

Team 32

Casper Collet and Chu Li

**Abstract:**

Everyone probably knows someone around them who has been tackled by divorce, it is simply inevitable. What if I told you that divorce can be predicted and evitable if dealt with the correct issues? In this paper, we are going to try and predict if couples are divorced using different machine learning methods. We will use a dataset with different questions for couples that were answered using a scale from 0-4, 0 being not applicable and 4 being applicable. We will try categorizing the questions into groups, finding the most important variables, and using all variables on their own so we can find out what method results in the highest accuracy and predictability of divorce.

Course: Data Science Lab (ECB3DSL), Academic year 2024/2025

Team: 32

Casper Collet, 2860422, c.b.collet@students.uu.nl

Chu Li

Supervisor(s): Dr. Ulrich Zierahn-Weilage, Dr. Tina Dulam

Word count:

**Statement of Originality**

We, the aforementioned students, herewith declare to have written this document and that we are responsible for the content of it. We declare that the text and the work presented in this document is original and that no sources other than those mentioned in the text and its references have been used in creating it.

Utrecht University School of Economics is responsible solely for the supervision of completion of the work, not for the content.

**Disclosure Statement**

In this project, we have made use of the following Generative AI tools:

We have used these tools for: Helping write the code, grammar checks in the paper and brainstorming.

**Division of Work**

We, the aforementioned students, herewith declare that we have divided the work on this project and this project paper as stated in the following table:

|  |  |
| --- | --- |
| Section and Content | Name and Student Number |
| 1 Introduction | All students |
| 2 Data | [Student Name & Number 1] |
| … | … |
| 5 Conclusions | All students |

Signatures

[copy scans of your signatures here, add your student name and number]

# Introduction

[Your introduction section should discuss the following things: What is the overall topic and what is your specific research question within the topic? Why does it matter, i.e. what do you contribute to the societal and scientific debate (link to the literature!)? This also implies that you give a very short overview over your main findings already here. How do you achieve this (i.e. brief overview over your methodology)?]

# Data

The dataset used in this paper is the divorce\_data dataset from Kaggle.com. It consists of 55 questions ranging from “If one of us apologizes when our discussion deteriorates, the discussion ends” to “I'm not afraid to tell my spouse about her/his incompetence”. Each question is answered with a number from 0 to 4, with 0 being that this question is not true for the couple and 4 being very true for the couple.

First of all, we started cleaning the data. This was not an enormous job to do because the dataset was quite clear and did not contain a lot/none of the N/A variables. Secondly, we tried different methods to see which of them resulted in the highest predictability of the divorce. The questions were divided into 3 different groups with the help of ChatGPT and critical thinking. The groups were named Communication\_data, value\_allignment\_data, and knowledge\_of\_spouse\_data. Communication included questions concerning communication skills, value alignment included questions concerning if the life values of the couples were aligned, and the knowledge of spouse included questions concerning knowledge of the spouse, pretty straight forward.

We also used random forest and logistic regression models to select the most important questions of the dataset. We figured that maybe if we select only the best variables, predictability might go up.

# Descriptive Analyses

[Provide those descriptive analyses, which are necessary to understand the relevant features of your data that matter for the research project]

# Empirical Approach

[Explain the empirical approach of your project – how this should look like strongly depends on your specific type of project]

# Results

[Show, explain, and discuss your results]

# Conclusion

[Conclude: What do you find? How does that contribute to the research question? How does that contribute to the relevant scientific and societal debate? What remains open?]

# References

Last Name 1, First Name 1, and Last Name 2, First Name 2 (year): “Title”, Journal Name, Volume(Issue):PageX-PageY.

# Appendix

[The appendix is optional and you do not necessarily need it. How to decide whether to put something in the appendix or main text? That is a subjective decision. Broadly speaking, the main text should be self-contained and everything that matters for the main line of arguments should be in the main text. The appendix instead is for smaller arguments and checks that are not necessary for the main story of your project.]