Divorce Prediction with Machine Learning: Insights and LIME Interpretability

Md Manjurul Ahsan
School of Industrial and Systems Engineering
University of Oklahoma
Norman, OK 73019, USA
ahsan@ou.edu

Abstract—Divorce is one of the most common social issues in developed countries like in the United States. Almost 50% of the recent marriages turn into an involuntary divorce or separation. While it is evident that people vary to a different extent, and even over time, an incident like Divorce does not interrupt the individual's daily activities; still, Divorce has a severe effect on the individual's mental health, and personal life. Within the scope of this research, the divorce prediction was carried out by evaluating a dataset named by the 'divorce predictor dataset' to correctly classify between married and Divorce people using six different machine learning algorithms- Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Classification and Regression Trees (CART), Gaussian Naïve Bayes (NB), and, Support Vector Machines (SVM). Preliminary computational results show that algorithms such as SVM, KNN, and LDA, can perform that task with an accuracy of 98.57%. This work's additional novel contribution is the detailed and comprehensive explanation of prediction probabilities using Local Interpretable Model-Agnostic Explanations (LIME). Utilizing LIME to analyze test results illustrates the possibility of differentiating between divorced and married couples. Finally, we have developed a divorce predictor app considering ten most important features that potentially affect couples in making decisions in their divorce, such tools can be used by any one in order to identify their relationship condition.

Index Terms—Machine Learning, Statistical Analysis, SVM, KNN

I. INTRODUCTION

Divorce also is known as the dissolution of marriage, which is the process of terminating a marriage or marital union [1]. While it seems as simple as a separation between two people, a divorce may have a significant effect on an individual. A study conducted by the Institute for Population Research [2] showed that around 40% of American children usually experience parental divorce/separation during their childhood and has a long-term effect on their mental health. Parallel research identifies that children from divorced families received less financial support compared to other children during their lifetime. Additionally, those children also felt more distress due to a lack of emotional support from the families [3]. As a result, parental divorce influences child behavior in a negative manner, which leads to anger, frustration, and depression. A divorce could significantly hamper couples, personal and social life. Duncan & Hoffman (1985) and Morgan (1989) [4], [5], showed that individuals' separation

also confronted a variety of stress, financial crisis, reduction in social networks and moving on (Amato, 2000;McLanahan & Sandefur, 1994) [6], [7]. Augustine (2000) found that men were nearly 4.8 times as likely to commit suicide as women. They have used the NLMS data for their study. Their study demonstrated a potential relationship between marital status with suicidal risk and,according to the Centers for Disease Control and Prevention (2017), On average, 1.4 million American attempted suicide. The combined medical and work loss are observed as 69 billion dollars [2]. Apart from this, the relationship between divorce and the individuals driving performance such as accident rate, speeding, failure to yield etc. are also highly correlated [8]. In a nutshell, divorce has a significant effect, which brings damage to various aspects such as social, economic, physical, mental and so on. During the past few decades, the effect of divorce and the reason behind the divorce/marriage, have been studied thoroughly, while very few studies focused on forecasting the divorce based on some specific criteria [9]–[12].

Irfan et al. (2018), uses Naïve Bayes and K-Nearest Neighbor methods to predict whether the couple are more likely to get divorce or not. The study took into account several attributes including the plaintiff's age, age of the defendant, marriage age, child age and so forth. Using those attributes, their computational results show that Naïve Bayes and KNN algorithms can differentiate between divorce and married couples with an accuracy of 72.5% and 57.5% respectively. However, the study negates the importance of feature selection procedure, which is one of the most significant drawbacks of this research. In addition, the lower accuracy of the model constitutes the probability of a less accurate model as a decision making tool. Another study conducted by Wolfinger (2018) uses a parametric sickle model in order to divorce prediction [10].

However, none of the referenced literature [9], [10] mentioned regarding the use of standard scale of divorce prediction, that potentially applied in the medical and psychology field. Mustafa et al. (2019) have used correlation-based feature selection and ANN networks in their research. They used the Divorce Predictor Scale (DPS), which is based on Gottman couples' therapy, which includes two class people: divorce (49%) and married (51%). Their analysis demonstrated

98.23% accuracy with Rbf. One of the biggest drawbacks of this research is that the dataset is highly balanced, which ultimately results in the possibility of higher accuracy with high overfitting risk [11].

Li (2018) uses the Markov logic Networks Based Method to predict Judicial Decision cases [12]. However, the study ignored divorce prediction combined with psychological situations, which is thoroughly studied by some other studies [11].

Ibrahim (2020) uses artificial neural network (ANN) methods to predict whether a couple is going to get divorce or not. The model went through multiple learning-validation cycles until it achieved an accuracy of 100%. Jue et al. (2020) uses three different machine learning algorithms– support vector machines (svm), random forest (rf), and natural gradient boosting (NGBoost)—on divorce prediction dataset. And their preliminary computational result indicates 98.33% accuracy on predicting whether the marriage is reliable or not. Ranjitha et al. (2020) uses particle swarm optimization (PSO) in order to reduce the redundant features that do not contribute to the divorce prediction. However, none of the study results explain in details of feature selection or discuss why their proposed model should be trusted.

While data mining is often employed by the majority of the study in psychology and psychiatry, one cannot conclude that it is adequate for a phenomenon with long-term consequences, such as divorce. Fortunately, if divorce predictors can be used in advance, many of the divorces would be avoided. Taking these opportunities into account, firstly, this paper analyzed six different machine learning algorithms that evaluate whether a couple might get divorce or not, using "divorce predictor dataset" Secondly, top 10 features out of 54 features were identified that highly correlate with the divorce prediction. Thirdly, the prediction probabilities are analyzed using Local Interpretable Model-Agnostic Explanations (LIME). Finally, an desktop based graphical user interface (GUI) application is developed using tkinter and machine learning model that can be accessed and used by individual to determine whether he/she is more likely get divorce in the future, ultimately presenting the opportunity to take necessary steps in saving conjugal life.

II. METHODOLOGY

Figure 1 represents the overall research methodology used in this study.

A. Data collection and preprocessing

We adapted the dataset named by "Divorce predictor dataset" from UCI machine learning Repository which is one of the most popular open sources for dataset [11]. Table I summarizes the data set characteristics in details, that contains 170 instances and 54 attributes. All attributes have been added to index A.

Figure 2 shows a histogram of first attribute with 170 instances. All 54 histograms have been added to Appendix A.
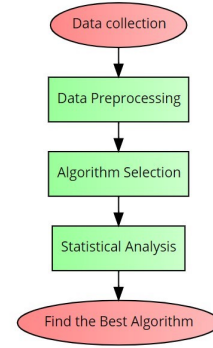

Fig. 1. Experiment Methodology.

TABLE I
DIVORCE PREDICTOR DATASET CHARACTERISTICS

Dataset Characteristics	Multivariate, Univariate
Attribute Characteristics	Integer
Associated Tasks	Classification
Area	Life
Number of Web Hits	40288
Number of Instances	170
Number of attributes	54
Missing Values	N/A
Date Donated	2019-07-24

We dedicated 80% of the data for the training and remaining 20% for the testing purpose. Figure 2 presents a set of representative data used in this study.

We applied six widely used machine learning algorithms: Logistic Regression(LR), Linear Discriminant Analysis(LDA), K-Nearest Neighbors(KNN), Classification and Regression Trees(CART), Gaussian Naive Bayes(NB) and Support Vector Machines(SVM) on "Divorce Predictor Dataset". The experiment was carried out using an office grade laptop with the specification as follows: Intel Core I7, 16 GB RAM. We have used 10 fold cross validation and the computational results are presented as the average score of all folds.

B. Results

Table II-B presents a summary of the performance of all algorithms on "Divorce predictor dataset". SVM outperformed all algorithms in terms of higher accuracy and lower error rate. In contrast, CART showed the worst performance across all measures.

TABLE II
DIVORCE PREDICTOR DATASET CHARACTERISTICS

Algorithm	Accuracy	Error
LR	0.978	0.33
LDA	0.985	0.028
KNN	0.985	0.028
CART	0.963	0.048
NB	0.978	0.04
SVM	0.99	0.02

```
In [47]: df['Atr1'].plot.hist().autoscale(enable=True,axis='both',tight=True)
```
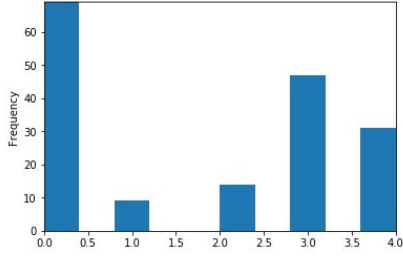
Fig. 2. A sample example of Histogram of first attribute with 170 instances.

Fig. 3. Sample Example Divorce Predictor Dataset.

We further explore the SVM algorithms performance with newly developed dataset by computing accuracy, precision, recall and f-1 score, which presents in Table 3.

TABLE III
DIVORCE PREDICTOR DATASET CHARACTERISTICS

	Precision	Recall	f1-score	Support
0	0.90	1.00	0.95	19
1	1.00	0.87	0.93	15
Avg	0.95	0.95	0.94	34(Total)

C. Explainable AI

Local interpretable model agnostic (LIME) being used to identify important features that helps SVM model to make decision during the prediction. Figure 4 presents top 10 features that helps SVM algorithm to correctly identify the married couples. Using sklearn tool such as selectkbest methods we have reevaluated the top 10 features, which are similar as found using explainable LIME.
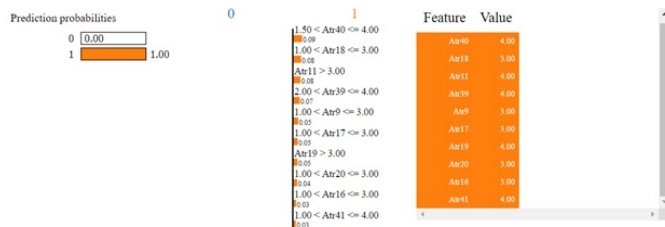

Fig. 4. Divorce predictor application.

Finally, a graphical user interface (GUI) based divorce predictor app is developed using top 10 important features from Gottman couples therapy which can be accessed by individual who are more likely identify their pre-existing relationship status.
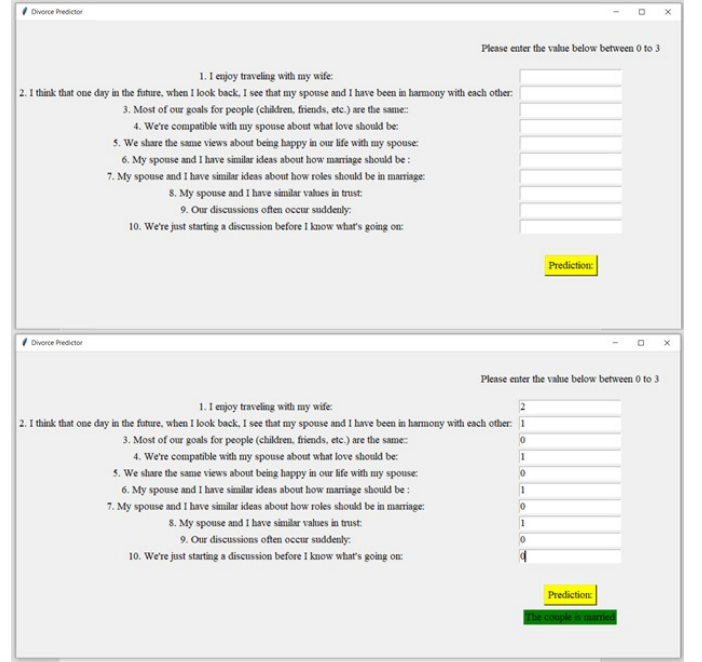

Fig. 5. Gottman couple therapy based app for divorce prediction using Tkinter GUI.

III. DISCUSSION AND POTENTIAL REMARKS

When the overall results of the study are examined, it was observed that the Divorce Predictors Scale developed within the scope of Gottman couples therapy can predict divorce rates with an accuracy of around 99% using SVM. Gottman and Silver (2015) stated that the decisive factor in satisfaction of couples with sexuality, love and passion in their marriages depends on the quality of friendship between husband and wife by 70%. In addition, factors such as value systems of couples, their views of love, perceptions and attitudes about trust are important for creating a shared meaning cite-gottman2014predicts.

We investigated a dataset which contains 54 attributes with two different classes (Divorced, married couples) with 170 instances. . We have used 6 different algorithms in order to predict the probability of whether the couple is more likely getting divorce or not. Our study shows that it is possible to achieve 99% accuracy with SVM. Using scikit-learn se-lectkbest methods and LIME, top 10 features are identified. Finally, a GUI based application is developed for individuals using the python Tkinter tool. Researcher, practitioner and psychiatrist can use such tools to evaluate the patient's current state regarding divorce and can take necessary steps to prevent the divorce. However, the dataset used in this study is highly balanced, ultimately resulting in higher accuracy during the

preliminary computations. In future the current research will focus on applying other machine learning and deep learning approaches to evaluate the study results more in details, and experiment with high dimensional data.

REFERENCES

[1] T. E. of Encyclopaedia Britannica. Divorce. [Online]. Available: https://www.britannica.com/topic/divorce

[2] D. o. S. R. a. t. o. a. A. P. Dr. Yongmin Sun. Short and long term effects of parental divorce. [Online]. Available: https://ipr.osu.edu/short-and-long-term-effects-parental-divorce

[3] S. L. Braver, I. M. Ellman, and W. V. Fabricius, "Relocation of children after divorce and children's best interests: New evidence and legal considerations." *Journal of Family Psychology*, vol. 17, no. 2, p. 206, 2003.

[4] G. J. Duncan and S. D. Hoffman, "A reconsideration of the economic consequences of marital dissolution," *Demography*, vol. 22, no. 4, pp. 485–497, 1985.

[5] M. J. Durcan and P. F. Morgan, "Evidence for adenosine a2 receptor involvement in the hypomobility effects of adenosine analogues in mice," *European journal of pharmacology*, vol. 168, no. 3, pp. 285–290, 1989.

[6] P. R. Amato, "The consequences of divorce for adults and children," *Journal of marriage and family*, vol. 62, no. 4, pp. 1269–1287, 2000.

[7] S. McLanahan and G. Sandefur, *Growing up with a single parent. What hurts, what helps.* ERIC, 1994.

[8] L. McMurray, "Emotional stress and driving performance: The effect of divorce." *Behavioral Research in Highway Safety*, 1970.

[9] M. Irfan, W. Uriawan, O. Kurahman, M. Ramdhani, and I. Dahlia, "Comparison of naive bayes and k-nearest neighbor methods to predict divorce issues," in *IOP Conference Series: Materials Science and Engineering*, vol. 434, no. 1. IOP Publishing, 2018, p. 012047.

[10] N. H. Wolfinger, "An alternative method of forecasting divorce rates based on the parametric sickle model," 2018.

[11] M. K. Yöntem, A. Kemal, T. Ilhan, and S. KILIÇARSLAN, "Divorce prediction using correlation based feature selection and artificial neural networks," *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, vol. 9, no. 1, pp. 259–273, 2019.

[12] J. Li, G. Zhang, H. Yan, L. Yu, and T. Meng, "A markov logic networks based method to predict judicial decisions of divorce cases," in *2018 IEEE International Conference on Smart Cloud (SmartCloud).* IEEE, 2018, pp. 129–132.

APPENDIX A
1. ATTRIBUTE

1) When one of our apologies apologizes when our discussions go in a bad direction, the issue does not extend.
2) I know we can ignore our differences, even if things get hard sometimes.
3) When we need it, we can take our discussions with my wife from the beginning and correct it.
4) When I argue with my wife, it will eventually work for me to contact him.
5) The time I spent with my wife is special for us.
6) We don't have time at home as partners.
7) We are like two strangers who share the same environment at home rather than family.
8) I enjoy our holidays with my wife.
9) I enjoy traveling with my wife.
10) My wife and most of our goals are common.
11) I think that one day in the future, when I look back, I see that my wife and I are in harmony with each other.
12) My wife and I have similar values in terms of personal freedom.
13) My husband and I have similar entertainment.
14) Most of our goals for people (children, friends, etc.) are the same.
15) Our dreams of living with my wife are similar and harmonious.
16) We're compatible with my wife about what love should be.
17) We share the same views with my wife about being happy in your life .
18) My wife and I have similar ideas about how marriage should be.
19) My wife and I have similar ideas about how roles should be in marriage.
20) My wife and I have similar values in trust
21) I know exactly what my wife likes.
22) I know how my wife wants to be taken care of when she's sick.
23) I know my wife's favorite food.
24) I can tell you what kind of stress my wife is facing in her life.
25) I have knowledge of my wife's inner world.
26) I know my wife's basic concerns.
27) I know what my wife's current sources of stress are.
28) I know my wife's hopes and wishes.
29) I know my wife very well.
30) I know my wife's friends and their social relationships.
31) I feel aggressive when I argue with my wife.
32) When discussing with my wife, I usually use expressions such as are you always or are you never.
33) I can use negative statements about my wife's personality during our discussions.
34) I can use offensive expressions during our discussions.
35) I can insult our discussions.
36) I can be humiliating when we argue.
37) My argument with my wife is not calm.
38) I hate my wife's way of bringing it up.
39) Fights often occur suddenly.
40) We're just starting a fight before I know what's going on.
41) When I talk to my wife about something, my calm suddenly breaks.
42) When I argue with my wife, it only snaps in and I don't say a word.
43) I'm mostly thirsty to calm the environment a little bit.
44) Sometimes I think it's good for me to leave home for a while.
45) I'd rather stay silent than argue with my wife.
46) Even if I'm right in the argument, I'm thirsty not to upset the other side.
47) When I argue with my wife, I remain silent because I am afraid of not being able to control my anger.
48) I feel right in our discussions.
49) I have nothing to do with what I've been accused of.
50) I'm not actually the one who's guilty about what I'm accused of.
51) I'm not the one who's wrong about problems at home.
52) I wouldn't hesitate to tell her about my wife's inade-

quacy.

53) When I discuss it, I remind her of my wife's inadequate issues.

54) I'm not afraid to tell her about my wife's incompetence.

The above dataset can be downloaded from http://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set