distributions were skewed left, with the highest count of productivity scores being four. Differences in productivity scores were visible when comparing patterns, with Pattern 1 having an average productivity score 6.7% higher than Pattern 3. As for the Friedman test, an average p-value of 8.794e-05 was found when comparing the productivity scores of the three patterns whilst random sampling a thousand times. Because the p-value is below 0.05, we can state that the productivity scores between patterns is significantly different.

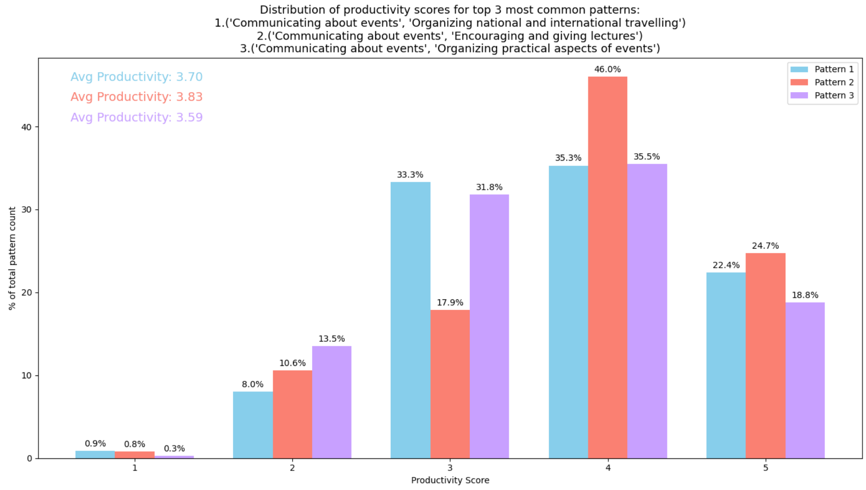


Figure 25: Distribution of productivity scores for top three most common patterns including work activity 'Communicating about events' for participant one.

For participant two, the distribution of productivity scores among the three most common patterns containing the work activity 'Discussing the progress of research with students' can be seen in Figure 26. Again, the distribution was skewed to the left, with the highest count of productivity scores being four. The distributions between patterns did seem quite similar, which was also expressed in the average productivity scores, with the biggest difference in average productivity, which is between Pattern 1 and 2, only being 1.9%. We found an average p-value of 0.51 when performing the Friedman test on the productivity scores related to these patterns. This p-value is well above 0.05 and therefore we can state that the productivity scores between these patterns are not significantly different.

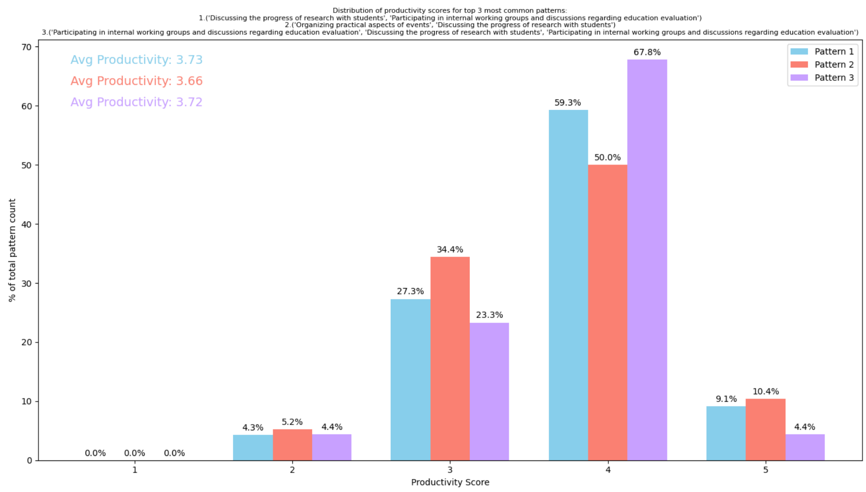


Figure 26: Distribution of productivity scores for top three most common patterns including work activity 'Discussing the progress of research with students' for participant two.

For participant three, the distribution of productivity scores among the three most common patterns containing the work activity 'Participating in group meetings' can be observed in Figure 27. The distribution skewed towards higher productivity scores, with no scores in the lowest category and minimal representation in the highest category. The most noticeable differences between patterns were in productivity categories three and four. Despite these variations, the average productivity scores were quite similar, with the largest difference being 3% between Pattern 2 and Pattern 3. The resulting average p-value from performing the Friedman tests on the productivity scores related to these patterns was 0.19. With this value being higher than 0.05, we can state that the productivity scores between these patterns are not significantly different.

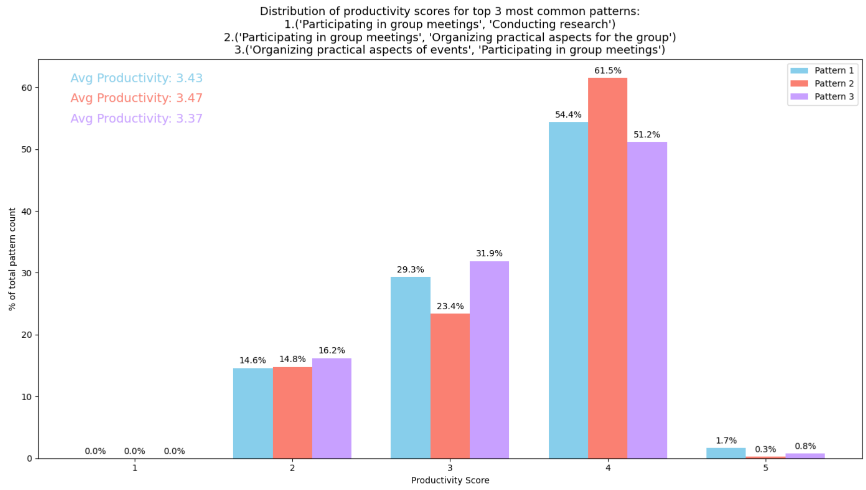


Figure 27: Distribution of productivity scores for top 3 most common patterns including work activity 'Participating in group meetings' for participant three.

Finally, the top three patterns of participant four containing the work activity 'Organizing practical aspects for the group' can be seen in Figure 28. The distributions of all patterns were skewed to the left again, although we did find more instances in the productivity category five. For these patterns, we can see differences in distribution among all productivity categories between them, except the productivity category one. The average productivity scores only show a substantial difference between Pattern 3 and the other patterns, with the biggest difference being 5.5% between Pattern 2 and 3. The Friedman tests, in combination with random sampling, resulted in an average p-value of 0.006, which is lower than 0.05. Therefore, we can state that the difference in productivity scores for these patterns are significant.

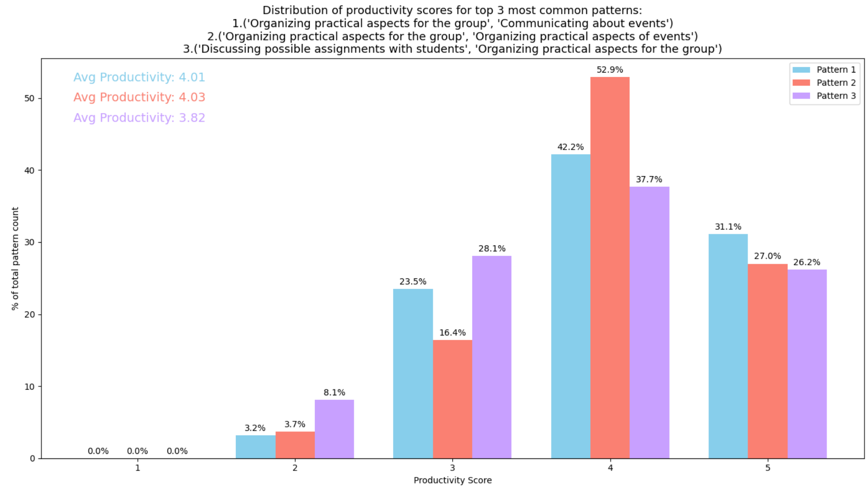


Figure 28: Distribution of productivity scores for top 3 most common patterns including work activity 'Organizing practical aspects for the group' for participant four.

In this section, we took an in-depth look at comparing similar sequential patterns to investigate possible significant differences between productivity scores among them. We do realise that these are quite specific cases among a vast amount of sequential patterns. Therefore, in the next section 4.4, we also visualized different aspects of these mined sequential patterns of work activities, whilst relating them to other factors, like pattern duration for instance.

4.4 WORKFLOW SUPPORT SYSTEM DESIGN

In this final section of our results, we will present interesting results of the different visualizations showed by the WorkflowAId tool. After this we will discuss the results of the MEM questionnaire regarding the tool. This questionnaire provides feedback on the tool and ultimately gives us insights into the attitude towards the intention to use the tool.

General metrics

Average productivity scores varied among all participants, ranging from 3.45 for participant three to a highest productivity score of 3.99 for participant four. The most common productivity score across all participants was four, although distributions varied; for instance, participant four had only three days with a score of two, while participant one had 22.

Regarding different days of the week we see different trends, with some participants experiencing higher productivity early in the week and at the end of the week. Participant one, for example, rated Saturday with an average score of 4.52 compared to 3.16 on Tuesdays, indicating a 43% higher productivity. In general, it is interesting to see that three out of four participants rated their weekends as more productive than weekdays.

We can also see that participants had different activities which they experienced as being productive and performed most of their time in. Regarding different parts of the day, we can see different characteristics for all participants. For instance, participant three spent more time working in the morning on unproductive days compared to productive days, whilst participant two spent more time in the morning on productive days compared to unproductive days. With regards to applications switches, we also see different characteristics per person. For instance, participant two used multiple applications the most on unproductive days (1) and one application the most on productive days (5). This was flipped for participant three, who used one application the most on unproductive days (1) and multiple applications the most on productive days (5). Although we do noticed here that, in general, differences were small.

Best versus worst weekday

For each participant, we identified different work activities with big differences in time spent. For example, participant two dedicated more time to 'Drafting conference papers' on unproductive days, whereas participant three allocated more time to this task on productive days.

Regarding division of work throughout the day, similarities were generally observed across participants. For application usage, most participants seemed to use one application and multiple applications a similar amount of time on both productive and unproductive days. However, participant four spent 2.3 times more time on a single application during unproductive periods compared to 1.5 times during productive periods.

Regarding work duration, participant one and four seem to have spent a similar amount of time working on their most productive day compared to the least productive day. However, participant two and three, on average, spent less time working on their most productive days compared to the least productive day.

Work activity

Regarding different days of the week, participant three and four show that the day of the week that the work activity is performed the most on does not necessarily yield the highest average productivity, as shown in the second plot. Across all participants, we can see some clear differences in time spent working on the work activity during different parts of the day when comparing productive and unproductive days. For instance, participant four spent approximately 23% more morning time on 'Organizing practical aspects for the group' on unproductive days compared to productive days. Similarly, there are differences in productivity when using one application versus multiple applications during work activities. For example, participant three spent about 38% more time 'Organizing practical aspects of events' whilst using multiple applications on productive days (5) compared to unproductive days (1).

Furthermore, we can see big differences in productivity scores between the most and least productive sequential patterns, with some patterns being experienced as 2.5 times more productive than others. We can also see how small differences in the composition and order of the pattern resulted in big differences in average productivity. For instance, participant one experienced the sequence 'Communicating about events', 'Conducting research', and 'Encouraging and giving lectures' as productive, averaging 4.83 over six occurrences. In contrast, when 'Encouraging and giving lectures' and 'Conducting research' are followed by 'Periodically discussing research results with fellow researchers and supervisor or co-supervisor' the participant experienced a low average productivity of 2.25 over eight occurrences.

Sequential pattern metrics

When looking at different parts of the day and its relation to sequential work activity patterns we can see some interesting insights. For instance, participant one achieved an average productivity of 5 when executing the pattern in the afternoon, but only did so six times. However, performing the same pattern in the evening, which occurred 14 times, resulted in an average productivity of 3.36.

Differences in experienced productivity for different duration's of the pattern also show interesting insights. For instance, participant two executed a pattern six times for a medium duration, achieving an average productivity of 3.5. However, when extending the duration to long, with only three occurrences, the participant experienced an average productivity of 5.

Regarding application switches, participant four repeated a pattern seven times using only one application, achieving an average productivity of 3.57. However, using multiple applications for the same pattern, with five occurrences, led to an average productivity of 4.8.

When looking at the most and least productive patterns, we found patterns consisting of different work activities and lengths for all participants. For instance, four of the top five productive sequential patterns consisted of two work activities for participant four, whereas the least productive ones involved at least three activities. Additionally, the order and composition also seems to influence average productivity. For instance, the top three sequential patterns of participant two all contained the work activity 'Organizing practical aspects of events', whilst all the least productive sequential patterns also contained this work activity.

Sequential pattern metrics - work activity insights

When focusing on sequential patterns consisting of specific work activities, we also find interesting results. An example of an interesting result here is that of participant one. In the analysis of the most and least productive sequential patterns, 'Organizing practical aspects of events' combined with 'Conducting research' and 'Drafting a research plan' appeared in four out of five of the least productive patterns. Interestingly, 'Drafting a research plan' did not feature in the top five most productive patterns, indicating the possible impact of that specific work activity on the sequential pattern as a whole.

4.4.1 MEM questionnaire results

In this final section of the results, we will discuss the MEM questionnaire data collected from the four participants in our study. As mentioned earlier, the results of this questionnaire will give us a good insight on if users would adopt our workflow support system design. The results of the questionnaire can be found in Appendix Q.

First of all, we asked the participants for their opinion on all the individual pages of the dashboard. We did this by asking if the different visualizations on the dashboard were clear and understandable to them and which visualization, if any, would improve their productivity. If they thought any would improve their productivity, we also asked why they thought it would. After evaluating all the different pages of the dashboard, we asked the participants closed questions regarding perceived usefulness (PU), perceived ease of use (PEoU) and intention to use (ItU). Finally, we asked the participants their final opinion about the WorkflowAId tool. The resulting averages for perceived usefulness, perceived ease of use and intention to use can be seen in Table 11.

	PU	PEoU	ItU
Average score	2.5	3	2.75

Table 11: MEM questionnaire averages for perceived usefulness (PU), perceived ease of use (PEoU) and intention to use (ItU).

First of all we can see an average PU that is inbetween disagree and neutral (2.5/5), meaning that the participants slighly disagreed about the perceived usefulness that WorkflowAId offered. When further investigating the individual questions in this category, it is interesting to see that all participants were negative or neutral about WorkflowAId improving their productivity. Next to this, all participants disagreed on the statement that WorkflowAId would make it easier to perform their work and perform tasks more quickly. The participants did mostly agree on the statements about WorkflowAId helping them understand their workflow, identify personal patterns and having the ability to useful during their work.

For the PEoU we see that the average is a little higher than PU, being neutral (3/5). When further investigating the individual questions in this category, it is interesting to see that the questions regarding the ease of changes that can be made by using the tool to improve productivity were answered negatively. Next to this, on average, questions regarding the rate of learning of using the tool and the ease of getting valuable insights from it were positive. The answers on whether the tool would be easy to use, flexible and understandable were mixed, containing positive, neutral and negative answers.

The final category of questions was regarding the ItU, which had an average score slightly below neutral (2.75/5). From the questions in this category, we can see that three participants did not think WorkflowAId could improve their productivity and one think it could. Additionally, three out of four participants agreed on the statement that they would use the tool to identify personal characteristics. All in all, we can see some interesting results from this questionnaire regarding the users' attitude towards adoptation of the WorkflowAId tool. We will discuss the reasoning of the participants behind this in more detail in our discussion, using the answers for the participants to the open questions.

5 DISCUSSION

The main goal presented in this paper was to improve productivity of university staff members by identifying personal patterns and characteristics that significantly improve productivity and aiding university staff members by optimizing and visualizing their workflow with the use of a workflow support system. We aimed to fulfill this goal by designing a workflow support system that utilizes a combination of a personalized neural network, data analytics, and pattern mining which effectively integrate both low-level and high-level variables.

5.1 INTERPRETATION OF THE RESULTS

5.1.1 Identifying work activities

The first research question constructed to support our main research goal was:

> *RQ1: Can we identify the work activity performed by a participant at any given moment in time by studying the low-level variables and patterns within the log?*

We successfully constructed personalized neural networks which were able to identify the work activity the participant was working on at any given time, using the raw log data as input. These neural networks all outperformed baseline methods, with the lowest performing neural network performing roughly 20 times better than the baseline and the best performing neural network performing 33 times better. The neural networks successfully integrated low-level variables and patterns in the raw log data by utilizing application and title information of the log in overlapping sequences with a Bidirectional LSTM structure. This approach enabled the prediction of work activities by analyzing a combination of low-level variables and patterns within the log. Thus, we can confidently answer our first research question positively.

5.1.2 Variables impacting productivity

The third research question we constructed to support our main research goal was:

> *RQ3: Do specific features, such as time-related features, application switches and the nature of the work activity, significantly impact personal productivity?*

Although this is our third research questions, we investigated it before RQ2, and therefore it will be discussed second.

Time-related features

The first time-related feature and its impact on personal productivity we studied was the day of the week. For certain participants, we found days where productivity was rated significantly higher compared to their day with the lowest average productivity rating. However, when including the total work duration of the day into analysis, these differences in productivity ratings did not reach statistical significance. This suggests that the productivity ratings are not only influenced by the day of the week, but also factors like the amount of time spent working.

When investigating the visualizations in the dashboards, showcasing the relationship between the day of the week and productivity, we observed interesting differences in productivity scores across different days, without considering work duration. For instance, one participant rated Saturday as 43% more productive on average than Tuesday. While these differences lacked statistical significance, they still showcase the potential effect of the day of the week on productivity ratings, which appeared to be personal as all participants identified a different day as their least productive. Another interesting finding was that three out of four participants rated work during their weekend as more productive than weekdays. One possible explanation for this could be that an individual might experience less pressure and stress during the weekend whilst working from home, resulting in better focus and more efficient task completion. But this is out of our scope and would be interesting for future research in this area.

The second time-related feature and its impact on personal productivity we studied was the part of the day. We found significant results for three out four participants, showing that they spent more time working in the afternoon compared to other parts of the day on productive days, compared to neutral or unproductive days. This suggests that these individuals feel more productive when spending more of their time working in the afternoon. But what if an individual spends more time working on productive days compared to neutral or unproductive days in general? This would make the initial statement less valid. Additionally, when comparing the least and most productive weekday, we can see that three out of the four participants, on average, worked less on their most productive day compared to their least productive day. We believe that these findings provide solid evidence to counteract the suggestion that individuals are only more productive in the afternoon because they spend more time working on productive days in general.

Work activity

To investigate the impact of the nature of the work activity on productivity, we investigated the top five work activities performed. Similar to our analysis of time of day, we investigated if significant differences existed in the time spent performing these activities across productive, neutral, and unproductive days. These top five activities accounted for between 50% and 78.5% of total work time for each participant, therefore providing a solid insight into the effect on productivity.

Although noticing differences across all work activities for each participant, we only found one significant results for three out of four participants. For all three participants there was a certain work activity that was performed significantly more on productive days compared to neutral or unproductive days. Interestingly, although seeing some overlap in work activities in the top five across participants, these significant work activities were all unique. This indicates the personal aspect, showing that participants all experience certain work activities in a different personal way. But again, what if an individual spends more time working on productive days compared to neutral or unproductive days in general? The visualizations provide clarity here, because when inspecting the percentage of time spent on these work activities in relation to the total time spent, we also see that all of these work activities are performed relatively more on the individual's most productive weekday compared to their least productive weekday.

The personal aspect of productivity is further supported by our visualization showing the different work activities in relation to their time spent and productivity scores, showing the difference among participants in work activities which they experience as being productive and unproductive.

Variable interactions

We decided not to significantly test the impact of application switches and work activity duration levels on productivity, as we saw no use for individuals to improve their productivity with this knowledge alone. Our visualizations express this by showing minimal differences between time spent using one application versus multiple applications in general. Additionally, our data showed that individual instances of work activities were very short in duration, with some lasting only seconds. Knowledge on if an individual is more productive performing a work activity for a couple of seconds versus a minute is not very useful in our opinion, and therefore we focused more on this variable when looking at chains of work activities in sequential order.

However, visualizing the interaction between application switches with other features and their combined effect on productivity did show some interesting results. First of all, when comparing the least and most productive weekdays, we found that one individual used one application relatively more on the least productive day compared to the most productive day. This suggest that, in general, application switches might not have an effect on productivity, but when focusing on a specific day of the week it does affect productivity. However, since this combined effect was observed for only one participant, it remains speculative. Furthermore, when analyzing specific work activities across participants, we identified interesting combined effects. For example, one participant spent 30% more relative time using multiple applications when performing a particular work activity on productive days compared to unproductive days. These larger differences in application switches levels were noticed when focusing on a certain work activity for all participants, indicating that in general application switches might not have an effect on productivity, but when focusing on a specific work activity it does affect productivity.

Whilst focusing on a specific work activity for all participants, we also found interesting effects of interactions with time-related features on productivity. First of all, for two participants, the day of the week they felt most productive performing the work activity was not the day they actually performed it the most on. This indicates working less on a work activity does not mean it will result in a lower feeling of productivity. Lastly, we noticed some clear differences among all participants in time spent performing a work activity on specific parts of the day on productive versus unproductive days. This shows that an individual might be better of performing certain work activities in the morning, and others in the afternoon.

All in all, we believe that we have collected enough evidence to answer our third research question in the following manner. We have found cases where the part of the day and the nature of the work activity significantly impact personal productivity, which is also supported by the visualizations in our dashboards. Although we found significant effects for day of the week on productivity, when accounting for time, the significance of the day of the week on productivity diminished. Whilst investigating the relationship between day of the week and productivity in our dashboards, we did find some interesting effects, which would be interesting for future research. As for 'personal productivity', we believe that we showcased the variety of relationships between features and productivity based on the individual through statistical analysis. These individual differences were further highlighted by our dashboards, indicating the importance of a personalized approach.

5.1.3 Patterns impacting productivity

The second research question constructed to support our main research goal was:

> *RQ2: Can we identify workflow patterns by examining different work activities performed throughout the day, and do these patterns significantly impact personal productivity?*

For this research question we will discuss the statistical findings of the effect of sequential patterns of work activities on productivity. Next to this we will discuss interesting patterns showcased by our dashboard.

Sequential work activity patterns

Earlier, we showed the vast number of patterns found for each participant through the mining of their sequential work activity patterns of lengths two through four. Because of this vast number, we attempted to pinpoint similar patterns that were common and investigated if there were significant differences in average productivity. With the patterns consisting of similar work activities, we could study the effect of the sequential order of the pattern on productivity.

When looking at the distributions of different patterns with similar work activities, across days with different productivity scores, we found differences for all participants. For two participants, these differences were significant. For one participant we found that when the individual was 'Communicating about events' and followed this directly up by 'Encouraging and giving lectures' the average productivity was 3.89. However, if the individual followed this work activity directly up by 'Organizing practical aspects of events' the average productivity was 3.59. This demonstrates that the sequence the individual decides to perform their work activities can impact productivity significantly.

Although we did not find significance for all cases, our dashboard did show interesting differences in productivity scores regarding sequential patterns containing a similar work activity. For all participants, we can see sequential patterns consisting of the same work activity but having very different productivity scores. With some sequential patterns showing differences of 150%. We do have to note that, whilst including only patterns with a minimum occurrence count, some of these sequential patterns were not common. This weakens the strength of the evidence, as these patterns might only have occurred on a couple of days, making the insights less robust and generalizable. However, it is still interesting to see these big differences in average productivity scores of sequential patterns that are similar in work activities, but differ in sequence.

Although not contributing to our research question, we still want to discuss the relationship between the sequential patterns and other features, which were showed in the dashboards. These features were part of the day, pattern duration and application switches. For these features, we attempted to pinpoint sequential patterns that were performed mostly in the feature category in which it was performed the most, but did not result in the highest productivity. Again, the occurrences of these specific sequential patterns were small, making them less robust and generalizable. However, we still showed interesting opportunities of behaviour change, which may impact productivity positively, like performing a specific sequential pattern in the morning, rather than in the afternoon.

Other patterns

Finally, although we focused on sequential patterns of work activities in our study, we still want to discuss other types of work related patterns impacting productivity showed by our dashboard. When looking at the productivity scores and time spent throughout the week, we saw different types of patterns for each individual, with some experiencing a higher productivity at the beginning of the week whilst working less and some at the end of the week whilst working less. The interesting fact here is that there are substantial differences in experienced productivity between the different days of the week. In our study, we focused on measuring the effect of shorter sequential patterns on productivity, but for future research it would be interesting to also study these longer, weekly patterns and their effect on productivity.

To answer our final research question, we managed to identify sequential patterns consisting of similar work activities, that significantly differed in productivity. Next to this, although not significant, we found that each individual had unique sequential patterns of work activities, which they experienced as productive and unproductive. Furthermore, the differences in experienced productivity between some of these patterns were substantial for all individuals, even though they consisted of similar work activities. These findings indicate the effect of the sequence of work activities on productivity. Additionally, these sequential patterns of work activities represent a specific kind of workflow pattern, and future research should explore other types of workflow patterns and their effect on productivity.

5.1.4 System design feedback

Finally, we wanted to discuss the results of the MEM questionnaire. These results gave us insights into the users' opinion on possible adoption of our WorkflowAId tool design. Overall, we saw scores that were averaging around neutral per category, and this was reflected in the spread of answers to the questions for perceived usefulness, perceived ease of use and intention to use regarding the WorkflowAId tool.

When looking at the opinions of the users on the visualizations, there were some visualizations that were seen as having potential to improve the users' productivity. For instance, two participants thought that seeing the average productivity score per day of the week could improve their productivity. With one participant stating that it could help "Reorganzing my work on the unproductive day(s)". These participants also voiced this opinion when reviewing the visualization which showed the productivity for different days of the week regarding a specific work activity, where one participant found it useful to "See what is the best day to perform that activity". Another visualization which could improve productivity was the overview of the different work activities and their corresponding productivity score, for which one participant stated: "I am interested in the overview of all my activities and to find out which ones are associated with lower scores.".

There were also visualizations regarding sequential patterns of work activities that were found to have the ability to improve productivity. One participant voiced that, whilst at first having some problems understanding the visualization, seeing for which part of the day a pattern was the most productive could improve productivity. Another participant thought the visualization showing the top five most productive patterns could improve productivity, as the individual stated: "I can reorganize my work according to the productive sequential patterns.".

Although these results show positive opinions on visualizations and their ability to help improve productivity, the overall consensus was that the WorkflowAId would not aid the participants in being more productive. We believe that this is due to the fact that participants had a hard time understanding some visualizations and did not really know how they could use the information to improve their productivity. When asking the participants about ability of the tool to help understand their workflow and personal patterns they were more positive. Overall, the participants found the WorkflowAID tool useful to identify characteristics about themselves, but lacked the ability to actually improve their productivity overall. This can also be seen in the distribution of the scores of the individual questions regarding perceived usefulness, perceived ease of use and intention to use. The questions about the ability of the tool to improve productivity were answered more negatively, whilst the questions about the usability and the ability of the tool to help individuals understand their personal characteristics were answered more positively. This overall consensus can also be seen in the final two questions regarding the intention to use, which is solely positive when asking the participants if they would use the WorkflowAId tool to identify personal characteristics, but negative two out of three times when asking if WorkflowAId would be used to improve productivity during work.

Now that we have discussed all the results in our study, we can state if we managed to reach our main research goal, being:

> **RG:** *To design a workflow support system that utilizes a combination of a personalized neural network, data analytics, and pattern mining which effectively integrate:*
>
> - *low-level variables, such as application usage.*
> - *high-level work activities.*
>
> *With the goal of improving productivity of university staff members by:*
>
> - *identifying personal patterns and characteristics that significantly improve productivity.*
> - *aiding university staff members by optimizing and visualizing their workflow with the use of a workflow support system.*

We believe that we partly reached our goal by achieving the actual design of a personalized tool that uses a combination of a personalized neural network, data analysis and pattern mining to show the user visualizations about their workflow in relation to their productivity. Although not always significant, we managed to show the impact that different low-level and high-level variables can have on productivity. Our study also showed the variety of personal characteristics and patterns that can impact productivity, varying per individual. We do believe that we could have provided more aid to the user by creating more simple visualizations, guided by better advice on how to actually improve on productivity. Therefore, future research should focus more on how to convey the information to the user in a way that it can be used to improve productivity.

5.2 IMPLICATIONS

This research presents a novel approach at designing a personalized workflow support system that helps individuals improve their productivity whilst working by visualizing and giving them advice about their workflow. The research method used builds upon state-of-the-art research regarding personal informatics by using a combination of machine learning, pattern mining and data analytics to analyze log data about work activities to gain unique personal insights into their work activity patterns and overall behaviour in relation to their self-assessed productivity. The results show the variety of factors and patterns that can have an impact on productivity, differing per individual. Although this study successfully identifies personal characteristics that impact productivity, it also shows the challenges of conveying this information in a user-friendly workflow support system design. Therefore, this research can serve as a foundation for developing personalized workflow support tools. However, it emphasizes the need for more research on the design of such tools to convey the information effectively, ensuring that the user can improve their productivity using the insights shown by the tool.

Secondly, a lack of research has been done on measuring the effect of a combination of personal factors and patterns on a self-assessed productivity score which is measured over a solid period of time [9, 11, 12, 14, 16, 17, 25]. This study contributes to this by studying the personal effects of a broad set of work activities, time-related features, application behavior and patterns on a self-assessed productivity score recorded over a period of nearly six months. In today's world, surveillance software used by companies to keep tabs on their employees can cause power imbalance between the individual and management [10]. This personalized approach could serve as a foundation of solutions for this by relying on a self-assessed productivity score which is made for personal use, giving control back to the individual by serving as a helping hand, focusing on improving the feeling of being productive.

The existing literature also shows a lack of research for studying the effect of patterns on a self-assessed productivity score. This research contributes to this by studying the effect of sequential patterns of work activities on this self-assessed productivity score. The findings of these effects could help individuals optimize their workflow by giving them insights into how to order their work activities in sequences which result in the patterns that are associated with a feeling of being productive.

The results of studying the personal characteristics and patterns are being showed to the individual with the use of a dashboard containing personalized visualizations and advice. Most research is focused on studying the effects of certain factors on productivity or visualizing work activity without any actionable advice [8]. This research contributes to this by not only analyzing the effects of personal characteristics and work patterns, but also by developing visualizations with accompanying personal advice in an attempt to aid the user in being more productive. The established literature shows a call for this type of research that not only studies factors that impact productivity, but also tries to aid the user by providing visualizations of the findings with textual aid to the user to help them in being more productive in a personalized manner [2, 8, 9, 14, 24]. The results in our study showed that these visualizations can aid the users in improving their productivity. On the other hand, this study also revealed the difficulties in designing such a tool from a set of visualizations, as it is challenging to create a tool that users will intend to use. This problem is also voiced in the existing literature, which showed that these kind of systems only get used for a minute per day on average [13].

Finally, by providing full Python code regarding the whole process, from anonymization of the raw Tockler data to constructing the visualizations displayed on the dashboard, this research is reproducible and therefore can serve as a foundation for further research [13].

5.3 LIMITATIONS

First of all, our study was performed on the data of four university staff employees. Although this still resulted in large datasets spanning over a time period of almost six months, this is quite a specific setting, and therefore might limit the generalizability of our results. We tried to mitigate this by also including feedback of the employees themselves into the study.

Additionally, compared to the whole dataset, we only had labeled datasets of two employees at our disposal. Therefore the data of the other two employees was labeled based on a neural network which was trained on a set of work activities used by the other employees, which might not capture all the work activities they performed throughout the whole time period. In general, the data used for training the neural networks and creating our dashboard design probably did not capture all the work activities performed, as it only consisted of data captured by their computer. This was also voiced by one of our participants, who stated the following about the dashboard design: "Interesting to see, but I do not see how it will improve productivity. My most productive days are probably whilst I'm giving lectures or had a convention. These days I'm not on my laptop that much but I was indeed very productive". As we only analyzed data captured by the users' computer, this is out of our scope, but it would be interesting for future research to capture and analyze more data than just work performed on the computer.

We realized that the anonymization process did not filter out all the sensitive data regarding person names whilst using the Named Entity Recognition technique. Because of the vast amount of data, we did not realize this. Future research should combine several techniques to make sure all sensitive data is filtered out.

Due to time constraints, we decided to only test and compare a set of Bidirectional Long Short-Term Memory networks and their results, which is a specific type of neural network. There might have been different types of neural networks that could have performed better. Although, we do think, looking at our time series data, that we had valid reasons to choose this type of neural network.

The self-assessed productivity score, resulting from the survey data used in our study, contained one productivity score for each day. We related this score to all the data collected on that day. There might have been certain parts during the day which weighed heavier to that score than other parts. To address this potential bias, we used a time-weighted average in our statistical analysis and dashboards.

With regards to patterns, we only investigated a set of sequential patterns of work activities. This is a specific type of pattern, and although we saw interesting signs of other types of patterns in our dashboards, we did not further investigate these. In some cases, the amount of times sequential patterns with very low and high productivity occurred was minimal, which might limit the robustness of the findings for these specific patterns. Additionally, some participants in our study did not see any value in the sequential patterns and could not use them to reorganize their work, as one participant stated: "I cannot really see these patterns as something that I could easily recognize in my work or that I could organize differently. The concept of work pattern that is used here seems detached from something that I can apply myself in organizing my work.". Future research should also explore the impact of non-sequential patterns on productivity.

Finally, our workflow support system dashboards were only designs. However, the Python code for the visualizations and the process to acquire the necessary data is already established. This means that creating an actual interactive dashboard would not require significant additional work, compared to the research that already has been done.

5.4 VALIDITY THREATS

While we showed that our study provided valuable insights, it is important to critically acknowledge certain threats we faced during this study that may affect the validity of our findings. In this section we will discuss the different threats of research design, as discussed by Chong-ho Yu and Barbara Ohlund [25]. These threats to validity are internal validity, external validity and construct validity.

5.4.1 Internal validity

Internal validity concerns whether the outcomes in our study were actually due to the treatment we used [25]. First of all, we solely focused on Bidirectional Long Short-Term Memory (Bi-LSTM) networks, whilst other types of neural networks might have performed better at classifying work activities. To mitigate this, we conducted an extensive parameter search to find the best performing models for our datasets. Additionally, we found solid reasoning to use this type of neural network to capture the temporal dependencies in our time series data.

Secondly, whilst significant effects of features and patterns were found on productivity, these tests all focused on specific relations. When we tested the impact of the different days of the week on productivity, we found that the significance diminished when considering the working time. This suggests that productivity might be influenced by a possibly complex combination of factors. To mitigate the risk of finding this spurious significance, we focused on investigating common relationships and patterns with many instances, with the aim of increasing the reliability of our results. Next to this, we provided personal dashboards that showed further details on these relationships and patterns. The visualizations in these dashboards showcased the individual differences in productivity scores present in many cases, supporting the significant findings in our study.

5.4.2 External validity

External validity relates to the generalizability of the findings in our study to other work areas or populations [25]. As our study was based on a specific type of work, namely that of university staff members on their computers, the results might not be applicable to other work areas or populations. Although this study is quite specific, the dashboards provided individual insights regarding the work behavior of the individual's themselves. This type of self-assessment helps to investigate if insights were meaningful and applicable within the context of similar type of work environments, battling this external validity to some extent.

5.4.3 Construct validity

Construct validity concerns whether the measures used in our study actually represent the concepts they are supposed to represent [25]. First of all, our productivity scores were measured at the end of each day for all participants. This type of measurement might cause some bias as participants take one moment to asses their whole day, which they might not remember in detail. Next to this, their assessment might be influenced by their mood at the specific time of assessment. The productivity score also does not capture the fluctuations in productivity throughout the day. These concerns were mitigated by having a clear survey, containing easily understandable Likert scale scores. Next to this, we used a time-weighted average in our statistical analysis and dashboards in an attempt to capture possible fluctuations throughout the day.

Secondly, our neural networks were trained on labeled data of two out of four participants regarding a time period of a month. Next to this, a certain set of work activities were performed during this time period which might not represent all the work activities performed throughout the study. We addressed these concerns by enhancing the robustness and generalizibilty of our neural networks in multiple ways. Firstly, we validated all results of our neural networks against a hold-out validation set. Additionally, for the two participants without any labeled data, we used a combination of data splits to ensure a high variety of possible work activities was included into the models.

Finally, we realise that by only using data captured by the participant's computer, activities impacting productivity but not captured by the computer are overlooked, which was also voiced by one of our participants. To mitigate this concern, we examined as much data as possible within the scope of our study. However, future research should focus on incorporating additional data capture methods to enhance construct validity.

5.5 FUTURE RESEARCH

In this research, we showed the potential of utilizing a combination of machine learning, pattern mining and data analytics techniques to find productive work characteristics and patterns, with the goal of aiding the individual in working in a more productive manner, as assessed by themselves. Our research can serve as a solid foundation for future research in this area.

Future research should focus on improving the generalizability in this area by studying the applicability of our approach in different contexts. Whilst exploring other contexts, future research should focus on the following:

- Exploring the performance of a variety of neural networks for identifying personal patterns and work characteristics for the individual.
- Exploring a bigger variety of features and the combined effect of them on productivity.
- Expanding on the number of individuals that are studied, resulting in more extensive and varied data. This will strengthen the analysis by gaining a deeper understanding of similarities and differences both between individuals and in the behavior of individual participants themselves.
- Gaining a more detailed understanding of productivity by measuring the metric in more detail without causing annoyance for the user [25]. This could be one productivity measure per hour instead of per day for instance.
- Exploring different types of patterns next to sequential patterns of work activities to gain a more overall understanding of patterns and their effect on productivity.

- Implementing and measuring the user experience of a workflow support system to assess its usability and effectiveness in improving productivity in a real-world scenario.

By focusing on these areas, future research endeavors can deepen our understanding of the personal characteristics and patterns that impact productivity, with the ultimate goal of developing tools that support the individual in achieving an optimal sense of productivity.

6 CONCLUSION

This research presents the design of a personalized workflow support system tool that effectively integrates low-level and high-level variables by utilizing machine learning, data analysis, and pattern mining techniques to reach the following main goal:

> **RG:** *To design a workflow support system that utilizes a combination of a personalized neural network, data analytics, and pattern mining which effectively integrate:*
>
> - *low-level variables, such as application usage.*
> - *high-level work activities.*
>
> *With the goal of improving productivity of university staff members by:*
>
> - *identifying personal patterns and characteristics that significantly improve productivity.*
> - *aiding university staff members by optimizing and visualizing their workflow with the use of a workflow support system.*

In the design, the different personal characteristics in relation to the self-assessed productivity are visualized, in some cases accompanied by textual aid. The results showed that these types of tools are able to provide insights to the user which can be seen as useful to gain valuable insights about personal behavior and behavioral changes that can help the user to improve their productivity.

This research contributes to a deeper understanding into not only personal characteristics and patterns that impact productivity, but it also shows the process of implementing such knowledge into a support tool design, for which feedback is given by the subjects themselves. It includes the actual design of the dashboard in this study which provides valuable information on whether the aspects being analyzed are meaningful and useful to the end user. By attempting to implement our findings into a dashboard design, we were able to gather feedback and gain insights into the actual usability of the relationships we identified for the end user, eventually showcasing the intent of using such a tool.

This strength in our research also showcases how difficult it is to voice these findings to the user in a user-friendly manner. Therefore, future research should focus on studying a bigger variety of personal characteristics and patterns, while also incorporating user feedback to improve the applicability and presentation of these insights in a user-friendly way.

REFERENCES

[1] Artificial neural networks and its applications. https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/, 2023.

[2] Utrecht university library - searching for literature. https://www.uu.nl/en/university-library/searching-for-literature, 2023.

[3] David H Autor, Lawrence F Katz, and Alan B Krueger. Computing inequality: have computers changed the labor market? *The Quarterly journal of economics*, 113(4):1169–1213, 1998.

[4] Alberto Bayo-Moriones, Amaya Erro-Garcés, and Fernando Lera-López. Computer use and pay for performance. *Human Resource Management Journal*, 32(2):341–363, 2022.

[5] Iris Beerepoot, Daniël Barenholz, Stijn Beekhuis, Jens Gulden, Suhwan Lee, Xixi Lu, Sietse Overbeek, Inge van de Weerd, Jan Martijn van der Werf, and Hajo A Reijers. A window of opportunity: Active window tracking for mining work practices. In *2023 5th International Conference on Process Mining (ICPM)*, pages 57–64. IEEE, 2023.

[6] Duncan P Brumby, Christian P Janssen, and Gloria Mark. How do interruptions affect productivity? In *Rethinking productivity in software engineering*, pages 85–107. Springer, 2019.

[7] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[8] Fred D Davis et al. Technology acceptance model: Tam. *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption*, pages 205–219, 1989.

[9] Annu Haapakangas, David M Hallman, Svend Erik Mathiassen, and Helena Jahncke. Self-rated productivity and employee well-being in activity-based offices: The role of environmental perceptions and workspace use. *Building and Environment*, 145:115–124, 2018.

[10] Will Douglas Heaven. This startup is using ai to give workers a "productivity score", December 10, 2020. October 21, 2023.

[11] Matthew Honnibal and Ines Montani. spacy. https://spacy.io/, 2015.

[12] Roslina Ibrahim, NS Leng, RCM Yusoff, GN Samy, Suraya Masrom, and ZI Rizman. E-learning acceptance based on technology acceptance model (tam). *Journal of Fundamental and Applied Sciences*, 9(4S):871–889, 2017.

[13] Luuk Janssen. Personalized workflow optimization for university staff empowerment: Introducing workflowaid. https://github.com/Lookszz/master_thesis, 2023.

[14] MayGo. Tockler. https://tockler.io/, 2023.

[15] Wes McKinney. Pandas: a python data analysis library. https://pandas.pydata.org/, 2010.

[16] Timothy Meline. Selecting studies for systemic review: Inclusion and exclusion criteria. *Contemporary issues in communication science and disorders*, 33(Spring):21–27, 2006.

[17] André N Meyer, Laura E Barton, Gail C Murphy, Thomas Zimmermann, and Thomas Fritz. The work life of developers: Activities, switches and perceived productivity. *IEEE Transactions on Software Engineering*, 43(12):1178–1193, 2017.

[18] Daniel L Moody. The method evaluation model: a theoretical model for validating information systems design methods. 2003.

[19] Hamaad Rafique, Alaa Omran Almagrabi, Azra Shamim, Fozia Anwar, and Ali Kashif Bashir. Investigating the acceptance of mobile library applications with an extended technology acceptance model (tam). *Computers & Education*, 145:103732, 2020.

[20] Nicholas Rescher. Methodological pragmatism. *Philosophy of Science*, 45(3), 1978.

[21] Mike Schuster. On supervised learning from sequential data with applications for speech regognition. 1999.

[22] Utrecht University. Regulations 2020 wp flow iii utrecht university. https://www.uu.nl/sites/default/files/regulations_2020_wp_flow_iii_utrecht_university.pdf, 2020.

[23] Yuli Vasiliev. *Natural language processing with Python and spaCy: A practical introduction*. No Starch Press, 2020.

[24] Roel J Wieringa. *Design science methodology for information systems and software engineering*. Springer, 2014.

[25] Chong-ho Yu and Barbara Ohlund. Threats to validity of research design, 2010.

A | APPENDIX A

Please fill in these questions for today:

1. Productive: How productive do you feel you were today? 1 = not at all and 5 = extremely
2. Energy: Today, I felt bursting with energy. 1 = strongly disagree and 7 = strongly agree
3. Enthusiasm: Today, I was enthusiastic about my job. 1 = strongly disagree and 7 = strongly agree
4. Immersion: Today, I was immersed in my work. 1 = strongly disagree and 7 = strongly agree

Provide your answer in one message (comma-separated), e.g.: 5, 2, 3, 2

B | APPENDIX B

Universiteit Utrecht

Consent form for participation in the master thesis research project

Optimizing Productivity: A Personalized Workflow Approach for Empowering Workers

Introduction

This project is constructed by Luuk Janssen (l.j.janssen1@students.uu.nl), supervised by Prof. dr. ir. Hajo Reijers (h.a.reijers@uu.nl) and built on preliminary research constructed by Dr. Iris Beerepoot (i.m.beerepoot@uu.nl) . In this research, data collected by the AWT software enriched with labelling regarding Work Practices is combined with daily surveys that gauge how workers perceive their productivity, among other factors. The goal is to create a personalized workflow model, with the aim of helping workers achieve maximum productivity in a personalized manner. During this research, the AWT application Tockler is being used. As a participant, you are asked to install Tockler from www.tockler.io. The duration of data collection will last from 01-12-2022 until 15-12-2023. Participants will not receive any compensation.

Research goal

As said, this research is focused on analysing personalized workflows to study their impact on productivity. The AWT software Tockler is being used to realise this. This software tracks what applications are being used and when. For example, in the history data of Tockler someone can see that Google Chrome was being used for 45 minutes, whereafter Adobe Acrobat was being used for 30 minutes. The titles of web pages are saved. However, Tockler does not save content on websites, mouse movement or keyboard actions. Next to used programs, Tockler records idle, offline and online state. The idle state activates after one minute. An interactive timeline chart is shown to visualize the tracking. The data is locally saved on your device and is not shared with or by Tockler. Next to the data collected by Tockler, at the end of each day, the participants complete a survey. In this survey, the participants give a Likert scale like score on how they feel their day was regarding Productivity, Energy, Enthusiasm and Immersion.

Data

Data used for this research can be divided into three different categories

1) Labelled data
2) Unlabelled data with Productivity, Energy, Enthusiasm and Immersion scores
3) Unlabelled data without scores

For participants that labelled their data, it is important to know that your data will be used to train a model which then can be applied to the unlabelled data with the goal of labelling all the data with 'WP flow activity' labels. This is because in this research I want to investigate activities, both on a low and higher level.

For participants that filled in the survey at the end of each day, it is important to know that your data will also be used to train a model which also then can be applied to the data without scores. This is because in this research I want to investigate the effect of workflows on the metrics resulting from the survey.

In order to protect your privacy and confidentiality, we employ advanced data anonymization techniques using the Python library called spaCy. This Natural Language Processing (NLP) tool helps us substitute specific identifying information such as names, addresses, and other personal identifiers with generic

C APPENDIX C

Example summary of short-list article by Mark et al. [16]:
This study explores the possibility of being in an attentional state that makes one susceptible to communications typically associated with distraction. They explore the confluence of multitasking and workplace communications from prior to an interaction, when tasks and communications are interleaved and at the end of the day. They observed logs of 32 employees for this study, revolving around F2F, email and Facebook communication. They found that certain attentional states lead people to be more susceptible to particular types of interaction.

Main points:

- Rote work is followed by more Facebook or face-to-face interaction.
- Focused and Aroused states are followed by more email, Arousal is predictive of subsequent email use, Arousal and Attentional state interact to predict a higher level of email use.
- The more time in email and face-to-face interaction, and the more total screen switches, the less productive people feel at the end of the day.
- Facebook use following a 'Rote' state was longer than when in a 'Bored' or 'Focus' state, Email use following a 'Focus' state was longer than when in a 'Rote' or 'Bored' state, F2F interaction following a 'Rote' state were more than when in a 'Bored' or 'Focus' state.
- The more switches one does, the longer time one spends in FB and email.
- The more frequently one switches Internet sites, the more time one spends in all communication types.
- The more projects one has, the more switching between projects one does, and the more opportunities there are for checking FB and email while switching.
- The more time spent on email or the more switches or the more F2F interactions, the less productive one feels.
- **Research Method**: in situ study at a large US corporation. They used a mixed-methods approach with logged people's digital activity along with using experience sampling. Next to this they did surveys to measure for subjective and demographic measures.The data was logged for 5 work days.They also used webcam data to determine F2F interactions. A pop-up window was used to capture the participants' perspective in situ.
- **Similarities**: This study uses software to track similar log data metrics, like time spent on an application to study its impact on productivity. They also investigate patterns (switches).

- **Differences**: In this study they specifically focus on Facebook, Email and F2F interactions, while we focus on a broad set of work activities. Next to this, they also investigate different 'States' of people, whilst we only focus on the subjective productivity of a person. The biggest difference in this study is that they constructed a model based on all 32 participants, whilst we create personalized neural networks.

D APPENDIX D: LONG-LIST

INCLUDED

[1] Inge Alberts. Challenges of information system use by knowledge workers: The email productivity paradox. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–10, 2013.

[2] Bon Adriel Aseniero, Charles Perin, Wesley Willett, Anthony Tang, and Sheelagh Carpendale. Activity river: Visualizing planned and logged personal activities for reflection. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 1–9, 2020.

[3] Gulnar Balakayeva, Mukhit Zhanuzakov, and Gaukhar Kalmenova. Development of a digital employee rating evaluation system (deres) based on machine learning algorithms and 360-degree method. *Journal of Intelligent Systems*, 32(1):20230008, 2023.

[4] Gabriel Barata, Hugo Nicolau, and Daniel Gonçalves. Appinsight: What have i been doing? In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 465–472, 2012.

[5] Jocelyn Cranefield, Michael Winikoff, Yi-Te Chiu, Yevgeniya Li, Cathal Doyle, and Alex Richter. Partnering with ai: The case of digital productivity assistants. *Journal of the Royal Society of New Zealand*, 53(1):95–118, 2023.

[6] Elena Di Lascio, Shkurta Gashi, Juan Sebastian Hidalgo, Beatrice Nale, Maike E Debus, and Silvia Santini. A multi-sensor approach to automatically recognize breaks and work activities of knowledge workers in academia. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–20, 2020.

[7] Anton N Dragunov, Thomas G Dietterich, Kevin Johnsrude, Matthew McLaughlin, Lida Li, and Jonathan L Herlocker. Tasktracer: a desktop environment to support multi-tasking knowledge workers. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 75–82, 2005.

[8] Casey Dugan, Werner Geyer, Michael Muller, Abel N Valente, Katherine James, Steve Levy, Li-Te Cheng, Elizabeth Daly, and Beth Brownholtz. "I'd never get out of this !?$% office" redesigning time management for the enterprise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1755–1764, 2012.

[9] Daniel A Epstein, Daniel Avrahami, and Jacob T Biehl. Taking 5: Work-breaks, productivity, and opportunities for personal informatics for knowledge workers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 673–684, 2016.

[10] Michael Freed, Jaime G Carbonell, Geoffrey J Gordon, Jordan Hayes, Brad A Myers, Daniel P Siewiorek, Stephen F Smith, Aaron Steinfeld, and Anthony

Tomasic. Radar: A personal assistant that learns to reduce email overload. In *AAAI*, volume 8, pages 1287–1293, 2008.

[11] Harmanpreet Kaur, Alex C Williams, Daniel McDuff, Mary Czerwinski, Jaime Teevan, and Shamsi T Iqbal. Optimizing for happiness and productivity: Modeling opportune moments for transitions and breaks at work. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2020.

[12] Young-Ho Kim, Eun Kyoung Choe, Bongshin Lee, and Jinwook Seo. Understanding personal productivity: How knowledge workers define, evaluate, and reflect on their productivity. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2019.

[13] Young-Ho Kim, Jae Ho Jeon, Eun Kyoung Choe, Bongshin Lee, KwonHyun Kim, and Jinwook Seo. Timeaware: Leveraging framing effects to enhance personal productivity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 272–283, 2016.

[14] Gloria Mark, Mary Czerwinski, and Shamsi T Iqbal. Effects of individual differences in blocking workplace distractions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[15] Gloria Mark, Shamsi Iqbal, and Mary Czerwinski. How blocking distractions affects workplace focus and productivity. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, pages 928–934, 2017.

[16] Gloria Mark, Shamsi Iqbal, Mary Czerwinski, and Paul Johns. Focused, aroused, but so distractible: Temporal perspectives on multitasking and communications. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 903–916, 2015.

[17] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, Paul Johns, Akane Sano, and Yuliya Lutchyn. Email duration, batching and self-interruption: Patterns of email use on productivity and stress. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1717–1728, 2016.

[18] André N Meyer, Thomas Fritz, Gail C Murphy, and Thomas Zimmermann. Software developers' perceptions of productivity. In *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 19–29, 2014.

[19] Andre N Meyer, Gail C Murphy, Thomas Zimmermann, and Thomas Fritz. Design recommendations for self-monitoring in the workplace: Studies in software development. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW):1–24, 2017.

[20] Stig Nyman, Mads Bødker, and Tina Blegind Jensen. Reforming work patterns or negotiating workloads? exploring alternative pathways for digital productivity assistants through a problematization lens. *Journal of Information Technology*, page 02683962231181602, 2023.

[21] John Rooksby, Parvin Asadzadeh, Mattias Rost, Alistair Morrison, and Matthew Chalmers. Personal tracking of screen time on digital devices. In *Proceedings*

of the 2016 CHI conference on human factors in computing systems, pages 284–296, 2016.

[22] Koustuv Saha and Shamsi T Iqbal. Focus time for wellbeing and work engagement of information workers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2023.

[23] Danielle E Thompson and Adam Molnar. Workplace surveillance in canada: A survey on the adoption and use of employee monitoring applications. *Canadian Review of Sociology/Revue canadienne de sociologie*, 2023.

[24] Vincent W-S Tseng, Matthew L Lee, Laurent Denoue, and Daniel Avrahami. Overcoming distractions during transitions from break to work using a conversational website-blocking system. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–13, 2019.

[25] Steve Whittaker, Vaiva Kalnikaite, Victoria Hollis, and Andrew Guydish. 'don't waste my time' use of time information improves focus. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 1729–1738, 2016.

EXCLUDED

[1] Kaveh Abhari and Isaac Vaghefi. Screen time and productivity: an extension of goal-setting theory to explain optimum smartphone use. *AIS Transactions on Human-Computer Interaction*, 14(3):254–288, 2022.

[2] Osumah Obaze Agbonluae, GO Omi-Ujuanbi, and Mabel Akpede. Coping strategies for managing occupational stress for improved worker productivity. *IFE PsychologIA: An International Journal*, 25(2):300–309, 2017.

[3] Vandana Babshetti and Nihar Ranjan. Machine learning algorithm for work from home analysis during epidemic (2022). *Grenze International Journal of Engineering & Technology (GIJET)*, 8(2), 2022.

[4] Yuki Ban, Sho Sakurai, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. Improving work productivity by controlling the time rate displayed by the virtual clock. In *Proceedings of the 6th Augmented Human International Conference*, pages 25–32, 2015.

[5] Miroslav Behan and Ondrej Krejcar. Concept of the personal devices content management using modular architecture and evaluation based design. In *Context-Aware Systems and Applications: First International Conference, ICCASA 2012, Ho Chi Minh City, Vietnam, November 26-27, 2012, Revised Selected Papers 1*, pages 151–159. Springer, 2013.

[6] Atze C Boerstra, Marcel GLC Loomans, and Jan LM Hensen. Personal control over indoor climate and productivity. In *13th International Conference on Indoor Air Quality and Climate, Indoor Air 2014*, 2014.

[7] Jeremy W Bray, Jesse M Hinde, David J Kaiser, Michael J Mills, Georgia T Karuntzos, Katie R Genadek, Erin L Kelly, Ellen E Kossek, and David A Hurtado. Effects of a flexibility/support intervention on work performance: Evi-

dence from the work, family, and health network. *American Journal of Health Promotion*, 32(4):963–970, 2018.

[8] Beatrice Brunner, Ivana Igic, Anita C Keller, and Simon Wieser. Who gains the most from improving working conditions? health-related absenteeism and presenteeism due to stress at work. *The European Journal of Health Economics*, 20:1165–1180, 2019.

[9] Scott A Cambo, Daniel Avrahami, and Matthew L Lee. Breaksense: Combining physiological and location sensing to promote mobility during work-breaks. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3595–3607, 2017.

[10] Adela Chen and Elena Karahanna. Life interrupted: The effects of technology-mediated work interruptions on work and nonwork outcomes. *MIS quarterly*, 42(4):1023–1042, 2018.

[11] Ming Chen, Bin Ran, Xiaoying Gao, Guilan Yu, Jing Wang, and J Jagannathan. Evaluation of occupational stress management for improving performance and productivity at workplaces by monitoring the health, well-being of workers. *Aggression and Violent Behavior*, page 101713, 2021.

[12] Emily IM Collins, Anna L Cox, Jon Bird, and Cassie Cornish-Tresstail. Barriers to engagement with a personal informatics productivity tool. In *Proceedings of the 26th Australian Computer-Human interaction Conference on Designing futures: The future of design*, pages 370–379, 2014.

[13] Laura Dabbish, Gloria Mark, and Víctor M González. Why do i keep interrupting myself? environment, habit and self-interruption. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3127–3130, 2011.

[14] Vedant Das Swain, Shane Williams, Adam Fourney, and Shamsi T Iqbal. Two birds with one phone: The role of mobile use in the daily practices of remote information work. In *2022 Symposium on Human-Computer Interaction for Work*, pages 1–8, 2022.

[15] François-Xavier De Vaujany, Aurélie Leclercq-Vandelannoitte, Iain Munro, Yesh Nama, and Robin Holt. Control and surveillance in work practice: Cultivating paradox in 'new'modes of organizing, 2021.

[16] Luminița-Mihaela Dumitrașcu, Liliana Feleagă, and Bogdan-Ștefan Ionescu. Time management and time utilization for urology surgeons, a step in implementing social responsibility, a theoretical and a practical approach. In *Business Revolution in a Digital Era: 14th International Conference on Business Excellence, ICBE 2020, Bucharest, Romania*, pages 319–329. Springer, 2021.

[17] Heljä Franssila. Work fragmentation, task management practices and productivity in individual knowledge work. In *Engineering Psychology and Cognitive Ergonomics: 16th International Conference, EPCE 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 29–38. Springer, 2019.

[18] Heljä Franssila, Jussi Okkonen, and Reijo Savolainen. Email intensity, productivity and control in the knowledge worker's performance on the desktop. In *Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content & Services*, pages 19–22, 2014.

[19] Christian Frisson, Sylvain Malacria, Gilles Bailly, and Thierry Dutoit. Inspectorwidget: A system to analyze users behaviors in their applications. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1548–1554, 2016.

[20] Dimitrios Giakoumis, Konstantinos Votis, Efthymios Altsitsiadis, Sofia Segkouli, Ioannis Paliokas, and Dimitrios Tzovaras. Smart, personalized and adaptive ict solutions for active, healthy and productive ageing with enhanced workability. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 442–447, 2019.

[21] Xingzhou Guo, Hongyue Wu, Yunfeng Chen, Yuan Chang, and Yibin Ao. Gauging the impact of personal lifestyle, indoor environmental quality and work-related factors on occupant productivity when working from home. *Engineering, Construction and Architectural Management*, 30(8):3713–3730, 2023.

[22] Amy Hackney, Marcus Yung, Kumara G Somasundram, Behdin Nowrouzi-Kia, Jodi Oakman, and Amin Yazdani. Working in the digital economy: A systematic review of the impact of work from home arrangements on personal and organizational performance and productivity. *Plos one*, 17(10):e0274728, 2022.

[23] M Anaam Hashmi, Abdullah Al Ghaithi, and Khaled Sartawi. Impact of flexible work arrangements on employees' perceived productivity, organisational commitment and perceived work quality: A united arab emirates case-study. *Competitiveness Review: An International Business Journal*, (ahead-of-print), 2021.

[24] Sam Irrinki. Personal tools for becoming a more successful engineer. *Journal-American Water Works Association*, 97(6):68–71, 2005.

[25] Mohammad Hossein Jarrahi, Dorothy Lee Blyth, and Cami Goray. Mindful work and mindful technology: Redressing digital distraction in knowledge work. *Digital Business*, 3(1):100051, 2023.

[26] Johanna Kallio, Elena Vildjiounaite, Jani Koivusaari, Pauli Räsänen, Heidi Similä, Vesa Kyllönen, Salla Muuraiskangas, Jussi Ronkainen, Jari Rehu, and Kaisa Vehmas. Assessment of perceived indoor environmental quality, stress and productivity based on environmental sensor data and personality categorization. *Building and Environment*, 175:106787, 2020.

[27] Johanna Kallio, Elena Vildjiounaite, Vesa Kyllönen, Jussi Ronkainen, Jani Koivusaari, Salla Muuraiskangas, Pauli Räsänen, Heidi Similä, and Kaisa Vehmas. Classifying teachers' self-reported productivity, stress and indoor environmental quality using environmental sensors. In *European Conference on Ambient Intelligence*, pages 27–40. Springer, 2019.

[28] Shun Kawakubo and Shiro Arata. Study on residential environment and workers' personality traits on productivity while working from home. *Building and environment*, 212:108787, 2022.

[29] Soban Ahmed Khan, Asma Ahmad Farhan, Labiba Gillani Fahad, and Syed Fahad Tahir. Personal productivity monitoring through smartphones. *Journal of Ambient Intelligence and Smart Environments*, 12(4):327–341, 2020.

[30] Cheryl Koopman, Kenneth R Pelletier, James F Murray, Claire E Sharda, Marc L Berger, Robin S Turpin, Paul Hackleman, Pamela Gibson, Danielle M Holmes, and Talor Bendel. Stanford presenteeism scale: health status and employee productivity. *Journal of occupational and environmental medicine*, pages 14–20, 2002.

[31] Itaru Kuramoto, Kazumasa Kashiwagi, Yu Shibuya, Yoshihiro Tsujino, and Shigeki Ohtsuka. How can entertainment improve workers' motivation and their productivity? In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, pages 24–31, 2004.

[32] Amanda E Lackey, Mariam Moshiri, Tarun Pandey, Chandana Lall, Neeraj Lalwani, and Puneet Bhargava. Productivity, part 1: getting things done, using e-mail, scanners, reference managers, note-taking applications, and text expanders. *Journal of the American College of Radiology*, 11(5):481–489, 2014.

[33] Laura C Larsson. Improving your productivity with a pda: some suggestions. *Library hi tech*, 21(4):426–439, 2003.

[34] Jemin Lee, Sihyeong Park, Taeho Kim, and Hyungshin Kim. Time-invariant features-based online learning for long-term notification management: A longitudinal study. *Applied Sciences*, 12(11):5432, 2022.

[35] Cuauhtémoc LÓPEZ, Arturo CHAVOYA, and Maria Elena MEDA-CAMPAÑA. Software development productivity prediction of individual projects applying a neural network.

[36] Cuauhtémoc López-Martín, Rosa Leonor Ulloa-Cazarez, and Andrés García-Floriano. Support vector regression for predicting the productivity of higher education graduate students from individually developed software projects. *IET Software*, 11(5):265–270, 2017.

[37] Yuhan Luo, Bongshin Lee, Young-Ho Kim, and Eun Kyoung Choe. Notewordy: Investigating touch and speech input on smartphones for personal data capture. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):568–591, 2022.

[38] Zhanna Lyubykh, Duygu Gulseren, Zahra Premji, Timothy G Wingate, Connie Deng, Lisa J Bélanger, and Nick Turner. Role of work breaks in well-being and performance: A systematic review and future research agenda. *Journal of Occupational Health Psychology*, 27(5):470, 2022.

[39] Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 107–110, 2008.

[40] Gloria Mark, Shamsi T Iqbal, Mary Czerwinski, and Paul Johns. Bored mondays and focused afternoons: the rhythm of attention and online activity in the workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3025–3034, 2014.

[41] Vanda C Marques and Gregory R Berry. Enhancing work-life balance using a resilience framework. *Business and Society Review*, 126(3):263–281, 2021.

[42] Anika L McGrath, Katerina Dodelzon, Omer A Awan, Nicholas Said, and Puneet Bhargava. Optimizing radiologist productivity and efficiency: Work smarter, not harder. *European Journal of Radiology*, 155:110131, 2022.

[43] Ray M Merrill, Steven G Aldana, James E Pope, David R Anderson, Carter R Coberley, Whitmer, and R William the HERO Research Study Subcommittee. Presenteeism according to healthy behaviors, physical health, and work environment. *Population Health Management*, 15(5):293–301, 2012.

[44] Kazuo Misue. Chronoview: A space-efficient method for visualizing temporal patterns. In *2014 11th International Conference on Computer Graphics, Imaging and Visualization*, pages 1–4. IEEE, 2014.

[45] Kazune Miyagi, Kotaro Oishi, Kosuke Uchiyama, Hirotake Ishii, and Hiroshi Shimoda. Proposal of intellectual productivity model based on work state transition. In *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition: 10th International Conference, EPCE 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part I 10*, pages 335–343. Springer, 2013.

[46] Zbakh Mourad, Aknin Noura, and Chrayah Mohamed. Towards a new method for classifying employee performance using machine learning algorithms. In *2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pages 1–5. IEEE, 2022.

[47] Muhanna Muhanna, Brian Tackitt, and Sergiu Dascalu. Prototype details of the smartphone-based researcher's companion software (rcs). *Journal of Computational Methods in Sciences and Engineering*, 9(s2):S191–S200, 2009.

[48] Helga Guðrún Óskarsdóttir, Guðmundur Valur Oddsson, Jón Þór Sturluson, and Rögnvaldur Jóhann Sæmundsson. A soft systems approach to knowledge worker productivity: A purposeful activity model for the individual. *Administrative Sciences*, 11(4):110, 2021.

[49] Miikka Palvalin, Maiju Vuolle, Aki Jääskeläinen, Harri Laihonen, and Antti Lönnqvist. Smartwow–constructing a tool for knowledge work performance analysis. *International Journal of Productivity and Performance Management*, 64(4):479–498, 2015.

[50] Aitor Iriondo Pascual, Dan Högberg, Dan Lämkull, Estela Perez Luque, Anna Syberfeldt, and Lars Hanson. Optimization of productivity and worker wellbeing by using a multi-objective optimization framework. *IISE Transactions on Occupational Ergonomics and Human Factors*, 9(3-4):143–153, 2021.

[51] Arturo Peralta, Eduardo Conde, Manuel García, and Francisco P Romero. Behavioral pattern recognition and knowledge extraction for decision-making in software project management. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 865–868. IEEE, 2016.

[52] Arturo Peralta, Francisco P Romero, Jose A Olivas, and Macario Polo. Knowledge extraction of the behaviour of software developers by the analysis of time recording logs. In *International Conference on Fuzzy Systems*, pages 1–8. IEEE, 2010.

[53] Ilaria Pigliautile, Sara Casaccia, Nicole Morresi, Marco Arnesano, Anna Laura Pisello, and Gian Marco Revel. Assessing occupants' personal attributes in relation to human perception of environmental comfort: Measurement procedure and data analysis. *Building and Environment*, 177:106901, 2020.

[54] Tye Rattenbury and John Canny. Caad: an automatic task support system. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 687–696, 2007.

[55] Yara Rizk, Vatche Isahagian, Merve Unuvar, and Yasaman Khazaeni. A snooze-less user-aware notification system for proactive conversational agents. *arXiv preprint arXiv:2003.02097*, 2020.

[56] Ruhi Sarikaya. The technology behind personal digital assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine*, 34(1):67–81, 2017.

[57] Benjamin Schooley, Steven Walczak, Neset Hikmet, and Nitin Patel. Impacts of mobile tablet computing on provider productivity, communications, and the process of care. *International journal of medical informatics*, 88:62–70, 2016.

[58] Nandita Sharma and Tom Gedeon. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer methods and programs in biomedicine*, 108(3):1287–1301, 2012.

[59] Nirvan Sharma and Patrick Hosein. A comparison of data-driven and traditional approaches to employee performance assessment. In *2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, pages 34–41. IEEE, 2020.

[60] Jianqiang Shen, Lida Li, Thomas G Dietterich, and Jonathan L Herlocker. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In *Proceedings of the 11th international conference on Intelligent user interfaces*, pages 86–92, 2006.

[61] Anya Skatova, Ben Bedwell, Victoria Shipp, Yitong Huang, Alexandra Young, Tom Rodden, and Emma Bertenshaw. The role of ict in office work breaks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 3049–3060, 2016.

[62] Mauricio Soto, Chris Satterfield, Thomas Fritz, Gail C Murphy, David C Shepherd, and Nicholas Kraft. Observing and predicting knowledge worker stress, focus and awakeness in the wild. *International Journal of Human-Computer Studies*, 146:102560, 2021.

[63] Fabian J Stangl and René Riedl. Interruptions in the workplace: An exploratory study among digital business professionals. In *International Conference on Human-Computer Interaction*, pages 400–422. Springer, 2023.

[64] A Sumukha, Ritwik Reddy, et al. Machine learning based personality classification using clustering algorithm. In *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)*, pages 571–575. IEEE, 2022.

[65] Alexander Sunk, Peter Kuhlang, Thomas Edtmayr, and Wilfried Sihn. Developments of traditional value stream mapping to enhance personal and organisational system and methods competencies. *International Journal of Production Research*, 55(13):3732–3746, 2017.

[66] Ish A Talati, Pranay Krishnan, and Ross W Filice. Developing deeper radiology exam insight to optimize mri workflow and patient experience. *Journal of digital imaging*, 32:865–869, 2019.

[67] Maki Tominaga, Takashi Asakura, and Tsuyoshi Akiyama. The effect of micro and macro stressors in the work environment on computer professionals' subjective health status and productive behavior in japan. *Industrial health*, 45(3):474–486, 2007.

[68] Catherine Wideman and Jacqueline Gallet. Analog to digital workflow improvement: A quantitative study. *Journal of Digital Imaging*, 19:29–34, 2006.

[69] Alex C Williams, Shamsi Iqbal, Julia Kiseleva, and Ryen W White. Managing tasks across the work–life boundary: Opportunities, challenges, and directions. *ACM Transactions on Computer-Human Interaction*, 30(3):1–31, 2023.

[70] K-L Wu, Philip S. Yu, and Allen Ballman. Speedtracer: A web usage mining and analysis tool. *IBM Systems Journal*, 37(1):89–105, 1998.

[71] Laura Zapata, Gerardo Ibarra, and Pierre-Henri Blancher. Engaging new ways of work: the relevance of flexibility and digital tools in a post-covid-19 era. *Journal of Organizational Effectiveness: People and Performance*, 2023.

E APPENDIX E: GRID-SEARCH PARTICIPANT ONE

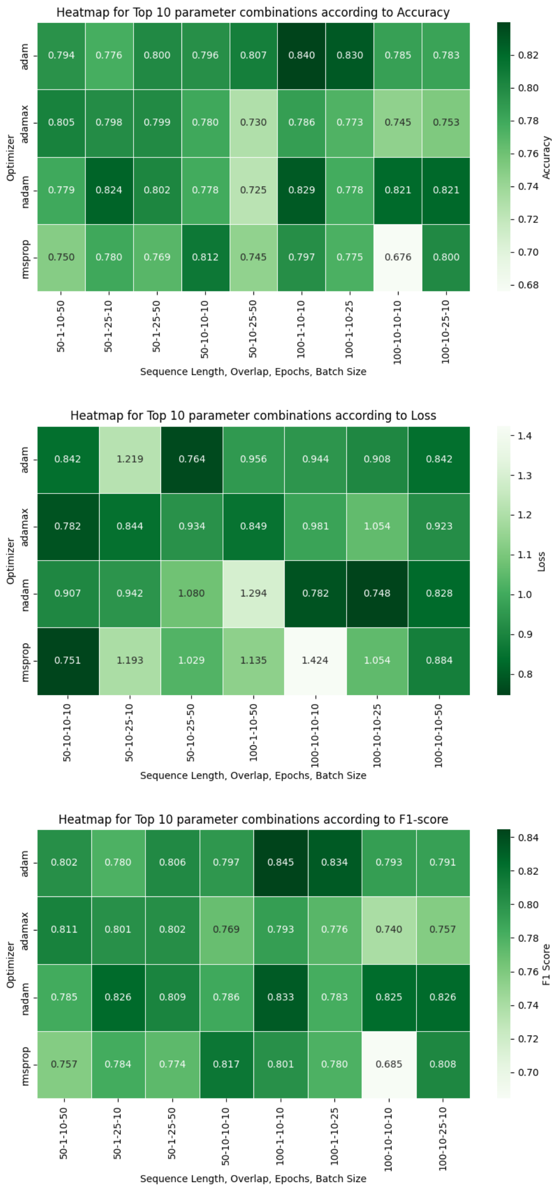


Figure 29: Heatmap of top 10 grid search combinations based on accuracy, loss and f1-score for p1.

F APPENDIX F: GRID-SEARCH PARTICIPANT TWO

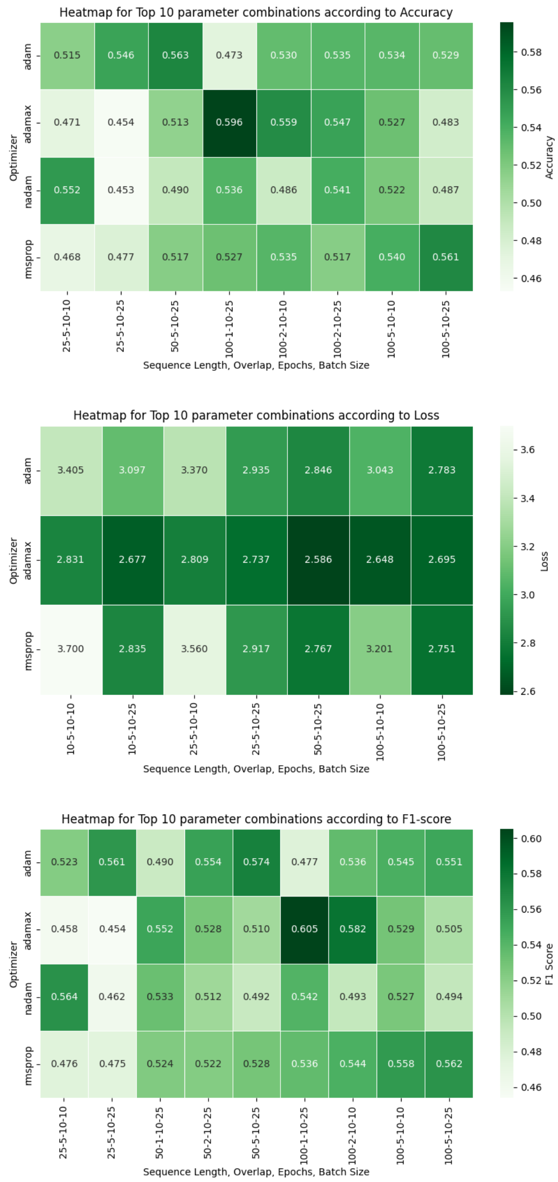


Figure 30: Heatmap of top 10 grid search combinations based on accuracy, loss and f1-score for p2.

G APPENDIX G: GRID-SEARCH PARTICIPANT THREE AND 4

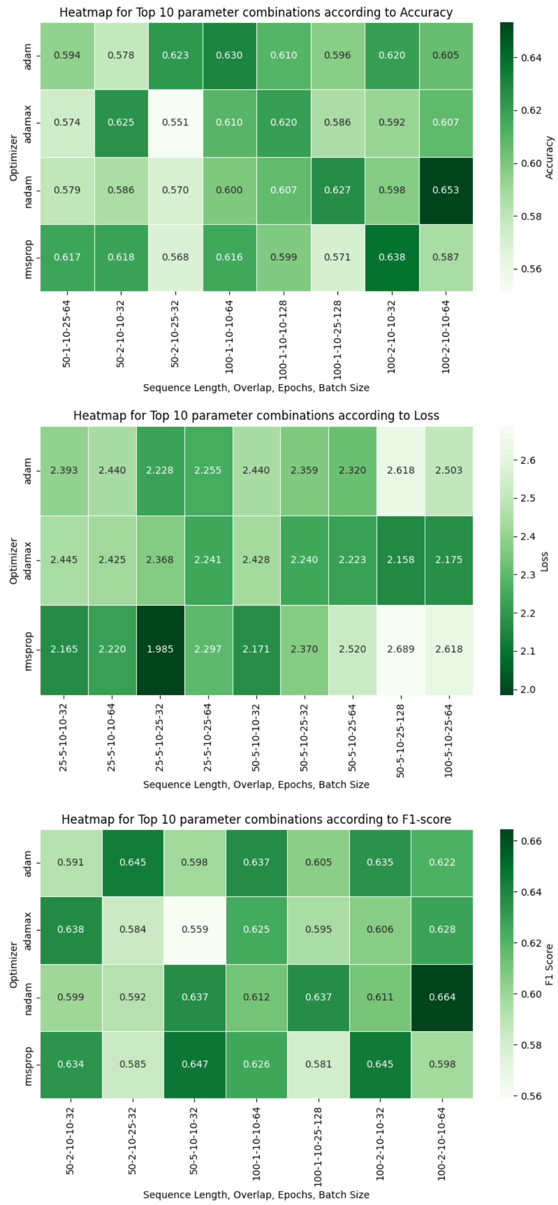


Figure 31: Heatmap of top 10 grid search combinations based on accuracy, loss and f1-score for p3 and p4.

H APPENDIX H: DESCRIPTION OF SUB-CATEGORIES FOR APP FEATURE VALUES

App sub-categories	Description
Web Browser Application	Application used for browsing the world wide web (e.g. Google Chrome)
Reading Application	Application used for reading text (e.g. Adobe Acrobat Reader)
Helper Application	Application used to provide support with tasks (e.g. Calculator)
Data Analysis Application	Application used for analyzing data (e.g. Microsoft Excel)
File Storage Application	Application used for storing files (e.g. Google Drive)
Task Management Application	Application used to provide support with managing tasks (e.g. Agenda)
Text Editing Application	Application used to create and edit textual content (e.g. Microsoft Word)
Media and Entertainment Application	Application used to watch and/or listen to media (e.g. Spotify, VLC media player)
Communication Application	Application used for communication by video, audio or text (e.g. WhatsApp, Microsoft Teams)
System Application	Application used to facilitate system operation and user interaction within the operating system (e.g. Finder, .exe installers)
Development Application	Application used for the development purposes (e.g. Notepad++)
Visual Editing Application	Application used for video and/or photo editing (e.g. Adobe Premiere Pro, Adobe Lightroom)

Table 12: Overview of App sub-categories with examples.

I APPENDIX I: FEATURE LEVELS FOR DATA ANALYSIS AND PATTERN MINING

Participant	Short (s)	Medium (s)	Long (s)
p1	<4	4-57	>57
p2	<60	60-120	>120
p3	<3	3-39	>39
p4	<6	6-66	>66

Table 13: Work activity duration values in seconds per participant corresponding to the levels.

Participant	morning (6am-12am)	afternoon (12am-6pm)	evening (6pm-12pm)	night (12pm-6am)
p1	✓	✓	✓	0.0066 (h)
p2	✓	✓	✓	4.02 (h)
p3	✓	✓	3.59 (h)	0.0033 (h)
p4	✓	✓	6.83 (h)	5.93 (h)

Table 14: Part of the day levels for each participant and values in hours if the level had too little data.

Participant	Pattern length	Low	Medium	High
p1	2	0	1	>1
p1	3	0	1	>1
p1	4	0	1-2	>2
p2	2	0	1-2	>2
p2	3	0	1-2	>2
p2	4	0	1-3	>3
p3	2	0	1	>1
p3	3	0	1-2	>2
p3	4	0	1-2	>2
p4	2	0	1	>1
p4	3	0	1	>1
p4	4	0	1-2	>2

Table 15: Total switches level values of each participant for each pattern length.

Participant	Pattern length	Short (s)	Medium (s)	Long (s)
p1	2	<18	18-102	>102
p1	3	<33	33-156	>156
p1	4	<52	52-208	>208
p2	2	<42	42-180	>180
p2	3	<54	54-240	>240
p2	4	<69	69-300	>300
p3	2	<15	15-79	>79
p3	3	<30	30-118	>118
p3	4	<45	45-160	>160
p4	2	<18	18-102	>102
p4	3	<33	33-147	>147
p4	4	<52	52-192	>192

Table 16: Total work duration level values of each participant for each pattern length.

J APPENDIX J: COEFFICIENTS ESTIMATED BY THE REGRESSION MODELS FOR ALL PARTICIPANTS.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.0829	0.515	4.041	0.000	1.063	3.103
C(day_of_week, Treatment("Tuesday"))[T.Monday]	0.4348	0.722	0.602	0.548	-0.994	1.864
C(day_of_week, Treatment("Tuesday"))[T.Wednesday]	1.2773	0.690	1.852	0.066	-0.087	2.642
C(day_of_week, Treatment("Tuesday"))[T.Thursday]	0.3060	0.785	0.390	0.697	-1.248	1.860
C(day_of_week, Treatment("Tuesday"))[T.Friday]	-0.0571	0.734	-0.078	0.938	-1.510	1.396
C(day_of_week, Treatment("Tuesday"))[T.Saturday]	1.9420	0.766	2.537	0.012	0.427	3.457
C(day_of_week, Treatment("Tuesday"))[T.Sunday]	1.5092	0.705	2.141	0.034	0.115	2.904
duration_m_tot	0.0032	0.002	1.932	0.056	-7.82e-05	0.007
C(day_of_week, Treatment("Tuesday"))[T.Monday]:duration_m_tot	0.0001	0.002	0.068	0.946	-0.004	0.004
C(day_of_week, Treatment("Tuesday"))[T.Wednesday]:duration_m_tot	-0.0027	0.002	-1.316	0.191	-0.007	0.001
C(day_of_week, Treatment("Tuesday"))[T.Thursday]:duration_m_tot	-0.0001	0.002	-0.042	0.967	-0.005	0.005
C(day_of_week, Treatment("Tuesday"))[T.Friday]:duration_m_tot	0.0022	0.002	0.932	0.353	-0.002	0.007
C(day_of_week, Treatment("Tuesday"))[T.Saturday]:duration_m_tot	-0.0014	0.003	-0.494	0.622	-0.007	0.004
C(day_of_week, Treatment("Tuesday"))[T.Sunday]:duration_m_tot	-0.0010	0.003	-0.354	0.724	-0.007	0.005

(a) Coefficients of p1.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.5152	0.211	11.903	0.000	2.093	2.938
C(day_of_week, Treatment("Sunday"))[T.Monday]	0.3504	0.541	0.647	0.520	-0.732	1.433
C(day_of_week, Treatment("Sunday"))[T.Tuesday]	0.8097	0.446	1.817	0.074	-0.081	1.701
C(day_of_week, Treatment("Sunday"))[T.Wednesday]	0.5055	0.438	1.155	0.253	-0.370	1.381
C(day_of_week, Treatment("Sunday"))[T.Thursday]	0.3349	0.487	0.687	0.494	-0.639	1.309
C(day_of_week, Treatment("Sunday"))[T.Friday]	0.5146	0.612	0.840	0.404	-0.710	1.739
C(day_of_week, Treatment("Sunday"))[T.Saturday]	-8.278e-16	2.59e-16	-3.195	0.002	-1.35e-15	-3.1e-16
duration_m_tot	0.0055	0.009	0.626	0.533	-0.012	0.023
C(day_of_week, Treatment("Sunday"))[T.Monday]:duration_m_tot	-0.0020	0.009	-0.222	0.825	-0.020	0.016
C(day_of_week, Treatment("Sunday"))[T.Tuesday]:duration_m_tot	-0.0030	0.009	-0.333	0.740	-0.021	0.015
C(day_of_week, Treatment("Sunday"))[T.Wednesday]:duration_m_tot	-0.0017	0.009	-0.187	0.852	-0.019	0.016
C(day_of_week, Treatment("Sunday"))[T.Thursday]:duration_m_tot	-0.0011	0.009	-0.124	0.902	-0.019	0.017
C(day_of_week, Treatment("Sunday"))[T.Friday]:duration_m_tot	-0.0017	0.009	-0.182	0.857	-0.020	0.017
C(day_of_week, Treatment("Sunday"))[T.Saturday]:duration_m_tot	0	0	nan	nan	0	0

(b) Coefficients of p2.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.8223	0.533	5.298	0.000	1.742	3.903
C(day_of_week, Treatment("Thursday"))[T.Monday]	0.5622	0.829	0.679	0.502	-1.118	2.243
C(day_of_week, Treatment("Thursday"))[T.Tuesday]	-0.2707	1.101	-0.246	0.807	-2.503	1.962
C(day_of_week, Treatment("Thursday"))[T.Wednesday]	1.7353	1.449	1.198	0.239	-1.203	4.673
C(day_of_week, Treatment("Thursday"))[T.Friday]	-0.3378	0.737	-0.458	0.650	-1.833	1.158
C(day_of_week, Treatment("Thursday"))[T.Saturday]	-4.575e-15	2.11e-14	-0.216	0.830	-4.74e-14	3.83e-14
C(day_of_week, Treatment("Thursday"))[T.Sunday]	0.0001	9.91e-05	1.112	0.274	-9.09e-05	0.000
duration_m_tot	0.0014	0.003	0.401	0.691	-0.006	0.008
C(day_of_week, Treatment("Thursday"))[T.Monday]:duration_m_tot	-0.0018	0.005	-0.333	0.741	-0.012	0.009
C(day_of_week, Treatment("Thursday"))[T.Tuesday]:duration_m_tot	0.0021	0.005	0.412	0.683	-0.008	0.012
C(day_of_week, Treatment("Thursday"))[T.Wednesday]:duration_m_tot	-0.0051	0.007	-0.734	0.468	-0.019	0.009
C(day_of_week, Treatment("Thursday"))[T.Friday]:duration_m_tot	0.0020	0.004	0.497	0.622	-0.006	0.010
C(day_of_week, Treatment("Thursday"))[T.Saturday]:duration_m_tot	0	0	nan	nan	0	0
C(day_of_week, Treatment("Thursday"))[T.Sunday]:duration_m_tot	0.0107	0.010	1.112	0.274	-0.009	0.030

(c) Coefficients of p3.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.7297	0.337	11.064	0.000	3.060	4.399
C(day_of_week, Treatment("Wednesday"))[T.Monday]	-0.3787	0.594	-0.637	0.525	-1.559	0.801
C(day_of_week, Treatment("Wednesday"))[T.Tuesday]	-0.2327	0.452	-0.515	0.608	-1.130	0.664
C(day_of_week, Treatment("Wednesday"))[T.Thursday]	0.1306	0.502	0.260	0.795	-0.866	1.127
C(day_of_week, Treatment("Wednesday"))[T.Friday]	-0.3662	0.536	-0.683	0.496	-1.431	0.698
C(day_of_week, Treatment("Wednesday"))[T.Saturday]	0.9166	1.461	0.628	0.532	-1.984	3.818
C(day_of_week, Treatment("Wednesday"))[T.Sunday]	0.0811	0.509	0.159	0.874	-0.930	1.092
duration_m_tot	-4.828e-05	0.002	-0.027	0.979	-0.004	0.004
C(day_of_week, Treatment("Wednesday"))[T.Monday]:duration_m_tot	0.0034	0.004	0.935	0.352	-0.004	0.011
C(day_of_week, Treatment("Wednesday"))[T.Tuesday]:duration_m_tot	0.0015	0.003	0.581	0.563	-0.004	0.007
C(day_of_week, Treatment("Wednesday"))[T.Thursday]:duration_m_tot	0.0017	0.003	0.594	0.554	-0.004	0.007
C(day_of_week, Treatment("Wednesday"))[T.Friday]:duration_m_tot	0.0043	0.003	1.388	0.169	-0.002	0.011
C(day_of_week, Treatment("Wednesday"))[T.Saturday]:duration_m_tot	0.0013	0.017	0.078	0.938	-0.033	0.035
C(day_of_week, Treatment("Wednesday"))[T.Sunday]:duration_m_tot	0.0020	0.003	0.618	0.538	-0.004	0.008

(d) Coefficients of p4.

Figure 32: Output of linear regression models of all participants.

K APPENDIX K: WORK ACTIVITY DISTRIBUTION AND PRODUCTIVITY FOR ALL PARTICIPANTS.

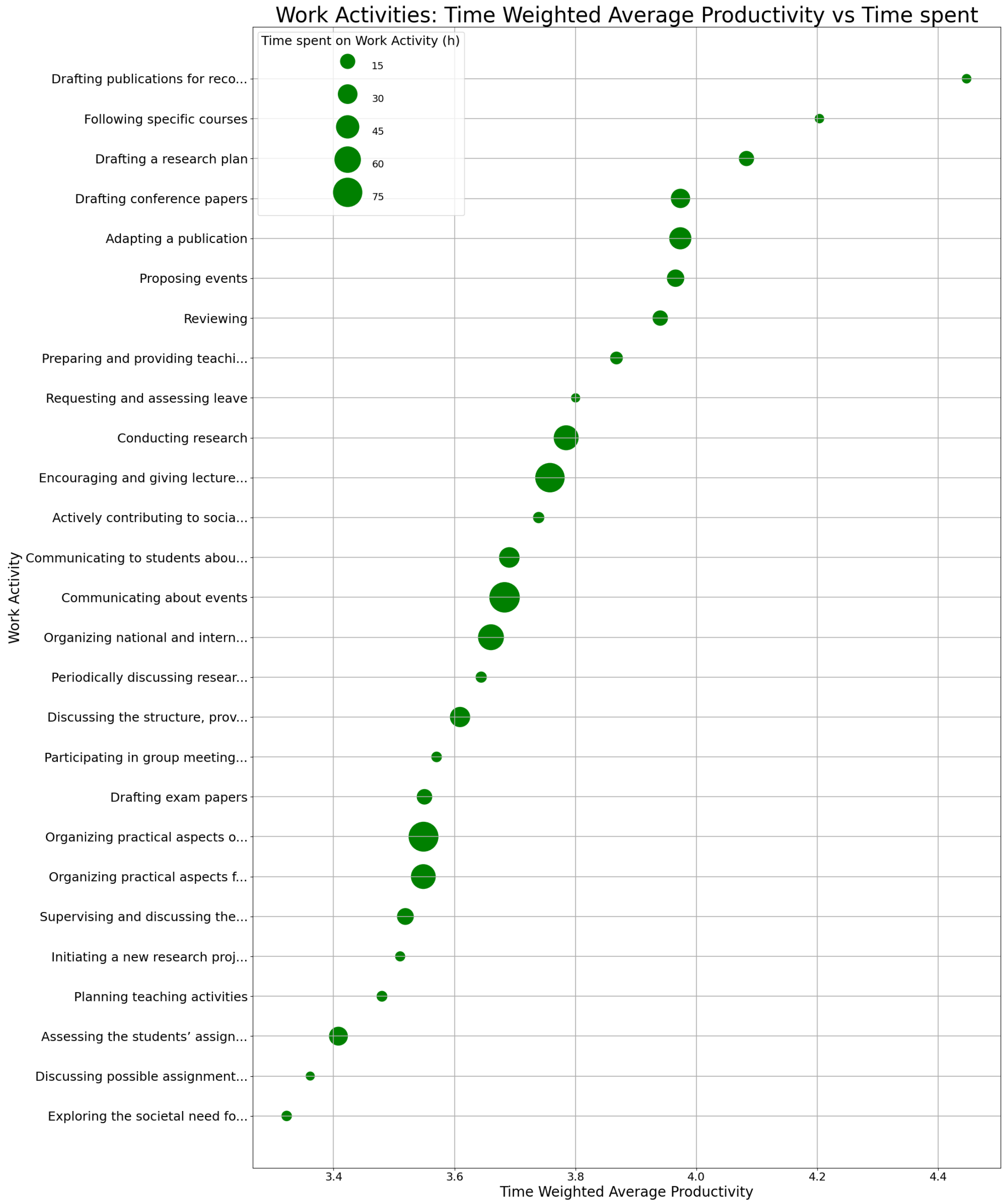


Figure 33: Total time spent in work activities vs time weighted average productivity for participant one.

Figure 34: Total time spent in work activities vs time weighted average productivity for participant two.

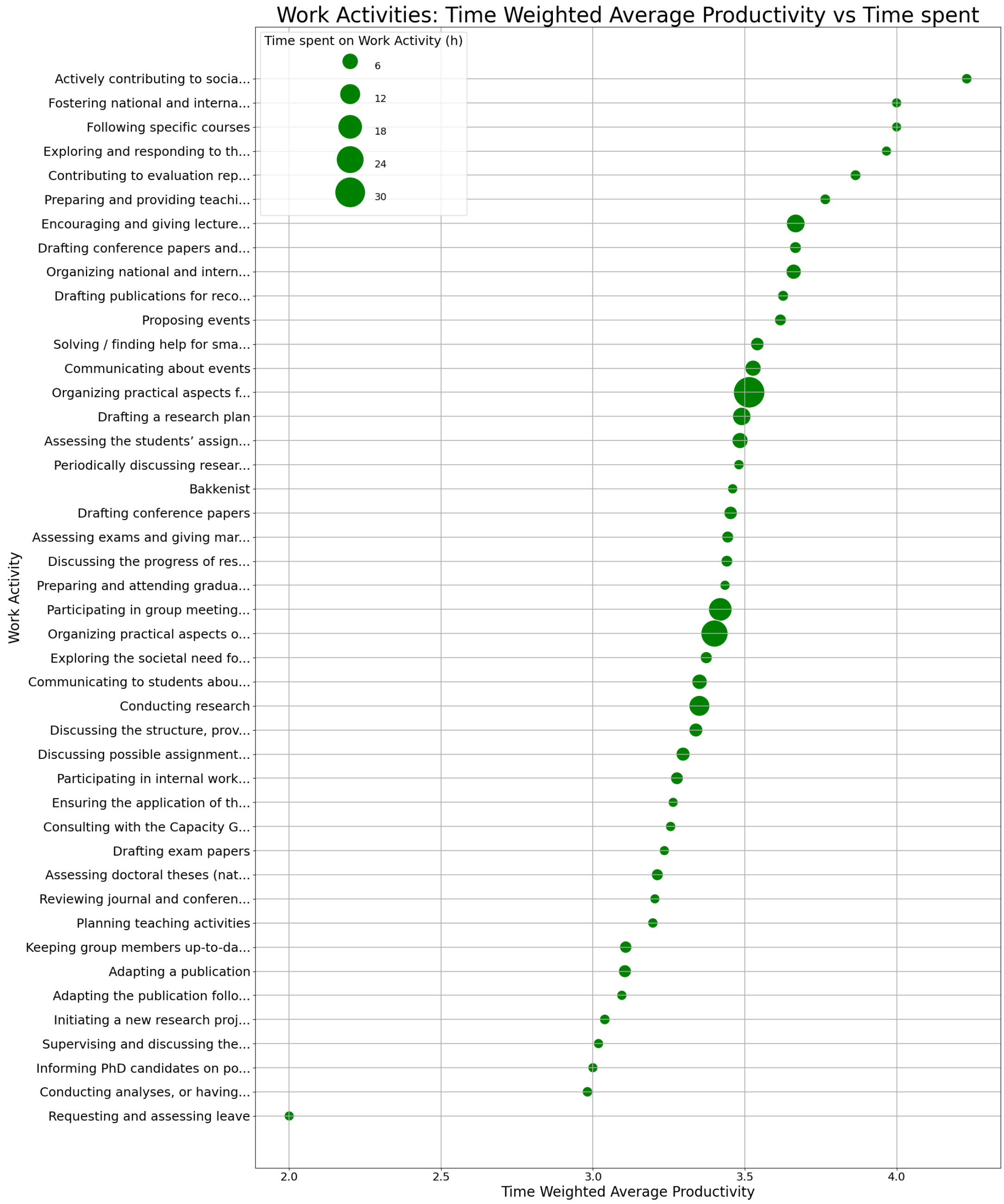
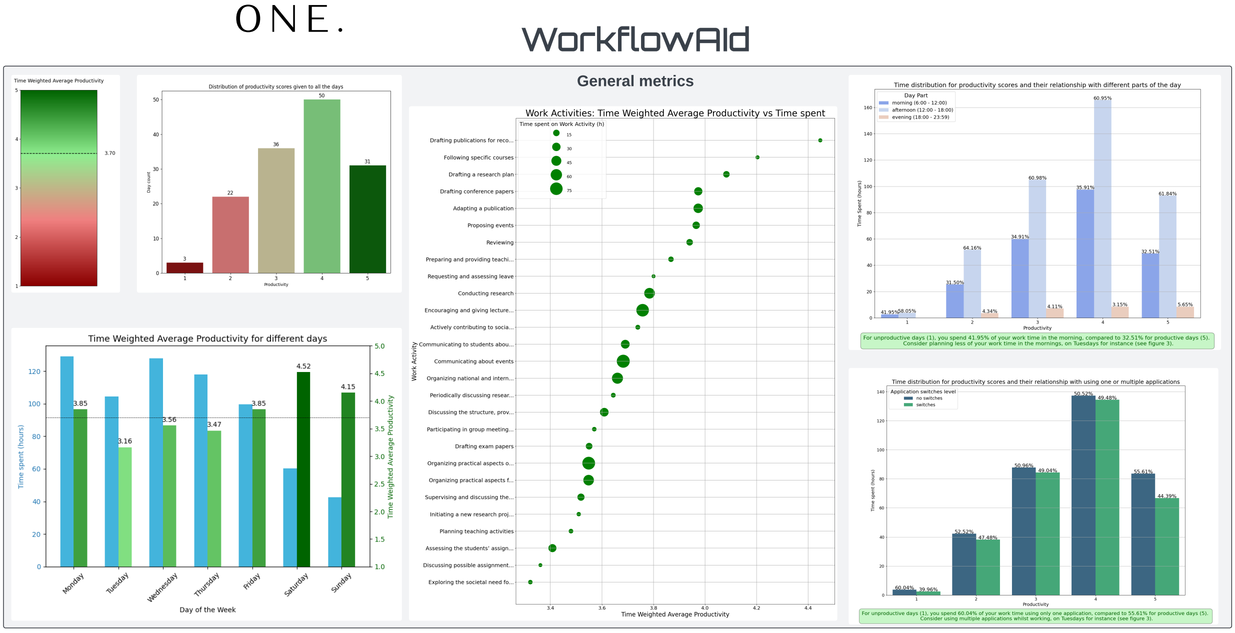
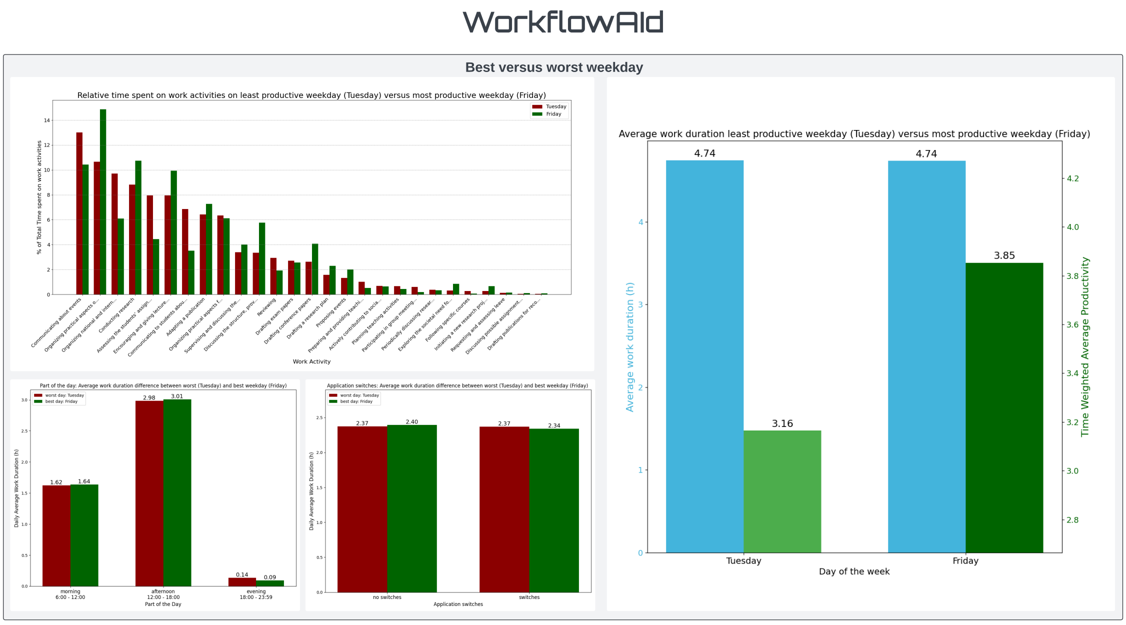


Figure 35: Total time spent in work activities vs time weighted average productivity for participant three.

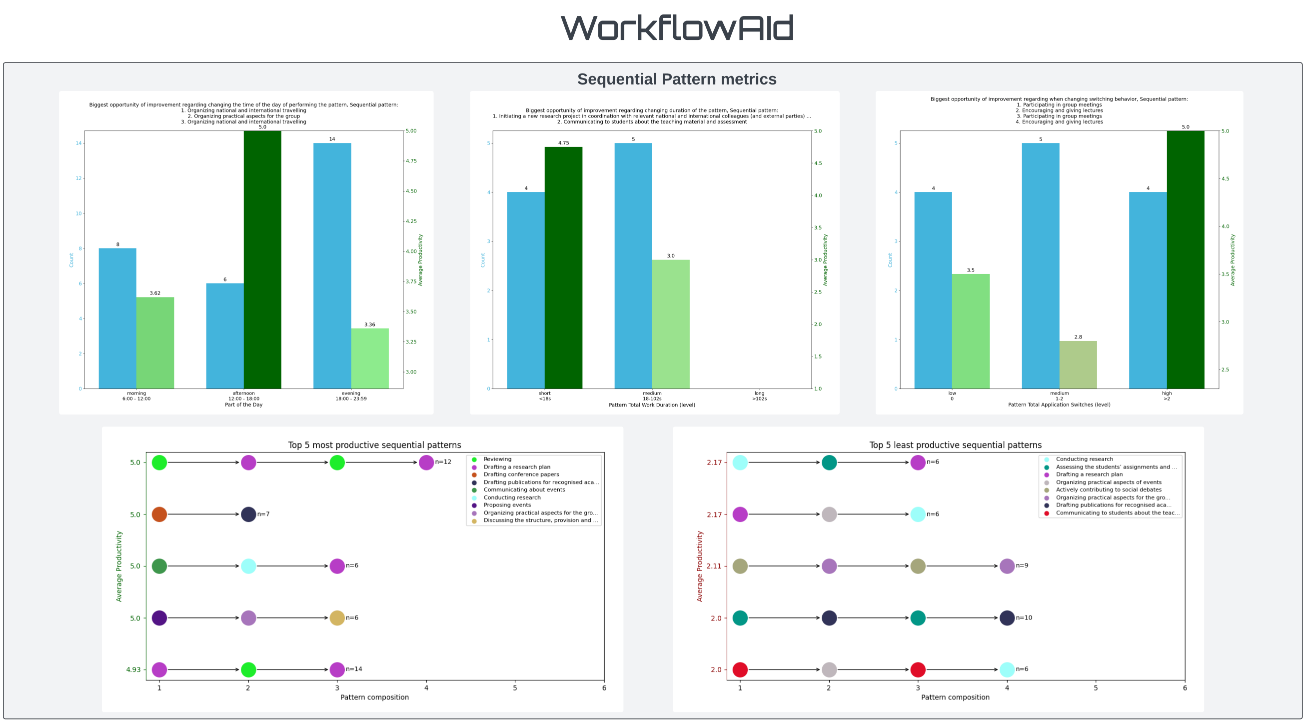
Figure 36: Total time spent in work activities vs time weighted average productivity for participant four.

L | APPENDIX L: WORKFLOW SUPPORT SYSTEM DESIGN FOR PARTICIPANT ONE.






WorkflowAId
Work Activity: Conducting Research
Conducting research: Time Weighted Average Productivity
Conducting research:
Time distribution for productivity scores and their relationship with different parts of the day
Day Part
morning (6:00 - 12:00)
afternoon (12:00 - 18:00)
evening (18:00 - 23:59)
For unproductive days (1), you spend 87.15% performing this work activity in the afternoon, compared to 69.22% for productive days (5).
Consider performing this work activity less in the afternoons, on Tuesdays for instance (see figure 2).
Conducting research: Top 5 most productive sequential patterns
Conducting research:
Time Weighted Average Productivity for different days
Time spent (hours)
Time Weighted Average Productivity
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week
Conducting research:
Time distribution for productivity scores and their relationship with using one or multiple applications
Application switches level
no switches
switches
For unproductive days (1), you spend 43.16% performing this work activity using only one application, compared to 26.59% for productive days (5).
Consider using multiple applications whilst performing this work activity, on Tuesdays for instance (see figure 2).
Conducting research: Top 5 least productive sequential patterns
Average Productivity
Pattern composition

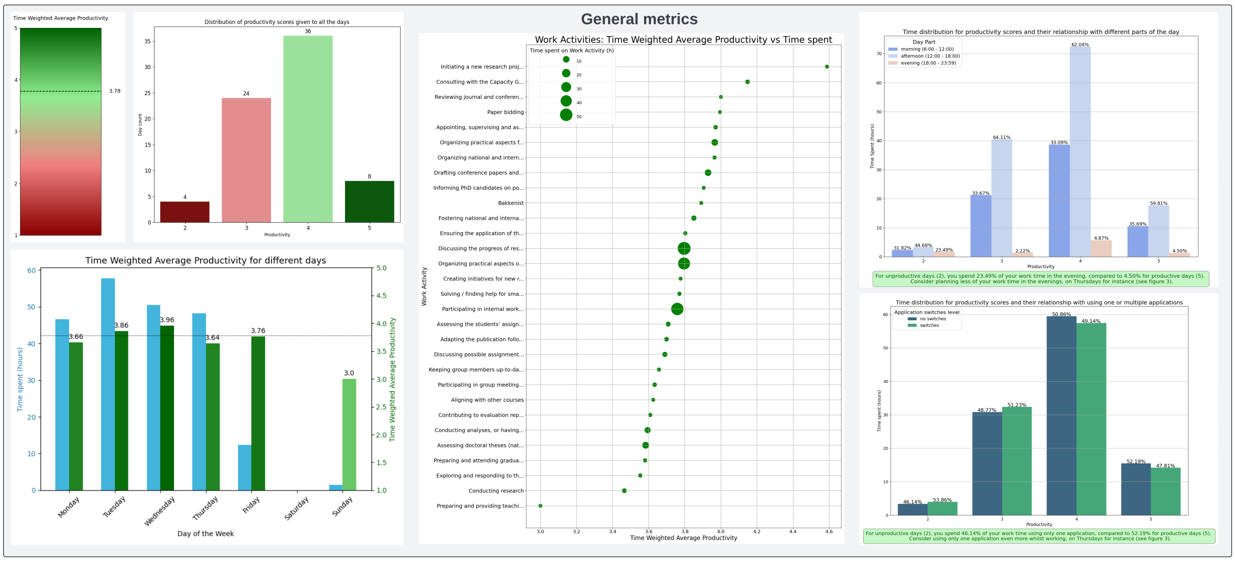

WorkflowAId
Sequential Pattern metrics
Top 5 most productive sequential patterns
Top 5 least productive sequential patterns
Average Productivity
Pattern composition


WorkflowAId
Sequential Pattern metrics - 'Organizing practical aspects of events' and 'Conducting research'
Biggest opportunity for improvement regarding changing the time of the day of performing the pattern, Sequential pattern:
1. Organizing practical aspects of events
2. Conducting research
3. Communicating to students about the teaching material and assessment
Count
Average Productivity
morning 6:00 - 12:00
afternoon 12:00 - 18:00
Part of the Day
10
3.8
4
4.5
Biggest opportunity for improvement regarding changing duration of the pattern, Sequential pattern:
1. Organizing practical aspects of events
2. Organizing practical aspects for the group
3. Conducting research
short <33s
medium 33-156s
long >156s
Pattern Total Work Duration (level)
0
3
4.67
6
3.0
Biggest opportunity for improvement regarding when changing switching behavior, Sequential pattern:
1. Organizing practical aspects of events
2. Conducting research
3. Organizing practical aspects of events
low 0
medium 1
high >1
Pattern Total Application Switches (level)
53
3.28
22
3.77
34
3.41
Organizing practical aspects of events + Conducting research: Top 5 most productive patterns
Conducting research
Organizing practical aspects of events
Communicating to students about the teac...
Supervising and discussing the progress ...
Organizing practical aspects for the gro...
Communicating about events
Periodically discussing research results...
4.75
4.5
4.25
4.25
4.25
n=4
Average Productivity
Pattern composition
Organizing practical aspects of events + Conducting research: Top 5 least productive patterns
Conducting research
Drafting a research plan
Organizing practical aspects of events
Communicating to students about the teac...
2.71
2.5
2.25
2.17
2.0
n=7
n=6
n=4
n=6
n=6
Average Productivity
Pattern composition

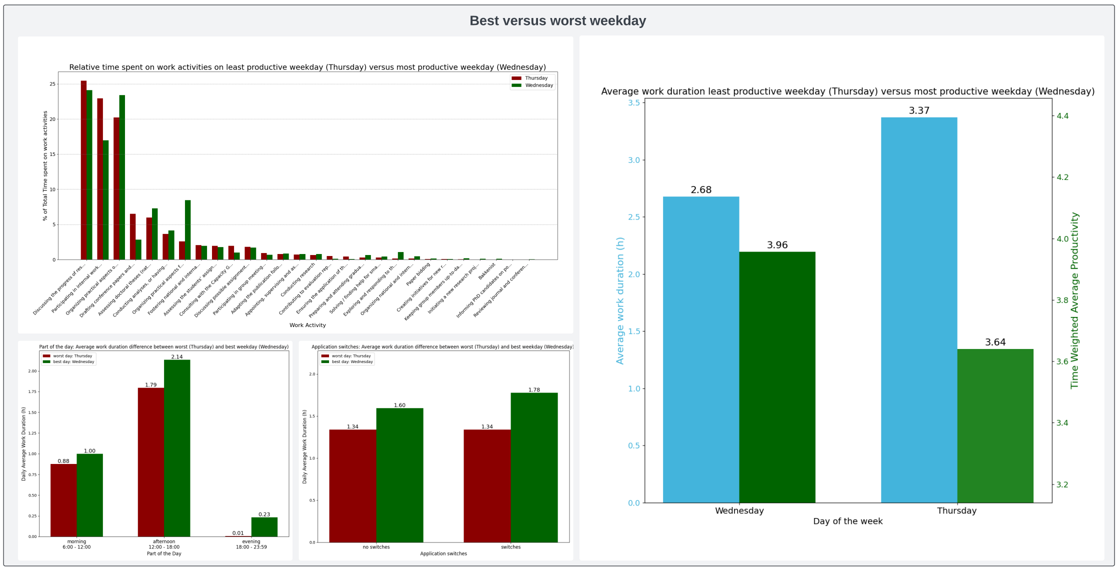
M

APPENDIX M: WORKFLOW SUPPORT SYSTEM DESIGN FOR PARTICIPANT TWO.

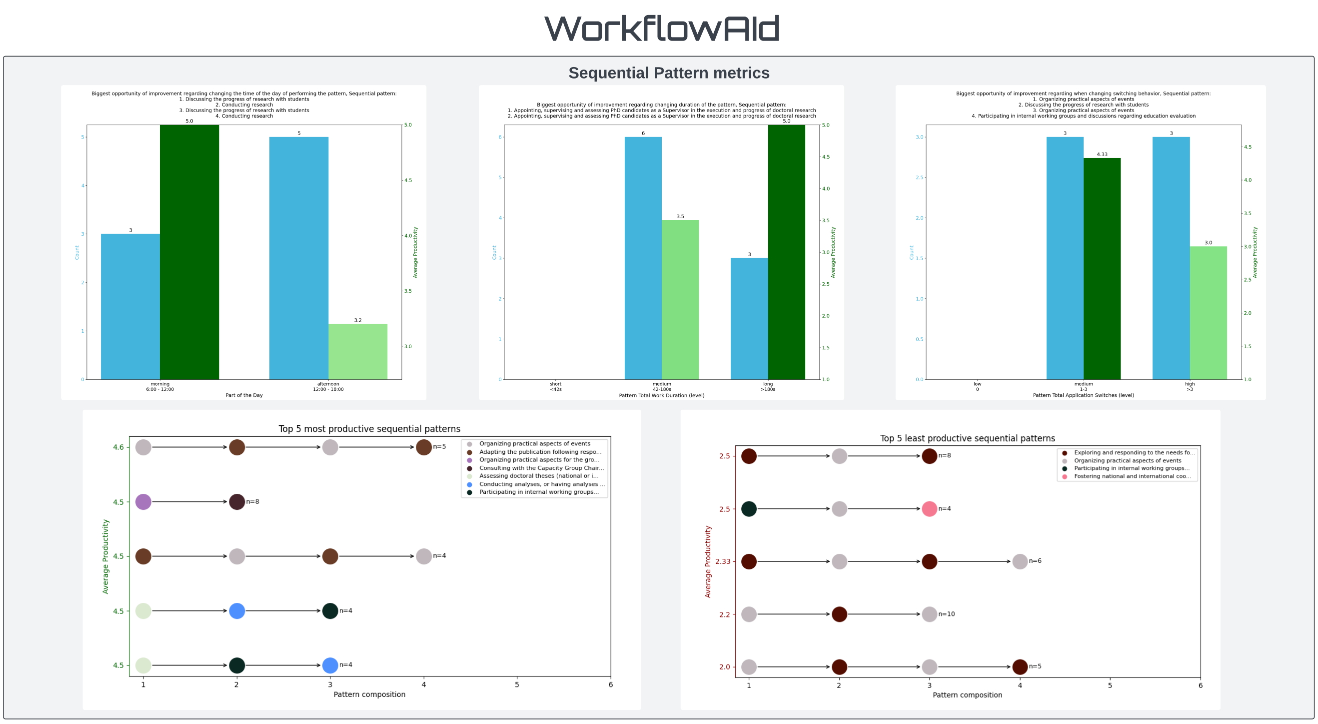
WorkflowAId

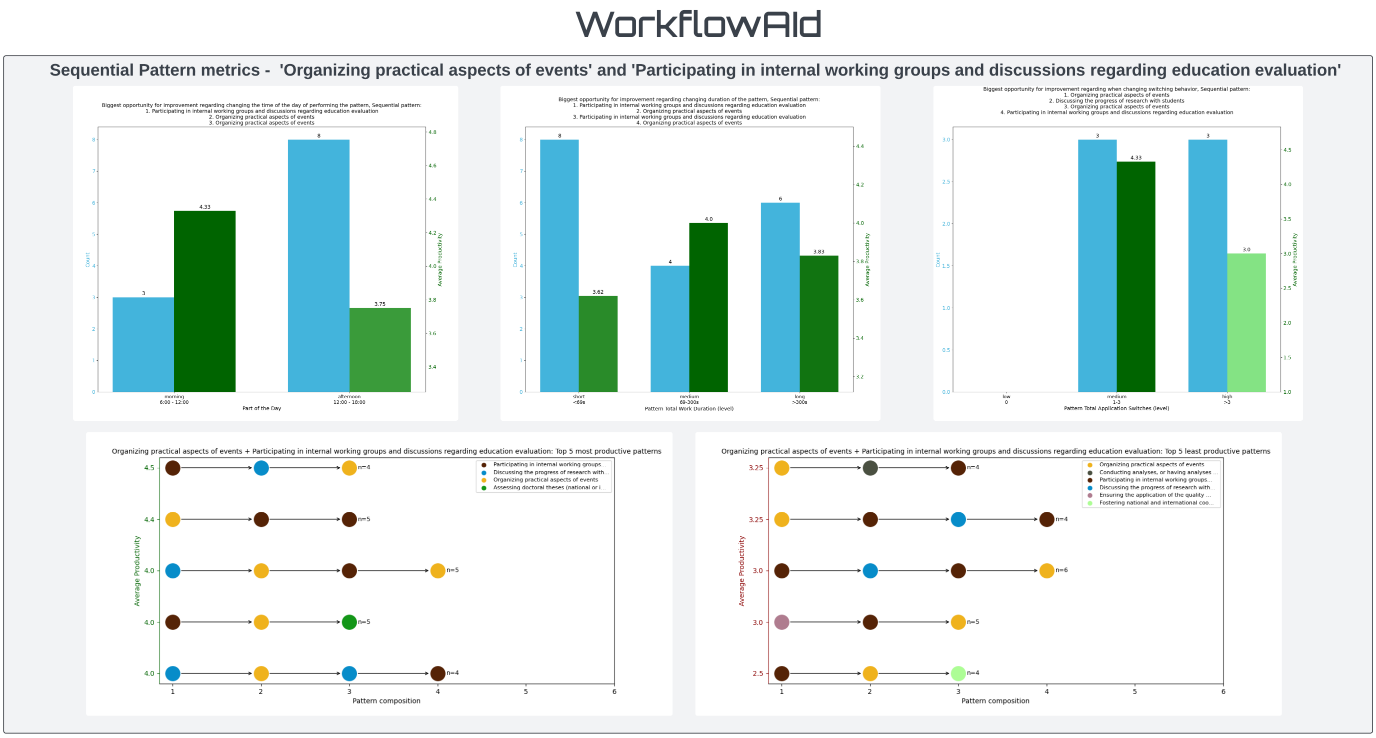


WorkflowAId



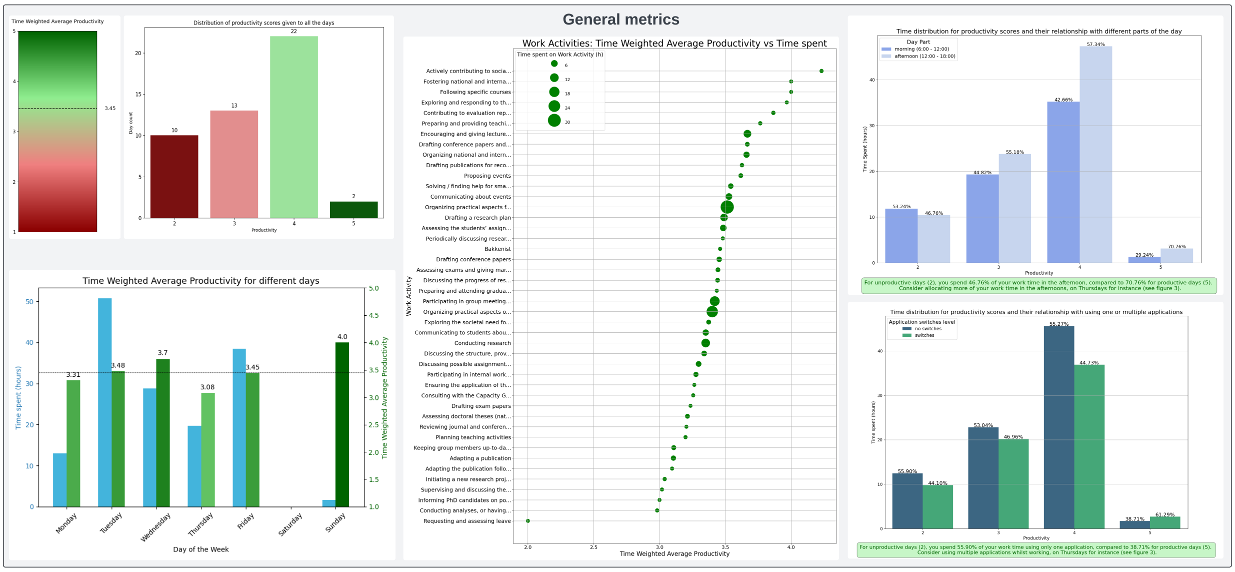

WorkflowAId
Work Activity: Organzing practical aspects of events
Organizing practical aspects of events: Time Weighted Average Productivity
3.80
Organizing practical aspects of events: Time distribution for productivity scores and their relationship with different parts of the day
Organizing practical aspects of events: Top 5 most productive sequential patterns
Organizing practical aspects of events: Time Weighted Average Productivity for different days
Day of the Week
Organizing practical aspects of events: Time distribution for productivity scores and their relationship with using one or multiple applications
Organizing practical aspects of events: Top 5 least productive sequential patterns
Pattern composition
Average Productivity


WorkflowAId
Sequential Pattern metrics
Part of the Day
Pattern Total Work Duration (level)
Pattern Total Application Switches (level)
Top 5 most productive sequential patterns
Top 5 least productive sequential patterns
Pattern composition
Average Productivity

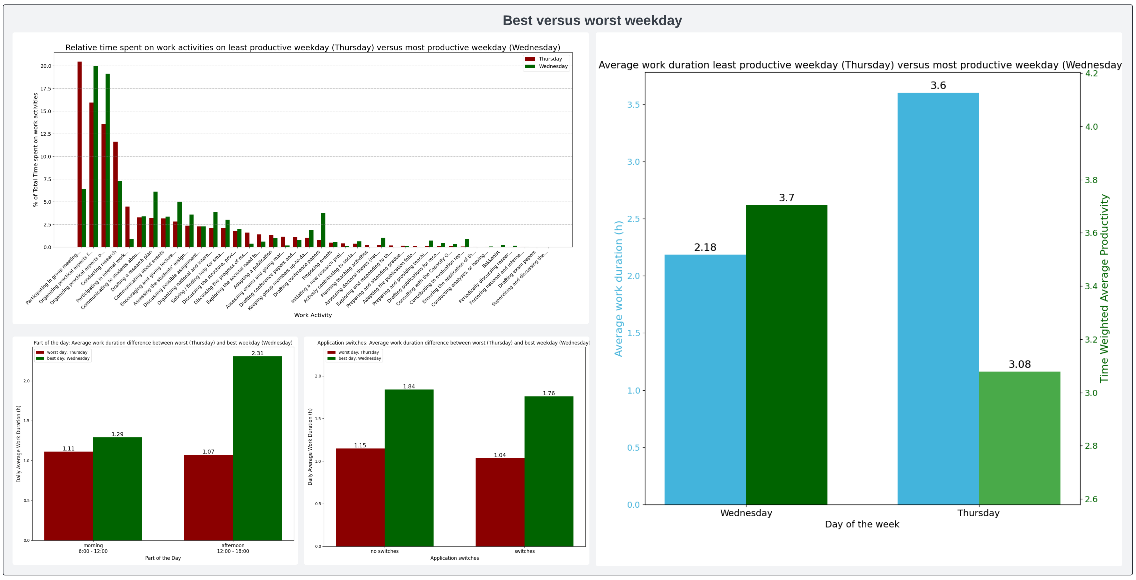

WorkflowAId
Sequential Pattern metrics - 'Organizing practical aspects of events' and 'Participating in internal working groups and discussions regarding education evaluation'
Biggest opportunity for improvement regarding changing the time of the day of performing the pattern, Sequential pattern:
1. Participating in internal working groups and discussions regarding education evaluation
2. Organizing practical aspects of events
3. Organizing practical aspects of events
Part of the Day
Biggest opportunity for improvement regarding changing duration of the pattern, Sequential pattern:
1. Participating in internal working groups and discussions regarding education evaluation
2. Organizing practical aspects of events
3. Participating in internal working groups and discussions regarding education evaluation
4. Organizing practical aspects of events
Pattern Total Work Duration (level)
Biggest opportunity for improvement regarding when changing switching behavior, Sequential pattern:
1. Organizing practical aspects of events
2. Discussing the progress of research with students
3. Organizing practical aspects of events
4. Participating in internal working groups and discussions regarding education evaluation
Pattern Total Application Switches (level)
Organizing practical aspects of events + Participating in internal working groups and discussions regarding education evaluation: Top 5 most productive patterns
Average Productivity
Pattern composition
Organizing practical aspects of events + Participating in internal working groups and discussions regarding education evaluation: Top 5 least productive patterns
Average Productivity
Pattern composition

N | APPENDIX N: WORKFLOW SUPPORT SYSTEM DESIGN FOR PARTICIPANT THREE.

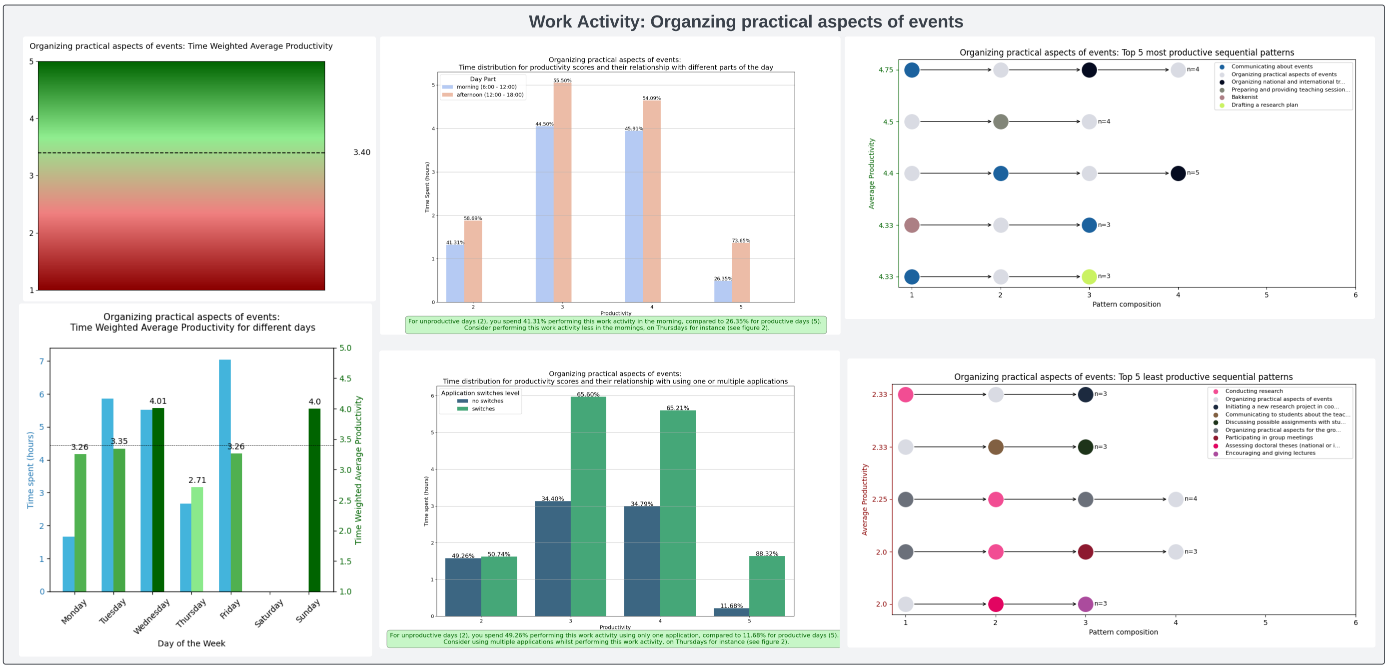
WorkflowAId



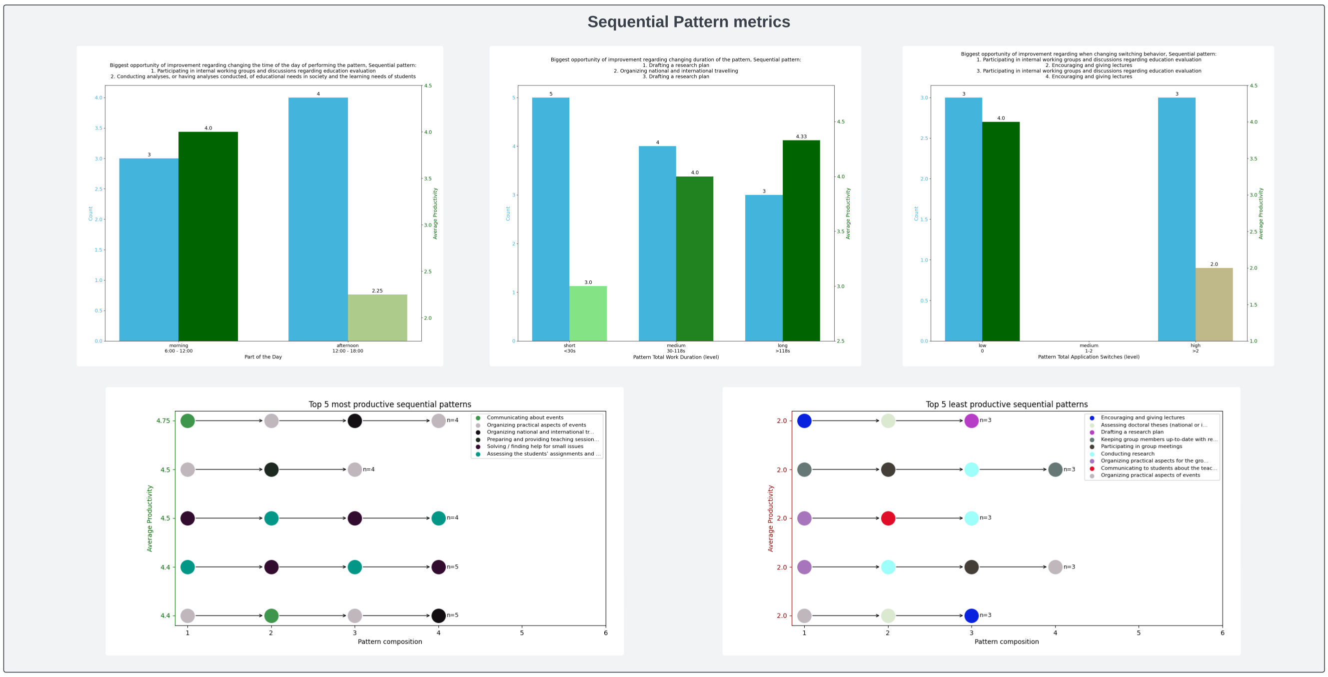
WorkflowAId



WorkflowAId



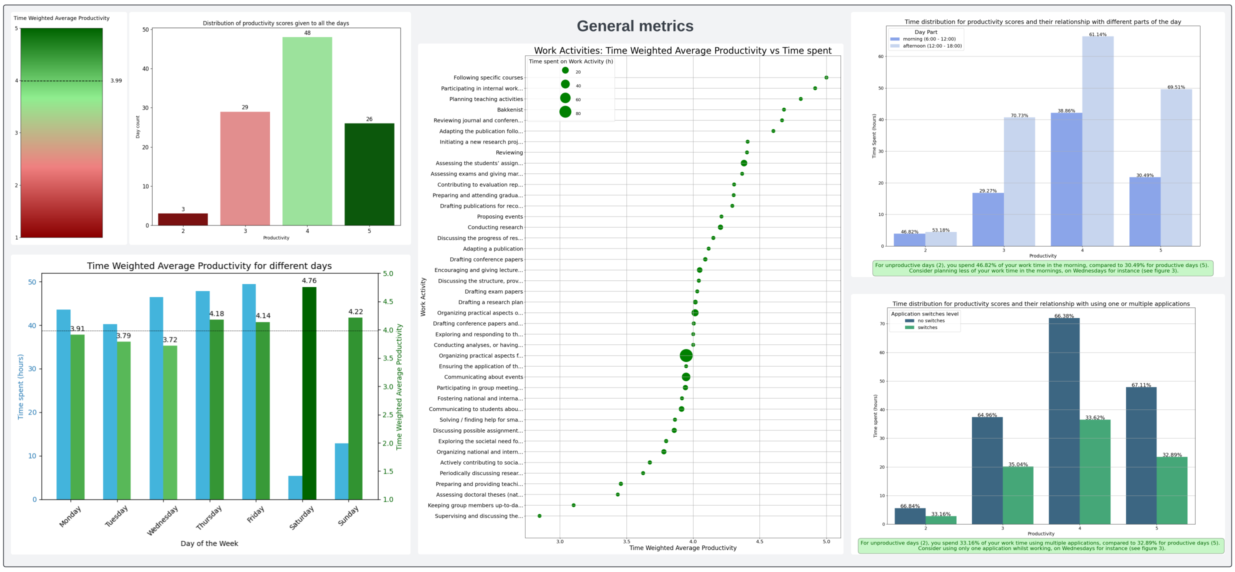
WorkflowAId



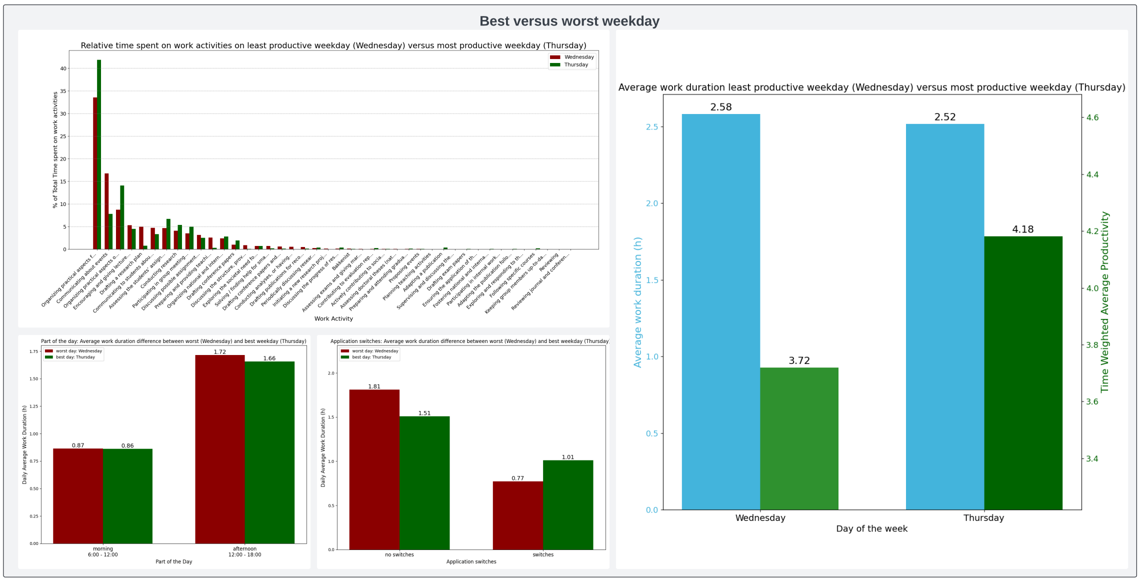
WorkflowAId

O APPENDIX O: WORKFLOW SUPPORT SYSTEM DESIGN FOR PARTICIPANT FOUR.

WorkflowAId

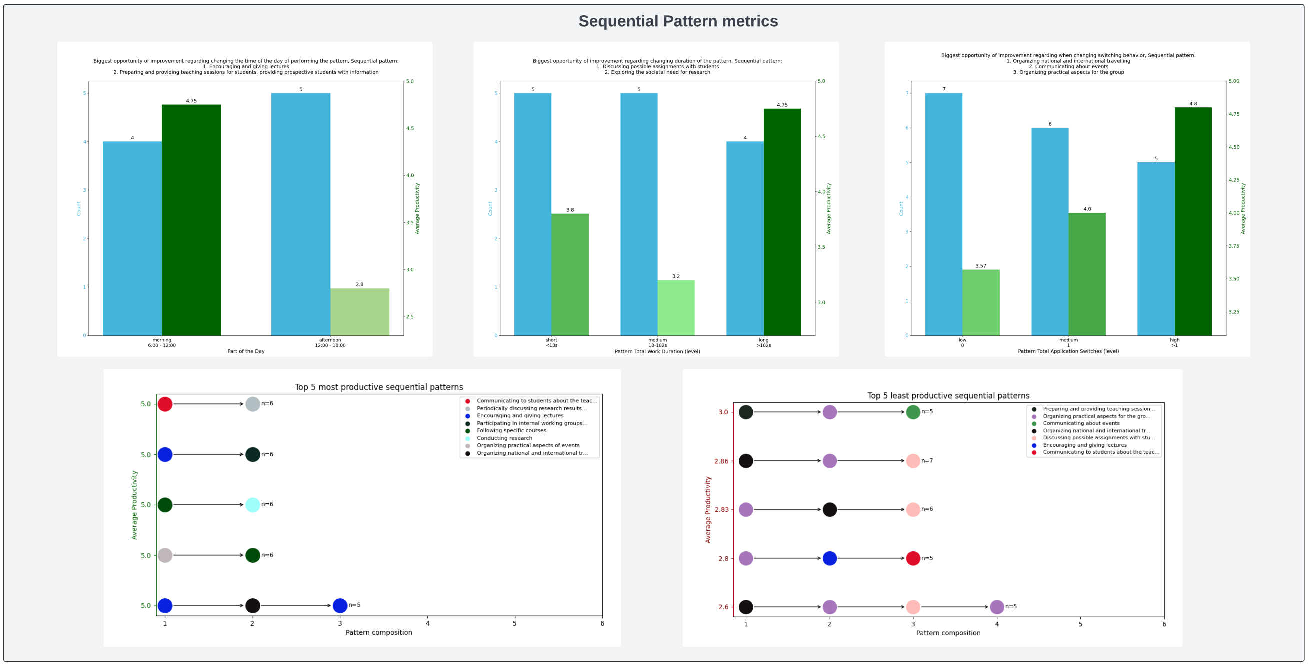


WorkflowAId

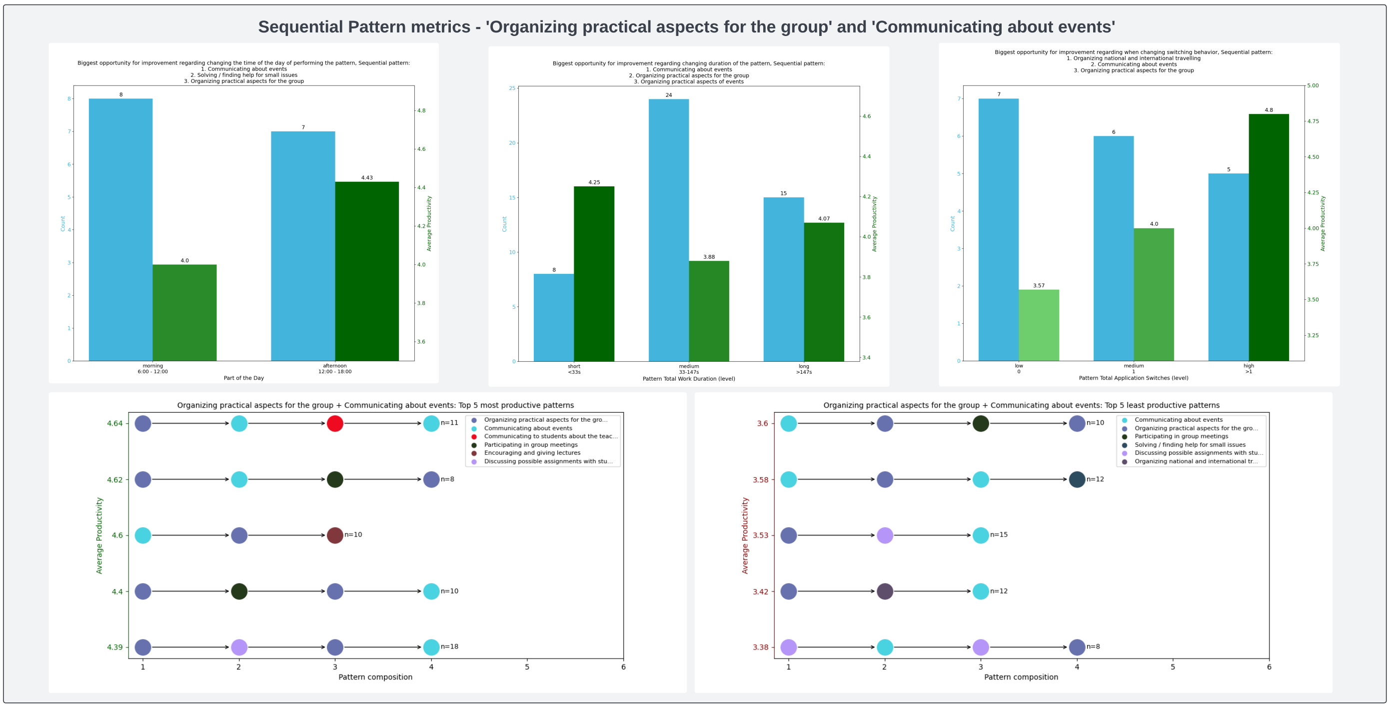


WorkflowAId

WorkflowAId



WorkflowAId



P APPENDIX P: LITERATURE RESEARCH PROTOCOL

Our research is positioned in the field of personal productivity, personal informatics system, machine learning and workflow management. Therefore, the search terms used to find relevant papers will be based on these topics and combinations of them. An overview of the sources used for finding the relevant literature can be seen in Table 17. The specific sources were chosen based on recommendations from the University of Utrecht for searching for literature [2], the fact that queries resulted in loads of relevant literature and the ease of use of the search engine.

Source	Description
Scopus	The Scopus search engine is a comprehensive and citation database and was used to find relevant articles by keyword combinations.
Google Scholar	The Google Scholar search engine supplies full text and/or metadata of literature across many publishers and was used to find related literature to the articles found on the Scopus search engine.

Table 17: Web sources used for literature research.

In the first phase of our literature research protocol, a long-list of papers is created by entering combinations of keywords in the Scopus search engine we elaborated on in Table 17. These search terms can be seen below.

Search terms:

- 'improving AND personal AND productivity AND work'
- 'personal AND productivity AND machine AND learning'
- 'productivity AND personal AND informatics AND systems'
- 'digital AND workflow AND optimization AND productivity'
- 'process AND mining AND machine AND learning AND productivity'
- 'work AND practices AND productivity AND digital AND work'

For each combination, resulting literature is judged based on the title and meta data, like publishing year. The judgement is made on multiple inclusion and exclusion criteria. These criteria determine the scope and validity of the literature review results [16]. The inclusion and exclusion criteria are listed below.

Inclusion criteria:

- IC1: The paper is an academic paper
- IC2: The paper is published before December 2023

- IC3: The paper focuses on one or more of the fields this research is positioned in

Exclusion criteria:

- EC1: The paper is not written in English
- EC2: The paper focuses on enterprise productivity as a whole, not on a personal level
- EC3: The paper does not focus on any of the fields this research is positioned in

The resulted long-list will be enriched by scanning their 'Related articles' in Google Scholar. This process is also known as 'Snowballing'. After this, the long-list is reduced to a short-list based on reading the abstract and conclusion. By reading these parts, we will be able to pick out the articles that were most relevant to our research by evaluating their connection to the field our study is positioned in, on which we elaborated earlier in this section. Literature from this short-list will be fully read and a summary will be made for each article. The summary always starts with an overview of the article by condensing the abstract. After this the article is fully read and the most important parts of the article are listed as bullet points. The articles are always read with the following questions in mind:

- What are the main findings in this research?
- What research methods were used in this research?
- How are the findings connected to our research?
- What is different about our research?

An example of such a summary can be seen in Appendix C. After the summaries are created, a topic structure is constructed and the literature review will be written. An overview of the articles found during different searching phases can be seen in Table 18 and the resulting long- and short-list can be found in Appendix D.

Phase	Resulting articles
Keyword search	852
Long-list criteria met - keyword search	59
Long-list criteria met - Snowballing	37
Long-list - total	96
Short-list	25

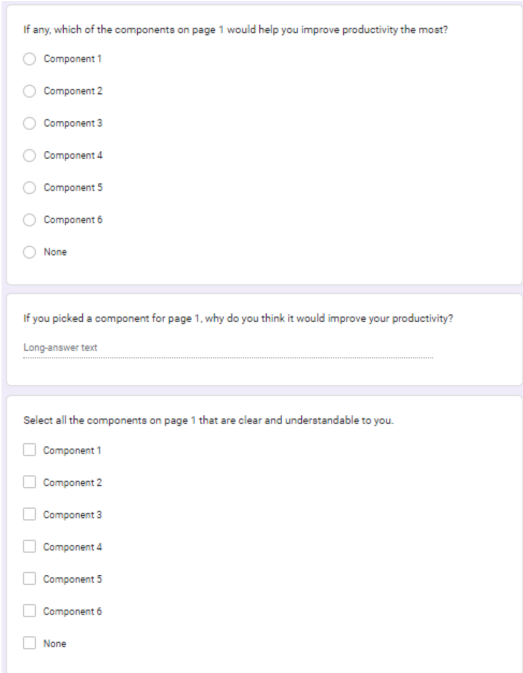
Table 18: Articles found during the different phases of our Literature Research Protocol.

Q APPENDIX Q: MEM QUESTIONNAIRE

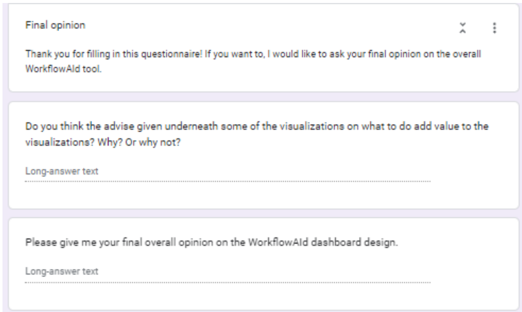
	1	2	3	4	5
Perceived Usefulness					
PU1: WorkflowAId would make me more productive	1/4	2/4	1/4		
PU2: WorkflowAId would make it easier to perform my work.	1/4	3/4			
PU3: WorkflowAId would enable me to perform tasks more quickly.		4/4			
PU4: The visualizations used in WorkflowAId help me understand my workflow.		1/4	1/4	2/4	
PU5: WorkflowAId would enable me to identify personal patterns.		1/4	1/4	2/4	
PU6: I would find WorkflowAId useful during my work.		2/4	1/4	1/4	
Perceived Ease of Use					
PEoU1: Learning to operate WorkflowAId would be easy for me.			2/4	1/4	1/4
PEoU2: WorkflowAId would make it easy for me to get valuable insights.		1/4	1/4	2/4	
PEoU3: My interaction with the WorkflowAId tool would be clear and understandable.		2/4	1/4	1/4	
PEoU4: I would think WorkflowAId would be flexible to interact with.		1/4	3/4		
PEoU5: It would be easy for me to make useful changes to improve my productivity with the use of WorkflowAId.		3/4		1/4	
PEoU6: WorkflowAId would be easy to use.		1/4	2/4	1/4	
Intention to Use					
ItU1: I would use WorkflowAId during my work to improve productivity.	2/4	1/4		1/4	
ItU2: I would use WorkflowAId during my work to identify personal characteristics.		1/4		3/4	

1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree

Table 19: MEM Questionnaire results of all participants.



(a) Questions about components dashboard.



(b) Questions about overall opinion on whole dashboard.

Figure 37: MEM questionnaire extra questions.