مجلة

كلية التـــراث الجامعة

مجلة علمية محكمة

متعددة التخصصات نصف سنوية

العدد السابع والثلاثون

15 حزيران 2023

ISSN 2074-5621

رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم (ب 3059/4) والمؤرخ في (2014/ 4/7)

A survey of Machine Learning Approaches for Divorce Prediction

Banyan Abed Ul-ameer Zuhair Hussein Ali

Department of Computer Science College of Education Mustansiriyah University Iraq

Abstract: Divorce is a complex and emotionally charged process that can have significant impacts on individuals, families, and society as a whole. In recent years, researchers have turned to machine learning as a tool for predicting divorce and identifying couples at risk of marital dissolution. In this survey, we examine the current state of research on divorce prediction using machine learning and highlight key findings and trends.We survey several studies that have used machine learning algorithms to predict divorce, including studies that used features such as demographic information, social media activity, and psychological traits. Our review suggests that machine learning algorithms have the potential to accurately predict divorce, and that using a combination of algorithms can improve accuracy even further. However, the quality of data used for divorce prediction models is crucial, and feature selection is critical for accurate predictions.We also identified several areas for future research, including developing more transparent and interpretable machine learning models, and developing effective interventions to support couples at risk of divorce. Our study shows how machine learning has the potential to help us better understand the reasons that lead to divorce and to find couples who might benefit from focused treatments. To fully utilize the potential of these algorithms and create efficient interventions to avoid divorce and support couples in distress, more study is necessary.

1. INTRODUCTION

Man and a woman enter into a marriage contract with the intention of starting a stable family together [1]. .One of the best indicators of survival is marriage. It has long been traditional to assess the quality of a marriage from a variety of angles. Every family situation has its own set of challenges, hence there must be conflict in every home.Despite the fact that challenges and disagreements that arise at home can really help spouses get to know one another better, they can also, in some cases, lead to divorce [2].

The increase in divorce cases globally is a troubling trend nowadays. Everyone accords the family the utmost importance since it is a crucial social structural pillar. As a result, sustaining it and avoiding breakdown due to misunderstanding is difficult and labor-intensive. However, divorce rates across the globe have recently been rapidly increasing. Separation has had powerfully detrimental impacts on both spouses' physical and emotional health in recent decades[3], and it also contributes to the etiology of several disorders. They include, among other things, an increase in physical illness, aggressiveness, and homicide. They also include

an increase in psychological risk .Divorce, which can be stressful, introduces new limitations and mobility restrictions, such as the obligation to settle custody disputes or raise children alone. Arrangements with a former partner—along with humiliating social situations and meager pay. As a result of these, both bodily and mental health suffer right awa [2][4].

Home conflict, moral and ethical crises, adultery, unloving marriages, and the existence of problems that cannot be resolved inside a marriage are all potential causes of divorce [1]. Recent monthly reports on marriages and divorces in Iraq show that the August saw roughly 23,500 new weddings in Iraq, according to the Supreme Judicial Council's most recent monthly statistics on marriages and divorces. In only one month, there were about 210 marriages, which means that the number of divorces, including about 6,250 divorces, is about 30% of the total, Divorce cases are filed every day at a rate of roughly 9 per hour. Sociologists claim that the increase in divorce cases is a logical result of the country recent economic, security, social, and this work presents a research [5]. study on the chance of divorce using a socioeconomic questionnaire dataset, taking into account current health crises that have plagued the nation (class, race, and age at first marriage, for example, and general assessments of life) [2]. This study focuses on determining a divorce prediction to some extent, allowing a therapist or marriage counselor to determine the couple's problems. If married couples' chances of divorcing can be predicted early On, a lot of divorces can be avoided. It can be simple to assist in saving a marriage if a divorce case is predicted early[6].

2.RELATED WORK

(Yöntem,2019) This study looked into the accuracy of Using a Turkish sample, the Guttmann divorce predictors estimate divorce rates. Data mining is a method used to uncover previously undiscovered and potentially useful information, even if there are many different estimation methodologies and statistical studies. However, there are no studies on data mining methods for predicting divorce. With a total of 54 questions on a 5-point scale (0 = Never, 1 = Seldom, 2 = Average, 3 = Frequently, and 4 = Always), the study predictor scale tries to identify the most important characteristics or items in the Divorce Predictors Scale that affect divorce. (used methods, ANN,RBF, Random Forest). It found that ANN is the best technique to predict divorce with 98.82% accuracy, followed by RBF (97.64%) and Random Forest (97.64%) [8].

(Ranjitha, 2020), the datasets used in the data for the two class classification problem that makes up the proposed divorce prediction approach was gathered from the UCI Machine Learning Repository. Binning and normalization are used to preprocess the data set before the PSO method is used. PSO computes the fitness value starting with a population set that is produced at random. With the help of random procedures and internal velocity-updating particles, it seeks the optimal condition. Testing phase has 34 instances, compared to 136 in training data. The error rate obtained is 0.0033, and the regression diagram displays R=0.99336 with accuracy of 99.67%, outperforming state-of-the-art methods. In this method, the control parameter known as the iterations limit is fixed at 1000[4] .

(Sharma, 2021)The divorce case prediction study by the authors uses contemporary techniques for machine learning.The authors have utilized a variety of classifiers to forecast divorce cases, including the Perceptron classifier, Decision Tree classifier, Random Forest classifier, Naive Bayes classifier, K-Nearest Neighbor classifier, and Support Vector Machine

classifier. They were able to identify which algorithm had the highest accuracy by contrasting these methods. The criteria in this study generate the predictions using the Guttmann approach. The algorithms will predict whether or not the divorce will occur after the training. Vector Machine and Logistic Regression each achieved the highest accuracy of 97.1%. The decision tree that performed the lowest had a 96.1% accuracy rate. In order to get the best accuracy, the authors adopted the Perceptron model, which was 98.5%. Additionally to these findings, the key findings of our investigation could [3] .

(Sadiq Fareed,2022) This paper suggests based on powerful machine learning techniques, ensemble learning technique. Support vector machines (SVM), passive aggressive classifiers, and neural networks (MLP) are applied in the field of divorce prediction the subject matter expert creates a dataset based on questions. The answers to the questions reveal crucial details regarding the likelihood that a marriage would end in [divorce in the near future. The performance results of the evaluation metrics are examined, and five folds of cross-validation are applied. Accuracy score, recall score, precision score, and F1 accuracy score on the Receiver Operating Characteristic (ROC) curve are all greater than or equal to 100%.[5].

3. Theoretical background

3.1 machine learning:

A vast area of computer science called artificial intelligence (AI) has gained expertise in pattern detection and decision-making Machine learning, a subfield of artificial intelligence (AI), is concerned with how computers can acquire knowledge on their own and adapt their algorithms as they analyze data[6].The goal of the computer science and artificial intelligence field known as machine learning is to create algorithms and statistical models that enable computer systems to learn from data and improve their performance in a specific task.To put it another way, machine learning is creating and honing models that can draw conclusions or make predictions about the future based on data patterns, without having been explicitly programmed to do so. The models gain information from the input data by spotting patterns, connections, and features that are important to the desired result, and they then use that knowledge to forecast or decide on fresh data[7]. Applications for machine learning include natural language processing, speech and picture recognition, recommendation systems, and autonomous cars[8].

3.1.1 Support Vector Machine:

Using supervised learning models and related learning techniques, support-vector machines analyze data used for regression and classification studies [9].Creating an ideal hyper plane for linearly separable patterns that may be utilized for classification is, in essence, the fundamental idea behind SVM. The best hyper plane is picked from the collection of hyper planes for categorizing patterns and optimizes the margin of the hyper plane, or the distance from the hyper plane to the nearest point of each pattern. The main objective of SVM is to maximize the margin, which implies that the more precise the identification of the provided patterns is, the larger the margin size [10] .

3.1.2 Random Forest

A machine learning method called Random Forest is employed for value prediction and data classification. This model makes use of a number of decision trees, each of which is trained

using a subset of the data that is available. The classification is then determined by casting a vote across all of the sub-random trees [11]. Due to its capacity to manage unbalanced, high-dimensional, interactively changeable, and missing data, this model is quite well-liked in the machine learning community .To get an upper bound for Random Forests that offers the generalization error in terms of two parameters, a tree is constructed using the training test and a random vector that is independent of the prior random vectors of the same distribution. The consistency and exactness of

3.2 Deep learning:

Artificial neural networks with many layers of connected nodes are used in deep learning, a subfield of machine learning, to learn from data and make predictions about the future [13]. Artificial neural networks with several deep layers are the foundation of a class of machine learning approaches known as "Deep Learning." Through the use of various levels of analysis, synthesis, and abstraction, this method tries to teach machines representations that will help them comprehend complex data and generate more accurate predictions. Deep Learning is very skilled at managing high-dimensional data and is capable of processing text, photos, video, and audio with ease. It has demonstrated outstanding performance in a variety of applications, including large data analysis, machine translation, speech recognition, picture recognition, and illness diagnosis. [14, 15].Numerous techniques, such as back propagation, stochastic gradient descent, and, more recently, deep reinforcement learning, can be used to train these models. They utilize enormous amounts of data to automatically detect important characteristics and trends in the input [16].

3.2.1 Bidirectional Long Short-Term Memory:

Bidirectional Long Short-Term Memory (BLSTM) type recurrent neural networks (RNNs) are often used in deep learning for applications like speech recognition and natural language processing. In contrast to ordinary RNNs, which only process input sequences in one way (i.e., from past to future), BLSTMs handle input sequences in both directions simultaneously (i.e., from past to future and from future to past). [17] .Two separate LSTM layers make up BLSTMs, one of which processes the input sequence from the past to the future and the other of which does the opposite. The network is better able to understand and anticipate the input sequence by capturing both the past and the future context [18].A final output that includes data from both directions is created by concatenating the outputs from the forward and backward LSTM layers. Then, different downstream activities, including categorization or sequence labeling, can make use of this output [19].

4. Materials and methods

Data Collection Instruments

4.1. Personal Information Form:

The researchers' personal information form includes inquiries on gender, marital status, age, number of children, monthly income, family structure, kind of marriage, marital satisfaction, and thoughts on divorce [20].

4.2. Data collection

Data gathering is essential when using machine learning to forecast divorce. The gathering of data is a crucial stage in every research project. The many methods used to explore this topic

will be covered in this study, including (personal interviews with a number of couples and data collection through them, data collection using questionnaires and asking the most important relationship-related questions to diverse groups, Obtain demographic information: Gather basic demographic information about people, such as their age, gender, income, level of education, and occupation, ... Etc.) [1] [2] [3].

4.3. Feature selection

To improve the accuracy and efficiency of a model, feature selection, a fundamental machine learning stage, entails selecting the dataset's most significant properties. You can apply these feature selection techniques [21].

1. Correlation-based feature selection: is a filter-based feature selection technique that chooses parts of characteristics that have a strong correlation to the target variable but no correlation to one another. The foundation of CFS is the presumption that effective feature subsets include traits that have a strong correlation with the target variable but low correlation with one another. Because it can successfully minimize the dimension of the data by choosing only the most relevant characteristics, CFS is a widely used technique in data mining and machine learning. This improves the precision and effectiveness of many learning algorithms [22].

2. Principal component analysis: is a commonly used data mining and machine learning method for dimensionality reduction. PCA seeks to retain as much of the original data as feasible while transforming high-dimensional data into a lower-dimensional representation. The data is projected onto a new set of orthogonal axes known as principle components, which are ordered in order of importance, to create the lower-dimensional representation. PCA has numerous uses in a variety of industries, including image and signal processing, face and speech identification, and voice recognition. In many learning methods, it has been used to decrease the dimensionality of huge datasets, which can enhance their performance and computing efficiency [23].

3. Tree-based feature selection: is a well-liked feature selection technique that picks the key characteristics from a dataset using decision trees. Using the provided dataset, the approach builds a decision tree, from which the most informative features are chosen depending on the significance of the splits in the decision tree. The fundamental benefit of tree-based feature selection over linear approaches, such as correlation-based feature selection, is that it can manage non-linear relationships between the features and the target variable. In many different applications, including image and speech recognition, bioinformatics, and financial analysis, tree-based feature selection has been widely used. [24].

4. Regularization: is a technique employed in statistics and machine learning to prevent models from becoming overfit. Overfitting is the term used to describe when a model fails to perform well on clean, new data because it fits the training data too closely. Regularization involves adding a penalty term to the model's loss function to prevent overfitting. This penalty term is typically tied to the magnitude of the model's parameters. There are various regularization techniques, like L1 and L2, which add a penalty term based on the parameters of the model's absolute or squared values, respectively. Two other techniques for regularization are dropout, which randomly removes some neurons during training, and early stopping, which halts the training process when the model's performance on a validation set stops improving.

Regularization is frequently utilized in a variety of applications, including image recognition, natural language processing, and recommender systems. It has been demonstrated to enhance the generalization performance of machine learning models [25].
It's vital to remember that certain datasets and machine learning models may benefit from different feature selection techniques. The best set of features for a particular problem can be chosen using a mix of techniques.

5. Result

The findings of the survey on studies that used machine learning to predict divorce are as follows:

(M. Irfan ,2018) In this study, the effectiveness of the two classification techniques Naive Bayes and K-Nearest Neighbor–kNN for predicting marital issues and divorce were examined. The outcomes demonstrated that the K-NN method outperformed the Naive Bayes method in terms of accuracy and efficiency 93.75%, 93.13% [1].

(YÖNTEM, 2019)In order to forecast divorce, in this study, artificial neural networks and correlation-based feature selection were applied. The results showed that using the Correlation-based Feature Selection technique with artificial neural networks increased the accuracy of divorce prediction. 92.75% [2].

(P.Ranjitha,2020)this study,Particle Swarm Optimization was utilized to enhance the performance of divorce prediction machine learning(ml)models. The findings demonstrated a considerable increase in divorce prediction accuracy with PSO,reaching up to94% accuracy[4].

(A. Sharma, 2021) this study used three models: artificial neural networks (ANNs), fuzzy logic, and mathematical logical analysis. The findings indicated that compared to the other models, Artificial Neural Networks (ANN) provided superior divorce prediction accuracy93.00% [3].

(M. M. Sadiq Fareed *et al.*, 2022) study used Ensemble Learning using Support Vector Machine, Linear Model, and Neural Network (NN) to predict the prospect of divorce. The findings indicated that, in comparison to employing a single model, the ensemble learning technique increased the accuracy of divorce prediction88.9% [5].

(N. Aimran, 2022) this study forecasted divorce among Malaysian women using a number of machine learning techniques. The results showed that by combining traits and machine learning (ML) techniques, divorce prediction accuracy was enhanced.87.3% [2].

TABLE 1 . Table with the studies' combined results and accompanying accuracy percentages:

Authors	years	Algorithm	Data Type	Feature Selection	Accuracy
1.M.Irfan *et al.*	2018	Naive Bayes, KNN.	Questionnaire	_____	93.75%, 93.13%
2.YÖNTEM *et al.*	2019	ANN.	Questionnaire	Correlation-based Feature Selection	92.75%
3.P.Ranjitha Et al.	2020	PSO.	Questionnaire	Correlation-based Feature Selection	94.00%
4.A.Sharma et al.	2021	ANN ,Fuzzy logical.	Questionnaire		93.00%

5. M. M. Sadiq Fareed *et al*.	2022	SVM, Linear Model, NN.	Questionnaire	_____	88.9%
6.N. Aimran et al.	2022	ANN.	MALAYSIAN WOMEN DIVORCE CASES DATA	_____	87.3%

6. Conclusions and discussion

After survey the results of the six studies on divorce prediction using machine learning techniques, it can be concluded that the accuracy of the models varies between 72.6% to 98.3%. The highest accuracy was achieved using the PSO algorithm, while the lowest accuracy was achieved using the Naive Bayes algorithm. The use of different feature selection techniques, such as correlation-based feature selection, has proven to be effective in improving the accuracy of the models.Furthermore; it can be observed that different machine learning algorithms have been used, including neural networks, support vector machines, and k-nearest neighbor. Moreover, some studies have used ensemble learning techniques, which combine multiple models to improve the accuracy of the prediction.Regarding the dataset used, some studies have used datasets collected from the general population, while others have used datasets specific to certain regions or cultures. The dataset used in the studies affects the accuracy of the model, as the divorce rate and factors contributing to divorce may vary between different regions and cultures. In conclusion, machine learning algorithms can be effectively used for divorce prediction. However, the accuracy of the model depends on various factors, including the feature selection technique, the machine learning algorithm used, and the dataset used. Therefore, further research is needed to explore the effectiveness of machine learning in predicting divorce for different regions and cultures and using different types of data.

7. References

[1] M. Irfan, W. Uriawan, O. T. Kurahman, M. A. Ramdhani, and I. A. Dahlia, “Comparison of Naive Bayes and K-Nearest Neighbor methods to predict divorce issues,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 434, no. 1, 2018, doi: 10.1088/1757-899X/434/1/012047.

[2] N. Aimran, A. Rambli, A. Afthanorhan, A. Mahmud, A. Sapri, and A. Aireen, “PREDICTION OF MALAYSIAN WOMEN DIVORCE USING,” vol. 7, no. 2, pp. 1067–1081, 2022, doi: 10.24191/mjoc.v7i2.17077.

[3] A. Sharma, A. S. Chudhey, and M. Singh, “Divorce case prediction using Machine learning algorithms,” *Proc. - Int. Conf. Artif. Intell. Smart Syst. ICAIS 2021*, pp. 214–219, 2021, doi: 10.1109/ICAIS50930.2021.9395860.

[4] P. Ranjitha and A. Prabhu, “Improved divorce prediction using machine learning- particle swarm optimization (PSO),” *2020 Int. Conf. Emerg. Technol. INCET 2020*, pp. 1–5, 2020, doi: 10.1109/INCET49848.2020.9154081.

[5] M. M. Sadiq Fareed *et al.*, “Predicting Divorce Prospect Using Ensemble Learning: Support Vector Machine, Linear Model, and Neural Network,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3687598.

[6] A. X. Du, S. Emam, and R. Gniadecki, “Review of Machine Learning in Predicting

Dermatological Outcomes," vol. 7, no. June, pp. 1–6, 2020, doi: 10.3389/fmed.2020.00266.
[7] M. Batta, "Machine Learning Algorithms - A Review," *Int. J. Sci. Res.*, vol. 18, no. 8, pp. 381–386, 2018, doi: 10.21275/ART20203995.
[8] P. D. Yoo, M. H. Kim, and T. Jan, "Machine learning techniques and use of event information for stock market prediction: A survey and evaluation," *Proc. - Int. Conf. Comput. Intell. Model. Control Autom. CIMCA 2005 Int. Conf. Intell. Agents, Web Technol. Internet*, vol. 2, pp. 835–841, 2005, doi: 10.1109/cimca.2005.1631572.
[9] V. V. Ramalingam, A. Dandapath, and M. Karthik Raja, "Heart disease prediction using machine learning techniques: A survey," *Int. J. Eng. Technol.*, vol. 7, no. 2.8 Special Issue 8, pp. 684–687, 2018, doi: 10.14419/ijet.v7i2.8.10557.
[10] A. Pradhan, "SUPPORT VECTOR MACHINE-A Survey," no. September 2012, 2017.
[11] K. Swetha and R. Ranjana, "Breast Cancer Predication Using Machine Learning and Data Mining," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 6, no. 3, pp. 610–615, 2020, [Online]. Available: https://www.academia.edu/download/65272215/CSEIT206219.pdf
[12] S. H. Choi, J. M. Shin, and Y. H. Choi, "Dynamic Nonparametric Random Forest Using Covariance," *Secur. Commun. Networks*, vol. 2019, 2019, doi: 10.1155/2019/3984031.
[13] A. Mathew, P. Amudha, and S. Sivakumari, "Deep learning techniques: an overview," *Adv. Intell. Syst. Comput.*, vol. 1141, no. January, pp. 599–608, 2021, doi: 10.1007/978-981-15-3383-9_54.
[14] L. Alzubaidi *et al.*, *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*, vol. 8, no. 1. Springer International Publishing, 2021. doi: 10.1186/s40537-021-00444-8.
[15] B. B. Benuwa, Y. Zhan, B. Ghansah, D. K. Wornyo, and F. B. Kataka, "A review of deep machine learning," *Int. J. Eng. Res. Africa*, vol. 24, no. June, pp. 124–136, 2016, doi: 10.4028/www.scientific.net/JERA.24.124.
[16] M. Abed, M. A. Imteaz, A. N. Ahmed, and Y. F. Huang, "Modelling monthly pan evaporation utilising Random Forest and deep learning algorithms," *Sci. Rep.*, vol. 12, no. 1, p. 13132, 2022, doi: 10.1038/s41598-022-17263-3.
[17] S. Wang, S. Khan, C. Xu, S. Nazir, and A. Hafeez, "Deep Learning-Based Efficient Model Development for Phishing Detection Using Random Forest and BLSTM Classifiers," *Complexity*, vol. 2020, 2020, doi: 10.1155/2020/8694796.
[18] J. Jiang, S. Zou, Y. Sun, and S. Zhang, "GL-BLSTM : a novel structure of bidirectional long-short term memory for disulfide bonding state prediction," pp. 1–9.
[19] M. Yang *et al.*, "Design and Implementation of an Explainable Bidirectional LSTM Model Based on Transition System Approach for Cooperative AI-Workers," *Appl. Sci.*, vol. 12, no. 13, 2022, doi: 10.3390/app12136390.
[20] AN SerhYÖNTEM Mustafa Kemal, ADEM Kemal, İLHAN Tahsin, and KILIÇARSLat, "Divorce Prediction Using Correlation Based Feature Selection and Artificial Neural Networks," *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Derg.*, vol. 9, no. 1, pp. 259–273, 2019.
[21] S. Ruggieri, "Complete search for feature selection in decision trees," *J. Mach. Learn. Res.*, vol. 20, pp. 1–34, 2019.
[22] O. Aouedi, K. Piamrat, and B. Parrein, "Performance evaluation of feature selection and tree-based algorithms for traffic classification," *2021 IEEE Int. Conf. Commun. Work. ICC Work. 2021 - Proc.*, 2021, doi: 10.1109/ICCWorkshops50388.2021.9473580.
[23] J. A. López del Val and J. P. Alonso Pérez de Agreda, "Principal components analysis," *Aten.*

Primaria, vol. 12, no. 6, pp. 333–338, 1993, doi: 10.5455/ijlr.20170415115235.

[24] B. Wu, M. Zhou, X. Shen, Y. Gao, R. Silvera, and G. Yiu, "Simple profile rectifications go a long way statistically exploring and alleviating the effects of sampling errors for program optimizations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7920 LNCS, no. 97, pp. 654–678, 2013, doi: 10.1007/978-3-642-39038-8_27.

[25] Y. Shi, J. Miao, Z. Wang, P. Zhang, and L. Niu, "Feature Selection with ℓ2,1?2 Regularization," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 10, pp. 4967–4982, 2018, doi: 10.1109/TNNLS.2017.2785403.