Utrecht University

Master's Thesis

Predicting Relevant Key Performance Indicators for Roles in Organizations

Author:
Jan Vrieler

Supervisors:
Prof.dr.ir. Hajo A. Reijers
Dr.ing. Georg Krempl
MSc. Ünal Aksu

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Business Informatics
Department of Information and Computing Sciences

January 6, 2021

UTRECHT UNIVERSITY

Abstract

Faculty of Science
Department of Information and Computing Sciences

Master of Science

Predicting Relevant Key Performance Indicators for Roles in Organizations

by Jan VRIELER

Business Intelligence (BI) and Business Process Monitoring (BPM) have become staples of modern management. BPM and BI keep track of business processes and capture them in Key Performance Indicators (KPIs). KPIs are visualizations of metrics that measure the status of business processes. By consulting KPIs, the status of a business process can be easily assessed and acted upon. Organizations provide KPIs to their employees to improve their decisions relating to the captured business processes. Creating KPIs is a time and cost-intensive process. To reduce the effort required for this process, some research has been done to automatically determine relevant KPIs. These research efforts have yielded approaches that either require manual labor or focus on organizations instead of end-users. However, predictions that fit organizations might not fit end-users. To provide KPIs relevant to end-users, their roles in their organizations should be taken into account when predicting relevant KPIs. By taking the characteristics of roles into account when predicting relevant KPIs, predictions can be specified for information needs specific to a role. This specification can increase the relevance of KPIs to end-users, improving their decision making. There is currently no approach available for automatically predicting relevant KPIs for roles in organizations. This study shows that the relevance of KPIs can be predicted based on role characteristics. The relevance of a KPI to a role is shown to be determined by the role's responsibilities and its place in the organizational structure. The approach presented in this study was applied in a real-life setting and evaluated with experts. The results show that role characteristics can be used to predict relevant KPIs for roles in organizations. This research presents an approach that is able to specify the prediction of relevant KPIs from organizations to roles while decreasing manual effort required.

Acknowledgements

I would like to thank my supervisors Rob Audenaerde, Ünal Aksu, Georg Krempl, and Hajo Reijers, for their help during this research. Rob, for encouraging me to write everything down and keep the final goal in sight, Ünal for the countless tips, advice, and expertise, and Georg and Hajo for their revisions and feedback. Second, I would like to thank ValueCare for supporting my research. Finally, a big thanks to my friends and family, and Anouk in particular, for their help.

Contents

List of Figures

List of Tables

List of Abbreviations

ARHR	**A**verage **R**eciprocal **H**it **R**ate
AZW	**A**rbeid, **Z**org en **W**elzijn (Labor, Care, and Welfare)
BI	**B**usiness **I**ntelligence
BRC-2014	**B**e**R**oepen **C**lassificatie 2014 (Dutch Classification of Occupations 2014)
CBS	**C**entraal **B**ureau voor de **S**tatistiek (Dutch Central Bureau of Statistics)
CEO	**C**hief **E**xecutive **O**fficer
CFO	**C**hief **F**inancial **O**fficer
CHR	**C**umulative **H**it **R**ate
CIO	**C**hief **I**nformation **O**fficer
CRISP-DM	**CR**oss-**I**ndustry **S**tandard **P**rocess for **D**ata **M**ining
DBC	**D**iagnose **B**ehandel **C**ombinatie (Diagnosis Treatment Combination)
GGZ	**G**eestelijke **G**ezondheids**Z**org (Mental Healthcare)
HIS	**H**ospital **I**nformation **S**ystem
HR	**H**it **R**ate
ISCO-2008	**I**nternational **S**tandard **C**lassification of **O**ccupations 2008
KPI	**K**ey **P**erformance **I**ndicator
KRD	**K**PI **R**elevance **D**ata
LOOCV	**L**eave-**O**ne-**O**ut **C**ross-**V**alidation
MAE	**M**ean **A**verage **E**rror
RMSE	**R**oot **M**ean **S**quared **E**rror
SVD	**S**ingular **V**alue **D**ecomposition

Chapter 1

Introduction

Business Process Monitoring aims to capture the performance of business processes using performance indicators [52]. It is a very common practice in many organizations. Organizations use Business Process Monitoring to increase the availability of information and understanding of ongoing business processes. [32]. Using information gained from monitoring business processes, organizations can react more quickly and accurately than they could have done without monitoring their business processes. Key performance indicators (KPIs) are often used to keep track of business process performance [44]. KPIs are defined as metrics that embed performance targets so that organizations can chart progress towards goals [8]. A standard method of displaying KPIs is by combining them in a dashboard.

Dashboards consolidate key performance indicators into a single visual display that offers an at-a-glance window into overall business performance [8]. The information in dashboards is usually visualized in a figure or a graph. Shaping the contained information in a figure or graph catches user attention, enabling decision making [7]. The visualization into a figure or graph not only grabs user attention but also simplifies interpreting the underlying data. Organizations need dashboards to supply their employees with the information they need to carry out their daily tasks. Providing relevant information to employees supports their decision making, improving its efficiency and quality.

Organizations obtain KPIs and dashboards either by developing them in-house or by outsourcing the production to specialized software vendors [7]. Such vendors typically offer predetermined sets of commonly used KPIs or create custom-made KPIs for their customers [44]. Finding relevant KPIs is a very time-consuming process, which often requires considerable human effort [6]. To support decision making, the content of a dashboard should relate to the decision to be made [22]. For instance, a dashboard on patient health will not assist in making financial decisions. The effectiveness of a dashboard can be hindered by providing irrelevant, or wrongly scoped, KPIs [22]. Thus, decision making can be supported more effectively by providing relevant KPIs. Many studies [6, 7, 26, 28, 40, 46] on determining relevant KPIs have been conducted. Generally, these studies either provide a template that can be customized for each situation or create KPIs from scratch, which requires considerable effort [6].

Moreover, in the studies that have been conducted, the focus lies on finding relevant KPIs for organizations. The scope of organizations might be too broad for employees that need to examine data on a more detailed level, however. For example, consider a dashboard on monthly revenue over an entire organization. A chief financial officer (CFO) is tasked with overseeing the organization's global financial trends, and the level of detail in the dashboard might suit him perfectly. However, to a division manager, the dashboard might already be too general to provide any detailed information on his division. Finally, an employee at the 'ground-level' can get no actionable information from the dashboard because its scope is much too broad for his daily activities. If the data specified until it is relevant to the employee, the data is too detailed for the CFO to see a picture of the entire enterprise. Even though all three employees need financial information, one size does not fit all. Currently, available approaches do not take the requirements of roles into account when predicting relevant KPIs to them.

1.1 Problem Statement

Contemporary literature on the prediction of relevant KPIs, or relevant KPI prediction, does not provide a method focusing on predicting relevant KPIs for roles in organizations. Many studies have been conducted on selecting relevant KPIs [6, 7, 26, 28, 40, 46], but most of these studies focus on organizations. Studies that do focus on selecting relevant KPIs for roles in organizations have produced approaches that require significant manual effort to implement. There are no available studies that both focus on selecting relevant KPIs for roles in organizations and reducing the effort needed during the selection of relevant KPIs. Aksu et al. [6] predicted the relevance of KPIs to organizations based on organizational characteristics. These characteristics include domain, location, and number of employees. By training prediction models using these characteristics as features and KPI relevance as the outcome, Aksu et al. [6] were able to predict relevant KPIs for organizations. The same approach can be taken towards the prediction of relevant KPIs for roles. However, roles and organizations are very different entities. Therefore, the characteristics that determine the relevance of KPIs to roles should be researched. By using role characteristics as a predictor of KPI relevance, roles can be taken into account when predicting relevant KPIs. By predicting relevant KPIs based on user needs or preferences, the resulting dashboards that contain them will be more relevant, making them more useful [21, 22]. There is a knowledge gap in the literature concerning the prediction of relevant KPIs for roles in organizations and the characteristics that make KPIs relevant to roles. This research, therefore, adopts the following main research question:

RQ: *How can relevant KPIs be predicted for roles in organizations?*

1.2 Research Goal

The goal of this research is to develop an approach that enables the prediction of relevant KPIs for roles in organizations based on role characteristics. The contribution of this research is twofold. First, this research develops an approach for predicting relevant KPIs for roles in organizations. Second, this research creates a technique for eliciting, comparing, and grouping roles in organizations. This section will first discuss the prediction of relevant KPIs for roles in organizations. Then, the goal of finding and comparing roles in organizations will be addressed. Finally, these two contributions will be combined to show how this research will produce an approach for predicting relevant KPIs for roles in organizations.

This research will use an approach developed by Aksu et al. [6] as a foundation to build upon. The use of organizational characteristics as features can also be applied to roles in organizations. In the case of roles, the set of defining characteristics is different. The next paragraphs will discuss the nature and characteristics of roles in organizations.

Organizations are groups of people working towards the same goal [15]. The people in these organizations have specific sets of responsibilities that contribute to reaching that common goal. Roles are grouping mechanisms for employees with the same responsibilities. In short, employees have roles, which are collections of responsibilities. These responsibilities determine the tasks employees must perform in the course of their work. In this sense, roles are defined by their responsibilities [15, 48, 47]. Since dashboards and KPIs exist to assist in the execution of daily tasks in organizations, responsibilities may determine which KPIs are relevant to a role. For example, the responsibilities of a nurse are related to taking care of patients. To support the daily activities of the nurse, KPIs should show information relating to the status of patients. Information on patient health can assist in making decisions that improve the patient's recovery, which is one of the nurse's responsibilities. If the KPIs showed information about finance, they would be of no use. This example shows that KPIs are only relevant when they can be used to support the tasks and responsibilities related to a role. Therefore, responsibilities are assumed to determine the relevance of KPIs to roles.

In addition to responsibilities, employees can be grouped by their place in the organizational structure. Organizational structure refers to an enduring configuration of tasks and activities [60]. This means the long-term form of an organization. There are multiple levels within organizational structures. Mintzberg developed a theory on the structure of organizations [39]. He proposed that all organizations consist of five distinct organizational levels: strategic apex, middle-line, operative core, technostructure, and support staff. A visualization of this structure can be found in Figure 1.1.

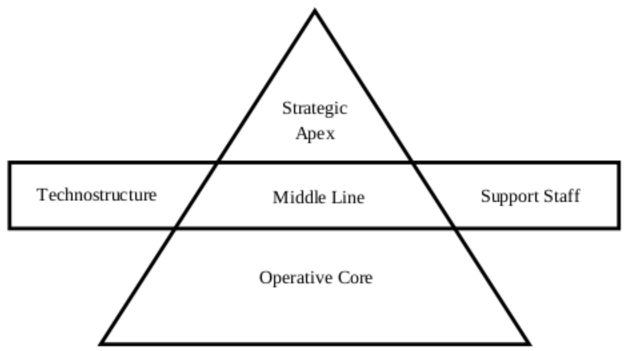

FIGURE 1.1: Mintzberg's model of organizational structure

Mintzberg's theory suggests a chain of command, or a sense of verticality, in each organization. The operational core is managed by and reports to the middle line, which reports to the strategic apex. The technostructure and the support staff are dispersed throughout the different levels of the organization. This verticality necessitates the development of KPIs and metrics at various organizational levels. An employee in the strategic apex has other responsibilities than an employee in the operative core. The strategic apex is not concerned with detailed data that enables proper functioning in the operative core. Reversely, the highly aggregated data needed for the entire organization's decisions are too broad to support the operative core's daily activities.

Therefore, the organizational level might determine which KPIs are relevant to an employee along with responsibilities.

This research aims to create a method for the prediction of relevant KPIs for roles in organizations. It will do so by using role characteristics as predictors for the relevance of KPIs. Before the relevance of KPIs to roles can be determined, the existing roles in organizations should be found, compared, and grouped. These goals pose the following challenges:

- Finding roles in organizations
- Comparing and grouping users with similar roles
- Finding role characteristics
- Predicting KPI relevance based on role characteristics

1.3 Thesis Outline

The remainder of this thesis is structured as follows. The following chapter, Chapter 2, will discuss the current literature related to the prediction of relevant KPIs for roles in organizations. Chapter 3 will provide the background information required for understanding the prediction models used in this research. Chapter 4 will introduce the employed research methods. Thereafter, Chapter 5 will provide an approach for the prediction of relevant KPIs for roles in organizations. The evaluation of the approach will be presented in Chapter 6. The results of the evaluation will be discussed in Chapter 7. Chapter 8 will conclude this thesis.

Chapter 2

Literature Review

In this section, previous works relating to the prediction of relevant KPIs will be discussed. The literature on roles in organizations, organizational structure, role-specific KPIs, and role characteristics will be discussed.

2.1 Organizational Structure

Organizational structure refers to an enduring configuration of tasks and activities [60]. Quite literally, this means the long-term form of an organization. There are multiple levels in organizational structures. Mintzberg developed a theory on the structure of organizations [39]. He proposed that all organizations consist of five distinct organizational levels: strategic apex, middle-line, operative core, technostructure, and support staff. A visualization of this structure can be found in Figure 1.1. The strategic apex contains employees that concern themselves with the strategic management of their organization. Middle-line management concerns itself with the management of practical operations while the operative core carries them out. The technostructure contains all employees that analyze data, like analysts or accountants. Finally, the support staff provides all services that are not essential to their organization's core activities, like human resource management, legal counsel, or maintenance [37].

In addition to the five organizational levels, Mintzberg describes two dimensions for organizations: coordinating mechanism and degree of centralization. Coordinating mechanism describes how organizations control their employees. Centralization describes the distribution of power within the organization. [37] Using organizational levels, coordination method, and degree of centralization, Mintzberg defines five standard organizational configurations. These configurations determine which organizational level is the organization's focus and how it is organized internally.

Mintzberg's theory suggests a chain of command, or a sense of verticality, in each organization. The operational core is managed by and reports to the middle line, which reports to the strategic apex. The technostructure and the support staff are dispersed throughout the different levels of the organization. This verticality necessitates the development of KPIs and metrics at different organizational levels. An employee in the strategic apex has different responsibilities than an employee in the operative core. The strategic apex is not concerned with detailed data that enables proper functioning in the operative core. Reversely, the highly aggregated data needed for the entire organization's decisions is too broad to support the daily activities of the operative core. Therefore, the organizational levels are adopted as a relevant factor for predicting relevant KPIs because they directly influence responsibilities.

One of the advantages of Mintzbergs's theory is the simple categorization of organizational structure it provides. An employee can fall under one of five categories, which is very easy to represent as a machine learning feature. However, there has been some critique on Mintzberg's theorem. It has been argued that Mintzbergs' theory does not fit with modern 'self-managing' organizations that provide autonomy to their employees. These organizations do not exhibit the verticality that is prominent in Mintzbergs' model. Critics argue that Mintbergs' ideas were interesting at the time of their conception but that they cannot cope with the structure of more 'modern' organizations [38, 43, 20]. However, it is expected that most organizations will still have a structure that fits with Mintbergs' theory. If Mintzbergs' theory does not prove a useful feature, it might be wise to see whether the examined organization fits the structure assumed by the theory.

2.1.1 Roles in Organizations

Organizations are groups of people working towards the same goal [15]. This definition can be applied to an international corporation and something on a smaller scale, like a sports team. In the international corporation, the goal is to create profit, whereas the sports team works toward a good performance on the pitch. In both cases, the people within the organization have specific tasks to achieve their goals. These specific tasks are also called 'jobs' or 'roles'.

There are two scientific fields that deal with the concepts of roles within organizations: organizational structure and resource-aware business process modeling. Both provide similar definitions of what a role is, with some specifics relating to the field's subject. In both fields, employees are engaged in organizations and work towards common goals. Roles are grouping mechanisms for employees, or resources, with the same responsibilities. In this sense, roles are defined by their responsibilities.

The organizational structure field defines roles as the behavior and responsibilities that come with the appointment of specific tasks within an organization [15]. Organizations are defined as groups of people working towards a common goal. There can be smaller organizations within organizations, like an accounting department within an enterprise.

The field of resource-aware business process modeling defines roles with domain-specific terms. Here, tasks are completed by resources. These resources can be either human or robotic. Human resources are grouped into organizations [48]. They have positions and possibly privileges that come with the position. They can also be part of organizational units, which are permanent groups within the organization. [47] Resources may have one or more associated roles. These roles are groups of resources with the same responsibilities [48]. Thus, people with similar roles carry out similar tasks.

In short, employees have roles, which are collections of responsibilities. These responsibilities are related to the tasks they must perform in the course of their work. Dashboards and KPIs exist to assist in making decisions related to performing daily tasks. These tasks are determined by the responsibilities that are associated with a role. Therefore, the responsibilities of a role may determine which KPIs are relevant for supporting decision-making.

2.2 Role-Specific Key Performance Indicators

Much research has been done on providing relevant KPIs [6, 7, 26, 28, 40, 46]. These studies either focus on relevant KPIs for organizations or require considerable manual effort to implement. Studies that focus on KPIs for organizations provide KPIs that are likely too broad to apply their individual employees. KPIs that are relevant to an organization are not useful to everyone within that organization. Approaches that require manual effort either create a large repository of available KPIs and match those with the needs of organizations or create KPIs from scratch [6]. Both the matching and creation of relevant KPIs require effort, which increases cost. There are no approaches available that both focus on relevant KPIs for roles in organizations and do not require a lot of effort to implement. However, there is an approach available that automatically provides relevant KPIs for organizations [6]. Aksu et al. [6] predicted the relevance of KPIs to organizations based on the characteristics of the examined organizations. These characteristics include domain, location, and the number of employees. By training prediction models using these characteristics as features and KPI relevance as the outcome, Aksu et al. were able to predict relevant KPIs for organizations.

There is a gap in the current literature regarding the prediction of relevant KPIs for roles in organizations. The study by Aksu et al. [6] will be extended to use roles. The concept of using organizational characteristics as features for a prediction model will be used to create role characteristics. These role characteristics will consist of role responsibilities and organizational level. The found gap in the literature can be filled by developing a method for the prediction of relevant KPIs for roles in organizations. The following chapter will discuss literature that is not directly related to the topic of relevant KPI prediction, but is required for a thorough understanding of this thesis.

Chapter 3

Theoretical Background

This chapter will provide the required background knowledge on recommender systems needed for a thorough understanding of this report. This research aims to predict relevant KPIs for roles in organizations. Recommender systems were chosen as the preferred prediction model due to the available data during the implementation of the proposed approach. The following sections will discuss recommender systems. First, a general introduction to recommender systems will be provided. After that, different kinds of recommender systems will be discussed, along with their advantages and disadvantages. Finally, methods for evaluating recommender systems will be addressed.

3.1 Recommender Systems

Recommender systems are widely used in modern business applications. Examples of recommender systems being used are streaming companies like Netflix or Amazon Prime. These companies recommend new movies or series to their users based on their users' past viewing behavior [13, 49]. The concept of using past viewing behavior to predict items for users can also be applied to KPIs. Relevant KPIs for roles can be predicted using records of the KPI-viewing behaviors of roles. Recommender systems can use this data to predict KPIs that are relevant to roles in organizations. There are different kinds of recommender systems that require different kinds of data. The rest of this section will discuss the different types of recommender systems.

Recommender systems come in all shapes and sizes. However, some general categories can be found. Recommender systems are usually either item-based or user-based. There is also a difference between memory-based and model-based recommender systems [41]. As their names suggest, item-based and user-based recommender systems respectively make recommendations based on item characteristics or user characteristics. User-based recommender systems are also called collaborative filtering algorithms. Item-based recommender systems use the similarities between items to identify other items to recommend [34]. This approach requires information on items. For movie recommendations, this information could include genre, year, director, cast, or reviews by a particular critic. The item-based recommender system looks which other movie is most similar to the watched movie and recommends it [34].

User-based recommender systems make their recommendations based on user behavior. Usually, the needed data is a collection of positive indications of interest in a particular product or service [54]. Users can express this interest either implicitly or explicitly. Implicit feedback data is created by recording clicks or views. Explicit feedback, like ratings or reviews, has been consciously left by a customer. Explicit feedback inherently contains more useful information because it shows a direct opinion. A bad rating clearly shows disinterest, while a click is difficult to interpret without context [36]. Some caution should be taken when exclusively working with implicit feedback because of the uncertainty regarding its context. For instance, if a website is openly available, traffic could be generated by bots or web-crawlers. These crawlers and bots can intentionally or unintentionally create associations between unrelated items [30].

Model-based and memory-based recommender systems differ in the way in which they are computed. Memory-based recommender systems construct matrices containing similarities like Pearson correlation or Cosine similarity between items or users and make their recommendations based on those similarities. They are called memory-based because these matrices are kept in working memory at all times. Keeping all data in memory makes them computationally intensive but also able to adapt to changes in the data [4]. Model-based recommender systems make use of machine learning models to construct a model that can make recommendations. Summarizing relations in the data in a model makes them much less computationally intensive but also makes them unable to adapt to changes in the data [4]. Generally speaking, memory-based models produce a higher accuracy but are less scalable than model-based systems due to their memory requirements.

One more distinction can be made between recommender systems. There is a difference between predicting the relevance of all items and predicting the top-N items most likely to interest the user. [19]. When relevance is predicted for all available items, there are different criteria than when only the top 10 most interesting items need to be recommended. For instance, the recommendation for all available items requires a minimal error across all predictions, whereas a top-N prediction only needs to minimize the error on the top-N predictions. The goal of this study is to predict the KPIs that are relevant to roles in organizations. In this research, the end-user is only interested in relevant KPIs. Therefore, the recommendation task to be completed is a top-N recommendation task.

3.1.1 Evaluating Recommender Systems

Recommender systems need to be evaluated after their construction to verify whether they produce relevant predictions. Different recommender systems require different kinds of data, and not all data is equally fit for use in recommender systems. Therefore, the predictions made by a recommender system should be tested to assess their quality. Recommender systems are mainly evaluated using metrics specific to the field of recommender systems [18]. Accuracy is often measured with the Mean Average Error (MAE) [14] or the Root Mean Squared Error (RMSE) [14]. Additionally there are Hit-Rate (HR) [19], Cumulative Hit Rate (CHR) [19], Average Reciprocal Hit Rate (ARHR) [19], coverage [14], diversity [16], and novelty [16]. Table 3.1 provides an overview of these metrics and their purpose.

MAE and RMSE both measure the overall accuracy of predictions made by the recommender system. RMSE punishes large errors more severely than MAE. Therefore RMSE should be used when large errors are undesirable, while MAE should be used to minimize overall error.

Metric	Abbreviation	Purpose
Mean Average Error	MAE	Accuracy of predictions
Root Mean Squared Error	RMSE	Accuracy of predictions
Hit-Rate	HR	Performance on top-N prediction tasks
Cumulative Hit-Rate	CHR	Performance on top-N prediction tasks
Average Reciprocal Hit-Rate	ARHR	Performance on top-N prediction tasks
Coverage	-	Percentage of users the model can predict for
Diversity	-	Average similarity of recommended items
Novelty	-	Average relevance of recommended items

Table 3.1: Overview of metrics for recommender systems

Hit-Rate is a measure for performance on top-N prediction tasks. It is calculated by using Leave-One-Out Cross-Validation (LOOCV). While training a recommender system, one of the items a user has watched is intentionally left out [18]. The results are then checked on whether the recommender system has predicted the omitted item. If it has, this is called a 'hit.' Hit rate is a percentage gained by dividing 'hit' predictions by the total amount of predictions [19]. Cumulative Hit-Rate only counts hits above a certain rank. Thus, CHR shows the number of hits that end up in the top-N recommendations [42]. ARHR puts more focus on the top recommendations because people often focus on the top items in a list of recommendations. Instead of just registering a hit, ARHR takes the reciprocal of the rank of a prediction. This means that it favors predictions with a higher popularity rank [19].

Coverage shows the percentage of users for which a top-N recommendation could be made [29, 14]. Diversity is the average similarity between all items recommended for a user [16, 29]. Novelty shows the average rank of recommended items. The higher the novelty, the more 'obscure' the recommendations are [16, 29].

Recommender systems are designed to predict interesting items based on behavior. This makes them very suitable for predicting relevant KPIs based on role characteristics. By identifying the KPIs that are found relevant by similar users, relevant KPIs can be predicted for other users. During the conducted case study, recommender systems were used to evaluate the approach for predicting relevant KPIs for roles in organizations in a real-life situation. These will be discussed in Chapter 6. The following chapter will provide the details of the research methods used for the creation of the approach aimed at the prediction of relevant KPIs for roles in organizations.

Chapter 4

Research Approach

This chapter contains the method for creating an approach for the prediction of relevant KPIs for roles in organizations. First, additional research questions will be motivated and presented. Thereafter, employed research methods are shown.

4.1 Research Questions

The main research question can be broken down into several, more specific, subquestions. In this section, sub-questions will be defined that specify different aspects of the main research question.

Based on recent literature about KPI relevance prediction [6], it is expected that there are characteristics of roles that determine whether a KPI is relevant. The literature on roles in organizations and organization structure has yielded that role responsibilities and organizational level are potential characteristics that determine the relevance of KPIs to roles [22, 37, 39, 48, 47]. When characteristics that influence the relevance of a KPI to a role are found, they can be used as features to predict relevant KPIs for roles in organizations. The relation between KPI relevance and role characteristics should be researched. Therefore, the first sub-question is formulated as follows:

How is the relevance of a KPI to a role related to its role characteristics?

Before the relevance of a KPI to a role can be predicted based on role characteristics, the characteristics themselves will have to be found. There is currently no method available for finding the responsibilities associated with roles in organizations. Organizations may keep a list of all employees and their roles somewhere in their administration. However, it is unlikely that this list also includes the details of the responsibilities in their daily operations. Therefore, it is necessary to investigate how the data on roles and responsibilities can be obtained. The second research question is:

How can roles and their responsibilities be collected from an organization?

It is expected that the responses given by the employees about their roles are extremely varied. Many roles and even more responsibilities can be thought of. It is also expected that each organization has a slightly different definition of the same role or its responsibilities. To be able to group these responses in some role, their similarity should be assessed. There is currently no specific metric for finding the similarity between roles. This metric should be defined to be able to group roles with similar responsibilities. The final sub-question is therefore formulated as follows:

How can the similarity between roles be determined?

4.2 Research Methods and Techniques

This section will discuss the research methods and techniques employed in this research. First, a literature study will be discussed. Thereafter, the research method used for designing an approach for predicting relevant KPIs for roles in organizations will be provided.

4.2.1 Literature Study

To get a firm grasp on the literature regarding relevant KPI prediction for roles in organizations, a literature study will be conducted. Several subjects will be examined in the literature study: prediction and selection of relevant KPIs for organizations and roles, structure of organizations, roles in organizations, and recommender sytems will be researched. Table 4.1 contains an overview of the articles that were found, checked and read during the literature study.

This research uses an approach created by Aksu et al. [6] as a foundation to build upon. Therefore, the most logical point for starting the literature is their study. Forward and backward snowballing techniques [57, 55] were used to locate the relevant publications on prediction and selection of relevant KPIs for organizations and roles. These techniques cover each other's weaknesses. It is only possible to find studies older than the study used as the starting point in backward snowballing. When snowballing forward, only newer studies can be found. By combining the methods, both older and newer relevant studies can be found, which creates an overview of all related articles. The study by Aksu et al. [6] contains many references to other studies in KPI relevance prediction. Since it is a relatively recent publication (2019), backward snowballing was mostly applied. The snowballing process on relevant KPI prediction was stopped after the snowballing yielded no new relevant studies. 'Relevant KPI Prediction,' 'Automated Prediction of KPIs,' 'KPI Relevance,' and 'KPI prediction' are the keywords that were are used in the search process.

Backward and forward snowballing were also be applied to studies in the field of recommender systems. The keywords used for finding relevant studies in this research field were: 'recommender systems,' 'collaborative filtering,' 'user-based collaborative filtering,' 'item-based collaborative filtering,' 'memory-based collaborative filtering,' 'model-based collaborative filtering,' 'k-nearest neighbors in collaborative filtering,' and 'singular value decomposition in recommender systems.' In some cases, recommender systems and collaborative filtering can be used interchangeably. Therefore, each keyword containing 'collaborative filtering' was also used with 'recommender systems.'

Each study was analyzed using the following steps. First, the abstract and keywords were read to obtain an overview of the contents of the study. If the keywords or abstract contained terms relevant to the research, the article was read more thoroughly. If a piece of the article was found relevant for this research, it was analyzed for information directly usable in the research or references that might point to other relevant articles. There were two stopping criteria when searching for the answer to a question in the available literature. First, the search was stopped if no new articles could be found and the second stopping criterion was not yet fulfilled. Second, the search was stopped when more than one article answered the defined question. If an article that answered the question was found, more articles that supported its claim were searched

Subject	Checked	Read	Used
KPI Relevance	27	20	12
Roles	21	10	4
Organizational Structure	18	12	5
Recommender Systems	40	30	16

TABLE 4.1: Counts of articles checked on research subjects

for to ensure that the answer was supported by multiple sources.

Studies were found using Google Scholar as search engine, which offers a list of related articles and a list of articles that cited a study. This functionality makes it ideal for both forward and backward snowballing.

4.2.2 CRISP-DM

This section will discuss the research method employed to design an approach for predicting relevant KPIs for roles in organizations. The prediction of relevant KPIs for roles in organizations is a data mining problem. The Cross-Industry Standard Process Model for Data Mining (CRISP-DM) has been specifically designed to provide a framework for data mining projects. It is designed to improve the reliability, replicability, and speed of data mining projects [56]. As it is specially tailored to data mining projects, it has a good fit with the goal of this research. Thus, CRISP-DM is expected to provide a solid basis that can be used to structure the creation of an approach for predicting relevant KPIs for roles in organizations.

Figure 4.1 contains an overview of the phases of the CRISP-DM model. The CRISP-DM model is a cycle of six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. CRISP-DM is an iterative model, which means that after phase 5, evaluation, it will be decided whether the results are satisfactory. If the results are accepted, the model moves on to the final phase, deployment. If the results are not satisfactory, the cycle begins anew. The phases of the model are executed again with the knowledge of the previous iterations. Each phase creates an output that is an input for the next. In addition to the feedback loop that is built into CRISP-DM, some phases contain overlap. There is some 'back-and-forth' between business understanding and data understanding and between data preparation and modeling. When business understanding grows, so might the understanding of available data. Reversely, insights gained in the data understanding phase might add to an understanding of the business. When preparing data for the modeling phase, it is important to know which model will be employed because different models take different inputs. If a modeling technique is not as successful as expected, this might require some more data transformation [56].

The remainder of this section will describe the phases of CRISP-DM in further detail. In particular, each step and how it will be applied in this research will be elaborated.

Business Understanding

The first phase of CRISP-DM is focused on creating an understanding of the business goals related to the problem that is to be solved. More specifically, requirements should be gathered to convert the problem into a data mining problem definition [56].

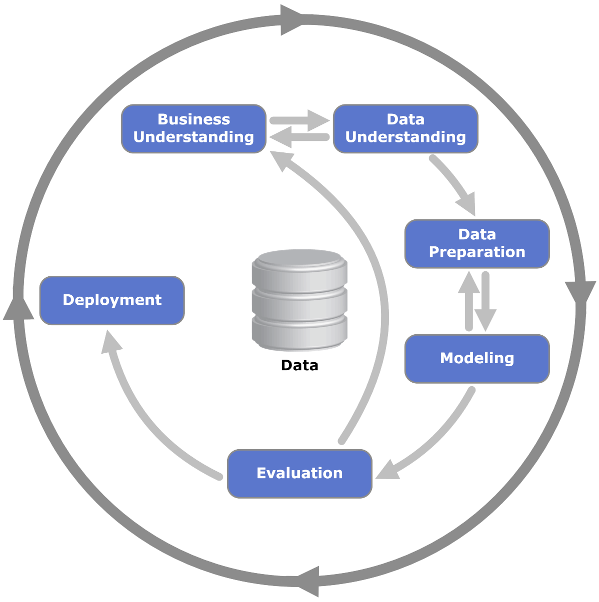

Figure 4.1: Phases of the CRISP-DM model

Unless this phase is entered in an iteration later than the first, it takes no inputs. If the result of the first CRISP-DM iteration has been found lacking, then this step takes the lessons learned from previous iterations as input. The output is the understanding of the problem as a data mining problem.

In this study, the goal of the business understanding phase is to convert the identified gap in the literature, the prediction of relevant KPIs for roles in organizations, into a data mining problem. This entails gathering the requirements of a solution and designing an approach that fulfills those requirements. These steps have largely been discussed already in Chapter 1. The research goal has been transformed into a data mining problem by taking an approach by Aksu et al. [6] as a starting point and then extending it to use roles. By taking inspiration from an existing approach, the following steps in the CRISP-DM cycle can focus on extending the approach to predict relevant KPIs for roles. This allows the data understanding phase to focus on the collection of role characteristics. Moreover, using an existing approach provides a supporting foundation for executing each step in the CRISP-DM cycle. The adaptation of the approach by Aksu et al. [6] will be discussed in more detail in Chapter 5.

Data Understanding

Data understanding is closely linked to business understanding. Some knowledge of the available data is required to know the requirements and goals. The data understanding phase is meant for the collection, exploration, and preparation of data. This phase can already produce some preliminary insights into the data. After the collection of data, the data should be subjected to a preliminary analysis. With this analysis, the size, structure, and possible problems of the data can be discovered [56].

In this study, the data understanding phase will focus on the collection data on roles in organizations and KPI relevance. The following should be assessed: which data on roles is already available, which data is not available, and how the unavailable data should be collected. More specifically, methods for the collection of role data and organizational levels should be constructed and applied. Thereafter, the data should be analysed to get acquainted with it and to spot possible difficulties that it might bring. This phase takes the understanding of the research problem in the form of a data mining gained from the business understanding phase as input.

Two sets of data need to be collected in this phase. First, data on roles and their characteristics should be collected. This data can be used to compare and group roles in the data preparation phase. Second, data on KPI relevance should be obtained. This data should relate the relevance of KPIs to roles in organizations.

Required Data

Role characteristics will be used to compare roles to each other and group them based on similarities. It has been established that role responsibilities and organizational level are the characteristics that possibly determine whether certain KPIs are relevant to a role [15, 39]. Therefore, these characteristics should be collected. In addition to the established characteristics, role names and role descriptions should also be collected. The addition of role names and descriptions provides an extra point of reference for whether two roles are similar. If two roles have similar descriptions and names, it might be a signal that they are the same role.

When all data mentioned above is combined, each role will have the following information: role name, role description, role responsibilities, and organizational level. It is unlikely that all this information is readily available. It is possible that an organization keeps track of the role names of its employees and organizational level might be inferred from the role names. However, role descriptions and role responsibilities are not likely to be documented by organizations. Therefore, role descriptions and role responsibilities will need to be collected from end-users.

Data Collection

For the purposes of this research, there are a few challenges on data collection. In general, prediction models require a large amount of data to produce accurate results. Moreover, role characteristics are specific to each role. The most knowledgeable person on the characteristics of a role is the person who carries it out. Therefore, asking the employee that carries out a specific role will probably yield the most accurate information on role name, role description, role responsibilities and organizational level. Due to the need for a large dataset to enable accurate predictions, it is expected that a standardized survey is the best method of collecting this data. Another method for collecting this data would be to interview employees. However, this method is expected to be too time consuming. A survey can be sent to many employees simultaneously, which enables a more efficient method of gathering the required data.

In the survey, respondents will be asked to give their role name, a description of their role, list a description of their responsibilities and indicate their organizational level. Role name, role description and role responsibilities will be answered in an open text box, allowing respondents to enter their own descriptions and responsibilities. Therefore, this data will be collected in textual form. The organizational level can be collected by asking the respondent to choose one of Mintzberg's five organizational levels.

In addition to the data on role characteristics, data on the relevance of KPIs to roles should be collected. Some score for the relevance of KPIs to roles is needed to train a prediction model. This score will be called the KPI relevance score. Together with the role characteristics, KPI relevance score makes up KPI Relevance Data (KRD). The KPI relevance score should consist of a KPI, a role, and the relevance of the KPI to that role. This data should be collected for as many combinations of roles and KPIs as possible. The relevance of KPIs to roles can be collected in two ways. First, it is possible to ask the employees their most used KPIs in the aforementioned survey. Asking for the most relevant KPIs will provide an overview of which KPIs are seen as relevant by which roles. However, this does not indicate which KPIs are seen as irrelevant by the employees. Another method of collecting KPI relevance scores is by using KPI-use logs. Aksu et al. [6] used KPI-use logs to develop a create a relevance score for the available combinations of roles and KPIs.

Data Preparation

The data preparation phase is concerned with transforming the raw data provided by the previous phase into data that can be used to train prediction models. This phase contains the selection, cleaning, and transformation of data. [56]

In the context of this study, the data preparation phase entails the transformation of the collected role characteristics into a machine-readable format. This means that all the role characteristics and KPI relevance scores need to be transformed into machine learning features. First, roles and responsibilities will need to be compared and grouped. For instance, ‘Financial Manager’ and ‘Manager Finance’ are very similar. If their responsibilities are also similar, then they probably describe the same role. Grouping roles and responsibilities prevents the creation of very similar features. Since the role responsibilities, role names, and role descriptions are entered as free text, there can be large differences in word use, abbreviations, and writing style. By standardizing these titles and responsibilities in some way, they can be made comparable to each other. After the roles and responsibilities have been compared and grouped, they will be transformed into features.

Collected data on organizational level and KPI relevance to roles are likely not textual. There are five organizational levels, so it makes sense that it should be transformed into a categorical feature. The KPI relevance scores will be transformed into a numerical feature. Aksu et al. [6] transformed KPI relevance scores into a Likert scale item ranging from one to five. Scaling KPI relevance scores to the Likert scale would likely require some normalization of the data to fit it into this range.

Modeling

In the modeling phase, prediction models are constructed and applied to the data. The most effective methods of analysis are found, and their parameters tuned [56]. Several prediction models will be constructed and evaluated with the data created in the data preparation phase. At the end of this phase, a prediction model should be selected for each KPI to predict relevant KPIs for roles in organizations. During the modeling phase, different models will be tested on their performance. The output of the best performing model will be used as the predictions for evaluation.

Evaluation

Evaluating the approach is vital to ensure the success of the process. Several factors will be evaluated. It should be assessed whether there are business issues that are not being addressed and whether the approach meets its goals [56]. It should, for instance, be evaluated whether the approach produces relevant results. If the approach does not produce usable results, then its steps should be re-evaluated.

Case Study

The approach will be evaluated by conducting a case study wherein the approach will be applied in a real-life setting. The case study organization will be selected on a few criteria. First, the case study organization should be a provider or creator of KPIs. Second, it should be willing to support a research and allow for contact with its customers. Finally, the case study organization should have as much information available on the use and characteristics of their KPIs as possible, along with some data about their users. Using these criteria has several advantages. Selecting a provider or creator of KPIs ensures that there is some expertise available on selecting relevant KPIs, as that is the main business of such an organization. The contact with customers should benefit the selection of customers for the evaluation of results with end users. Finally, the collection of KPI and user data should make the collection of data much easier and less time-consuming. Selecting an organization that fits these criteria will ensure a suitable environment for the evaluation of an approach for the prediction of relevant KPIs for roles in organizations.

There are many advantages to employing a case study as a method of evaluation. One of these benefits is discovering omitted or missed variables that could be at play in the approach [12]. Since role characteristics have not been collected before, there could be aspects to it that were not identified by analyzing the literature. By implementing the approach, the missing aspects of role characteristics could be found and added to the approach. Moreover, the approach requires data on real employees with roles and their characteristics. The approach aims to recommend relevant KPIs for roles in organizations. The relevance of KPIs to a role can be evaluated best by asking the employees that have the roles that the approach predicts relevance for. The evaluation of predictions made by the approach can be done by asking employees about the predictions in semi-structured interviews.

End-User Interviews

These end-user interviews will verify whether the user role, role description, and responsibilities are correct according to the interviewee. With these questions, the grouping of responsibilities and roles can be tested. If the employee is assigned incorrect responsibilities or an incorrect role, the prediction model will most likely make wrong predictions. Thereafter, the interviewee will be asked about the relevance of KPIs that were recommended by the prediction models. The results of these questions can be used to test whether the approach produces practically usable results. The use of semi-structured interviews allows for a natural flow of conversation while providing support for the interviewer during the interview.

The end-user should evaluate the applicability of the produced results. If the model does not produce KPIs relevant to the end-user, it will not be usable in practice.

End-user validation can be used to test the quality of a software product without needing any knowledge of the product itself on the end user's part [35]. The practical usefulness model can be assessed by verifying whether the resulting KPIs are relevant to those who want to use them.

It is possible that some of the KPIs generated by the approach are useful to the end-user and that some are not. To create a metric for the approach's overall performance, the agreements and disagreements should be taken into account. The questions in Table 4.2 can all be answered with a *Yes* or *No*. The questions 'Is this relevant to your job?' and 'Would you use this KPI in the line of your work?' show whether the approach produces meaningful results. It can be assessed whether the approach predicts KPIs that are relevant to the end-user. The interview uses the descriptor 'job' instead of role because it is more commonly used in conversation.

The collected answers to the interview questions will be aggregated into percentages. These percentages will represent the amount of correct predictions over the total amount of predictions. For instance, if an interviewee rates 27 out of 30 predicted KPIs as relevant, then the percentage will be 90%. These percentages by themselves cannot be used to assess performance, as there are no other relevance percentages to compare then to. However, by combining these percentages with the opinions the interviewees express during the interview, it can be assessed whether the approach produces relevant results for end-users.

Number	Question	Expected Result
Part 1: Start of interview		
1	Do you agree with the job title?	Yes/No and Why
2	Do you agree with the job description?	Yes/No and Why
3	Do you agree with the shown responsibilities?	Yes/No and Why
Part 2: Repeats for each KPI		
4	Have you used this KPI before?	Yes/No and Why
5	Is this KPI relevant to your job?	Yes/No and Why
6	Would you use this KPI in your line of work?	Yes/No and Why

Table 4.2: Questions for end-user evaluation interview

Deployment

The final phase in the CRISP-DM method, deployment, is rather straightforward. If the approach has been successfully designed, created, and evaluated, it can deployed for use in a real-life situation [56]. This phase lies out of this research's scope and will, therefore, not be mentioned further.

This chapter has shown the research methods used to create an approach for recommending relevant KPIs for roles in organizations. The following chapter will show the approach for the prediction of relevant KPIs for roles in organizations.

Chapter 5

Approach: Predicting Relevant KPIs for Roles in Organizations

This chapter presents an approach for predicting relevant KPIs for roles in organizations. First, a general overview of the approach will be given. Then, each component in the approach will be discussed in more detail. Each component will be discussed regarding the CRISP-DM phases it contains, its required inputs and outputs, and the different steps within the component. After the relevant KPIs for roles have been predicted, the evaluation phase of the CRISP-DM model will take place. The evaluation of the approach will be discussed in the following chapter

Figure 5.1 provides a high-level overview of the approach for predicting relevant KPIs for roles in organizations. The approach uses data on role characteristics and predicts the relevance of KPIs for roles based on those characteristics. Data on characteristics and the relevance of KPIs to roles is collected and transformed into KPI Relevance Data (KRD) in the role selection component. KRD is then used to train a prediction model for each KPI. These prediction models can then predict the relevance of KPIs for roles based on role characteristics. The remainder of this section will discuss the inputs and outputs of each step in Figure 5.1. After that, role selection, prediction model development, and relevant KPI prediction components will be discussed in more detail.

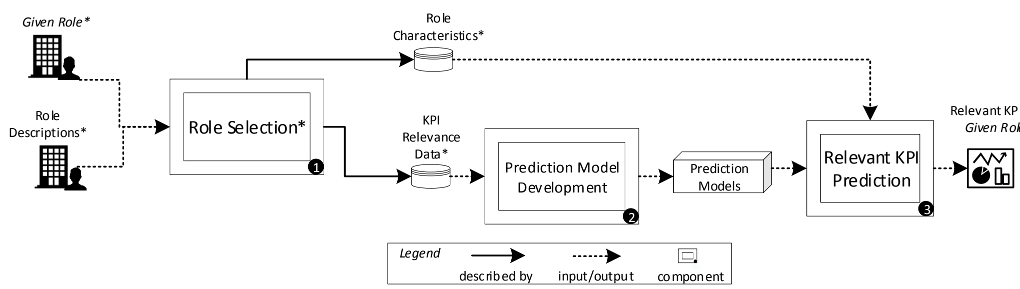

FIGURE 5.1: High level overview of proposed approach [1]

Role Selection transforms textual inputs from the collection of role data into a machine-readable format. It takes the descriptions of role names, role responsibilities, organizational level, and KPI relevance scores and transforms them into KPI Relevance Data. When predicting relevant KPIs for a new role, the characteristics of that role are transformed in this step to facilitate relevant KPI prediction.

[1]This figure is adapted from 'Automated Prediction of Relevant Key Performance Indicators for Organizations' by Aksu et al. [6]

Role Characteristics contain the values of role characteristics by which a role can be described (eg. responsibilities and organizational level) and the relevance of KPIs for roles.

KPI Relevance Data consists of two parts: (1) the relevance scores of a set of KPIs for a number of roles and (2) the role characteristics of those roles with their values. Table 5.1 contains an example of KPI Relevance Data. Each row of the KPI Relevance Data contains the combination of a role, KPI, organizational level (ol), responsibilities (r), and KPI relevance score for the KPI to the role.

KPI	Relevance	Role	ol1	...	ol5	r1	r2	...	r...
kpi01	2	Role 1	0	...	1	1	0	...	0
kpi01	1	Role 1	0	...	1	1	0	...	0
kpi01	4	Role 2	1	...	0	0	0	...	0
...	...	...	...	...	...	...	...	...	...
kpi ...	2	Role ...	0	...	0	0	1	...	1

TABLE 5.1: Example of KPI Relevance Data

Prediction Models are aimed at predicting relevant KPIs for roles in organizations. Each prediction model encodes a KPI, the role characteristics that make that KPI relevant to roles, and a prediction modeling technique that predicts the relevance of that KPI based on role characteristics.

5.1 Component 1: Role Selection

This section will discuss the transformation of role names, role responsibilities, organizational level, and KPI relevance scores to create KRD. After this transformation, the data can be used to train prediction models. First, the process of comparing role names and characteristics will be described. After that, the transformation of the data into usable features will be discussed.

The International Standard Classification of Occupations (ISCO-2008) [24] is a list of all occupations recognized by the United Nations. This standard contains descriptions of all roles occurring in specific occupational fields, along with the descriptions and responsibilities of those roles. By comparing role names, role descriptions, and role responsibilities from the survey with the roles provided in ISCO-2008, standardized roles, and responsibilities can be found. In this way, the roles found in the survey can be grouped into the standard role definitions with responsibilities defined by the ISCO-2008.

The roles and responsibilities can either be compared automatically to the ISCO-2008 standard with some text analysis method or be grouped by asking domain experts to compare the compared data with the ISCO-2008 standard. If enough data is available to train an algorithm that compares roles and responsibilities, this might be the most efficient option. After training the model, it can be reused without much extra effort, reducing the effort needed when replicating the steps of the approach. However, if there is no relevant data available to train a text analysis models, then interviewing experts on roles in the examined occupational field will yield more accurate results. The method that should be employed will depend on the available resources during the implementation of the method.

To create KRD, data obtained in during data collection should be transformed into features. Textual data is generally not used as a feature in prediction models; therefore, this textual data should be converted into numerical form. When transforming textual responsibilities to numerical features, each role responsibility can be transformed into a binary feature. A role either has a responsibility, or it does not. If a role has a responsibility, the feature's value is a 1; otherwise, it is a 0. In this way, each responsibility is encoded as a feature.

Organizational level is a categorical feature with five possible values: strategic apex, middle line, technocracy, support staff, and operative core. Organizational level can be transformed into a usable feature by applying one-hot encoding [17]. In one-hot encoding, a categorical value with n values is transformed into n - *1* binary features.

In addition to role characteristics, the relevance of KPIs to roles should be measured somehow. The relevance of KPIs to roles can be captured in KPI relevance scores. The data on KPI relevance to roles that is collected during data collection will have to be transformed into some machine readable feature. Aksu et al. [6] used a five-point Likert scale to represent the relevance of KPIs to roles where five means most relevant, and one signifies the least relevant KPIs.

5.2 Component 2: Prediction Model Development

Figure 5.2 shows the details of the prediction development component. This part of the approach is based on work by Aksu et al. [6]. Instead of using organizational characteristics, the proposed approach uses role characteristics. The parts that have been changed from approach by Aksu et al. have been marked with an asterisk. This section will discuss the elements shown in Figure 5.2 and explain how the prediction models are developed.

Figure 5.2 shows the process of developing prediction models. As can be seen from the figure, a separate prediction model is trained for each KPI. KPI Relevance Data is split into datasets containing all data relating to individual KPIs. This data split means that there is a different dataset for each KPI. Not every role uses the same KPIs. Therefore, the relevance data for a role may be present for some KPIs but not for others. The difference in available data per KPI is the reason that different prediction models can be applied. Some prediction models are better suited for a set of data than other prediction models. Thus, the most accurate predictions can be provided by selecting the prediction model that suits the available data best. Which model performs best can be assessed by training multiple prediction models on the available data and calculating performance metrics on the results.

KRD is used to train the models. The role characteristics are used as features, and the KPI relevance scores are used as the outcome feature. The available data for each KPI is split into a training set and a test set to facilitate evaluating the produced models. The training set is used to train the models. The test set is used to validate the models. The scores that the prediction models produce can be used to calculate performance metrics. These metrics can be used to compare the quality of the predictions.

Which metrics are best suited to assess the performance of applied prediction models might differ per model, as some types of models have distinct performance metrics. However, accuracy, precision, and recall are the commonly used metrics for assessing the performance of prediction models [25, 61]. To create a baseline for comparing performance between the developed prediction models, a model that creates random predictions should be included. This inclusion is done to provide a point of reference for the performance of the tested models. It is useful to know which model performs best and how much better it performs than just predicting randomly[30] The classifier that performs best on the test-set created from KRD. The model that performs best according on accuracy, precision, and recall will be chosen as the preferred model.

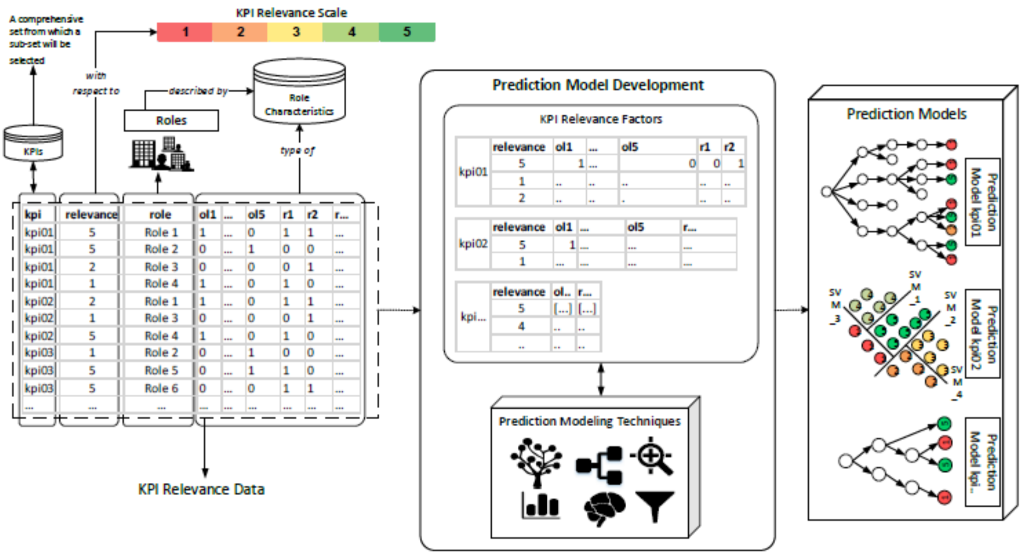

FIGURE 5.2: Details of prediction model development [2]

5.3 Component 3: Relevant KPI Prediction

After the best performing prediction models have been trained, a prediction model should be available for each KPI. When KPI relevance predictions are needed for a given role, a separate prediction is made for each KPI. The given role's characteristics are transformed into features. Thereafter, all prediction models are run with the given role, and predicted KPI relevance values are obtained.

Figure 5.3 shows an overview of the prediction process. This figure has been adapted from Aksu et al. [6]. All extension points have been marked with an asterisk. After all predicted relevance scores have been obtained, the list of scores can be ordered from high (5) to low (1). The KPIs with the highest predicted relevance are likely the most relevant to the given role.

[2] This figure is adapted from 'Automated Prediction of Relevant Key Performance Indicators for Organizations' by Aksu et al. [6]

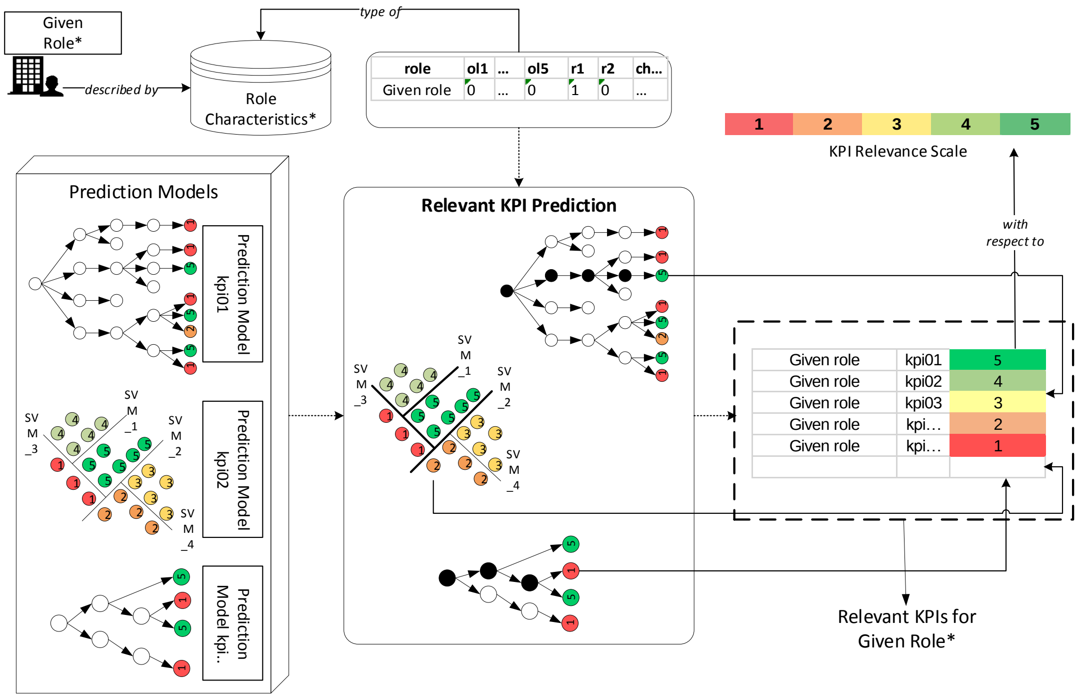

FIGURE 5.3: Details of Relevant KPI Prediction [3]

This chapter has formulated the proposed approach for the prediction of relevant KPIs for roles in organizations. In the next chapter, the case study evaluation the approach for the prediction of relevant KPIs for roles in organizations will be discussed.

[3] This figure is adapted from 'Automated Prediction of Relevant Key Performance Indicators for Organizations' by Aksu et al. [6]

Chapter 6

Evaluation

This chapter will discuss the evaluation of the proposed approach in this research. The approach was evaluated with a case study. In this case study, the approach was applied to a real-life situation. Relevant KPIs were predicted for the customers of an organization. First, the case study and its details will be discussed. Then, the data collection step will be discussed, followed by the application of the approach. Finally, the obtained results will be presented.

6.1 Case Study

The case study organization is called ValueCare. ValueCare is a BI and daily-auditing company based in the Netherlands. The majority of ValueCare's clients operate in the Dutch healthcare sector. Currently, ValueCare provides an online BI solution that allows its users to search for dashboards with a custom search engine. ValueCare's users are mainly employed in the Dutch Healthcare sector. These users are present in all organizational levels of the organization. The product is used by board members and medical specialists alike

ValueCare's product has grown into a considerable collection of KPIs and other data visualizations. The amount of available information has grown so large that the product's users experience *library anxiety*. Library anxiety can be explained as a type of information overload. If there are too many options to choose from, users might feel discouraged and refuse to use the product due to its complexity [10].

ValueCare aims to create a relevant starting point for each search to prevent its users from feeling discouraged when using their product. It is expected that by providing a relevant set of KPIs, the user will feel less anxious and more willing to use the product. ValueCare presumes that offering a set of relevant KPIs will reduce the user anxiety, which will improve the usability of the product, increasing user retention.

ValueCare was chosen as case study organization for multiple reasons. First, it has a predetermined set of KPIs that is available to its users. KPI use logs are kept for each user, indicating which KPIs are used and by whom. Together, these factors produce an environment that is well suited for testing this approach. Finally, ValueCare was willing to provide experts on the KPI use of their clients to evaluate results. An overview of the consulted experts can be found in Appendix B.

6.1.1 Data Collection

The goal of the data collection step was to gather the data that the approach requires. Role names, role responsibilities, organizational level, and KPI relevance scores were needed as input for the role selection component. None of these inputs were readily available. This section will first discuss the collection of role names, role responsibilities, and organizational level. Thereafter, the collection of KPI relevance scores will be discussed.

The initial plan for data collection was to contact the case study organization's customers with a questionnaire. These customers consist of hospital and mental healthcare institutions (GGZ) in the Netherlands. The case study organization's users are the employees of these organizations. Data on role characteristics and KPI relevance were to be collected with the questionnaire. Unfortunately some complications during the case study arose that prevented the use of questionnaires. The COVID-19 outbreak increased the pressure on the Dutch healthcare system to such a degree that there was little time for distractions such as questionnaires. After consultation with experts at the case study organization, it was concluded that it was unlikely that a large number of responses could be gathered through this method of data collection during the Corona crisis.

Role responsibilities, organizational level and KPI relevance scores were not collected due to the restrictions during data collection. Other possible sources of data for role characteristics or KPI relevance scores were discussed with experts at the case study organization. These experts pointed to two data sources: (1) the case study organization kept a list of role names for all users, (2) the case study organization recorded KPI-use logs for each user. By using this information, a dataset could be created with views per KPI and a role name for each user.

The collected use logs consist of two different varieties of logs: dashboard use logs, and KPI use logs. The KPI use logs contain information on individual KPIs. These logs document the use of specific KPIs and therefore offer detailed information on KPI use. The dashboard use logs document the use of dashboards. These dashboards logs provide less specific information on KPI use because they do not point towards specific KPIs. The dashboard use logs were not usable immediately and were therefore transformed into KPI use logs. Each dashboard contains several KPIs. If a user makes use of a dashboard, all contained KPIs are presented. Experts at the case study organization indicated that there is no way to find out which KPIs were used when looking at dashboard use logs. Moreover, they stated that it was best to make a one-to-one conversion between dashboard use and contained KPIs.

Two sets of data were collected during the data collection step: (1) a list of role names for users, (2) a list of KPI-views per user. Role characteristics were not available due to restrictions in data collection. This change in available data required a reassessment if the approach for grouping roles and predicting relevant KPIs. The steps taken during the case study will be discussed in the following chapter.

6.2 Applying the Approach for Predicting Relevant KPIs for Roles

This section will discuss the implementation of components in the approach during the case study. First, the comparison and grouping of role names for the creation of KRD will be discussed. After that, the development of the prediction models and the prediction of relevant KPIs for roles in organizations will be discussed.

6.2.1 Role Selection

The comparison of roles took a different form due to the changed role data. Instead of role name, role description, and role responsibilities being available, only role name was available. Experimentation with different string matching methods yielded that string matching was not a valid method when only role names are available.

Four text similarity metrics were used for the comparison of role names, Levenshtein distance, Cosine similarity, Cosine distance, and a combination of Cosine distance and Levenshtein distance [1, 9] were used. These metrics provided insufficient accuracy due to the absence of responsibility data. Thereafter, principal component analysis and factor analysis were applied to discover latent traits (roles) in the KPI Relevance Data [3, 58, 45, 11]. However, both Bartlett's test of Sphericity and the Kaiser-Meyer-Olkin test proved that the data was not fit for latent variable discovery [11].

After discussing the comparison of roles with experts at the case study organization, a more manual approach was taken. Instead of comparing all available roles with each other, the roles were compared to an established standard. The used standard is based on the International Standard Classification of Occupations (ISCO) from 2008 [24]. However, the obtained role names were written in Dutch. Therefore, a Dutch interpretation of ISCO was chosen. This standard was created by the research program for work and wellbeing (AZW) by the Dutch Bureau for Statistics (CBS) [50]. The standard is called the list of vocations for the healthcare sector and is taken from the Dutch adaptation of ISCO-2008, the BRC 2014 [23]. The BRC 2014 was chosen because the collected role names were all specific to the Dutch healhcare sector. The standard contains a list of specialized vocations in healthcare and a more generalized grouping for that vocation. For instance, a urologist and enterologist are different roles but are both medical specialists.

In addition to the specific medical roles that occur in the case study organization's data, some more general roles were added. Manager and secretary are not specifically medical roles and were therefore not included in the BRC 2014 standard. These roles were added in consultation with experts by analyzing all roles that did not fall under the BRC 2014.

The collected role names were entered as free text and therefore required some preprocessing. First, all special characters were removed, and all role names were made lowercase to prevent differences in capitalization and special characters from adding noise to the data. Thereafter, the role names were analyzed on whether they contained specific words or parts of words. An exception to this practice is the filtering of medical specialists. Medical specialist proved to be too generic to provide any meaningful information. Therefore, medical specialists were dissected further into medical specialisms. A full list of the medical specialisms can be found in Table C.1 in the Appendix.

The method for grouping role names was created in collaboration with experts at the case study organization. The grouping method works based on common elements found in role names for similar roles. For instance, secretary, secretary cardiology, or secretariat all contain the sequence 'secre-'. In collaboration with experts, identifying sequences in role names were defined and linked to roles found using the BRC 2014 standard and expert knowledge. If a role name contained none of the defined

Matching strings	Role category
Specialist	See Table C.1
psy, ologie, oloog, poli	Medical
verpleeg	Nursing staff
financ, fact	Financial staff
advi	Advisory staff
dbc, oplosser	DBC staff
behan	Practitioner
admin, secret	Administrative staff
consul	Consultants
assist	Medical assistants

Table 6.1: Overview of strings that point to a role

Role	Count
Medical specialist	3304
Manager	299
DBC staff	197
Advisory staff	135
Administrative staff	129
Healthcare control	80
Consultants	43
Financial staff	40
Nursing staff	24
Assistants	17
Practitioners	13

Table 6.2: Overview of categorized roles

sequences, the role name was marked as 'no result.' This method was conducted by iterating on the results gained from adding new sequences. After running the grouping method, all role names marked with 'no result' were examined again for sequences that might group them into a role. The results were iterated upon until there were no more role names that could be grouped. The ungrouped role names could either not be classified due to data entry errors or due to them being too specific to label accurately. The sequences and their resulting groups can be found in Table 6.1. Table 6.2 shows the total amount of role names assigned to roles.

6.2.2 Prediction Model Development

After the roles selection component, the KPI Relevance Data consisted of the following features: a user, the user's role name, and a count of views for each KPI. Recommender systems are designed to recommend items based on viewing behavior. Therefore, recommender systems were chosen as the best fitting prediction model.

Taking different variants of recommender systems into account, it was expected that a memory-based and user-based recommender system would produce the best results. Due to the lack of information on the KPIs themselves, an item-based approach seemed infeasible. The only recorded feedback was implicit. More specifically, only clicks in the portal were recorded. Fortunately, the main concerns with using implicit feedback data do not apply to this data. The case study organization's product is not publicly available and has separate instances for all its customers. There are no bots or crawlers active, so all traffic is generated by users.

It was expected that a memory-based recommender system would produce results with the highest accuracy. Memory-based recommender systems generally produce higher accuracy results at the cost of higher computational requirements [4]. Although much data was collected, it was not expected that the model's computational cost would be a problem in this research.

To test which prediction model was most successful in predicting relevant KPIs, all models were subjected to testing on the same dataset. The results of this test can be found in Table 6.3. The technical details on the used recommender systems can be found in Appendix A. In addition to the tested algorithms, an algorithm that makes random predictions was added for reference.

Metric	SVD - Tuned	User-Based KNN	User-Based-KNN with means	SVD-Untuned	Random
RMSE	14,897	14,857	14,819	14,897	15,369
MAE	6,354	5,927	5,603	6,354	6,467
HR	0,038	0,020	0,021	0,038	0,021
CHR	0,038	0,020	0,021	0,038	0,021
ARHR	0,022	0,014	0,014	0,022	0,013
Coverage	1,000	1,000	1,000	1,000	1,000
Diversity	0,316	0,311	0,314	0,316	0,385
Novelty	36,550	83,082	43,467	36,550	166,602

Table 6.3: Performance of tested models

Table 6.3 shows a few interesting things. First, all algorithms, including the random predictor, behave similarly. Both Singular Value Decomposition algorithms (SVD) have identical results. This similarity shows that the standard settings of the SVD provide the best results. The K-Nearest Neighbour approaches perform slightly better on accuracy, while SVD performs much better on hit rate than its competitors. The ARHR shows that the SVD performs best on top-N recommendation tasks. Because this research is concerned with predicting the most relevant KPIs, this is the most important metric to look at. Therefore, SVD was chosen as the final prediction model.

6.2.3 Relevant KPI Prediction

After Singular Value Decomposition was selected as the best prediction model, it was rerun to obtain the results. Instead of only entering the users with role and viewing numbers as features, the roles found during the role comparison step were also entered as users. Including roles in the prediction model has two advantages. First, it reinforces the similarities between members of the same role. Theoretically, members of the same role should also have similar viewing patterns. By adding roles as a feature along with the view data, this connection is artificially enhanced.

The second and largest benefit of adding roles to the users is preventing the cold start problem. The cold start problem emerges when there is no data to extrapolate from [33]. The selected prediction model works by measuring similarities in viewing behavior and then recommending items popular with similar users. If no behavior has been recorded yet for a new user, then there is no way to calculate similarity to other users. This problem is known as 'cold start.' By assigning a role to a new user, it is ensured that there is at least one similarity with other users in the same role. Moreover, because role is the only similarity, the model will predict the most popular KPIs for people with the same role, providing a recommendation purely for that role. The addition of roles as 'users' both creates a solution for the cold start problem and a role-based prediction method.

6.3 Results

The output obtained by running the SVD contains n KPI recommendations per role. The value of n is determined by the number of recommendations that could be made for a role. Exploratory analysis of the results in preparation for the evaluation showed that the quality of predictions trailed off around the 10^{th} prediction. Therefore, the maximum amount of recommendations obtained per role was 10.

Experts at the case study organization were interviewed because end-users were not available for evaluation. These experts are employees at the case study organization and specialize in the creation of new KPIs or are contact points for new KPI requests from client organizations. A list of the interviewed experts can be found in Appendix B. During each interview, experts were asked to evaluate predictions for roles they were familiar with. Then, each prediction made for that role was discussed in turn. For each prediction, the experts were asked whether the predicted KPI was relevant and whether it was useful for a user with that role.

As a rule of thumb, the percentage of the evaluation set to the training set is between 5 and 10 in evaluating typical machine learning prediction problems. This percentage was 9 for the created evaluation set in this research, which is sufficient to show that the developed prediction models can capture the knowledge in real-life scenarios.

Overall, at least 88% of predictions made were found both relevant and useful for roles. More detailed information on the evaluation can be found in Table 6.4. The number in the 'Expert'-column refers to the experts mentioned in Table B.1 in Appendix B. One prediction pointed to a KPI that contained information specific to one hospital and was therefore found irrelevant. A KPI that is not relevant to a role is also seen as not useful. Additionally, one prediction pointed to a KPI with relevant but unverified data. This prediction was therefore deemed relevant but not useful, bringing the percentage of useful predictions to 96%.

Expert	**% Relevant KPIs**	**% Useful KPIs**
1	100	100
4	100	100
7	100	90
8	87,5	87,5

Table 6.4: Evaluation of predictions for roles

Chapter 7

Discussion

This section will discuss multiple aspects of the results shown in the previous chapter. First, the results will be interpreted and their implications will be discussed. Second, the limitations of the proposed approach will be given. Finally, possibilities for future research will be provided.

7.1 Interpretation and Implications of Results

The predictions made by the approach were overall received positively by the experts. When asked whether they thought the predictions would help guide users to relevant data, they all answered affirmingly. This response indicates that the approach can predict relevant KPIs for roles. The interviewed experts mentioned that some roles might benefit from more specification, especially where the users' position in the organizational structure was mentioned as an essential factor for determining whether the KPI was relevant. The absence of data organizational level might have decreased the quality of some of the predictions. Adding Mintzberg categories as a feature could have made an additional distinction in the similarity between roles. Moreover, the expert that evaluated the recommendations for a lung specialist stated that the predictions were good, but generic. Along with the fact that the lung specialist only received 8 recommendations instead of 10, points to a possible lack of data for that role. However, even with a possible lack of data, the predictions were found to be relevant.

The results show that recommender systems can recommend relevant KPIs for roles in organizations. The KPIs that were predicted for roles should be more relevant than KPIs that are predicted based on their organizations' characteristics. Moreover, the predictions for one organization should also apply to users with the same role in another organization, as long as both organizations use the same set of KPIs. The case study has shown that applying a recommender system to roles produces relevant KPI predictions. This approach should work in any sector, as long as role data is available. Therefore, it is expected that the approach shown in this research can be generalized to other sectors. If new users are added to the organizations used in producing this approach, the inclusion should take little time. The new organization should only provide a list of role names grouped into the roles defined using the BRC 2014 standard. Then, recommendations will be produced based on these roles. After some time, use logs on the new organizations will have been collected, which can further specify the predictions. Using this approach can decrease the time and resources needed for finding relevant KPIs, decreasing the cost of implementation.

7.2 Limitations

It is important to note that the approach has some limitations. The approach can not produce recommendations for roles that do not have enough data associated with them to make recommendations. The role 'Nursing Staff' did not receive any recommendations with 23 occurrences, whereas 'Financial staff' did with 40. This seems to indicate that the cutoff level for occurrences lies somewhere between 23 and 40 in this case. However, this might differ per dataset, as total number of occurences in the data is not the only factor that impacts whether the approach is able to generate recommendations. It is possible that the approach is not able to generate a full list of recommendations. This shortcoming can be seen in Table 6.4, where one role got eight recommendations. This can be solved by adding more users with use data relating to a role.

Furthermore, when a new user with a new or underrepresented role is added to the data, there is a possibility that the approach will not be able to make predictions. A solution could be to provide a standard set of the most used KPIs as a recommendation to get him started. After some time, the user should get more accurate recommendations for his role.

Finally, the evaluation of the approach does not cover all predictions. Moreover, there weren't enough experts available to sufficiently evaluate recommendations made for individual users even though the interviewed experts reviewed these recommendations as 'promising'. To increase the reliability of the evaluation, more interviews with experts should be conducted on roles and individual users. The current evaluations shows that the recommendations are relevant, but the sample size hinders the generalizability of this evaluation to other, unevaluated, roles.

7.3 Future Work

The main opportunity for further research is to apply the research without the limitations caused by COVID-19. Although the implemented approach yielded positive results, it is important to assess whether it can produce more accurate results using role descriptions, role responsibilities, and organizational level. Another avenue for further research is the integration of this method with the automated generation of engaging dashboards. Current research for the automated generation of dashboard leans on the prediction of relevant KPIs for organizations [7]. By specifying these predictions to roles, more engaging dashboards can be automatically generated for roles or even users instead of organizations.

Finally, the predictions made for individual users should be evaluated. Due to privacy constraints, it was not possible to evaluate the available predictions for individual users. Predictions for users should be more detailed than predictions for their role. Therefore, using predictions for users should yield more relevant predictions.

Chapter 8

Conclusion

This study has provided a method for the prediction of relevant KPIs for roles in organizations. The approach requires KPI Relevance Data that contains role characteristics and KPI relevance scores. The approach first groups users into roles using role characteristics. After that, the prediction model predicts relevant KPIs for roles in organizations based on the KPI Relevance Data.

To test the validity of the approach, it was implemented and applied in a real-life setting. KPI Relevance Data is a new type of data, and therefore not readily available in organizations. KPI use logs were used as a proxy for it. The approach produced provided relevant KPI predictions for roles for users in organizations. Experts that evaluated the results of the approach stated that the predictions made were both relevant and useful in most cases.

The approach can be applied in any occupational sector, as long as there is data on roles and data on KPI relevance for those roles. The ISCO-2008 standard is not restricted to one occupational field and can thus be used to group individuals into role groups. Then, relevant KPIs can be predicted for roles in organizations. By automating the determination of relevant KPIs, the proposed approach help organizations to reduce the effort needed for selecting relevant KPIs for their employees.

Bibliography

[1] V. I. Levenshtein. “ Binary Codes Capable of Correcting Deletions, Insertions and Reversals”. In: *Soviet Physics Doklady* 10 (Feb. 1966), p. 707.

[2] W. Mc Kinney. “ Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56 –61. DOI: 10.25080/Majora-92bf1922-00a.

[3] Williams L.J. Abdi H. “Principal Component Analysis”. In: *Wiley Interdisciplinary Reviews: Computational Statistics* (2010).

[4] P.H. Aditya, I. Budi, and Q Munajat. “A Comparative Analysis of Memory-based and Model-based Collaborative Filtering On the Implementation of Recommender System for E-commerce in Indonesia: A Case Study Pt X”. In: *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE. 2016, pp. 303–308.

[5] R. Ahuja, A. Solanki, and A. Nayyar. “Movie Recommender System Using K-means Clustering and K-nearest Neighbor”. In: *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE. 2019, pp. 263–268.

[6] Ü. Aksu, D. Schunselaar, and H.. Reijers. “Automated Prediction of Relevant Key Performance Indicators for Organizations”. In: *Business Information Systems*. Ed. by W. Abramowicz and . Corchuelo. Cham: Springer International Publishing, 2019, pp. 283–299. ISBN: 978-3-030-20485-3.

[7] Ü. Aksu, A. del Rio-Ortega, M. Resinas, and H. Reijers. “An Approach for the Automated Generation of Engaging Dashboards”. In: Springer International Publishing, Oct. 2019, pp. 363–384.

[8] J.D. Baker. “Language of Improvement: Metrics, Key Performance Indicators, Benchmarks, Analytics, Scorecards, and Dashboards”. In: *Aorn Journal* 102.3 (2015), pp. 223–227. DOI: 10.1016/j.aorn.2015.07.015. eprint: https://aornjournal.onlinelibrary.wiley.com/doi/pdf/10.1016/j.aorn.2015.07.015. URL: https://aornjournal.onlinelibrary.wiley.com/doi/abs/10.1016/j.aorn.2015.07.015.

[9] L. Baoli and H. Liping. “Distance Weighted Cosine Similarity Measure for Text Classification”. In: *Intelligent Data Engineering and Automated Learning – Ideal 2013*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 611–618.

[10] D. Bawden and L. Robinson. “The Dark Side of Information: Overload, Anxiety and Other Paradoxes and Pathologies”. In: *Journal of Information Science* 35 (2009), pp. 180–191.

[11] A. Beavers, J. Lounsbury, J. Richards, and S. Huck. “Practical Considerations for Using Exploratory Factor Analysis in Educational Research”. In: *Practical Assessment, Research, and Evaluation* 18.1 (2013), p. 6.

[12] A. Bennett. "Case Study Methods: Design, Use, and Comparative Advantages". In: *Models, Numbers, and Cases: Methods for Studying International Relations* (2004), pp. 19–55.

[13] J. Bennett and S. Lanning. "The Netflix Prize". In: *Proceedings of KDD cup and Workshop*. Vol. 2007. New York. 2007, p. 35.

[14] F. Berisha. "Quality of the Predictions: Mean Absolute Error, Accuracy and Coverage". In: *Preprint* (Sept. 2017). DOI: 10.13140/RG.2.2.26870.65606.

[15] D. Bridgeland and R. Zahavi. *Business Modeling A Practical Guide to Realizing Business Value*. Elsevier Inc, Jan. 2009, 1 online resource (xvii, 387 s.)

[16] P. Castells, N. Hurley, and S. Vargas. "Novelty and Diversity in Recommender Systems". In: *Recommender Systems Handbook*. Boston, MA: Springer US, 2015, pp. 881–918. ISBN: 978-1-4899-7637-6.

[17] P. Cerda, G. Varoquaux, and K. Bal ázs. "Similarity Encoding for Learning With Dirty Categorical Variables". In: *Machine Learning* 107.8-10 (2018), pp. 1477–1494.

[18] P. Cremonesi, R. Turrin, E. Lentini, and M. Matteucci. "An Evaluation Methodology for Collaborative Recommender Systems". In: *2008 International Conference on Automated Solutions for Cross Media Content and Multi-Channel Distribution*. Dec. 2008, pp. 224 –231. DOI: 10.1109/AXMEDIS.2008.13.

[19] M. Deshpande and G. Karypis. "Item-based Top-n Recommendation Algorithms". In: *Acm Transactions On Information Systems (tois)* 22.1 (2004), pp. 143–177.

[20] T. Dolan. "Revisiting Adhocracy: From Rhetorical Revisionism to Smart Mobs". In: *Journal of Futures Studies* 15 (Nov. 2010), pp. 33–50.

[21] S. Few. "Show Me the Numbers". In: *Analytics Press* (2004).

[22] S. Few. *White Paper: Common Pitfalls in Dashboard Design*. Tech. rep. 2006. URL: https://www.perceptualedge.com/articles/Whitepapers/Common_Pitfalls.pdf.

[23] D. Fouarge and S. Dijksman. *Beroepenindeling Roa-cbs 2014 (brc 2014)*. Dutch. WorkingPaper 005. Netherlands: Research Centre for Education and the Labour Market, Jan. 2015. DOI: 10.26481/umarot.2015005.

[24] D. Fouarge and S. Dijksman. *International Standard Classification of Occupations: Isco-08*. English. Tech. rep. Geneva: International Labour Office, Jan. 2012.

[25] P. Galdi and R. Tagliaferri. "Data Mining: Accuracy and Error Measures for Classification and Prediction". In: Jan. 2018, pp. 431–436. ISBN: 9780128096338. DOI: 10.1016/B978-0-12-809633-8.20474-3.

[26] A. Horenbeek and L. Pintelon. "Development of A Maintenance Performance Measurement Framework"using the Analytic Network Process (anp) for Maintenance Performance Indicator Selection". In: *Omega* 42 (Jan. 2014), pp. 33–46. DOI: 10.1016/j.omega.2013.02.006.

[27] N. Hug. "Surprise: A Python Library for Recommender Systems". In: *Journal of Open Source Software* 5.52 (2020), p. 2174. DOI: 10.21105/joss.02174. URL: https://doi.org/10.21105/joss.02174.

[28] S. Kaganski, J. Majak, Kristo Karjust, and S. Toompalu. "Implementation of Key Performance Indicators Selection Model As Part of the Enterprise Analysis Model". In: *Procedia Cirp* 63 (Dec. 2017), pp. 283–288. DOI: 10.1016/j.procir.2017.03.143.

[29] M. Kaminskas and D. Bridge. "Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-accuracy Objectives in Recommender Systems". In: *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7.1 (Dec. 2016). ISSN: 2160-6455. DOI: 10.1145/2926720. URL: https://doi.org/10.1145/2926720.

[30] F Kane. *Building Recommender Systems with Machine Learning and AI*. Online Course. 2020.

[31] J. Keller, M. Gray, and J. Givens. "A Fuzzy K-nearest Neighbor Algorithm". In: *IEEE Transactions On Systems, Man, and Cybernetics* 4 (1985), pp. 580–585.

[32] F. Koetter, M. Kochanowski, and M. Kintz. "Leveraging Model-driven Monitoring for Event-driven Business Process Control." In: *EMoV+ MinoPro@ Modellierung*. Citeseer. 2014, pp. 21–33.

[33] X. Lam, T. Vu, T. Le, and A. Duong. "Addressing Cold-start Problem in Recommendation Systems". In: *Proceedings of the 2nd International Conference On Ubiquitous Information Management and Communication*. ICUIMC '08. Suwon, Korea: Association for Computing Machinery, 2008, 208â€"211. ISBN: 9781595939937. DOI: 10.1145/1352793.1352837. URL: https://doi.org/10.1145/1352793.1352837.

[34] M. Li, B. Dias, W. El-Deredy, and P. Lisboa. "A Probabilistic Model for Item-based Recommender Systems". In: *Proceedings of the 2007 ACM conference on Recommender systems*. 2007, pp. 129–132.

[35] H. Liu, F. Kuo, and T.Y. Chen. "Teaching An End-user Testing Methodology". In: *2010 23rd IEEE Conference On Software Engineering Education and Training*. IEEE. 2010, pp. 81–88. DOI: 10.1109/CSEET.2010.28.

[36] N. Liu, E.. Xiang, M. Zhao, and Q. Yang. "Unifying Explicit and Implicit Feedback for Collaborative Filtering". In: *Proceedings of the 19th ACM international conference on Information and knowledge management*. CIKM '10. Toronto, ON, Canada: Association for Computing Machinery, 2010, 1445â€"1448. ISBN: 9781450300995. DOI: 10.1145/1871437.1871643. URL: https://doi.org/10.1145/1871437.1871643.

[37] F. Lunenburg. "Organizational Structure: Mintzberg's Framework". In: *International Journal of Scholarly, Academic, Intellectual Diversity* 14.1 (2012), pp. 1–8.

[38] F. Martela. "What Makes Self-managing Organizations Novel? Comparing How Weberian Bureaucracy, Mintzbergâ€™s Adhocracy, and Self-organizing Solve Six Fundamental Problems of Organizing". In: *Journal of Organization Design* 8 (Dec. 2019), p. 23. DOI: 10.1186/s41469-019-0062-9.

[39] Henry Mintzberg. *Structure in Fives: Designing Effective Organizations*. Prentice-Hall, Inc, 1993.

[40] S. Mufazzal and S.M. Muzakkir. "A New Multi-criterion Decision Making (mcdm) Method Based On Proximity Indexed Value for Minimizing Rank Reversals". In: *Computers & Industrial Engineering* 119 (2018), pp. 427 –438. ISSN: 0360-8352. DOI: https://doi.org/10.1016/j.cie.2018.03.045. URL: http://www.sciencedirect.com/science/article/pii/S0360835218301360.

[41] M. Nilashi, O. Ibrahim, and K. Bagherifard. "A Recommender System Based On Collaborative Filtering Using Ontology and Dimensionality Reduction Techniques". In: *Expert Systems With Applications* 92 (2018), pp. 507–520.

[42] X. Ning and G. Karypis. "Slim: Sparse Linear Methods for Top-n Recommender Systems". In: *2011 IEEE 11th International Conference on Data Mining*. IEEE. 2011, pp. 497–506.

[43] M. Parikh. "Move Over Mintzberg, Let Adhocracy Give Way to Ambidexterity". In: *Management Decisions* 54 (June 2016), pp. 1047–1058. DOI: 10.1108/MD-07-2014-0483.

[44] D. Parmenter. *Key Performance Indicators: Developing, Implementing, and Using Winning KPIs*. John Wiley & Sons, 2015.

[45] P.R. Peres-Neto, D.A. Jackson, and M. Somers K. "Giving Meaningful Interpretation To Ordination Axes: Assessing Loading Significance In Principal Component Analysis". In: *Ecology* 84.9 (2003), pp. 2347–2363. DOI: 10.1890/00-0634. eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/00-0634. URL: https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/00-0634.

[46] A. del Rio-Ortega, M. Resinas, C. Cabanillas, and A. Ruiz-Cortas. "On the Definition and Design-time Analysis of Process Performance Indicators". In: *Information Systems* 38 (June 2013), pp. 470–490. DOI: 10.1016/j.is.2012.11.004.

[47] N. Russell, W. van der Aalst, A. ter Hofstede, and D. Edmond. "Workflow Resource Patterns: Identification, Representation and Tool Support". In: *Proceedings of the 17th International Conference On Advanced Information Systems Engineering*. CAiSE'05. Porto, Portugal: Springer-Verlag, 2005, 216â€“232. ISBN: 3540260951.

[48] N. Russell, W. van der Aalst, A. Ter Hofstede, and D. Edmond. "Workflow Resource Patterns: Identification, Representation and Tool Support". In: *International Conference on Advanced Information Systems Engineering*. Vol. 3520. June 2005, pp. 216–232. DOI: 10.10071143185516.

[49] B. Smith and G. Linden. "Two Decades of Recommender Systems At Amazon.com". In: *Ieee Internet Computing* 21.3 (2017), pp. 12–18. DOI: 10.1109/MIC.2017.72.

[50] Centraal Bureau voor de Statistiek. *Website Centraal Bureau voor de Statistiek*. 2020. URL: https://www.cbs.nl/nl-nl/achtergrond/2020/17/beroepenlijst-arbeidsmarkt-zorg-en-welzijn.

[51] The Pandas Development Team. *Pandas-dev/pandas: Pandas*. Version latest. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[52] A. Van Looy and A. Shafagatova. "Business Process Performance Measurement: A Structured Literature Review of Indicators, Measures and Metrics". In: *SpringerPlus* 5.1 (2016), p. 1797.

[53] G. Van Rossum and F. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697.

[54] J. Wang, A. De Vries, and M. Reinders. "Unifying User-based and Item-based Collaborative Filtering Approaches By Similarity Fusion". In: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. 2006, pp. 501–508.

[55] R. Wieringa. *Design Science Methodology for Information Systems and Software Engineering*. Springer, 2014.

[56] R. Wirth and J. Hipp. "CRISP-DM: Towards A Standard Process Model for Data Mining". In: *Proceedings of the 4th International Conference On the Practical Applications of Knowledge Discovery and Data Mining* (Jan. 2000).

[57] C. Wohlin. "Guidelines for Snowballing in Systematic Literature Studies and A Replication in Software Engineering". In: *Proceedings of the 18th International Conference On Evaluation and Assessment in Software Engineering*. London, England, United Kingdom: Association for Computing Machinery, 2014. ISBN: 9781450324762. DOI: 10.1145/2601248.2601268. URL: https://doi.org/10.1145/2601248.2601268.

[58] S. Wold, K. Esbensen, and P. Geladi. "Principal Component Analysis". In: *Chemometrics and Intelligent Laboratory Systems* 2.1 (1987). Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists, pp. 37 –52. ISSN: 0169-7439. DOI: https://doi.org/10.1016/0169-7439(87)80084-9. URL: http://www.sciencedirect.com/science/article/pii/0169743987800849.

[59] S. Zhang, W. Weihong, J. Ford, F. Makedon, and J. Pearlman. "Using Singular Value Decomposition Approximation for Collaborative Filtering". In: *Seventh IEEE International Conference on E-Commerce Technology (CEC'05)*. 2005, pp. 257–264.

[60] W. Zheng, B. Yang, and G.N. McLean. "Linking Organizational Culture, Structure, Strategy, and Organizational Effectiveness: Mediating Role of Knowledge Management". In: *Journal of Business Research* 63.7 (2010), pp. 763 –771. ISSN: 0148-2963.

[61] T. Zimmermann, R. Premraj, and A. Zeller. "Predicting Defects for Eclipse". In: *Proceedings of the Third International Workshop On Predictor Models in Software Engineering*. PROMISE â€™07. USA: IEEE Computer Society, 2007, p. 9. ISBN: 0769529542. DOI: 10.1109/PROMISE.2007.10. URL: https://doi.org/10.1109/PROMISE.2007.10.

Appendix A

Technical Details of Implementation

A.1 Data Handling

This chapter will show the technical means used for the implementation in the case study. First, data transformation will be discussed. After that, the tools used in the prediction model development and relevant KPI prediction step will be addressed. For everything involving programming, Python 3.8 [53] was used. Jupyter Notebooks were used for the rapid prototyping of code. Jupyter Notebooks were used because of their integrated functionality for Python Pandas, enabling more easy data handling. Pandas is a python library designed for easy handling of data in data science [2, 51].

A.2 Prediction Model Development

The prediction models that were tested were created using Surprise for Python [27, 30]. Surprise is a Python package built for education on recommender systems. It has multiple built-in algorithms, including a random predictor and performance metrics, and is built for easy experimentation with recommender systems. Surprise is not very memory efficient, so it might not be the right fit for large datasets. However, for the amount of data used in this research, this was not an issue.

To allow for a fair comparison between the used recommender systems, a program was built in Python that automatically extracts, transforms, splits the data, calculates performance metrics, and outputs them to an Excel sheet. This setup ensures that all the evaluated algorithms use the same dataset for training and testing and provides a comprehensive overview of evaluation metrics. After the best model was found, that model was run a final time to obtain the recommendations.
After deciding that the system would be a memory-based model using implicit feedback data, three options were considered: K-nearest neighbors, K-nearest neighbors with means, and Singular Value Decomposition. K-Nearest neighbors works by identifying the k most similar users, known as neighbors, and makes a prediction based on what these neighbors have watched. K-Nearest neighbors searches for the things that the neighbors have marked as enjoyable, but the original user has not rated yet [5, 31, 30]. K-nearest neighbors with means is a variation of the original k-nearest neighbors algorithm that takes the user's average score into account when making a prediction. Taking the user's average score into account mitigates the effect of different rating behaviors between users. One user might structurally rate a point higher or lower than another, regardless of how he rates something. This difference is not based on preference but on a personal rating bias. By subtracting the mean of a user's average

score from his predictions, this bias can be mitigated [30].

Singular Value Decomposition (SVD) is a complicated mathematical procedure that predicts ratings or views. It does this by predicting the missing values in the data with the existing data. View data is often sparse and therefore has many unknown values. SVD reduces the original matrix into two smaller matrices that, when multiplied, return an approximation of the original data. During this transition, the algorithm tries to construct the two smaller matrices so that the differences between the already known values and their predictions are as small as possible. When the two created matrices are multiplied back together, not only the known values are calculated but also the values that were missing before. In this way, SVD predicts the expected values for unrated items [59, 30].

For the implementation of SVD, the optimal hyper-parameters had to be found. The optimal hyper-parameters were found by executing a gridwise cross-validation. The ranges chosen for the parameters can be found in table A.1. Each possible combination in the values of the parameters in table A.1 was checked to obtain the best performing combination. This combination yielded 0 epochs, a learning rate of 0.05, and 20 factors.

Parameters	**Range**
Number of epochs	[0, 10 ..., 60, 70]
Learning rate	[0.05, 0.10, 0,15, 0.20
Number of factors	[20, 30 ... , 90, 100]

Table A.1: Checked hyperparameters for SVD

Appendix B

Consulted Experts

Expert	**Profession**	**Expertise**	**Consulted on**
1	Managing Consultant	11 years of experience with ValueCare´s clients wants and needs in the product	Reality-check for assumptions, interpretation of data, evaluation of results
2	Director	12 years of experience with hospital organization and organizational processes	Reality-check for assumptions, interpretation of data, business goals for model implementation
3	Consultant	3 years of expertience developing, designing, and implementing HIS at Chipsoft	Validity of role data, structure of HIS, available data at ValueCare
4	Consultant	Experience as medical specialist, specialized in the needs and wants of medical staff	Reality-check for how ValueCare product is used, interpretation of data, and evaluation of results
5	Director	Expert on the organization of medical institutions and their habits. Contacts all levels of organizations.	User habits for interpretation of data, interpretation of results, available data
6	ICT-Specialist	Broad knowledge on software, development, data, visualization, analysis, and UI/UX	Thesis supervisor at ValueCare, advised on model selection, data handling, and database structure.Evaluation of results
7	Consultant	Contact point for KPI wants and needs of medical personnel	Evaluation of results
8	Consultant	Contact point for KPI wants and need of medical personnel	Evaluation of results

Table B.1: Experts consulted during research

Appendix C

Medical Specialisms

Roles	Number of Users
oncologie, mondziekten kaak en aangezichtschirurgie, revalidatie, ergotherapie, kaakchirurgie, kinderchirurgie, fysiotherapie, kindercardiologie, psychologie, geriatrieint, endocrinologie	10-19
psychiatrie, revalidatiegeneeskunde, geriatrie, keel- neus- en oorheelkunde, cardiothoracalechirurgie, keelneusoorziekten, klinische genetica	20-29
radiotherapie, gynaecologieobstetrie, gastroenterologie hematologie, inwendige geneeskunde	30-39
orthopaedie, keel- neus- oorheelkunde, geestelijke gezondheidszorg	40-49
verloskunde gynaecologie, maagdarm en leverziekten, plastische chirurgie, longgeneeskunde, reumatologie, longziekten	50-59
anesthesiologie, orthopedie, urologie	70-79
neurochirurgie	90-99
chirurgie	100-109
intensivecare, allergologie, oogheelkunde	110-119
dermatologie	130-139
medisch	140-149
gynaecologie	160-169
heelkunde	170-179
neurologie	190-199
kindergeneeskunde cardiologie	200-209
internegeneeskunde	260-269

Table C.1: An overview of all found medical specialisms