

Utrecht University



Faculty of Science

Nabil Chakroun – 2281104

July 03, 2024

Master Thesis Applied Data Science

Thesis commission:

First examiner: Dr. Paulina Pankowska

Second examiner: Prof. dr. ir. Vincent Buskens

Predicting fertility (as a part of the PreFer data challenge)

Abstract

Understanding and accurately predicting fertility rates is essential in today's rapidly changing world for addressing demographic shifts, healthcare planning and economic forecasting. Using the comprehensive Longitudinal Internet Studies for the Social Sciences (LISS) dataset, this research integrates, demographic, socio-economic and environmental factors to develop predictive models. Advanced machine learning models, including Logistic Regression, Decision Trees, Support Vector Machines, Random Forest and Gradient Boosting Machine, are applied. These models are evaluated with metrics such as F1-score, precision, recall, accuracy and AUC-ROC. Essential steps in the process include, handling missing data with Multiple Imputation by Chained Equations (MICE), feature selection and feature transformations. The results conclude that Decision Trees and Gradient Boosting Machines models provide the most accurate predictions; however, the Logistic Regression is overall the best performing model. This research enhances the understanding of fertility prediction by integrating demographic, socio-economic and environmental factors into advanced predictive machine learning models. It also contributes to existing literature by using the LISS dataset and advanced machine learning algorithms to create accurate fertility prediction models, offering new insights and practical tools for policy-making and healthcare planning.

Contents

1 Introduction	3
2 Data and Methods	3
2.1 Data	5
2.2 Features selection	6
2.3 Handling missing data	8
2.4 Variable transformations	11
2.5 Models	12
3 Results	14
3.1 Feature importance	14
3.2 Model's performance	16
4 Discussion	20
4.1 Evaluation of predictive models and key fertility features	20
4.2 Limitation and further research	21
4.3 Conclusion	22
Appendices	23
References	29

1. Introduction

Fertility, the ability to conceive a child, is a vital aspect of human existence with profound implications for nations, societies and individuals. The ability to predict fertility rates accurately is of great importance in addressing a numerous of challenges. These challenges involve demographic shifts, social development goals, healthcare planning, economic forecasting, family planning initiatives and environmental sustainability efforts (Bodin et al., 2021).

Predicting fertility is particularly important and challenging nowadays as fertility rates have been experiencing significant changes worldwide, presenting different confronts for healthcare providers, decision-makers and individuals alike (Bodin et al., 2021). For instance, the issue of ageing societies in South-Korea and Europe demonstrates this phenomenon strongly (D'Ambrogio & European Parliamentary Research Service, 2023). With declining fertility rates and increased life expectancy, European countries are grappling with the implications of an ageing population on retirement and pension systems (D'Ambrogio & European Parliamentary Research Service, 2023). The imbalance between the shrinking working age population and the growing elderly population, calls for a strategic planning to ensure the sustainability of social security programs, possible young migration-workers and to address the healthcare needs of ageing individuals (D'Ambrogio & European Parliamentary Research Service, 2023).

Traditionally, predicting fertility rates depend on limited datasets, manual data collection processes and simple statistical methods such as, basic linear regression models(Freedman et al., 1975). As a result, these methods lacked the ability to capture non-linear, complex relationships and were based on simplified assumptions about population dynamics. Data collection was time consuming and the analysis was constrained by computational limitations. However, recent advancements in data science, computational techniques and statistical modelling have revolutionized the field of predicting fertility(APA PsycNet, 2021). Modern approaches leverage advanced machine learning algorithms, such as neural networks and random forests. These methods are capable of handling large and diverse datasets. In addition, the advancements and introduction of cloud computing and parallel processing, have allowed access to powerful computing infrastructure. Therefore accelerating the development and launch of complex fertility prediction models.

In addition, the increased availability of large demographic datasets, coupled with sophisticated analytical tools, has aided the development of predictive models, and has opened new techniques for exploring and forecasting fertility trends. Enabling researchers to unravel the complex dynamics underlying fertility patterns, since these predictive models are able to capture the complex interaction of socio-economic, biological and environmental factors influencing fertility dynamics (APA PsycNet, 2021).

Building upon this foundation, recent studies delved into the use of advanced deep learning, machine learning and ensemble learning techniques in fertility prediction. In papers such as, Goyal et al. (2020), Ahinkorah et al. (2021) and Barnett-Itzhaki et al. (2020) researchers

studied a diverse set of machine learning, deep learning and ensemble learning techniques to predict fertility. These methods included logistics regression, k-nearest neighbours, 1-D neural network, random forest, AdaBoost and voting classifiers. Offering a range of approaches to address the complexities of fertility prediction tasks and potentially enhancing the accuracy and robustness of the predictive models.

This research aims to investigate fertility prediction using a comprehensive approach that integrates socio-economic, demographic and environmental factors. Building on existing research that has highlighted the importance of these factors in influencing fertility. This study utilizes the Longitudinal Internet Studies for the Social Sciences (LISS) dataset. By employing machine learning algorithms and statistical techniques, the goal is to develop robust predictive models capable of forecasting fertility rates with precision and accuracy.

Additionally, this research aims to advance the current understanding of fertility prediction and also take an innovative approach by leveraging the comprehensive and longitudinal nature of the LISS data. While there have been many studies on fertility prediction, the existing literature often focuses on isolate factors or uses traditional statistical methods, which may not capture the full complexity of fertility trends. This leaves a gap in studies that integrate multiple socio-economic, environmental and demographic factors using advanced predictive techniques. By utilizing machine learning models, this research addresses this limitation and makes a significant contribution to the literature. The integration of these diverse factors within predictive models may not only enhance the accuracy of fertility predictions but could also provide a more holistic understanding of the underlying dynamics. This comprehensive approach lays a robust foundation for further research, offering new insights and practical tools for policy-making and healthcare planning, especially in the context of demographic challenges and ageing populations.

2. Data and Methods

2.1. Data

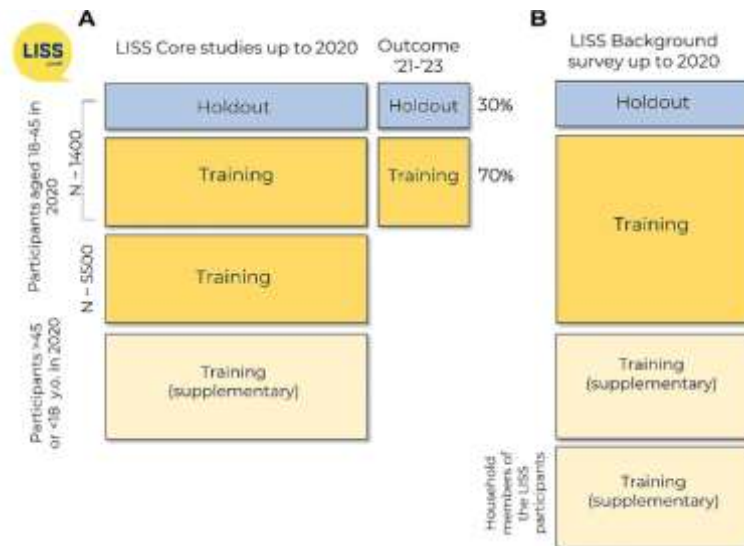


Figure 1: Survey data from the LISS panel used in the data challenge (Sivak et al., 2024).

This study will use data from the LISS panel survey from the years 2007-2020. The LISS panel survey is managed by the non-profit research institute Centerdata. It is a comprehensive online survey system derived from a probability sample of the Dutch population register, supplied by Statistics Netherlands. This panel is representative of the Dutch population and uses refreshment samples to correct for any initial selection biases. The LISS panel collect data through two main sources: the LISS Core Study and the Background surveys. The Core Study is an annual longitudinal survey that includes a variety of topics such as education, health, religion, values and personality. The Background survey, which is updated monthly, gathers basic socio-demographic data about households and all their members (LISS Panel - LISS Panel, 2023) .

For this data challenge, the goal is to predict fertility outcomes for individuals who were aged 18-45 in 2020, using data from the Core Study collected from 2007-2020. In addition, this dataset contains over 31,000 variables from these modules. The LISS panel began in 2007 with approximately 8000 individuals (around 5000 households), with an annual dropout rate of about 10%. However, this dropout rate is mitigated by recruiting new members every two years (Sivak et al., 2024).

The challenge involves predicting which participants will have a child between 2021 and 2023. The dataset is divided into a training set, containing 70% of the target group with known outcomes and is available to challenge participants. The holdout set, comprising the remaining 30% of the target group, is used for evaluation. Although the sample size of about 1400 respondents with known outcomes is relatively small, it is typical for social science research. The rich longitudinal data from the LISS panel still offers a valuable resource for this prediction task (Sivak et al., 2024).

2.2. features selection

Feature selection is a pivotal step in the process of constructing predictive models, particularly in the realm of data-driven research (Chowdhury & Turin, 2020). This process becomes even more critical when dealing with datasets that have many features but a limited number of observations such as the LISS dataset, because such scenarios can lead to two major challenges in machine learning namely the curse of dimensionality and overfitting. This section embarks on the crucial task of identifying and prioritizing a subset of informative variables from a large pool of potential predictors. By carefully selecting relevant features, the aim is to enhance the interpretability, efficiency and generalization capability of our predictive models (Chowdhury & Turin, 2020).

The initial dataset that was going to be used encompassed an extensive array of variables, approximately just over 31000. At first sight, approximately 40 variables were identified for potential use. The selection process was informed by domain expertise, extensive literature review, Exploratory Data Analysis (EDA) and statistical methodologies to ensure the inclusion of relevant features. However, a considerable proportion of these 40 features exhibited substantial missingness. Ranging from 50 % to 90% across the sample. Variables with more than 50% missingness were excluded from the final feature selection. Making the final count of features to be 18. The final set of 18 variables can be found in Table 1. Additionally, the rationale for including each variable is explained in detail in Appendix A.1.

This decision was made to ensure the robustness and reliability of the predictive model (Jäger et al., 2021). High levels of missing data can introduce bias, reduce statistical power and compromise the validity of the analysis (Kumar et al., 2017). By omitting variables with these high percentage of missingness, we enhance the accuracy and interpretability of the model, as it focuses on features with more complete and reliable information (Kumar et al., 2017). This approach may result in the loss of potentially relevant information, however it is still worth doing because the inclusion of variables with substantial missing data can lead to biased results and reduce the robustness of the model. By focusing on variables with more complete data, we ensure the model's reliability and validity. Reliability is ensured by minimizing the impact of missing or incomplete data, which can introduce biases and inconsistencies. In addition, validity is maintained by selecting variables that are well-documented and representative of the factors influencing fertility rates. Hence, it became essential to carefully select which features to include in the analysis. It was important to focus on variables that had both enough data available and were relevant for prediction.

In the following section, the Exploratory Data Analysis and statistical methodologies used in the selection process of the features will be explained.

Exploratory Data Analysis has been employed as a step in this research, allowing for a comprehensive understanding of the dataset's structure and characteristics. Through techniques such as summary statistics, data visualization (e.g., scatter plots, histogram plots), and correlation analysis, outliers have been identified, data distribution assessed and potential

relationships between variables explored. For example, histogram plots provided insights into the distribution of continuous variables, aiding in outlier detection. In addition scatter plots facilitated the identification of patterns or associations between variables. EDA offers the advantage of uncovering hidden patterns or anomalies in the data, guiding in the selection of relevant variables for further investigation.

Furthermore statistical methodologies have been employed in this research to potentially select final features. Techniques such as penalized regression (e.g., Ridge or Lasso regression) and stepwise regression were used to identify potential features that were not going to be used if they showed no significance in the final models. Penalized regression methods help address multicollinearity by shrinking the coefficients of less important variables, while stepwise regression simplifies the model by iteratively adding or removing features based on their statistical significance. Together, these methods complement each other by enhancing model stability and reducing overfitting, ensuring the robustness of the final models.

The application of these methods did not lead to a decrease in the number of features. Since they showed agreement with the initial features selected for logistic regression. Moreover, the models did not show any improvement or provide indications that further adjustments were necessary. In fact, these methods reduced the performance metrics compared to logistic regression, which is why they were ultimately not used in the final models.

Table 1: selected features

Variable	Label
nomem_encr	Number of household member encrypted
outcome_available	Whether the outcome is available for the respondent
cf20m024	Do you currently have a partner?
cf20m030	Are you married to this partner?
ci20m383	Primary occupation
ch20m133	How often did you have a drink containing alcohol over the last 12 months?
ch20m160	Soft drugs (such as hashish, marijuana)
ch20m161	XTC (such as MDMA)
ch20m163	Hard drugs (such as stimulants, cocaine, heroin)
ch20m270	Laughing gas
cr20m162	To what extent would you describe yourself as a religious person?
cw20m000	Respondent has paid work (according to household box data)
gender_bg	Gender
migration_background_bg	Origin
age_bg	Age of the household member on December 2020
nettohh_f_2020	Net household income in Euros
opmet_2020	Highest level of education with diploma
new_child	Whether respondent had a child in 2021-2023

2.3. Handling missing data

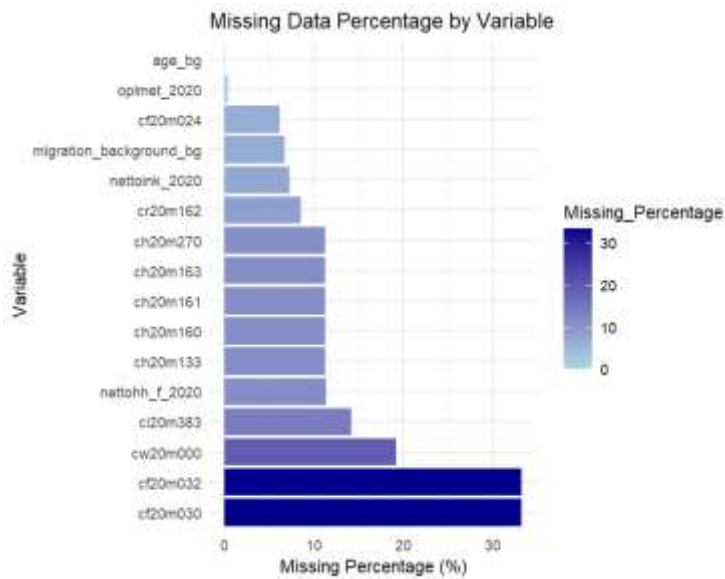


Figure 2: Missing percentages of selected features from 2020

Handling missing data is a critical step in any research, particularly when dealing with datasets that have varying degrees of missingness across different features, such as the LISS panel (Jäger et al., 2021). In our dataset, as can be seen in Figure 3, there are two variables: ‘married to partner’ and ‘partner’s gender’, which had substantial amount of missing data, with 33.23% missing values each. Other variables like ‘paid work presence’ and ‘primary occupation’ also showed significant missingness, with 19.15% and 14.18% respectively. To address these gaps effectively and ensure robust analysis, the MICE (Multiple Imputation by Chained Equations) package was used, which is well-regarded for its ability to handle different types of missing data by specifying appropriate imputation methods for each variable (Van Buuren & Groothuis-Oudshoorn, 2011).

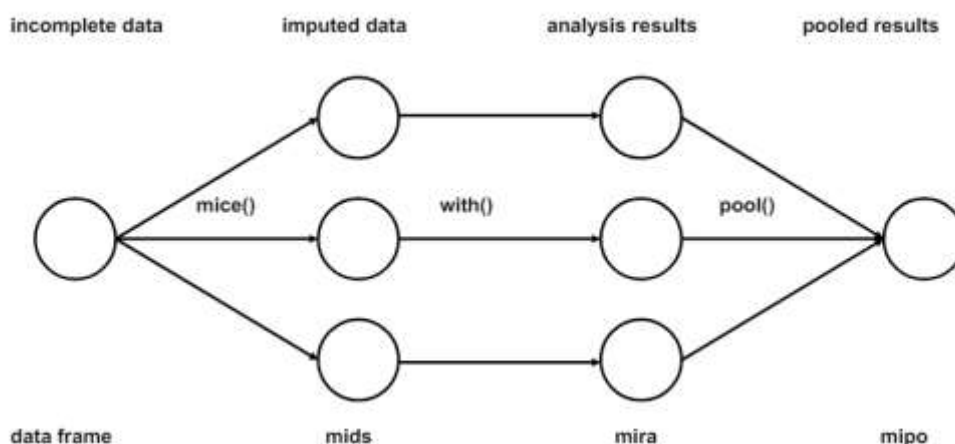


Figure 3: Main steps in MICE (Van Buuren & Groothuis-Oudshoorn, 2011)

The main steps in MICE are summarized in Figure 4. MICE operates by iteratively imputing missing values through a series of conditional models. It begins by initializing missing entries with placeholder values such as the median or mean. Then, it cycles through each variable with missing data, treating it as the dependent variable in a regression model with all other variables as predictors. For continuous variables, predictive mean matching is used, while logistic regression is applied for binary variables. Categorical variables, both ordinal and nominal, are imputed using multinomial logistic regression or predictive mean matching, depending on their nature (Van Buuren & Groothuis-Oudshoorn, 2011).

Mathematically, if X_j is the variable being imputed, MICE models $X_j = f(X_{-j}) + \varepsilon$, where X_{-j} represents all other variables in the dataset and ε is the error term. This process is repeated multiple times to create several imputed datasets, which are then analysed separately and pooled. The Pooling uses Rubin's Rules to combine estimates, accounting for both within-imputation and between-imputation variability¹:

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^m Q_i$$

$$T = \bar{U} + \left(1 + \frac{1}{\bar{m}}\right) B$$

Where \bar{Q} is the pooled estimate, m is the number of imputations, Q_i is the estimate from the i -th imputation, \bar{U} is the average within-imputation variance, and B is the between-imputation variance. This method provides robust and reliable estimates, ensuring that the imputed values reflect the underlying data structure and variability (Van Buuren & Groothuis-Oudshoorn, 2011).

Using MICE, the selected features were assigned to their corresponding imputation methods, ensuring that the imputed values were coherent with the nature of the data. For instance, continuous variables like 'nettohh_f_2020', which had a missingness rate of 11.35% were imputed using predictive mean matching. Binary variables such as 'Partner status' were imputed using logistic regression models. Categorical variables, both ordinal and nominal, are imputed using multinomial logistic regression. By carefully selecting the imputation methods based on variable types, we mitigated the potential biases and inaccuracies that could arise from improper handling of missing data (Van Buuren & Groothuis-Oudshoorn, 2011).

¹ Within-imputation variability arises from the random component of the imputation model. Even with a well-specified model, the imputed values will exhibit some natural variation around the estimated mean. Between-imputation variability captures the uncertainty introduced by the missing data itself. Since multiple imputed datasets are created, estimates such as the mean, will vary across them.

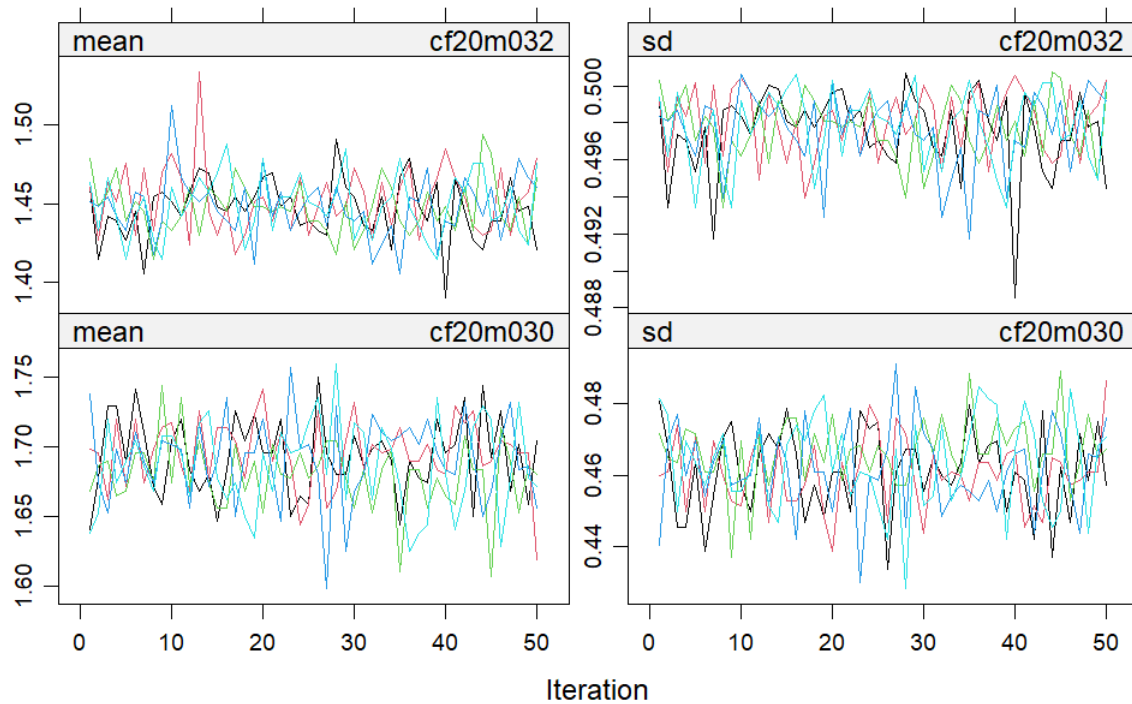


Figure 4: Trace Plots of MICE for Partner status (*cf20m030*) and Partner Gender (*cf20m032*)

In the context of multiple imputation, trace plots as seen in figure 3, serve as critical tools for assessing the reliability and convergence of imputed values across iterations. The criteria for a good trace plot, include minimal fluctuation and convergence of the imputed values and standard deviations across iterations.

This research, utilized trace plots to monitor the imputation process for all features that were employed for the machine learning models. In figure 3, the trace plots for 2 variables, which had the highest percentage of missingness, are illustrated. The trace plots illustrated consistent trends: as the number of iterations increased, the mean imputed values and the standard deviations for both variables stabilized, converging towards a steady estimate. Therefore, these variables demonstrated the criteria for a good trace plot, thereby indicating robustness in the imputation process. Such stability ensures that the imputed values accurately reflect the underlying patterns in the original data. Thus, enabling reliable subsequent analyses. Trace plots for all features can be found in Appendix A.3.

2.4. Variable transformations

In this section, binary encoding has been implemented as an essential technique for transforming categorical variables such as ordinal and nominal into a binary presentation. Binary encoding, which is a widely-used method in machine learning and data analysis, converts categorical variables into binary vectors (Dahouda & Joe, 2021). Effectively capturing their underlying information without introducing high dimensionality. For this research, the data already included categorical variables that were mostly integers. We used binary encoding to transform these categorical features into binary features. Specifically, for each original feature, we created one binary feature. For example, the 'educational level' feature originally had 9 values (integers ranging from 1 to 9). This feature then is transformed into a binary feature, where 0 represented a low educational level and 1 represented a high educational level. The split between 0 and 1 was based on theoretical knowledge and statistical analyses, examining the relationship between the original 9 values and the outcome. This transformation is particularly useful because many machine learning models, such as linear regression and support vector machines, cannot handle categorical data directly and require numerical input. By representing categories with binary values (0 or 1), the models can incorporate these features into their computations, thereby handling categorical data effectively without treating them as purely numerical values.

Compared to other methods like one-hot encoding, binary encoding creates a more compact representation of the data. One-hot encoding converts each category of a categorical variable into a separate binary feature, which can result in a large number of features and sparseness when dealing with many categories. This compactness from binary encoding improves the computational efficiency and performance of machine learning models by reducing the complexity and the amount of memory required. Furthermore, binary encoding helps preserve the ordinal relationships within the data, when applicable. For example, for the 'educational level' feature, where the values ranged from 1 (low education) to 9 (high education), binary encoding can group these into a binary feature that reflects an ordinal split, such as 0 for low education (1-4) and 1 for high education (5-9). This maintains the integrity of the underlying information and ensures that the encoded data still reflects the inherent order of the original categories.

An illustrative example supporting the use of binary encoding based on the religiousness variable and its relationship with the outcome variable (having a child) is provided in Appendix A.2

Additionally, the continuous predictor 'Net household income' undergoes Standardization Z-score normalization to ensure compatibility with statistical modelling techniques. Z-score normalization transforms the distribution of a feature to have a mean of 0 and standard deviation of 1 by subtracting the mean and dividing by the standard deviation (Nogueira & Munita, 2020). This process effectively scales the data, enabling different features to be compared on a common scale without being influenced by their original status. The advantages of this approach are particularly useful in this research, since all the other features are transformed into binary variables, ensuring that each feature contributes proportionately to

the analysis, regardless of its scale. Thereby enhancing the robustness and reliability of the statistical models used.

Furthermore, Z-score normalization facilitates the convergence of gradient-based optimization algorithms, such as those used in logistic regression and neural networks, which are utilized in this research by providing a more stable and consistent gradient. Overall, this approach ensures that the standardized continuous variable aligns with the binary-encoded categorical variables, contributing to a balanced and effective analytical framework.

2.5. Models

In this research, several advanced machine learning models have been developed to analyse and predict the occurrence of having a new child within 2021-2023. These machine learning models include Logistic Regression, Decision Trees, Support Vector Machine, Random Forest and Gradient Boosting Machine. Each model provides distinct benefits and perspectives on the data. Enabling a thorough examination of the factors affecting the outcome variable. In addition, performance metrics such as F1-score, precision, recall, accuracy and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) have been used to evaluate the effectiveness of these different machine learning models. These metrics offer valuable insights into the predictive abilities of the models and aid in evaluating their performance across various assessment criteria.

In this analysis the Logistic Regression is the baseline model due to its simplicity and interpretability. It simply models the probability of a binary outcome using a logistic function (Hosmer et al., 2013). Mathematically, the logistic regression model is expressed as:

$$P(Y = 1 | X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 + \dots + \beta_p x_p)}}$$

where P is probability of the target variable being 1, β_0 is the intercept and β_i are the coefficients for the predictor variables X_i (Hosmer et al., 2013).

Following the Logistic Regression, Decision Trees offer a non-parametric approach, providing a hierarchical model that splits the data based on certain variable values (Bonaccorso, 2018). Each node in the tree represents a decision rule, splitting the data into branches. Eventually leading to further nodes or terminal leaves. The training of the decision tree model is based on using the Cart algorithm, which minimizes the entropy or Gini impurity at each split (Bonaccorso, 2018). This research experimented with different possible hyperparameters such as complexity parameter (cp), maximum depth and minimum samples per split, to prevent overfitting and ensure generalization. For instance, this research experimented with maximum depth ranging from 3 to 30 and cp values from 0.001 to 100. The best performing model was selected based on accuracy and other metrics on the validation set.

In addition, Support Vector Machine has also been used for this research. Support Vector Machine is a powerful classification machine learning model that finds the optimal

hyperplane by separating the classes in a high dimensional feature space, which can be seen in figure 4 (Mohammed et al., 2016).

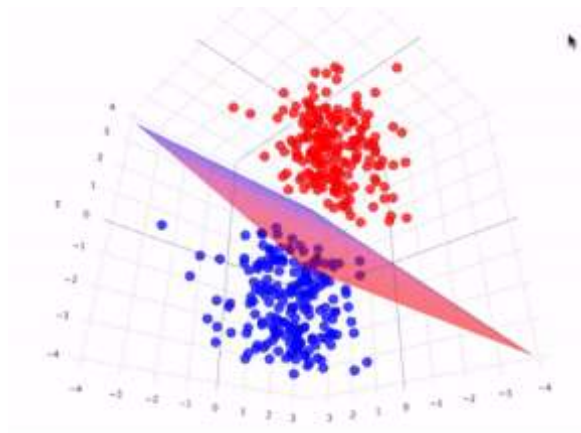


Figure 5: An illustration of Support Vector Machine (Elhamraoui, 2021)

It then uses a kernel function to transform the input features into a higher dimension where a linear separator between the features can be found (Mohammed et al., 2016). In this research, the radial basis function (RBF) kernel was used, which is defined as:

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right),$$

where γ is a hyperparameter that determines the influence of individual training examples (Mohammed et al., 2016). For this research, hyperparameters such as γ and the penalty parameter C , were tuned to balance the trade-off between minimizing the classification error and maximizing the margin.

Finally, this research employs two ensemble methods: Random Forest and Gradient Boosting Machine. These are both ensemble methods that leverage multiple models to enhance prediction accuracy and control overfitting.

Random Forest does this by building multiple decision trees and combining their results, each tree being trained on a bootstrap sample of the training data (Bonaccorso, 2018). Furthermore, a subset of the variables is randomly selected at each split to ensure diversity among the trees (Bonaccorso, 2018). In this research, 500 trees were used and $mtry$, which is the number of features considered at each split was set to 3, to optimize performance.

On the other hand, Gradient Boosting Machine builds models sequentially, with each new model correcting the errors of its predecessor (Natekin & Knoll, 2013). GBM uses gradient descent to reduce a differentiable loss function, typically the deviance in classification problems (Natekin & Knoll, 2013). Important hyperparameters include the depth of interactions, the number of trees, the learning rate and the fraction of data used for each base learner. This research selected an interaction depth of 3, a learning rate of 0.01 and 500 trees, which was based on cross-validation results. GBM's ability to capture complex relationships in the data makes it a robust choice for classification tasks (Natekin & Knoll, 2013).

3. Results

This research evaluated the performance of five machine learning models: Logistic Regression, Decision Tree, Support Vector Machine, Random Forest and Gradient Boosting Machine for predicting the target variable of having a child within 2021-2023. In doing so, the dataset was split into training and testing sets with a 70-30 ratio. Specifically, 70% of the data was used to train the model, while the remaining 30% was used to test and evaluate its performance. Each model's performance was assessed using accuracy, precision, recall, F1-score and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. Accuracy measures the proportion of correct predictions out of all predictions. Precision indicates the proportion of true positives among the predicted positives. Recall, or sensitivity, measures the ability to identify all actual positive cases. The F1-score is the harmonic mean of precision and recall, providing a balance between the two. The AUC of the ROC curve evaluates the model's ability to distinguish between positive and negative classes, with higher values indicating better performance. Table 2 summarizes the metrics for each model, while Figure 4&5 present the feature importance and figure 6 presents a combined Roc curve plot for visual comparison.

3.1 Feature importance

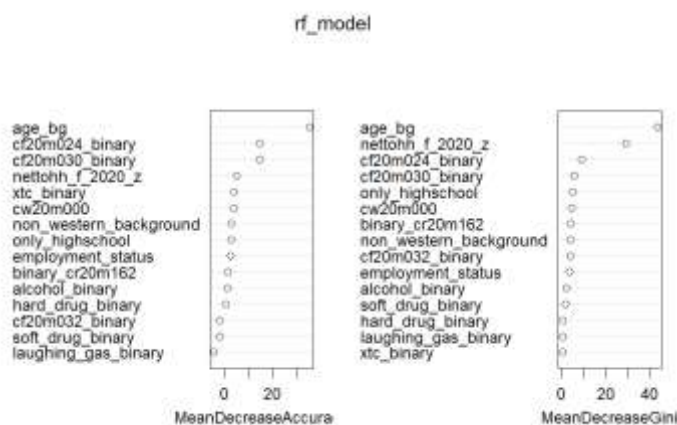


Figure 6: Random Forest Feature Importance

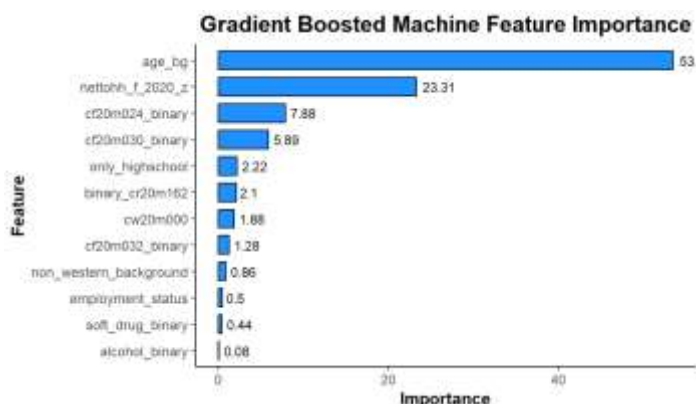


Figure 7: Feature importance GBM model

Feature importance in machine learning models provide insight into which variables are most influential, in this context, in making predictions. In Random Forests, feature importance can be explained as follows: it is assessed through the mean decrease in accuracy or the mean decrease in impurity, when the variable's values are permuted. In GBM, feature importance is calculated based on how often a feature is used for splitting the data across all the trees and the reduction in error it achieves. Higher importance scores indicate that the feature significantly contributes to the model's predictive power. In Support Vector Machines, feature importance revolves around the coefficients assigned to each feature in the hyperplane that separates different classes or predicts probabilities. Features with larger coefficients exert more influence on the decision boundary, therefore playing a crucial role in classification tasks. Lastly, in Decision Tree models, feature importance is evaluated based on how much each feature contributes to splitting nodes and organizing the data. Features that lead to significant reductions in impurity or entropy are considered more important as they better define different classes or outcomes.

For the GBM and Random Forest model, the features age, household income, partner status (cf20m024_binary) and partner gender (cf20m030_binary) emerged as the most important features. Suggesting that these variables play an important role in determining the likelihood of having a child within 2021-2023. This important role can be explained per feature. Age is often associated with fertility and life stage, while household income reflects financial stability and therefore also the ability to support a child. Partner status and partner gender most likely influence family planning decision and social-biological dynamics. Analysing the importance of these variables helps in interpreting the machine learning model's decisions and provide valuable visions into the underlying factors that influence the target variable. Since this reveals which features are prioritized and contribute most significantly to predicting the outcome.

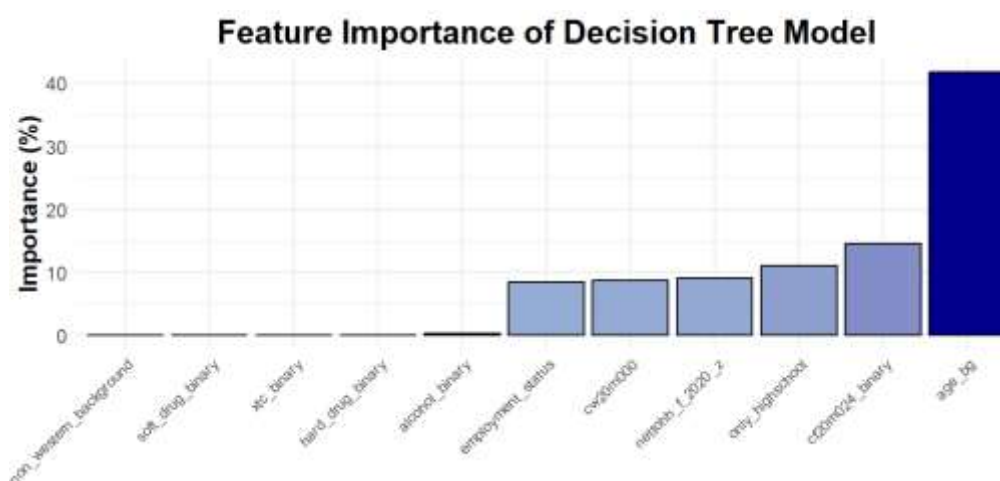


Figure 8: Feature importance of Decision Tree Model

Additionally, the feature importance of the Decision Tree model reveals several essential predictors influencing the likelihood of having a child between 2021-2023. Age emerges as the most significant feature, highlighting its dominant role. Features such as partnership status

(‘cf20m024_binary’) and educational attainment (‘only_highschool’) also exhibit substantial predictive power. Household-income (‘nettohh_f_2020_z’) and employment status also show a significant contribution to the model. These findings are also consistent with the logistic regression model and the Support Vector Machine model. Thereby emphasizing the robust impact of demographic and socio-economic features on predicting the outcome.

3.2 Model’s performance

Model	Accuracy	Precision	Recall	F1-Score	AUC
Logistic Regression	0.7643	0.7708	0.9823	0.8638	0.7713
Decision Tree	0.8283	0.4789	0.7083	0.5714	0.8461
SVM	0.7677	0.0423	0.7500	0.0800	0.7897
Random Forest	0.7912	0.2535	0.6667	0.3673	0.8314
GBM	0.8277	0.4507	0.7273	0.5565	0.8591

Table 2: Performance results per machine learning model

The models assessed for predicting the occurrence of having a child within 2021-2023 exhibited distinct strengths and weaknesses across different performance metrics. Both the Gradient Boosted Machine and Decision Tree models distinguished themselves with high accuracies of 0.8277 and 0.8282, respectively. The Decision Tree’s ability to capture complex data relationships contributed to its competitive accuracy, although with potential variability in F1-score and precision on unseen data. Also, the GBM’s ensemble learning approach iteratively improved predictions, illustrating strong performance in accuracy and AUC. However, other performance metrics such as F1-score and precision, indicate possible areas where further optimization could enhance performance.

This discrepancy of the GBM can be attributed to its iterative nature, which focuses on minimizing prediction errors overall rather than specifically optimizing for precision or minimizing false positives. As GBM build successive models to correct errors made by previous models, it may prioritize improving overall accuracy and AUC, which are crucial for general predictive performance.

Furthermore, the Random Forest model achieved an accuracy of 0.7912, with a precision of 0.2535, recall of 0.6667, and F1-score of 0.3673. Random Forest demonstrated trade-offs typical of ensemble methods, which are also mentioned for the GBM model. Namely, its ensemble of decision trees allows it to capture complex data interactions effectively, leading to competitive recall and accuracy. However, the lower precision suggests that while Random Forest identifies many actual positive cases, it also predicts a significant number of false

positives. This trade-off arises because individual decision trees within the ensemble may independently make errors, which are then aggregated in the final prediction.

Additionally, the Support Vector Machine model achieved an accuracy of 0.7677 with a recall of 0.7500 indicating its effectiveness in identifying individuals likely to have another child. However, its precision 0.0423 and F1-score of 0.0800 were notably lower compared to other models, highlighting SVM's challenge in simultaneously achieving high precision and recall due to its sensitivity to kernel selection and data distribution. SVM's approach involves separating classes with a hyperplane, making it effective in recall but less so in precision, due to its inherent difficulty in correctly classifying positive cases while avoiding false positives. Highlighting the importance of selecting the correct model based on specific application requirements and understanding the nuances of model performance across the different evaluation metrics.

The Logistic regression exhibited robust performance across multiple evaluation metrics. It produced an accuracy of 0.7643, precision of 0.7708, recall of 0.9823 and a F1-score of 0.8638. The high recall of 0.9823 showed that the model effectively identified almost all individuals who actually had a child. Also the high F1-score of 0.8638 reflects a balanced performance between the precision and recall. Logistic regression's ability to perform well across these metrics can be attributed to its linear nature and suitability for binary classification tasks. Furthermore, highlighting its capacity to balance between minimizing false positives and capturing all positive instances in prediction the outcome of having a(nother) child within 2021-2023.

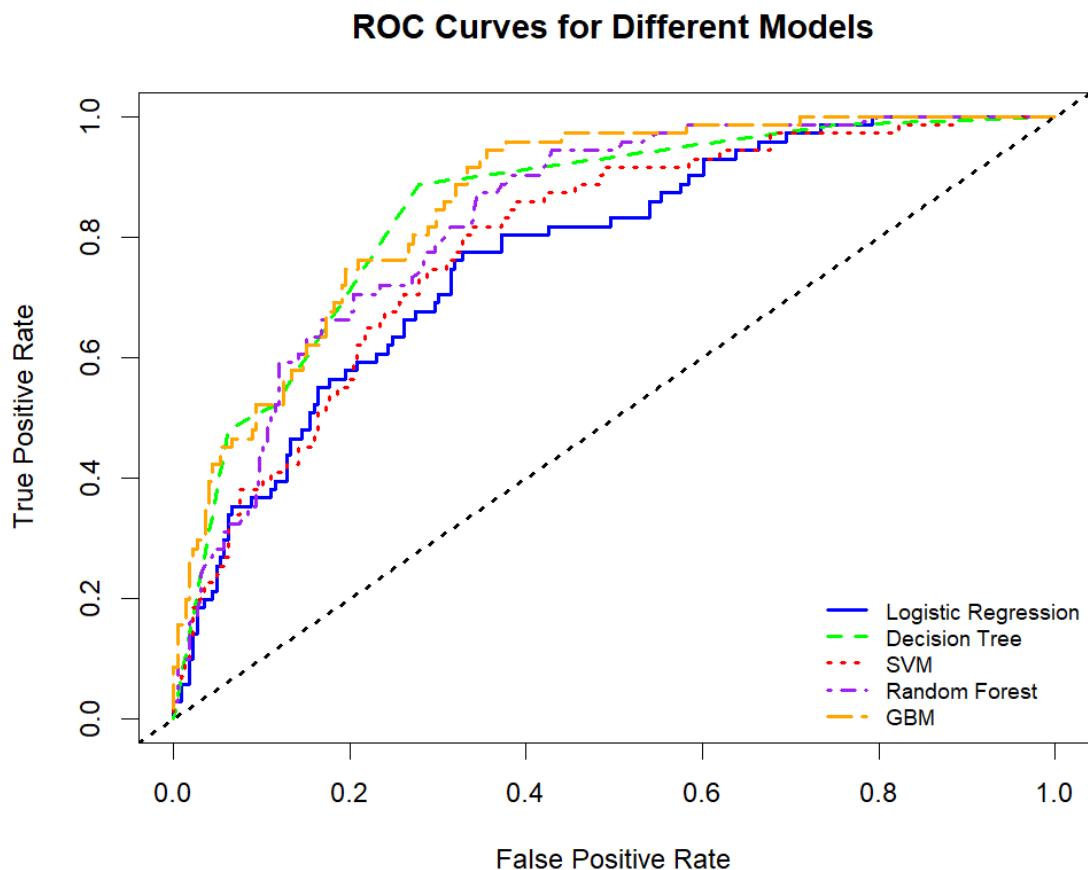


Figure 9: ROC curve for each machine learning model

In evaluating the model's predictive performance, the ROC curve and the AUC provide insightful metrics. The ROC curve shows the trade-off between sensitivity (True Positive Rate) and False Positive Rate (1-specificity) across different thresholds. Among the examined models, the GBM model illustrated the highest discriminatory ability with an AUC of 0.8591. Indicating that the GBM model is adept at distinguishing between individuals likely to have a child and those who are unlikely. Also, the Support Vector Machine demonstrated a competitive performance with an AUC of 0.7897.

Additionally, the Random Forest and Decision Tree exhibited competitive performance with AUC values of 0.8314 and 0.8461, respectively. These models leverage ensemble learning techniques, which enable them to capture potential complex interactions and nonlinearities in the data. Therefore, contributing to their strong discriminatory power. The Logistic Regression model, while more straightforward in its approach, also showed respectable performance with an AUC of 0.7713.

In summary, the models evaluated in this research each demonstrate unique strengths and considerations for predicting the occurrence of having a child within 2021-2023. For accuracy, the Decision Tree and Gradient Boosted Machine performed notably well, each achieving accuracies above 0.82. For precision, Logistic Regression outperformed other with a score of 0.7708, indicating its proficiency in minimizing false positives. Also, when

prioritizing recall, which measures the ability to identify all actual positive cases, Logistic Regression also demonstrated high performance of 0.9823. Once again, for the F1-score, Logistic Regression also outshone the other models, with a score of 0.8638. GBM emerged as the top performer with an AUC of 0.8591, indicating its effectiveness in distinguishing individuals likely to have a child from those who are not.

4. Discussion

4.1 Evaluation of predictive models and key fertility features

This research, aimed at predicting fertility outcomes, specifically the likelihood of having a child within 2021-2023 in the Netherlands, using advanced machine learning models. These models include Gradient Boosted Machines, Random Forests, Logistic Regression, Support Vector Machines and Decision Trees. Each model illustrated varying degrees of predictive capability, with feature importance analysis revealing key insights into the features influencing fertility.

Overall, while the GBM model exhibits the highest AUC, indicating strong overall discriminative ability, the Logistic Regression model demonstrates the best average performance across multiple metrics, particularly excelling in recall and F1-score. Therefore, considering both discriminative ability and balanced predictive performance, Logistic Regression would be the most reliable model for predicting whether an individual will have a child in this context.

The Random Forest and GBM models identified several critical features that significantly impacted the prediction having a child within 2021-2023. The Logistic Regression, SVM and Decision Tree models provided complementary perspectives on this feature importance. Among these, age, household income, partner status, and partner gender emerged as the most important variables.

Age was the most influential feature, illustrating its important role in fertility decisions. Recent studies have consistently highlighted age as a critical determinant influencing fertility choices. Adsera and Ferrer (2013) investigated the impact of age at migration on fertility decisions among immigrants in Canada, highlighting age as a pivotal factor in reproductive timing and family formation (Adsera & Ferrer, 2013). This resonates with this research's findings, which also emphasize the significance of age in shaping fertility preferences and outcomes.

Additionally, household income was another significant feature, highlighting the economic considerations in family planning. Income disparities have emerged as a key area of investigation in fertility research. Birdsall and Jamison (1983) examined income effects on fertility rates in China, illustrating how economic stability influences reproductive decisions (Birdsall & Jamison, 1983). Aligning with this research's exploration of socio-economic factors and their role in shaping fertility preferences among the individuals. Higher household income often correlates with better access to education, healthcare and resources which can influence the decision to have children.

Partner gender and partner status also played vital roles in predicting the outcome, reflecting the social and relational aspects of fertility decision. Research by Sturm et al. (2023) provides further validation, emphasizing how partnership status significantly shapes fertility intentions across various European countries. Their findings highlight that individual's decisions regarding family planning are linked to the quality and nature of a possible relationship.

Moreover, machine learning methodologies have revolutionized fertility research by integrating diverse socio-economic and behavioural predictors. Barnett-Itzhaki et al. (2020) demonstrated the efficacy of machine learning in predicting IVF outcomes, emphasizing variables such as age, lifestyle factors and ovarian reserve (Barnett-Itzhaki et al., 2020). These findings are similar to this research, where similar features emerged as influential factors in understanding fertility outcomes.

In line with these studies, this research confirms the relevance of these factors in predicting fertility outcomes. The consistency of this thesis findings with existing literature improves the validity of the results and emphasizes the utility of machine learning models in demographic research. Furthermore, this research extends the existing body of knowledge by incorporating a comprehensive set of features, including the use of laughing gas, partner status, partner gender, which have been less frequently examined in previous research.

Furthermore, the findings of this research have several implications for policy and practice, particularly in understanding the key factors of influencing fertility. The models illustrated age, household income, partner status and partner gender as significant predictors of fertility outcomes. For instance, policymakers could develop age-sensitive policies that provide tailored support, such as educational programs for younger couples and accessible fertility treatments for older future parents. Economic stability, indicated by the significance of household income, suggests the need for financial incentives, subsidies for childcare and accessible housing and healthcare, to lessen the economic burden of raising children. Also, recognizing the importance of partner status and gender, policies should support diverse family structures and provide if necessary, relationship counselling and education. Promoting work-life balance through parental leave, flexible working hours, and family-friendly workplace environments can help individuals manage career and family aspiration. By implementing these interventions, policymakers can create a supportive and more safe environment that addresses the complex challenges of fertility, promoting healthier and more stable family structures.

4.2 Limitation and further research

While this study provides important insights, it is not without limitations. Firstly, the data set used was relatively small, comprising only 987 individuals. This limited sample size may affect the generalizability of the results to a broader population. Also, the data used in this research may not capture all relevant factors influencing fertility. Which may lead to unobserved variables that could affect the outcomes and therefore bias the results.

Additionally, the machine learning models were used and trained on specific datasets, and their generalizability to other populations or contexts should be validated through further research.

In addition, the machine learning models used in this research, while powerful, are not without their limitations. For example, while Random Forests and GBM are robust in handling non-linear relationships and complex interactions, they can be computationally

intensive and may require substantial tuning to achieve optimal performance. Also, the interpretation of these models can be challenging compared to simpler models like Logistic Regression.

Future research should explore larger and more diverse data sets with a relative low amount of missingness, to validate these findings and improve their generalizability. Additionally, integrating more complete longitudinal data could provide deeper insights into the temporal dynamics of fertility decisions. Additionally, researchers should also consider developing more interpretable models or enhancing the interpretability of complex models to facilitate their application in policy-making and practical settings. Future research could also explore a wider range of machine learning models beyond those used in this research. Techniques such as neural networks or more sophisticated deep learning models might provide additional insights or improve predictive accuracy. By utilizing a broader spectrum of machine learning models, future studies can further enhance our understanding of the complex factors influencing fertility outcomes and develop more robust predictive frameworks.

Also incorporating advanced statistical techniques, such as causal inference methods or Bayesian approaches, could offer deeper insights into the causal relationships underlying fertility decisions. By systematically controlling for confounding variables and exploring causal mechanisms, researchers can strengthen the robustness of their findings and inform more effective policy interventions. Additionally, working together with experts in sociology, anthropology and public health, can help researchers better understand the social and cultural factors that affect fertility decisions. Using qualitative research methods like in-depth interviews or focus groups can also add valuable insights by capturing detailed personal experiences and perspectives that numerical values alone might miss.

Furthermore, using spatial analyses and geographic information systems (GIS) can show differences in fertility rates across different areas and identify regions that need specific policy actions. Mapping fertility trends at various levels can give important insights into where reproductive health issues are most pronounced and help direct resources to where they are needed most.

4.3 Conclusion

This research demonstrates that fertility can be predicted with a reasonable degree of accuracy using machine learning (ML) approaches. The performance of these models however, varies significantly depending on the specific algorithm employed. Among the models tested, the Gradient Boosted Machine (GBM) exhibited the highest Area Under the Curve (AUC), indicating strong overall discriminative ability, whereas the Logistic Regression model showed the best average performance across multiple metrics, particularly excelling in recall and F1-score. While this research is not without its limitations, it does offer interesting insights into the potential of Machine Learning models in demographic studies and public policy planning. The research highlights how different ML algorithms can capture complex patterns to predict fertility, providing a more nuanced understanding of demographic trends.

Appendices

A.1 Feature selection

Current Partner (cf20m024): This variable denotes whether the respondent is currently in a relationship, providing insights into their interpersonal dynamics and familial structure. Partnerships often symbolize a desire for companionship and stability. These factors significantly shape decisions concerning family planning and the prospect of raising children (Birdsall & Jamison, 1983).

Married to Partner (cf20m030): Marriage is a significant life event associated with family formation and long-term commitment. This variable captures respondent's marital status, reflecting long-term relationship commitment. It influences family planning decisions due to societal and legal expectations associated with marriage (Birdsall & Jamison, 1983).

Partner's Gender (cf20m032): Understanding respondents partners genders sheds light on relationship dynamics influencing family planning. Gender roles and expectations vary, affecting decision-making regarding parenthood and family roles. This variable adds depth to fertility prediction models by accounting for relational dynamics.

Primary Occupation (ci20m383): Respondents' primary occupations reveal socioeconomic status, stability and lifestyle choices. Career-focused individuals may delay parenthood for professional growth, while others prioritize family. Occupation influences fertility decisions, reflecting individuals' life goals and priorities (Birdsall & Jamison, 1983).

Alcohol Consumption (ch20m133): Alcohol habits impact various aspects of life, including family planning. Excessive consumption may signal lifestyle choices affecting fertility. Therefore monitoring alcohol use allows for a holistic assessment of factors influencing fertility intentions and outcomes (Sharma et al., 2013).

Soft Drugs Usage, XTC Usage, Hard Drugs Usage & Laughing Gas Usage (ch20m160, ch20m161, ch20m163 & ch20m270): Usage of substances like soft drugs, XTC, hard drugs, and laughing gas provides valuable insights into individuals decision-making processes and lifestyle choices, impacting family planning readiness and attitudes towards risk-taking (Sharma et al., 2013). Soft drug use, such as marijuana, may reflect a preference for personal freedom over traditional values. Thereby influencing fertility decisions. Similarly, the use of substances like XTC(MDMA) and hard drugs like heroin or cocaine may prioritize immediate pleasure-seeking tendencies over long-term commitments (Sharma et al., 2013). In addition, despite the perception of being less harmful than the before, laughing gas usage may indicate a propensity for altered states of consciousness, influencing fertility planning approaches (Van Amsterdam & Van Den Brink, 2022). Thus, understanding the collective impact of drug and alcohol usage adds nuance to fertility prediction models, reflecting the complexities of individual behaviours and preferences in family planning.

Religiousness (cr20m162): Religious beliefs shape values, morals, and life choices, including family planning. More religious individuals may adhere to specific doctrines, affecting fertility intentions. Religiousness reflects cultural norms and personal beliefs, influencing fertility decision-making (McQuillan, 2004).

Paid Work Presence (cw20m000): Employment status is closely linked to financial stability, independence, and social status, all of which affect family planning decisions. Stable employment provides security, making individuals more ready to start a family. Employment status also reflects societal norms and expectations, making it relevant for predicting fertility outcomes (Birdsall & Jamison, 1983).

Migration Background (migration_bg): Cultural and migration backgrounds shape values, beliefs, and attitudes towards family and parenthood. Immigrants and individuals from diverse backgrounds have unique perspectives on family planning. Understanding migration background contributes to a comprehensive understanding of fertility dynamics (Sharma et al., 2013).

Age on December 2020 (age_bg): Age is fundamental in fertility decisions, with rates varying across age groups. Younger individuals may prioritize education, while older ones face biological limitations. The age of an individual at a particular moment is indicative of their current life phase and where they stand in terms of their reproductive journey (Adsera & Ferrer, 2013).

Net Household Income (nettohh_f_2020): Household income reflects financial resources, stability and socio-economic status, influencing family planning. Higher incomes provide access to healthcare, education and childcare, impacting fertility outcomes. Lower incomes may constrain family planning options, affecting fertility intentions (Skakkebaek et al., 2016).

Education Level (oplmet_2020): Education is tied to socio-economic factors influencing fertility decisions, Higher education delays childbearing and influences family size preferences. It also reflects cognitive abilities and exposure to diverse perspectives, shaping attitudes towards family formation (Adsera & Ferrer, 2013).

A.2 Illustration binary encoding

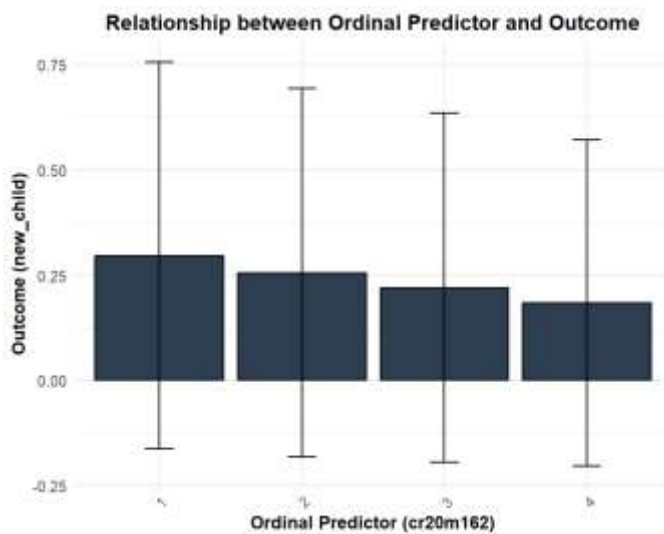


Figure 3: Relationship between Categorical Predictor and Outcome

Figure 5 illustrates the relationship between an ordinal predictor variable 'cr20m162' and the outcome variable 'new_child' by showing the average outcome value within each category of the predictor. Each category represents an amount of religiousness. Each bar represents the mean outcome value for a specific category of the predictor, with error bars indicating the variability around the mean. This visualization helps to assess whether there are significant differences in the mean outcome value across different categories of the predictor.

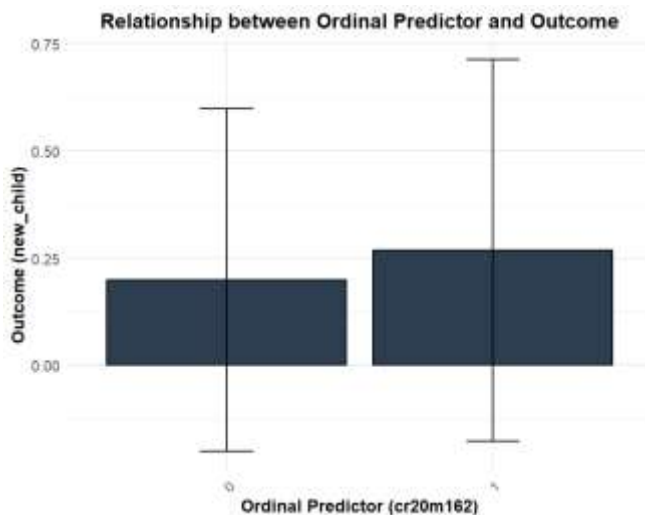


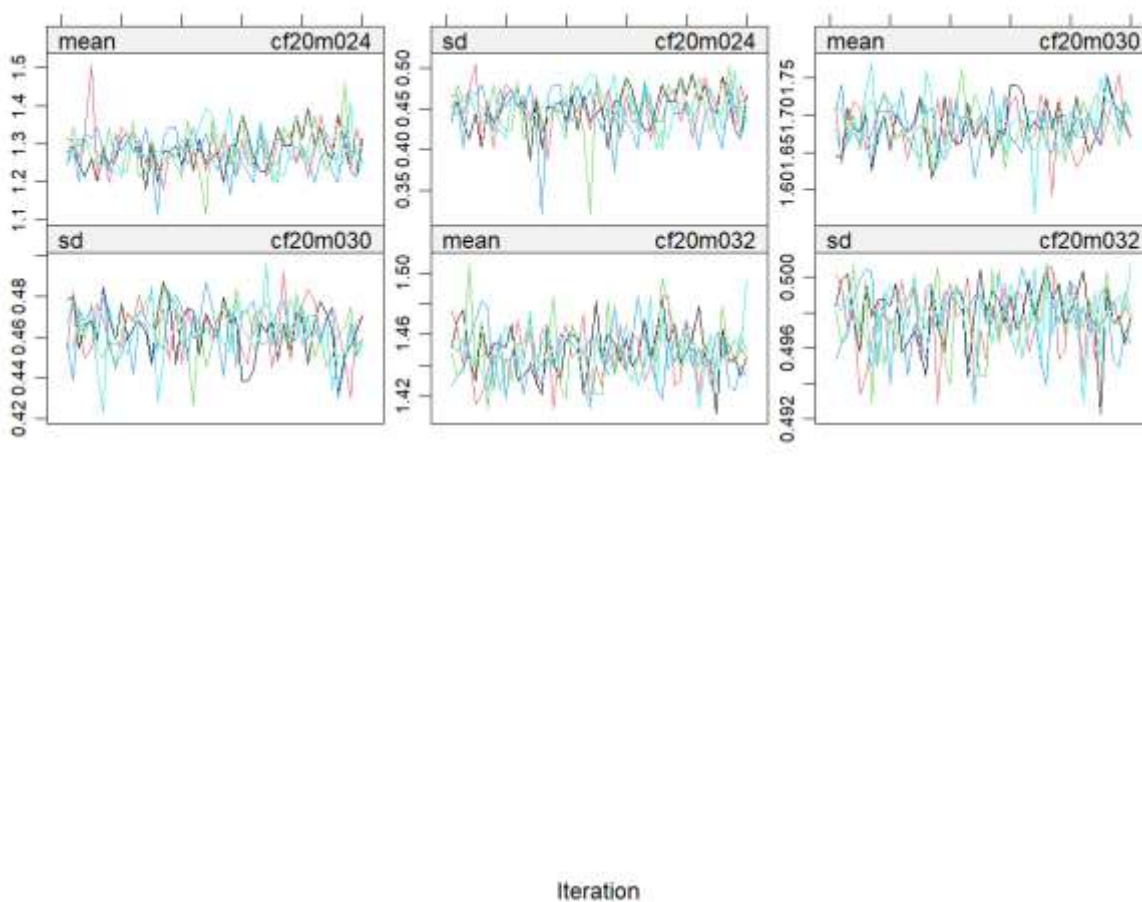
Figure 4: Visualizing the Effect of Binary Encoding on Categorical Predictors

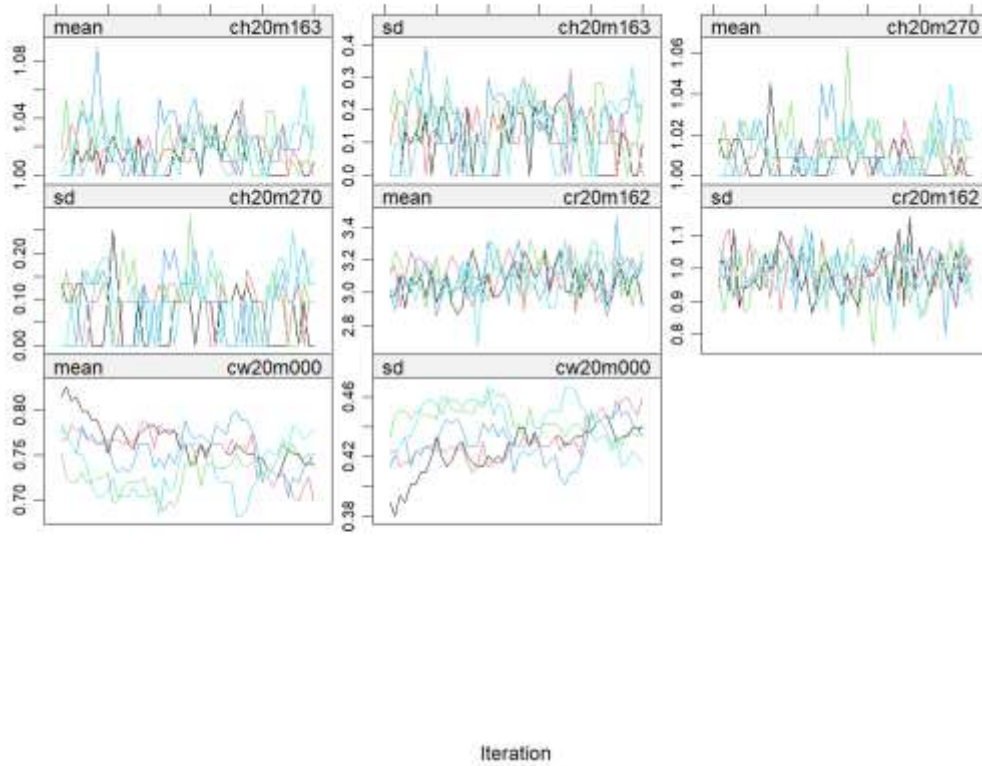
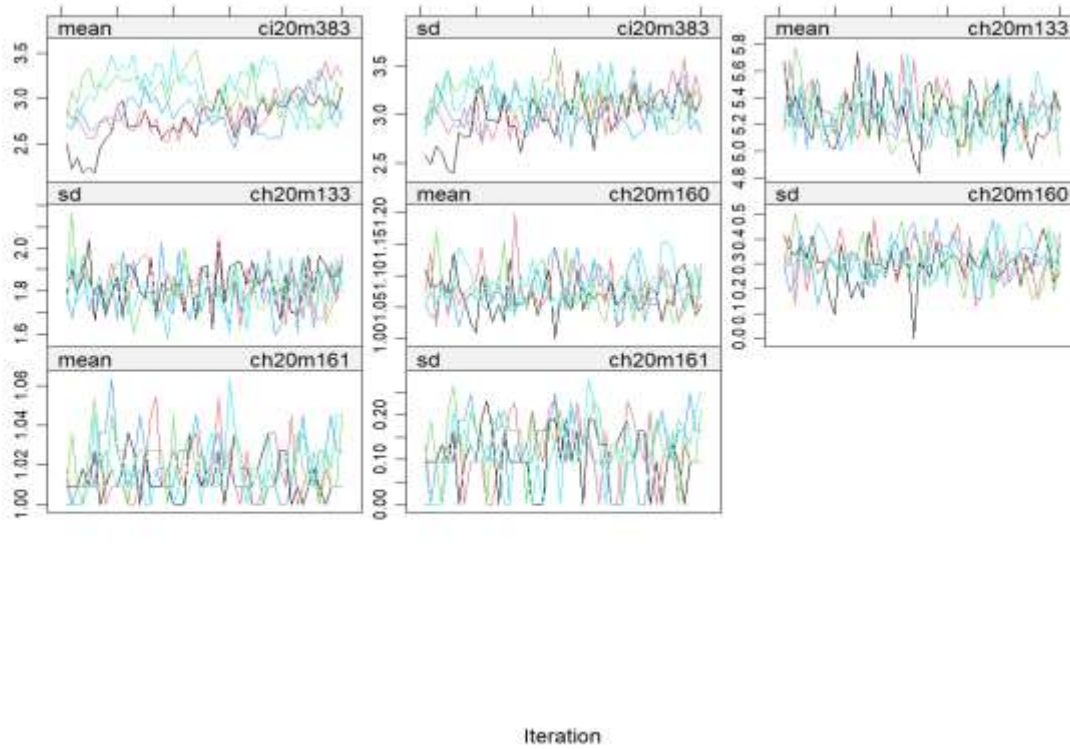
In Figure 6, the outcome of binary encoding applied to the ordinal variable 'religiousness' (Variable name: cr20m162) is presented, contrasting with Figure 5. This variable originally encompassed responses ranging from 1 to 4, each representing different degrees of self-described religiosity: 1 denoting 'certainly religious,' 2 indicating 'somewhat religious,' 3 representing 'barely religious,' and 4 signifying 'certainly not religious.' Through binary encoding, values 1 and 2 were transformed into 1, while values 3 and 4 were converted

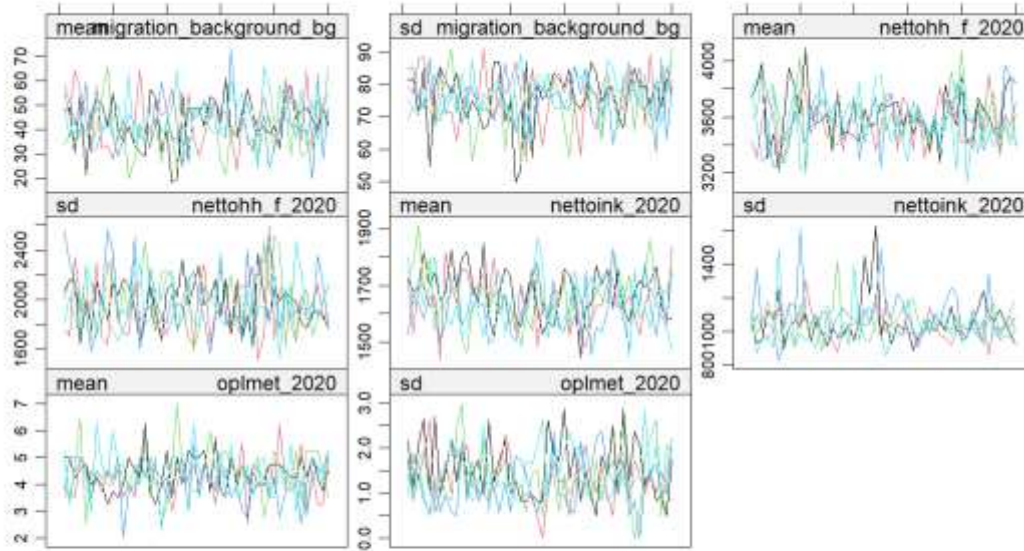
to 0. Thereby dichotomizing the variable based on the level of religiosity reported by the certain individual.

The resulting binary representation facilitates a simplified interpretation of religiousness, distinguishing between individuals identifying as religious (assigned a value of 1) and those identifying as less religious or non-religious (assigned a value of 0). This transformation enables clearer delineation of religious and non-religious categories for analytical purposes.

A.3 Trace plots of MICE







Iteration

References

- Adsera, A., & Ferrer, A. (2013). Factors influencing the fertility choices of child immigrants in Canada. *Population Studies/Population Studies*, 68(1), 65–79. <https://doi.org/10.1080/00324728.2013.802007>
- Ahinkorah, B. O., Seidu, A., Budu, E., Agbaglo, E., Adu, C., Dickson, K. S., Ameyaw, E. K., Hagan, J. E., & Schack, T. (2021). Which factors predict fertility intentions of married men and women? Results from the 2012 Niger Demographic and Health Survey. *PloS One*, 16(6), e0252281. <https://doi.org/10.1371/journal.pone.0252281>
- American Psychological Association (APA) PsycNet. (2021). <https://psycnet.apa.org/record/2022-13889-008>
- Barnett-Itzhaki, Z., Elbaz, M., Buttermann, R., Amar, D., Amitay, M., Racowsky, C., Orvieto, R., Hauser, R., Baccarelli, A. A., & Machtinger, R. (2020). Machine learning vs. classic statistics for the prediction of IVF outcomes. *Journal of Assisted Reproduction and Genetics*, 37(10), 2405–2412.
- Birdsall, N., & Jamison, D. T. (1983). Income and Other Factors Influencing Fertility in China. *Population And Development Review*, 9(4), 651. <https://doi.org/10.2307/1973544>
- Bodin, M., Plantin, L., Schmidt, L., Ziebe, S., & Elmerstig, E. (2021). The pros and cons of fertility awareness and information: a generational, Swedish perspective. *Human Fertility*, 26(2), 216–225. <https://doi.org/10.1080/14647273.2021.1968045>
- Bonaccorso, G. (2018). *Machine learning algorithms: Popular algorithms for data science and machine learning*, 2nd Edition. Packt Publishing Ltd.
- Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. *Family Medicine And Community Health*, 8(1), e000262. <https://doi.org/10.1136/fmch-2019-000262>
- Dahouda, M. K., & Joe, I. (2021). A Deep-Learned Embedding Technique for Categorical Features Encoding. *IEEE Access*, 9, 114381–114391. <https://doi.org/10.1109/access.2021.3104357>
- D'Ambrogio, E. & European Parliamentary Research Service. (2023). Cohesion and ageing society in the EU. In EPRS | European Parliamentary Research Service (Report PE 747.104). [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747104/EPRS_BRI\(2023\)747104_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747104/EPRS_BRI(2023)747104_EN.pdf)
- Elhamraoui, Z. (2021, 14 december). The Support Vector Machine Algorithm in a Fun & Easy Way. Medium. <https://medium.datadriveninvestor.com/support-vector-machine-svm-algorithm-in-a-fun-easy-way-fc23a008c22>
- Freedman, R., Hermalin, A. I., & Chang, M. (1975). Do statements about desired family size predict fertility? the case of Taiwan, 1967–1970. *Demography*, 12(3), 407–416. <https://doi.org/10.2307/2060824>

- Gibreel, A., Maheshwari, A., Bhattacharya, S., & Johnson, N. P. (2009). Ultrasound tests of ovarian reserve; A systematic review of accuracy in predicting fertility outcomes. *Human Fertility*, 12(2), 95–106. <https://doi.org/10.1080/14647270902896256>
- Goyal, A., Kuchana, M., & Ayyagari, K. P. R. (2020). Machine learning predicts live-birth occurrence before in-vitro fertilization treatment. *Scientific Reports*, 10(1).
- Hosmer, D. W., Jr, Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons.
- Jäger, S., Allhorn, A., & Bießmann, F. (2021). A Benchmark for Data Imputation Methods. *Frontiers in Big Data*, 4. <https://doi.org/10.3389/fdata.2021.693674>
- Kumar, A., Boehm, M., & Yang, J. (2017, May). Data management in machine learning: Challenges, techniques, and systems. In *Proceedings of the 2017 ACM International Conference on Management of Data* (pp. 1717-1722)
- LISS panel - LISS panel. (2023, 9 november). LISS Panel. <https://www.lissdata.nl/>
- McQuillan, K. (2004). When does religion influence fertility? *Population And Development Review*, 30(1), 25–56. <https://doi.org/10.1111/j.1728-4457.2004.00002.x>
- Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). Machine learning. In *CRC Press eBooks*. <https://doi.org/10.1201/9781315371658>
- Natekin, A., & Knoll, A. C. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7. <https://doi.org/10.3389/fnbot.2013.00021>
- Nogueira, A. L., & Munita, C. S. (2020). Quantitative methods of standardization in cluster analysis: finding groups in data. *Journal Of Radioanalytical And Nuclear Chemistry*, 325(3), 719–724. <https://doi.org/10.1007/s10967-020-07186-6>
- Sharma, R., Biedenharn, K. R., Fedor, J. M., & Agarwal, A. (2013). Lifestyle factors and reproductive health: taking control of your fertility. *Reproductive Biology And Endocrinology*, 11(1). <https://doi.org/10.1186/1477-7827-11-66>
- Sivak, E., Pankowska, P., Mendrik, A., Emery, T., Garcia-Bernardo, J., Höcük, S., Karpinska, K., Maineri, A., Mulder, J., Nissim, M., & Stulp, G. (2024). Combining the strengths of Dutch survey and register data in a data challenge to predict fertility (PreFer). *Journal Of Computational Social Science*. <https://doi.org/10.1007/s42001-024-00275-6>
- Skakkebaek, N. E., Meyts, E. R., Louis, G. M. B., Toppari, J., Andersson, A., Eisenberg, M. L., Jensen, T. K., Jørgensen, N., Swan, S. H., Sapra, K. J., Ziebe, S., Priskorn, L., & Juul, A. (2016). Male Reproductive Disorders and Fertility Trends: Influences of Environment and Genetic Susceptibility. *Physiological Reviews*, 96(1), 55–97. <https://doi.org/10.1152/physrev.00017.2015>
- Stulp, G., Top, L., Xu, X., & Sivak, E. (2023). A data-driven approach shows that individuals' characteristics are more important than their networks in predicting fertility preferences. *Royal Society Open Science*, 10(12). <https://doi.org/10.1098/rsos.230988>

Sturm, N., Koops, J. C., & Rutigliano, R. (2023). The Influence of Partnership Status on Fertility Intentions of Childless Women and Men Across European Countries. *European Journal Of Population*, 39(1). <https://doi.org/10.1007/s10680-023-09664-5>

Tan, J. (2024). Beyond fertility figures: towards reproductive rights and choices. *Humanities & Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-02608-2>

Van Amsterdam, J., & Van Den Brink, W. (2022). Nitrous oxide–induced reproductive risks: Should recreational nitrous oxide users worry? *Journal Of Psychopharmacology*, 36(8), 951–955. <https://doi.org/10.1177/02698811221077194>

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal Of Statistical Software*, 45(3). <https://doi.org/10.18637/jss.v045.i03>