

Divorce Prediction Using Machine Learning: An Experimental Study on Questionnaire-Based Relationship Data

Usha, Dr. Anupam Bhatia, Naveen Kundu

MCA Student, Associate Professor, Asst Prof.

Deptt. of Comp. Sc. and Application, Ch. Ranbir Singh University, Jind, Haryana, India

Abstract

In now days divorce is one the major problems in human life. Divorce can affect the social and personal life of a person. The reason of divorce are lack of communication and trust between them. These problems can make small issue and with time they turned into a major problem that lead to be get divorced. The purpose of the study is to make a model that will tell the probability of divorce between couples using machine learning algorithms based on the questionary based data set that have questions related to relationship. Dataset preprocessing The preprocessing of data included the deletion of repeated data and missing values from the dataset. The next step is to divide the data into traning set and testing test 80% and 20 % to improve the modal performance. We use following algorithms in machine learning to predict the divorce these algorithms are Logistic Regression, Decision Tree, and Random Forest. We will compare the results of these algorithms and then show a result that the couple will get divorce or not.

Keywords

Divorce Prediction, Machine Learning, Random Forest, Logistic Regression, Decision Tree, Questionnaire Dataset, Binary Classification

1. Introduction

Marriage is a important part of human life and plays a important role to make life better and happy. A marriage provides a stability in life with emotional support and togetherness with partner. But in today's time there are a lot of problmes in relationship such as less communication, financial problems, misunderstanding and arguments with each other. If these problems are not sorted with time it can convert into a major conflict between couples and lead to be get divorce. Divorce not only ends a relationship it also the effect the individuals of the couple and there children also because due to divorce children have to choose one from mother and father. Therefore, identify the issue in the starting phase will help the couple to solve them and not to get divorce.

In recent years, the divorce rate has increased so much. So the research want to make a platform to predict the divorce. Research are finding the recognizable patterns in between married couples that increase the risk of divorce between them. Machine learning is used to analyze these patterns using structured data and make predictions after analysing the data. We didn't use traditional methods that based on human judgment, machine learning finds hidden relationships factors that make a relationship stable. Previous research shows that the questions based data can give high accuracy uisng various algorithm like

Decision Tree, Logistics Regression, Random Forests, Support Vector Machine, Neural Network and Ensemble Learning method. In this research we used factors such as communication, trust, emotional support and misunderstandings to find the relationship rate of success and failure.

This research shows the use of machine learning in predicting divorce using relationship dataset. The dataset used in the research contains 54 features that describe the features of couples' relationship, such as communication, trust, emotional support and overall behaviour. The main motive of the project is to predict whether a couple will get divorce or not. To prepare the data for applying in the machine learning, duplicate values and missing values are removed from the data set. After that the data is divided into training and testing sets. Machine learning algorithms such as – Logistics Regression, Decision Tree and Random Forest were applied to the data set and then their results are compared with each other. This research follows a step-by-step process, starting from data collection and creation of model training and evaluation. This research shows how machine learning finds patterns and predicts the divorce.

The motivation for this study comes after seeing the increase in divorce rate with time. Increased divorce rate affects the upcoming generation also at this rate of divorce children have become very less interested to get in a relationship. This model will help to secure a relationship by finding the reason of their conflicts. If people will get early knowledge about their problems they will have a chance to save the relationship by working on these problems. This research shows how machine learning can be helpful to solve real life social issues by analyzing dataset and identifying patterns. This study highlights the computational techniques to understand and predict human behaviour.

The main objective of the study is to develop a machine learning model that predicts divorce and check the performance of different classification algorithms on the given dataset. By comparing the results coming from Logistics Regression, Decision Tree, and Random Forest models, this study aims to identify the most effective method for predicting divorce outcomes.

2. Literature Review

Previous research on divorce prediction and union dissolution shows that both social factors and machine learning methods can contribute to understanding marital instability. Fareed et al. proposed an ensemble learning approach for divorce prediction using a questionnaire-based dataset and tested support vector machine, passive-aggressive classifier, and neural network models. Their study reported very high predictive performance and showed that combining models can improve divorce classification results. This paper is especially relevant because it uses the same general problem setting as the present study: structured relationship data converted into a binary prediction task.

Aimran et al. studied divorce prediction among Malaysian women using Decision Tree, Logistic Regression, and Artificial Neural Network models. Their results showed that Decision Tree performed best among the tested algorithms, and they identified important

predictors such as wives' employment status, husbands' employment status, type of marriage, education level, age group, and religion. This study is useful because it confirms that divorce can be predicted from structured demographic and relationship-related variables and that different models may show different performance depending on the dataset.

Arpino et al. applied Random Survival Forests to union dissolution data from Germany and showed that machine learning can outperform conventional regression models in predictive accuracy. Their work also highlighted important predictors such as life satisfaction and the division of housework, while demonstrating that ML methods are able to detect nonlinear and complex interactions among variables. This supports the use of ensemble methods in divorcerelated prediction problems.

Beyond direct prediction studies, several papers explain broader factors associated with divorce. Kapelle found that divorce is linked to long-term wealth disadvantages, especially for housing wealth, and that the financial consequences may persist long after separation. This shows that divorce is not only a relationship issue but also an economic one, making early prediction potentially valuable from a preventive perspective.

Balatonyi examined marriage and divorce rituals during the COVID-19 period and showed that pandemic-related restrictions and changing social conditions affected both marriage and divorce patterns. Her work suggests that external social disruptions can influence family stability and relationship outcomes. Similarly, Hasanudin et al. discussed domestic violence against women and divorce in Indonesia and reported that domestic violence cases increased during the pandemic, with economic stress, jealousy, family interference, and early marriage identified as contributing factors. These studies are important because they show that divorce is often influenced by multiple social and emotional pressures rather than a single cause.

Anchesi et al. reviewed the relationship between parental divorce and children with autism or ADHD. Their findings suggest that divorce risk is not directly tied to a child's diagnosis itself, but rather to the family's ability to cope with stress and behavioral challenges. This adds another layer to the literature by showing that family strain, stress management, and adaptation strategies can also affect marital stability.

Overall, the literature shows that divorce prediction is a multi-dimensional problem. Some studies focus on machine learning performance, while others highlight social, emotional, financial, and family-related causes of divorce. The common conclusion is that relationship stability can be represented through structured predictors and analyzed effectively using machine learning. Based on this foundation, the present study compares Logistic Regression, Decision Tree, and Random Forest on a questionnaire-based divorce dataset to determine which method performs best for this task.

3. Motivation and Problem Statement

The motivation of the research comes from the increase of the divorce rate all around the world and the serious impact over the individuals, their families and children's life also. The

project report highlights the modern lifestyle, work load, financial problems, lack of communication, trust issues and emotional distance make the relationship weaker over time. These problems start as a small issues, but when they are ignored, they can convert into serious conflict between couples and may lead to be divorced.

Another important motivation is the impact of divorce on the children and families. Because after divorce children have to choose one from the parents. It is important for a child to live with both parents because both of them teach him different lesson of life, if one the parents are not there it will leave a great impact on the children thinking. It make the changes like sad, insecure, loss of confidence and behavioral changes. In this report we also notes many relationship issues that are difficult to detect through observation.

For a research perspective, divorce is a suitable problem for machine learning because behaviour of a person in relationship can be represented through input features through question over communication, trust, emotional understanding, satisfaction. Previous studies already shown that machine learning can predict divorce with high accuracy when trained on a survey based dataset. This make the problem both practically and technically suitable for machine learning model's.

The main problem come in this study is that old methods such as counselling, observation, or personal judgement will not show the hidden relationships problem at right time. Human relationships are very complex, and important issues will remain unnoticed until the relationship has already near to an end. Therefore, a model is needed that can learn from the previous relationship data and predict whether a couple will get divorce or not.

This leads to the main problem statement of the paper:

Problem Statement:

The main research question of the study is whether a machine learning model that was trained on relationship question-based dataset can be accurately predict whether a couple will divorce or not. It was important to make a right decision about divorce because it affects their individual also.

To make the right decision, the machine learning model compare the results of three machine learning algorithms: Logistics Regression, Decision Tree, and Random Forest. We will use a cleaned and preprocessed divorce dataset. Our aim is to identify the model that can predict more accurately and show how relationship related factors are used to for early prediction.

4. Methodology

This study will use supervised learning to predict whether a couple will divorce or not. The main aim is to train and compare different machine learning models using questionnaire based dataset based on relationship and find which model give the best accuracy for divorce prediction. The dataset the contain the data of couples who are in living or get divorce to know the prospective about marriage to make a perfect prediction about divorce. The

structure of the study including collection of data and loading, cleaning of data, data preprocessing and scaling, model training, prediction, and results.

4.1 Dataset Description

The data set contain answer related to relationship based questions that were converted in numerical values for analysis. The dataset contain 54 input variable and one output variable. The input variable represent various characteristics of a relationship including communication, trust, emotional understanding, and overall behaviour. The output variable is binary with one value is indicating divorce and other value is indicating no divorce. Because of its structure the dataset is well suited for supervised learning task. Similarly question based dataset were used in previous research for better performance and its is suitable for divorce prediction.

4.2 Data Collection and Preparation

We used a secondary set from a existing divorce prediction dataset. Before use the dataset in the modals we load the data set in python environment to examined its structure and quality. The data was cleaned by deleting the duplicate and missing values from the data set. It is important to give the clean data in the model because if the data was wrong it affect the modal performance and reduce the accuracy and reliability of the divorce prediction.

4.3 Feature Engineering and Feature Selection

After cleaning the data, improvements add to the data set to make better predictions. A new feature named AverageScore, was created to give overall score to the relationship based on the relationship related variables. In addition features section is used to identify the most important features for the prediction task and remove less important features that increase the modal complexity. These steps help in to improve modal efficiency and prediction accuracy. Similarly features engineering and features selection were used in previous studies for prediction studies to finding the most influential features made the model more accurate.

4.4 Data Splitting and Scaling

The dataset was divided into training and testing sets using an 80:20 ratio. The training set was used to train the machine learning models, on the other side the testing set was used to analyze the performance on unseen data. Before model standard scaler was used to make all features values follow a similar scale to reduce complexity in the modal. This steps helps to improve the sability and effectiveness on the learning progress, this helped classification algorithms that work better when all features are all on a similar scale. The data preprocessing method used in this study followed the project report and standard approaches used in divorce prediction research.

4.5 Algorithms Used

We use three diffrent supervised learning algorithms for comparison:

Logistic Regression

Logistic Regression was used because it can predict two possible outcomes: Yes or No.

Decision Tree

Decision Tree was used because it can identify relationship between variables and its results are easy to understand.

Random Forest

Random Forest was used in this as an ensemble method because it combines multiple decision trees and achieve better accuracy than a single tree. Previous research on divorce shows that ensemble method and tree based models are highly competitive for this model in predicting divorce of couples.

4.6 Model Training and Evaluation

I trained all three machine learning models on the training data and then tested them on the test data. I compared all the models using their accuracy percentages. Logistic Regression achieved 84% accuracy while Decision tree performed better than this with 87% accuracy. But Random Forest was the best at 90%. So, out of the three, Random Forest gave the highest accuracy. That means it was the best model in my project. It looks like Random Forest was better at finding patterns in the data and making correct predictions compared to the other two. Also, it handles complex relationship data and reduces overfitting.

4.7 Experimental Design Summary

Basically (In short), I just followed the normal Machine Learning Steps for this project. First I cleaned (processed) the questionnaire-based dataset, and created some new features to make it better (to improve the quality of the data). Then I split the dataset into two parts - one for training and one for testing. I also fixed the values so they were all in the same range. After that, I trained different models and tested them to see how well they worked. Then I compared their results (performance) to find the best one. This way of doing things is good because it's simple, anyone can do it again, it's based on data, and it helps in predicting divorce outcomes.

5. Results and Discussion

The experiment result reveals a high difference in the performance of the machine learning models. In these models, Random Forest achieved the highest accuracy of approx. 90%, Decision Tree shows accuracy of 87%, and Logistic Regression shows accuracy of 84%. This result shows that the Random Forest model is more effective in predicting divorce outcomes from the relationship-based dataset.

Random Forest performed better than any other model because it combines the predictions of many decision trees. This reduces the overfitting and helps the model to make accurate predictions on the new data. Relationship factors are connected with each other in complex

ways, Random Forest can find these patterns more effectively than simple models like Logistic Regression.

The result of this study is similar to previous researches. Early studies shown that tree based and ensemble learning models can perform better than traditional methods for predicting divorce. Researchers are also found that machine learning can find important patterns in question based dataset to give more accurate predictions of divorce.

The results shows that factors such as communication, trust, emotional understanding and overall relationship behaviour plays a important role in predicting the whether a marriage is stable or not. This suggest that divorce risk can be detect from relationship pattern and does not occur mistakenly. Detecting these pattern early can help couples, counselors and researchers to identify relationship problems before they become more serious.

Overall, the results shows that machine learning is an effective choice for divorce prediction when working with structured question based data. Compared to all models, Random Forest proved to be the most accurate model in predicting divorce, making it best choice for the dataset.

6. Conclusion and Future Scope

6.1 Conclusion

This study showed the machine learning can accurately predict divorce using relationship factors such as communication, trust, emotional understanding, and relationship satisfaction. This project follows all the main steps of machine learning such as loading and cleaning the data, selecting important features, scaling the data, train the models and analze their performance. The dataset contain 54 question that are related to relationship and the target variable show whether couple get divorce or not.

I will use three different models for this project . Random Forest was the best one with around

90% accuracy. Decision Tree got 87%, and Logistic Regression was at 84%.This show that Random Forest handles this problem better than other models. Divorce data is probably complicated and very hard to understand, so the ensemble method works well. Other researchers who did similar projects also got the same result .

This project show that using using machine learning to solve social problem is also usefull. The project also confirms the practical usefulness of applying machine learning to social and emotional problems. The system can find patterns in relationship data and help in detect the risk of divorce. This study is not only about technology but also has important for society problems.

6.2 Future Scope

Relationship issues are different according to their lifestyles and their social background. It is important to notice the issues with their reasons and then suggest a solution that will not change their lifestyles. So the more advanced models will notice all these problems then make a solution.

More advanced machine learning models can be used for the prediction of divorce in future, with the help of more advanced models we can get nearly 95% to 100% accuracy in divorce prediction. With more accuracy this prediction model can help to reduce the divorce rate. People will easily solve their relationship issues with the help of this model.

The system can be converted into a mobile application which helps people to solve issues at any time and suggest them how to solve these issues. This can help people in financial way also because counselors' fees were so high to pay for a normal person. But with the app, normal people will easily solve their relationship issues.

Future version of the model has more advanced features than today's models. In advanced models, people will directly ask about their issues and the model can suggest them how to solve these issues. Advanced versions will give more accurate ways to solve their issues and ask questions about their relationship issues and then make a perfect decision.

7. References

- [1] M. M. S. Fareed, et al., "Predicting Divorce Prospect Using Ensemble Learning: Support Vector Machine, Linear Model, and Neural Network," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 1509904, pp. 1–9, 2022.
- [2] N. Aimran, A. Rambli, A. Afthanorhan, A. Mahmud, A. Sapri, and A. Aireen, "Prediction of Malaysian Women Divorce Using Machine Learning Techniques," *Malaysian Journal of Computing*, vol. 7, no. 2, pp. 221–233, 2022.
- [3] B. Arpino, M. Le Moglie, and L. Mencarini, "What Tears Couples Apart: A Machine Learning Analysis of Union Dissolution in Germany," *Demography*, vol. 58, no. 5, pp. 1613–1638, 2021.
- [4] N. Kapelle, "Time Cannot Heal All Wounds: Wealth Trajectories of Divorcees and the Married," *Journal of Marriage and Family*, vol. 84, no. 5, pp. 1384–1405, 2022.
- [5] J. F. Balatonyi, "Ways to Get and End Marriage: Relationships between Marriage and Divorce Rituals during the Coronavirus," *Religions*, vol. 15, no. 1, pp. 1–17, 2024.
- [6] O. S. M. Hasanudin, et al., "Phenomena of Domestic Violence Against Women and Divorce in 2020–2022 in Indonesia: An Islamic Perspective," *Al-Manāhij: Jurnal Kajian Hukum Islam*, vol. 17, no. 2, pp. 271–286, 2023.
- [7] S. D. Anchisi, F. Corallo, M. Di Cara, A. Quartarone, R. Catalioto, F. Cucinotta, and D. Cardile, "Autism and ADHD: A Literature Review Regarding Their Impacts on Parental Divorce," *Children*, vol. 10, no. 8, pp. 1–16, 2023.