Proceedings of the International Conference on Artificial Intelligence and Smart Systems (ICAIS-2021)
IEEE Xplore Part Number: CFP21OAB-ART; ISBN: 978-1-7281-9537-7

Divorce case prediction using Machine learning algorithms

Aditya Sharma
Computer Science and Engineering
Jaypee University of Information Technology
Waknaghat, India
vivid.aditya@gmail.com

Arshdeep Singh Chudhey
Computer Science and Engineering
Jaypee University Of Information Technology
Waknaghat, India
arshdeepsingh31052001@gmail.com

Dr. Mrityunjay Singh
Dept. of Computer Science and Engineering
Jaypee University Of Information Technology
Waknaghat, India
mrityunjay.cse045@gmail.com

2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) | 978-1-7281-9537-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICAIS50930.2021.9395860

***Abstract*—The number of divorce cases are increasing very rapidly all over the world. In the last few decades, the number of divorces have gone up from 1 in 1000 to 13 in 1000 in India. Due to this reason, it is a major concern for marriage counsellors and therapists. Therefore, an effective divorce prediction technique is needed that helps a marriage counsellor or a therapist to identify how severe a case is. In this work, the authors present a study on divorce case prediction using the existing machine learning algorithms. The authors have applied the Perceptron classifier, Decision Tree classifier, Random Forest classifier, Naive Bayes classifier, K-Nearest Neighbour classifier and Support Vector Machine classifier for prediction of divorce cases, and determined the best accuracy by comparing these algorithms. The criteria employed in this study makes use of Gottman method to make the predictions.The algorithms after the training will predict whether the divorce will occur or not. This can help the therapist to analyse how tense the situation is between a couple and hence counsel them accordingly. The authors have achieved the highest accuracy of 98.5% with the Perceptron model.**

***Index Terms*—Machine learning,Divorce Prediction, K-Nearest Neighbours, Gottman Method of Relationship Therapy, Logistic Regression ,Support Vector Machine, Descision Trees.**

I. Introduction

Today, the increasing rate of divorce cases throughout the world is a concerning issue.The family is an integral unit of social structure and has utmost value for everyone. Therefore, helping it stay intact and avoiding breakdown due to misunderstanding is a problem which needs a lot of attention. But, recently the number of divorce cases are rising rapidly throughout the world. From last few decades, the number of divorces has gone up from 1 in 1000 to 13 in 1000 in India. [1]. This study lays emphasis on getting a prediction for divorce to an extent, so that it may helpful for a therapist or a marriage counselor to figure out the problem between the couple. The criteria used is the Gottman method of relationship therapy [2]. This method was developed by John Gottman, a Professor of psychology at the University of Washington. The method stated that the problems in a relationship are caused by the factors called "Four horseman". These factors are criticism, defensiveness, stonewalling and contempt [3]. The method aims to increase friendship conflict resolution in a productive manner and have a shared meaning in life. This theory consists of seven basic principles that are:

- *Love Maps* - This principle advocates about being aware about partners feelings and how they view the world and analyse what values, routines and goals they have.
- *Sharing Admiration and Fondness* - This principle suggests that learning to live with difference is an important part of marriage.
- *Turning towards and discussing* - The study suggests that the couples who resolve conflicts instead of running away are more stable.
- *Positive perspective* - This principle states that one should view his/her partner as a friend and not an adversary and be willing to compromise and adapt.
- *Solve problems together* - Couples who solve issues calmly and by accepting each other's vulnerabilities and conversational habits lead to a satisfactory decision.
- *Manging conflict* - Couples who are satisfied report that the majority of their conflicts are present throughout the course of time and are dealt with only as needed. The Gottman method provides a way to manage conflict not resolve them.[3]This is done by negotiating what couples expect from each other in a relationship and how their life goals can affect the relationship. This is done by not critising and taking into consideration that some things will remain the way they are and they need to be accepted by both the parties.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 17,2021 at 11:00:55 UTC from IEEE Xplore. Restrictions apply.

- *Shared meaning* - This principle states that when people in a marriage are able to create a common meaning in life and in their struggles they tend to progress in life , forming a happy relationship.

This paper presents a method to predict divorce cases by combining the relationship theory with machine learning algorithms. The proposed method helps a therapist or a marriage counselor during therapy or provides a way to identify the basic reason behind their divorce. In recent times, a lot of machine learning algorithms have been developed that provide different accuracy for different applications.The algorithms are used in accordance with the set of features which are based on the above mentioned principles .These features are rated by the couples on scale of 0 to 4 , with 4 indicating strong agreement and vice versa. These set of features are used to predict whether the couple will be divorced or not. For a given dataset, we can compare the different machine learning algorithms on the basis of their performance on a particular application. In this study, we have used the Perceptron classifier, Naive Bayes classifier, Logistic Regression classifier, K-nearest neighbours classifier and Support Vector machine classifier and compared their accuracy to find the best match for this scenario. This prediction can help the therapist to know how severe a case is and then suggest the necessary steps to aid the couple .
The organization of this paper is as follows: Section II describes the dataset used for this study, and section III presents the related work. Section IV discuss the proposed approach; the experimental work is discussed in section V. Section VI concludes the work and state its possible extension.

II. Dataset used

In this study, we use the Dataset that is downloaded from UCI Machine learning repository[4].This dataset provides all the necessary information needed for the prediction of divorce cases. The data set consist of 170 couples who answer a set of 54 questions. These questions are in accordance with the gottman couple theory and the answer is on scale of 0 to 4 , where 4 indicates strongly agree and vice versa. These questions act as a set of features for the algorithm and after training is complete the algorithm makes prediction based on answer to these questions. Figure 1. contains a subset of the dataset used. The output is provided in form of 0 and 1 where value 1 represents Divorced and value 0 represents not divorced.
The responses were gathered on a 5 point scale (0 = Never, 1 = Seldom, 2= averagely, 3 = frequently, 4 = Always).
Some of the following characteristics are as follows:

1) We are like two strangers who share the same environment at home rather than family.
2) I relish our vacations with my wife.
3) I relish traveling with my wife.
4) My wife and most of our goals are shared.
5) We don't have time as parents at home.

	Atr1	Atr2	Atr3	Atr4	Atr5	Atr6	Atr7	Atr8	Atr9	Atr10
0	2	2	4	1	0	0	0	0	0	0
1	4	4	4	4	4	0	0	4	4	4
2	2	2	2	2	1	3	2	1	1	2
3	3	2	3	2	3	3	3	3	3	3
4	2	2	1	1	1	1	0	0	0	0

Fig. 1: A small subset of dataset of couples for different attributes which is used in this study

III. Related work

This section presents the work related to our study. In the literature, the researchers have presented few techniques based on machine learning [5], [6], [7]. M. Irfan et al.[5] did a comparison of KNN and Naive Bayes classifiers to predict divorce cases. They used the dataset from the Cimahi Religious Court Office and they achieved an accuracy of 72.5%. Yontem et al. [6] used correlation-based feature selection and artificial neural networks for the prediction of divorce cases. They used the same dataset from UCI Machine Learning and they achieved the highest accuracy of 98.23% with classification. Somya Goel et al. [7] has proposed an algorithm Augur Justice that first identifies the religion and on the basis of that they predict the probability of either winning or losing the case described by the user. They have achieved an accuracy of 84.09% on Hindu Dataset, 55.56% on Christian Dataset and 100% on muslim Dataset.
Although there are no more studies on prediction of Divorce cases using Machine Learning Methods but Machine Learning methods such as classification and estimation are already in use in many studies of psychology and psychiatry. Baca-Garcia et[8] al used data mining techniques that predicted psychiatrists decision to hospitalize in 509 cases of adult suicide attemptors. They used 5 variables like consumption of drug during suicide attempt and family history of suicide attempts etc. and achieved an accuracy of 99% using Forward selection. Song[9] has proposed a method in which they investigated the psychological evaluation data of college students using Machine Learning models like KNN, Naive Bayes and SVM. The best result obtained by them was using SVM that was an accuracy of 79.1%. Erikson[10] in their study have used temporal data mining techniques to determine adverse drug reactions in the inpatient psychiatric population. It will help to detect adverse drug reaction by reducing risk of manual reporting.

IV. Proposed approach

In this study, we make use of different machine learning models which are Perceptron Classifier, Random Forest Classifier, Naive Bayes Classifier, Logistic Regression, K Nearest Neighbour, Support Vector Machine and Decision

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 17,2021 at 11:00:55 UTC from IEEE Xplore. Restrictions apply.

Tree for divorce case prediction as each model has its own pros and cons . After that we compared their accuracy to predict the best among these models. Since so far previous researchers have only compared few machine learning models providing nothing conclusive enough. Thus, comparing these many models can help us know the best possible accuracy of the models for this application. Figure 2. contains a flowchart to illustrate the proposed method for this study. For sake of our model we did various training and testing splits i.e (80-20, 70-30, 60-40) and results of different splits are given.

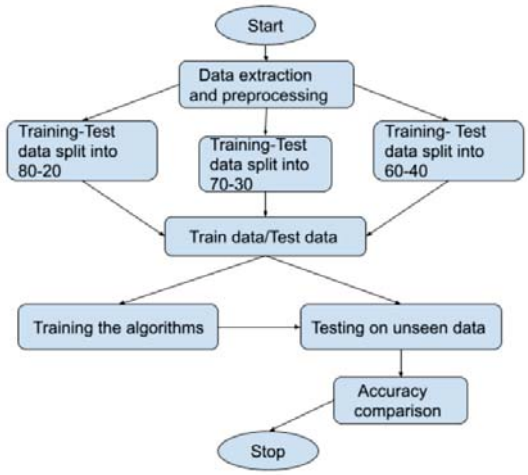


Fig. 2: Proposed methodology flowchart

A. PERCEPTRON

Perceptron is a single layer neural network which acts as a linear binary classifier.[11] The first layer is an input layer followed by a hidden layer and then an output layer.This algorithm takes the input and then multiplies it with their respective weights and these results are then added to get a weighted sum. At last an activation function is applied to the weighted sum to get the desired result. This utilizes supervised learning to train the network.

B. NAIVE BAYES

Naives Bayes is a classifier which uses probability to make predictions. It is based on Bayes Theorem.[12] The features are considered independant and their contribution is considered to be equal. It makes prediction based on whether features used favour a particular outcome or not, based on the result it is allocated a class. It makes use of the equation (1). Naive Bayes is faster as compared to other classification algorithms.

$$P(C/x) = \frac{P(x/C) * P(C)}{P(x)} \tag{1}$$

where (C/x)-Posterior probability
P(x/C)-Likelihood
P(C)-Class prior probability
P(C)- Prediction prior probability

C. LOGISTIC REGRESSION

Logistic Regression is a classification algorithm. It makes prediction of input into suitable classes by calculating a cost function which makes use of a sigmoid function.[13] Sigmoid function is represented by equation (2).The cost function makes use of a hypothesis and the provided output during the training. The hypothesis provides an estimated probability of occurence of an event. The cost function tells how different is the hypothesis from the output. The cost is then minimised using optimization techniques such as gradient descent, conjugate gradient etc . Logistic regression is very easy to interpret and implement and also very efficient to train .

$$S(x) = \frac{1}{1 + exponent^{-value}} \tag{2}$$

D. K-Nearest Neighbour

K-nearest neighbours is a supervised algorithm used for classification. It puts the new data in the category in which it finds a suitable match ,based on the training. It does not make any assumption on underlying data.[14] It allocates a class based on Euclidean distance with previously known data entries.This leads to closest neighbours being allocated the same class. It is a proximity search method between the new case and old case.

E. Support vector machine

Support Vector Machine (SVM) is a supervised machine learning algorithm with a given set of training examples with each of them belonging to some categories, in which it builds a model that assigns new examples to each category.[15] A support vector machine is an alternative view of logistic regression. It uses margin to create a decision boundary hence there is appropriate space between the objects (also known as Large margin classifier).The decision boundary seperates different data points into two classes. There can be many decision boundaries but SVM helps to find a boundary in which there is maximum distance between the data points and hence they are optimally seperated from each other. SVM implements the “one-versus-one” approach for multiclass classification. It performs well on a large dataset but the training time is on the higher side.

F. Decision Tree

A Decision tree as the name suggests has a structure like a tree which is used for classification and prediction.The leaf nodes or terminating nodes hold the class label, the internal nodes of the tree have testing conditions and the branches of trees have outcome for the testing condition.[16] Decision tree is a popular tool in machine learning which is made up of decisions and their probable conclusions including different outcomes for a particular situation, it also provides effective cost handling and resource utility. Decision trees can be very helpful when you have to choose from several courses of action.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 17,2021 at 11:00:55 UTC from IEEE Xplore. Restrictions apply.

V. Experimental setup

Python programming language is used to train the model and all of the experiments have been performed on Google Colaboratory. First of all, Perceptron learning rate was set to be 0.0001 and tolerance was set to be 0.001 and all the other parameters like constant by which updates are multiplied and numbers of CPUs used were set to be default. For Naive Bayes prior probability was set to be none meaning they are not adjusted according to the data and var smoothing was set to the default value as well i.e. of $1 * e^{-9}$. In case of Logistic Regression the tolerance was set to be 0.001 and all the parameters like weight associated with classes and maximum number of iterations taken for solvers to converge was set to default. In case of Support Vector Machine tolerance was set to be 0.001 and other parameters like degree of polynomial kernel function and regularization parameter were set to their default value. Now in the case of KNN, the number of neighbours was set to be 15 and all other parameters like leaf size and weight were set to be default. Lastly for Decision Tree all the parameters like maximum depth of tree and minimum number of samples required to be at leaf node all were set to be default.

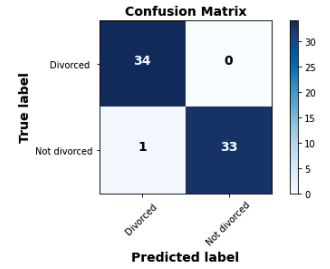
A. Performance metrics

A confusion matrix is a tool that can be used to evaluate the classification model to estimate whether an object is either correct or false. It is a matrix prediction that contains comparison of actual (ground truth) and predicted information. Confusion matrix is represented by 2 classes Positive and Negative. Positive class usually denotes abnormal behavior and negative class usually denotes normal behavior. There are 4 quadrants in Confusion matrix:
True Positive(TP) : It is the output when the model correctly labels a positive class.
True Negative(TN) : It is the output when the model correctly labels a negative class.
False Positive(FP): It is the output when the model incorrectly labels a positive class.
False Negative(FN): It is the output when the model incorrectly labels a negative class.
For calculation the equations used were:

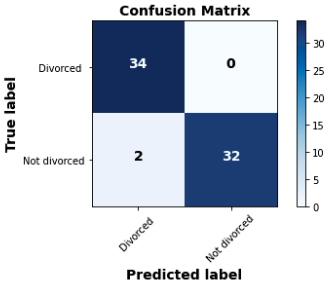
$$Accuracy = \frac{Number\ of\ correct\ values * 100\%}{total\ data} \quad (3)$$

B. Experimental Results

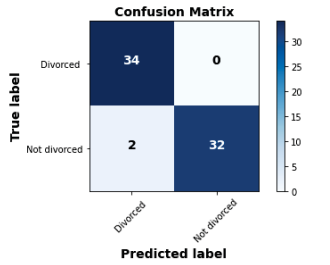
Perceptron, Decision Tree, Naive Bayes, K Nearest Neighbour, Logistic Regression and Support Vector Machine have been applied to the divorce dataset mentioned in the study. The highest accuracy was obtained in a Perceptron with 60-40 training test split. Followed by Perceptron, Naive Bayes, K nearest neighbour, Logistic Regression and Support vector machine performed equally well and the highest performance in all these cases were observed when we did a 60-40 training test split as well. After that, the least accuracy was obtained by the decision tree. We observed that in all the models except the decision tree with a decrease in training and test split ratio the accuracy of the model increased.We have observed that many of our models show somewhat the same kind of accuracy during various train and train splits. This could well and truly be because of lack of dataset as it only contains data of about 170 couples.



(a) Perceptron



(b) Naive Bayes

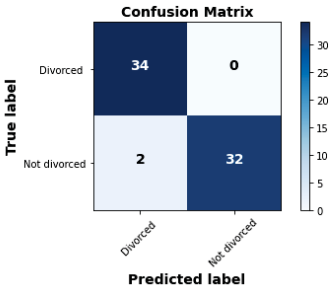


(c) Logistic Regression

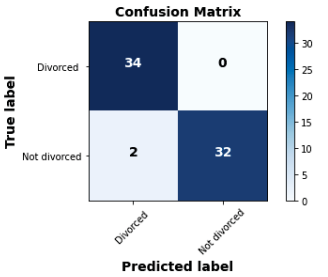
Fig. 3: Confusion matrix for Perceptron , Naive bayes and Logistic regression for 60 -40 split

In Figure 3. and Figure 4. Confusion Matrix of machine learning models used are provided for the results obtained. Firstly the Perceptron model labeled 34 of the couples as correctly divorced and 33 of couples as correctly not divorced. After that Logistic Regression, Naive Bayes, K Nearest Neighbour, Support Vector Machine all labeled 34 of the patients as correctly divorced and 32 of the couples as correctly not divorced. In Last, Decision Tree labeled 33 of the patients as correctly divorced and 32 of the patients as correctly not divorced.
In another detailed comparison performance of the entire machine learning models used for different training and test splits has been given in Table 1. The best performance was achieved by Perceptron with an accuracy of 98.5%. Followed by Naive Bayes, K Nearest Neighbour, Support

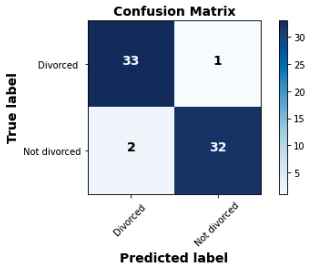
Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 17,2021 at 11:00:55 UTC from IEEE Xplore. Restrictions apply.



(a) K nearest neighbor



(b) Support Vector Machine



(c) Decision tree

Fig. 4: Confusion matrix for K-Nearest Neighbours, Support Vector Machine and Decision Tree for 60-40 training split.

Vector Machine, and Logistic Regression they all achieved the highest accuracy of 97.1%. The worst performance achieved was by a decision tree that was as an accuracy of 96.1%. As a result, we can say Perceptron provides superiority over other Machine learning models.
In addition to these result main outcomes of our study could be:

1) We have used 6 different machine learning models with different training and test splits as compared to other studies where very few models have been used.
2) Since we are working on divorce and married couples. One can easily avail dataset from courts.
3) In our model, we haven't used any feature selection making it very easy to use.

TABLE I: Accuracy of different machine learning models for different training and test splits.

(a) Accuracy for 80-20 split

Machine Learning Models	ACCURACY
Perceptron	0.9706
Naive Bayes	0.9412
K Nearest Neighbour	0.9412
Decision Tree	0.9412
Support Vector Machine	0.9412
Logistic Regression	0.9412

(b) Accuracy for 70-30 split

Machine Learning Models	ACCURACY
Perceptron	0.9608
Naive Bayes	0.9608
K Nearest Neighbour	0.9412
Decision Tree	0.9608
Support Vector Machine	0.9412
Logistic Regression	0.9608

(c) Accuracy for 60-40 split

Machine Learning Models	ACCURACY
Perceptron	0.9853
Naive Bayes	0.9706
K Nearest Neighbour	0.9706
Decision Tree	0.9559
Support Vector Machine	0.9706
Logistic Regression	0.9706

VI. Conclusion and Future work

Early prediction of a divorce case can make it easy to help save a marriage ,as divorce cases are increasing day by day their early prediction can be of great help to the therapist couselling the couple. In this study performance of our model was tested on different training and test splits and variable accuracy was obtained. Performance results show that Perceptron outperformed other machine learning models with the highest accuracy of 98.5%. In further studies classification performance of different machine learning models can be increased with an increase in the number of couples in the dataset. Feature selection can also be used in the future which can help to decrease training time and increase the accuracy of our model. A desktop based tool can also be made that can be used by marriage counsellors and court for prediction of divorce cases.

References

[1] S. Bhatt, "Happily divorced: Indian women are breaking the stigma around separation like never before," https://economictimes.indiatimes.com/magazines/panache/happily-divorced-indian-women-are-breaking-the-stigma-around-separation-like-never-before/articleshow/67704287.cms?from=mdr, January 2019.
[2] E. Lisitsa, "An introduction to the gottman method of relationship therapy," https://www.gottman.com/blog/an-introduction-to-the-gottman-method-of-relationship-therapy/, May 2013.
[3] G. Therapy, "Gottman method," https://www.goodtherapy.org/learn-about-therapy/types/gottman-method, April 2018.
[4] D. M. K. Yöntem, "Divorce predictors data set data set," http://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set, August 2019.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 17,2021 at 11:00:55 UTC from IEEE Xplore. Restrictions apply.

[5] 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.
[6] 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.
[7] S. Goel, S. Roshan, R. Tyagi, and S. Agarwal, "Augur justice: A supervised machine learning technique to predict outcomes of divorce court cases," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*. IEEE, 2019, pp. 280–285.
[8] E. Baca-García, M. M. Perez-Rodriguez, I. Basurte-Villamor, J. Saiz-Ruiz, J. M. Leiva-Murillo, M. de Prado-Cumplido, R. Santiago-Mozos, A. Artés-Rodríguez, J. De Leon *et al.*, "Using data mining to explore complex clinical decisions: a study of hospitalization after a suicide attempt," *Journal of Clinical Psychiatry*, vol. 67, no. 7, pp. 1124–1132, 2006.
[9] S.-M. Bae, S.-H. Lee, Y.-M. Park, M.-H. Hyun, and H. Yoon, "Predictive factors of social functioning in patients with schizophrenia: exploration for the best combination of variables using data mining," *Psychiatry investigation*, vol. 7, no. 2, p. 93, 2010.
[10] R. Eriksson, T. Werge, L. J. Jensen, and S. Brunak, "Dose-specific adverse drug reaction identification in electronic patient records: temporal data mining in an inpatient psychiatric population," *Drug safety*, vol. 37, no. 4, pp. 237–247, 2014.
[11] I. Stephen, "Perceptron-based learning algorithms," *IEEE Transactions on neural networks*, vol. 50, no. 2, p. 179, 1990.
[12] M. M. Saritas and A. Yasar, "Performance analysis of ann and naive bayes classification algorithm for data classification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 7, no. 2, pp. 88–91, 2019.
[13] G. Gasso, "Logistic regression," 2019.
[14] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, "Knn model-based approach in classification," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2003, pp. 986–996.
[15] V. Cherkassky and Y. Ma, "Practical selection of svm parameters and noise estimation for svm regression," *Neural networks*, vol. 17, no. 1, pp. 113–126, 2004.
[16] L. Rokach and O. Maimon, "Decision trees," in *Data mining and knowledge discovery handbook*. Springer, 2005, pp. 165–192.

Authorized licensed use limited to: Univ of Calif Santa Barbara. Downloaded on May 17,2021 at 11:00:55 UTC from IEEE Xplore. Restrictions apply.