International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

Prediction of the Probability of Divorce Using Machine Learning Algorithms

Ankur Kumar[1], Soumayadip Saha[2], Joyitree Mondal[3], Animesh Mitra[4], Dr. Avijit Kumar Chaudhuri[5]

[1,2,3,4] UG-Computer Science and Engineering, Techno Engineering College Banipur, Banipur, Habra

[5] Assistant Professor, Computer Science and Engineering, Techno Engineering College Banipur, Habra

[1]0009-0003-4897-8104,[2]0009-0006-2745-2869,[3]0000-0002-5924-6780, [4]0009-0002-2748-6023,[5]0000-0002-5310-3180

Abstract
For a happy and healthy life, a happy married life is very much essential. But nowadays, divorce cases are increasing rapidly day by day. According to a study, the divorce rate worldwide was 4.08% per 1000 married people in May 2022. For this reason, there is a need for effective prediction of divorce rate which helps the marriage counselor or therapist to understand how serious a case is. For this research purpose, a dataset was collected from UCI repository and contains some data based on the questions asked to the couple and the answers they gave. First, the dataset was cleaned using different ranking methods such as Information Gain, One R, Gain Ratio and ReliefF. Using these ranking methods, the most important fields that really affect the divorce are selected. Then, different classification algorithms such as Logistic Regression, Naïve Bayes, SGD, Decision Tree, Random Forest and Multilayer Perceptron were used and compared to find the accuracy. These algorithms are used first with all fields and then with 6 and 7 fields. These algorithms are used for 50:50, 66:34, 80:20 training/testing split and for 10-fold cross validation. When 7-fields are combined and checked with all algorithms for training-test and 10-fold cross validation, then 100% accuracy was found in Decision Tree, Random Forest and Multilayer Perceptron algorithms.
Keywords: Divorce, Feature Selection, Decision Tree, Random Forest, Multilayer Perceptron, 10-fold cross validation.

Introduction

Marriage is legally and socially sanctioned union that is regulated by laws, rules, customs, beliefs, and attitudes which determines the rights and duties of the couples. Marriage brings happiness because it is a lifelong commitment between two people who love and care for each other. It is a partnership that can weather storms of life and provide stability and support during the good times and bad. Marriage brings happiness because it is built on trust, mutual respect, and communication. When these things are present, couples can work through anything that comes their way otherwise there are chances of Divorce. Divorce may have a significant emotional, financial, and social impact on individuals and families. It can lead to feeling of sadness, anger, and grief, as well as a sense of loss and trauma. Children of divorced parents can also experience emotional difficulties, such as anxiety and depression. Financially, divorce can lead to a decrease in income and an increase in expenses, as well as the potential for legal fees. Socially, divorce can lead to changes in relationships with friends and family members, and can also impact one's sense of identity and self-worth. It is important for individuals going through a divorce to seek support from friends, family, and professionals to help them cope with these challenges. For this research purpose a data set has been selected from UCI repository which contain many question asked by the therapist or marriage counselors to the couple and answer of those questions. This paper will give a proper view of those fields which really affect a relationship or a marriage. In this research paper we have used different ranking algorithms to rank the variable first according to their importance. Then different classification algorithms such as Logistic Regression, Naïve Bayes, SGD, Decision Tree, Random

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

Forest and Multilayer Perceptron are used to obtain accuracy, kappa value and ROC area. These algorithms were used for 50:50, 66:34, 80:20 training/testing split and for 10-fold cross validation. When 7 fields are combined and checked with all algorithms for training-test and 10-fold cross validation, then 100% accuracy, a value of 1 for kappa and ROC area was obtained in Decision Tree, Random Forest and Multilayer Perceptron algorithms.

Literature Study

Author and Year	Data source	Method used	Accuracy
(A. Sharma, A. S. Chudhey & M. Singh,2021)	UCIMLR	Machine learning, K-Nearest Neighbours, Logistic Regression, Perceptron classifier, Decision Trees	98.5% (Perceptron classifier)
(M.S.Devi, D.Umanandhini, A.P.S. Anandaraj, S. Sridevi,2021)	UCIMLR	Random Forest, precision, accuracy, classification, cross validation	98% (Random Forest)
(P. Ranjitha & A. Prabhu,2020)	UCIMLR	Bio inspired optimization Algorithm, Particle Swarm Optimization	99.67% (Particle Swarm Optimization)
(Ibrahim M. Nasser,2019)	UCIMLR	Data Mining; Machine Learning; Deep Learning; Predictive Analysis; Artificial Neural Network; Divorce Prediction	100% (Artificial Neural Network)
(M.K.Yontem, K.Adem,T.Ilhan & S.Kilicarslan,2019)	UCIMLR	Data mining, artificial neural networks, divorce, divorce prediction	98.82% (Artificial Neural Network)

Methodology

Different ranking methods and classification algorithms have used for training/testing set to get the maximum accuracy. The data set contain 55 fields in which one is dependent (‘class’) and others are Variables. First all the variables are combined and checked with Logistic Regression, Naive Bayes, SGD, Decision Tree, Random Forest, Multilayer Perceptron for 10-fold cross validation, 50:50, 66:34, 80:20 training/testing sets, but 100% accuracy was not obtained. After that 6 fields are used obtained by different features selection methods like Information Gain, One R, Gain Ratio, Relief F

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

differently, and then 7 fields are used and achieved 100% accuracy in Decision Tree, Random Forest and Multilayer perceptron algorithms.
The Dataset "Divorce Predictors data set" was gathered for development of this research paper from "https://archive.ics.edu/ml/datasets/Divoce+Predicators+data+set".

Sl.no	Variable	Variable Description
1	If one of us apologizes when our discussion deteriorates, the discussion ends.	False(0), True(1)
2	I know we can ignore our differences, even if things get hard sometimes.	False(0), True(1)
3	When we need it, we can take our discussions with my spouse from the beginning and correct it.	False(0), True(1)
4	When I discuss with my spouse, to contact him will eventually work.	False(0), True(1)
5	The time I spent with my wife is special for us.	False(0), True(1)
6	We don't have time at home as partners.	False(0), True(1)
7	We are like two strangers who share the same environment at home rather than family.	False(0), True(1)
8	I enjoy our holiday with my wife.	False(0), True(1)
9	I enjoy travelling with my wife.	False(0), True(1)
10	Most of our goals are common to my spouse.	False(0), True(1)
11	I think that one day in the future ,when I look back,I see that my spouse and I have been in harmony with each other.	False(0), True(1)
12	My spouse and I have similar values in terms of personal freedom.	False(0), True(1)
13	My spouse and I have similar sense of entertainment.	False(0), True(1)
14	Most of our goals for people (children, friends, etc) are the same.	False(0), True(1)
15.	Our dreams with my spouse are similar and harmonious.	False(0), True(1)
16.	We're compatible with my spouse are similar about what love should be.	False(0), True(1)
17.	We share the same views about being happy in our life with my spouse.	False(0), True(1)
18.	My spouse and I have similar ideas about how marriage should be	False(0), True(1)
19.	My spouse and I have similar ideas about how roles should be in marriage.	False(0), True(1)
20.	My spouse and I have similar values in trust.	False(0), True(1)
21.	I know exactly what my wife likes.	False(0), True(1)
22.	I know how my spouse wants to be taken care of when she/he sick.	False(0), True(1)
23.	I know my spouse's favorite food.	False(0), True(1)
24.	I can tell you what kind of stress my spouse is facing in her/his life.	False(0), True(1)
25.	I have knowledge of my spouse's inner world.	False(0), True(1)
26.	I know my spouse's basic anxieties.	False(0), True(1)
27.	I know what my spouse's current sources of stress are.	False(0), True(1)
28.	I know my spouse's hopes and wishes.	False(0), True(1)
29.	I know my spouse very well.	False(0), True(1)

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

30.	I know my spouse's friends and their social relationships.	False(0), True(1)
31.	I fell aggressive when I argue with my spouse.	False(0), True(1)
32.	When discussing with my spouse, I usually use expressions such as 'you always' or 'you never'.	False(0), True(1)
33.	I can use negative statements about my spouse's personality during our discussion.	False(0), True(1)
34.	I can use offensive expression during our discussion.	False(0), True(1)
35.	I can insult my spouse during our discussions.	False(0), True(1)
36.	I can be humiliating during our discussions.	False(0), True(1)
37.	My discussions with my spouse in not calm.	False(0), True(1)
38.	I hate my spouse's way of open a subject.	False(0), True(1)
39.	Our discussions often occur suddenly.	False(0), True(1)
40.	We're just starting a discussion before I know what's going on.	False(0), True(1)
41.	When I talk to my spouse about something, my calm suddenly breaks.	False(0), True(1)
42.	When I argue with my spouse's way of open a subject.	False(0), True(1)
43.	I mostly stay silent to calm the environment a little bit.	False(0), True(1)
44.	Sometimes I think it's good for me to leave home for a while.	False(0), True(1)
45.	I'd rather stay silent than discuss with my spouse.	False(0), True(1)
46.	Even if I'm right in the discussion, I stay silent to hurt my spouse.	False(0), True(1)
47.	When I discuss with my spouse, I stay silent because I am afraid of not being able to control my anger.	False(0), True(1)
48.	I feel right in our discussion.	False(0), True(1)
49.	I have nothing to do with what I've been accused of.	False(0), True(1)
50.	I'm not actually the one who's guilty about what I'm accused of.	False(0), True(1)
51.	I'm not the one who's wrong about problems at home.	False(0), True(1)
52.	I wouldn't hesitate to tell my spouse about her/his inadequacy.	False(0), True(1)
53.	When I discuss, I remind my spouse of her/his inadequacy.	False(0), True(1)
54.	I'm not afraid to tell my spouse about her/his incompetence.	False(0), True(1)

The different ranking methods and algorithms are described below.
1) **Information Gain: -** The gain information value is generated from the entropy value that has not been separated and then reduced by the entropy value of the results after separation [1].
2) **One R: -** One R is a method for evaluating the worth of individual attributes (or features) in a dataset, with the goal of identifying which attributes are the most useful for a particular task. The method works by creating a simple decision tree called a "OneR" model, which uses only one attribute at a time to make predictions. The accuracy of the model is then calculated for each attribute, and the attribute with the highest accuracy is considered the most useful.[1]
3) Gain Ratio: - Gain Ratio (GR) is a modification of IG (Information Gain). IG is to form the induction of the decision tree (ID3), while the Gain Ratio is used in C4.5 is a transformation of ID3.[1]
4) **Relief F: -** Relief Feature Selection is a technique for sequence or attribute ranking based on instances that perform a random sampling process on an instance of data, then look for the closest

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

neighbors of identical and opposite classes. The value of the attribute in the nearest neighbor is then compared with the instance and then updates the relevant score for each attribute [1].
After this, we have used different algorithms for this research work. They are:-
a) **Logistic Regression: -** Logistic Regression is a type of classification algorithm that is used to predict the output class for a given input. This is achieved by using a cost function that incorporates a sigmoid function, which helps to determine the likelihood of the input belonging to a particular class [5]
The formula for logistic regression is represented by the logistic function or sigmoid function:

b) **Naive Bayes: -** It is a probabilistic machine learning algorithm that makes classifications based on Bayes' theorem. It is called "naive" because it makes the assumption that all features are independent of each other, which is often not the case in real-world data. In the method being referred to, the features are assumed to be independent and are given equal weight in making predictions. The algorithm determines whether the features used are more likely to lead to one outcome versus another, and assigns a class based on this prediction [5].

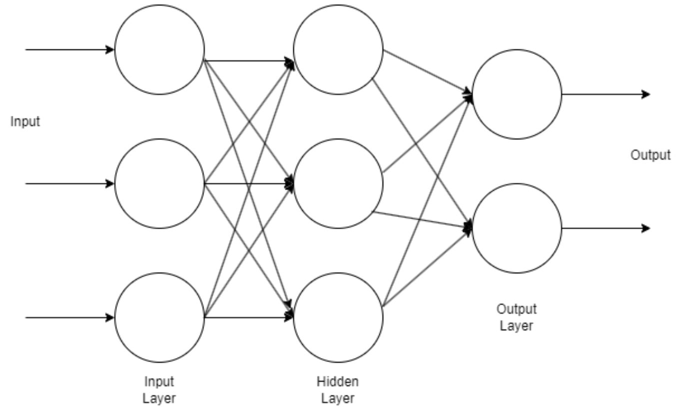
The formula for Naive Bayes is based on Bayes' theorem, which states that:

c) SGD: - SGD stands for Stochastic Gradient Descent, an optimization algorithm used to minimize an objective function by updating the parameters in the direction of the negative gradient of the loss function. Unlike batch gradient descent, which updates the parameters using the average gradient of the loss function with respect to the parameters over the entire training set, SGD updates the parameters after each training example. SGD is efficient for large datasets and is often used in training deep learning models, where the size of the dataset can make batch gradient descent computationally infeasible. [12]
The formula for Stochastic Gradient Descent (SGD) is:

d) **Decision Tree: -** A decision tree is a type of supervised machine learning algorithm that is commonly used in data analysis, classification, and prediction tasks [5]. A decision tree works by recursively partitioning a dataset into subsets based on the values of one or more input variables and also used heuristic values [5].

The decision tree algorithm uses a recursive approach to partition the data and create a tree-like structure.
e) **Random Forest: -** Random Forest is an ensemble literacy system for bracket and retrogression tasks. It combines multiple decision trees to make a prediction. In a Random Forest, each tree in the ensemble is built using a random subset of the training data and a random subset of the features. The final prediction is made by combining the predictions of each tree in the forest through a voting process or by taking the average of the predictions [25].
The formula for a Random Forest prediction can be expressed as a combination of the predictions made by individual decision trees in the forest.
f) Multilayer Perceptron (MLP):- A Multilayer Perceptron (MLP) is a type of artificial neural network that consists of multiple layers of interconnected nodes, or neurons [5]. It is also known as a

feed forward neural network, as the data flows through the network in a feed forward manner, from input layer to output layer, without looping back. The input layer receives the input data and passes it through the hidden layers, where the data is transformed and processed, before finally producing the output at the output layer. Each neuron in a hidden layer receives inputs from the neurons in the previous layer, performs a weighted sum, and then applies an activation function to produce an output [24].



Results and Discussions

With all the above algorithms, accuracy, ROC and kappa score are calculated. First, all the fields are taken from the dataset and combined with each other to calculate the accuracy, ROC, kappa score using all the above classification algorithms with 50-50, 66-34, 80-20 dataset and 10-fold cross validation.

Accuracy						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	83.52	91.76	85.88	89.41	85.88	84.70
66/34	78.57	85.71	81.25	82.14	82.14	82.14
80/20	73.52	80.14	77.94	77.94	81.61	73.52
10 Fold crossValidation	79.41	87.64	80.58	89.41	80	85.88
Kappa						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.46	0.67	0.54	0.67	0.29	0.53
66/34	0.24	0.46	0.32	0.53	0.07	0.43
80/20	0.25	0.34	0.21	0.47	0.14	0.25
10 Fold crossValidation	0.41	0.56	0.43	0.67	0.08	0.60
ROC Area						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.81	0.88	0.78	0.88	0.90	0.89
66/34	0.75	0.79	0.64	0.91	0.73	0.78
80/20	0.63	0.67	0.59	0.79	0.67	0.65
10 Fold crossValidation	0.79	0.87	0.72	0.89	0.90	0.88

After examining the result it is noticed 100% accuracy is not achieved by any of the algorithms. Then ranking method has been used to choose the most important fields from the dataset. Different ranking

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

methods are used then with the chosen fields the classification algorithm is tested. The first 6-fields are taken from each ranking method and tested with the above classification algorithms individually. After 7 fields are taken and tested individually. Then 100% accuracy was found in Decision Tree, Random Forest and Multilayer Perceptron algorithms.

Information Gain:-

When 6 – fields are combined together and tested.
Fields are - [25, 22, 27, 5, 13, and 10]

Accuracy						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	94.11	95.29	94.11	94.11	97.64	94.11
66/34	94.64	94.64	94.64	94.64	95.53	94.64
80/20	95.58	92.64	96.32	96.32	96.32	93.38
10 Cross Validation	91.17	92.35	94.11	95.29	98.23	95.88
Kappa						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.81	0.83	0.81	0.81	0.92	0.81
66/34	0.81	0.80	0.81	0.81	0.84	0.81
80/20	0.83	0.70	0.86	0.85	0.86	0.76
10 Cross Validation	0.67	0.65	0.76	0.81	0.93	0.84
ROC Area						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.96	0.94	0.89	0.95	0.99	0.90
66/34	0.91	0.95	0.84	0.84	0.97	0.92
80/20	0.91	0.93	0.90	0.90	0.97	0.91
10 Cross Validation	0.96	0.96	0.85	0.97	0.99	0.91

When 7 – fields are combined together and tested.
Fields are - [25, 22, 27, 5, 13, 10, and 33]

Accuracy						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	90.58	89.41	95.29	**100**	**100**	95.29
66/34	91.07	87.50	96.42	**100**	**100**	**100**
80/20	86.02	84.55	66.17	91.17	94.11	**100**
10 Cross Validation	85.88	86.47	94.70	**100**	**100**	**100**
Kappa Score						

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.74	0.71	0.88	**1**	**1**	0.88
66/34	0.74	0.61	0.90	**1**	**1**	**1**
80/20	0.59	0.55	0.06	0.74	0.83	**1**
10 Cross Validation	0.65	0.65	0.87	**1**	**1**	**1**
ROC Area						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.95	0.95	0.96	**1**	**1**	0.94
66/34	0.96	0.96	0.97	**1**	**1**	**1**
80/20	0.94	0.89	0.53	0.99	0.96	**1**
10 Cross Validation	0.93	0.92	0.96	**1**	**1**	**1**

One R:-

When 6 – fields are combined together and tested.
Fields are - 54, 20, 18, 17, 16, and 19

Accuracy						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	94.12	95.29	94.11	96.47	97.64	94.11
66/34	94.64	94.64	94.64	96.42	95.53	94.64
80/20	95.58	92.64	96.32	83.08	96.32	93.38
10 fold cross validation	91.17	92.35	94.11	93.52	98.23	95.88
Kappa Score						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.81	0.83	0.81	0.88	0.92	0.81
66/34	0.81	0.80	0.81	0.87	0.84	0.81
80/20	0.83	0.70	0.86	0	0.86	0.76
10 fold cross validation	0.67	0.65	0.76	0.79	0.93	0.84
Roc Area						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.96	0.94	0.89	0.97	0.99	0.90
66/34	0.91	0.95	0.89	0.91	0.97	0.92
80/20	0.91	0.93	0.90	0.50	0.97	0.91
10 fold cross validation	0.96	0.96	0.85	0.99	0.99	0.91

When 7 – fields are combined together and tested.
Fields are - 54, 20, 18, 17, 16, 19 and 21

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

Accuracy						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Tree	MLP
50/50	90.58	89.41	95.29	97.64	**100**	95.29
66/34	91.07	87.5	96.42	97.32	**100**	**100**
80/20	86.02	84.55	66.17	74.26	94.11	**100**
10 fold cross validation	85.88	86.47	94.70	98.23	**100**	**100**
Kappa Score						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.74	0.71	0.88	0.93	**1**	0.88
66/34	0.74	0.61	0.90	0.92	**1**	**1**
80/20	0.59	0.55	0.06	0.23	0.83	**1**
10 fold cross validation	0.65	0.65	0.87	0.95	**1**	**1**
Roc Area						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.95	0.95	0.96	0.99	**1**	0.94
66/34	0.96	0.96	0.97	0.97	**1**	**1**
80/20	0.94	0.89	0.53	0.54	0.96	**1**
10 fold cross validation	0.93	0.92	0.96	0.99	**1**	**1**

<u>Gain Ratio:-</u>
When 6 – fields are involved
Field attributes are - 25, 22, 27, 5, 13, and 10

Accuracy						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	94.11	95.29	94.11	96.47	97.64	94.11
66/34	94.64	94.64	94.64	96.42	95.53	94.64
80/20	95.58	92.64	96.32	83.08	96.32	93.38
10 fold cross validation	91.17	92.35	94.11	93.52	98.23	95.88
Kappa Score						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.81	0.83	0.81	0.88	0.92	0.81
66/34	0.81	0.80	0.81	0.87	0.84	0.81
80/20	0.83	0.70	0.86	0	0.86	0.76
10 fold cross validation	0.67	0.65	0.76	0.79	0.93	0.84
Roc Area						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.96	0.94	0.89	0.97	0.99	0.90
66/34	0.91	0.95	0.89	0.91	0.97	0.92

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

80/20	0.91	0.93	0.90	0.50	0.97	0.91
10 fold cross validation	0.96	0.96	0.85	0.99	0.99	0.91

When 7 – fields are involved
Field attributes are - 25, 22, 27, 5, 13, 10, and 33

Accuracy						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	90.58	89.41	95.29	97.64	**100**	95.29
66/34	91.07	87.5	96.42	97.32	**100**	**100**
80/20	86.02	84.55	66.17	74.26	94.11	**100**
10 fold cross validation	85.88	86.47	94.70	98.23	**100**	**100**
Kappa Score						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.74	0.71	0.88	0.93	**1**	0.88
66/34	0.74	0.61	0.90	0.92	**1**	**1**
80/20	0.59	0.55	0.06	0.23	0.83	**1**
10 fold cross validation	0.65	0.65	0.87	0.95	**1**	**1**
Roc Area						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.95	0.95	0.96	0.99	**1**	0.94
66/34	0.96	0.96	0.97	0.97	**1**	**1**
80/20	0.94	0.89	0.53	0.54	0.96	**1**
10 fold cross validation	0.93	0.92	0.96	0.99	**1**	**1**

Relief F:-
When 6 – fields are involved
Field attributes are - 25, 22, 27, 1, 16, and 33

Accuracy						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	94.11	95.29	94.11	96.47	97.64	94.11
66/34	94.64	94.64	94.64	96.42	95.53	94.64
80/20	95.58	92.64	96.32	83.08	96.32	93.38
10 Fold cross Validation	91.17	92.35	94.11	93.52	98.23	95.88
Kappa						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.81	0.83	0.81	0.88	0.92	0.81
66/34	0.81	0.80	0.81	0.87	0.84	0.81
80/20	0.81	0.70	0.86	0	0.86	0.76

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

10 Fold cross Validation	0.67	0.65	0.76	0.79	0.93	0.84
ROC Area						
Training-Testing Data	Logistic Regression	Naive Bayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.96	0.94	0.89	0.97	0.99	0.90
66/34	0.91	0.95	0.89	0.91	0.97	0.92
80/20	0.91	0.93	0.90	0.50	0.97	0.91
10 Fold cross Validation	0.96	0.96	0.85	0.99	0.99	0.91

When 7 – fields are involved
Field attributes are - 25, 22, 27, 1, 16, 33, and 37

Accuracy						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	90.58	89.41	95.29	97.64	**100**	95.29
66/34	91.07	87.5	96.42	97.32	**100**	**100**
80/20	86.02	84.55	66.17	74.26	94.11	**100**
10 fold cross validation	85.88	86.47	94.70	98.23	**100**	**100**
Kappa Score						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.74	0.71	0.88	0.93	**1**	0.88
66/34	0.74	0.61	0.90	0.92	**1**	**1**
80/20	0.59	0.55	0.06	0.23	0.83	**1**
10 fold cross validation	0.65	0.65	0.87	0.95	**1**	**1**
Roc Area						
Training-Testing Data	Logistic Regression	NaiveBayes	SGD	Decision Tree	Random Forest	MLP
50/50	0.95	0.95	0.96	0.99	**1**	0.94
66/34	0.96	0.96	0.97	0.97	**1**	**1**
80/20	0.94	0.89	0.53	0.54	0.96	**1**
10 fold cross validation	0.93	0.92	0.96	0.99	**1**	**1**

From the above results it is examined that –

- When all fields are combined together 100% accuracy is not achieved.
- When 6 fields are combined taken from each ranking method and tested with classification algorithms then also 100% accuracy is not achieved from any classification algorithms.
- Using Information Gain ranking method when 7 – fields are combined then 100% accuracy, kappa=1, and ROC=1 are achieved in decision tree and random forest algorithm for 50-50 data. For 66-34 data and 10-fold cross validation 100% accuracy, kappa=1, and ROC=1 are achieved in decision tree, random forest, and multilayer perceptron algorithm. For 80-20 data only MLP gave 100% accuracy, kappa=1, and ROC=1.
- When One R, Gain Ratio, and Relief F ranking method is used and 7 – fields are combined then 100% accuracy, kappa=1, and ROC=1 are achieved in random forest only for 50-50 data. For 66-34

data and 10-fold cross validation 100% accuracy, kappa=1, and ROC=1 are achieved in random forest, and multilayer perceptron algorithm. For 80-20 data only MLP gave 100% accuracy, kappa=1, and ROC=1.

Conclusion

In conclusion, this paper aimed to analyze the accuracy of the divorce rate prediction dataset through 50-50, 66-34, 80-20 and 10-fold cross validation. First different features selection methods are used to select the important fields. Then the most important 7 features are combined and tested with different classification algorithms. After testing results showed that the Random Forest classifier, MLP and Decision tree had an accuracy of 100%, kappa=1, ROC=1 for different cases. These results demonstrate the effectiveness of using the Random Forest, MLP, and Decision tree classifier in improving the accuracy of the divorce rate prediction dataset. This paper will help the marriage counselor or therapist to understand the most effective reason of the divorce cases. This paper will also help the couples to understand the different factors that really affect the relationships or marriages. Because a happy and healthy marriage keeps the couples happy and depression free.

Future References

- The same machine learning models can potentially be used to classify different data models from different fields. However, rigorous testing and validation are essential, and the approach should be justified by highlighting the novelty and unique challenges of the new domain. For example, a classification model that has been trained on medical data can potentially be used to classify data from a different field, such as financial data. However, it is important to keep in mind that the effectiveness of the model in a new domain will depend on several factors such as the similarity between the two domains, the availability and quality of data, and the suitability of the model architecture and parameters.
- This analysis has been made on secondary dataset collected from UCI ML repository. We have conducted an analysis of this dataset using a particular model or algorithm to predict the likelihood of divorce based on various demographic and social factors. However, we state that their next target is to collect real-life data from their own society and test it against the model derived from the analysis of the secondary dataset. This is because real-life data may contain additional or different variables or factors that could affect the prediction accuracy of the model. By testing the model on real-life data, we hope to validate the effectiveness of their model and improve its predictive accuracy for practical use in the future.

References

Here are some references for research papers on divorce prediction-
1. M. Simanjuntak, M. Muljono, G. F. Shidik and A. Zainul Fanani, "Evaluation Of Feature Selection for Improvement Backpropagation Neural Network in Divorce Predictions," 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2020, pp. 578-584, doi: 10.1109/iSemantic50169.2020.9234297.
2. "Divorce Prediction Using OneR Classifier Algorithm" by Farhan M. Alruban and Mohammed M. Nasseif.
3. Atkins, D. C., & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: A tutorial on count regression and zero-inflated models. Journal of Family Psychology, 21(4), 726–735.
4. Joshua R. Goldstein; The leveling of divorce in the united states. Demography 1 August 1999; 36 (3): 409–414. doi: https://doi.org/10.2307/2648063
5. A. Sharma, A. S. Chudhey and M. Singh, "Divorce case prediction using Machine learning algorithms," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 214-219, doi: 10.1109/ICAIS50930.2021.9395860.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 2 Volume No.7 March - April – 2023
DOI:10.46647/ijetms.2023.v07i02.024 ISSN: 2581-4621

6. Trent, Katherine, and Scott J. South. “Structural Determinants of the Divorce Rate: A Cross-Societal Analysis.” Journal of Marriage and Family, vol. 51, no. 2, 1989, pp. 391–404. JSTOR, https://doi.org/10.2307/352502.
7. "Divorce Prediction Using Decision Table Algorithm" by Sonali Wagh and Archana Shinde.
8. Yöntem, M , Adem, K , İlhan, T , Kılıçarslan, S. (2019). DIVORCE PREDICTION USING CORRELATION BASED FEATURE SELECTION AND ARTIFICIAL NEURAL NETWORKS. Nevşehir Hacı Bektaş Veli University SBE Dergisi, 9 (1), 259-273.
9. Peters, H. Elizabeth. “Marriage and Divorce: Reply.” The American Economic Review, vol. 82, no. 3, 1992, pp. 686–93.
10. P. Ranjitha and A. Prabhu, "Improved Divorce Prediction Using Machine learning- Particle Swarm Optimization (PSO)," 2020 International Conference for Emerging Technology (INCET), Belgaum, India, 2020, pp. 1-5, doi: 10.1109/INCET49848.2020.9154081.
11. Shankhdhar, A., Gupta, T., Gautam, Y.V. (2021). Divorce Prediction Scale Using Improvised Machine Learning Techniques. In: Suma, V., Chen, J.IZ., Baig, Z., Wang, H. (eds) Inventive Systems and Control. Lecture Notes in Networks and Systems, vol 204. Springer, Singapore. https://doi.org/10.1007/978-981-16-1395-1_57
12. Amato, Paul R. "Research on divorce: Continuing trends and new developments." Journal of marriage and family 72.3 (2010): 650-666.
13. Susan L. Brown, I-Fen Lin, The Gray Divorce Revolution: Rising Divorce Among Middle-Aged and Older Adults, 1990–2010, The Journals of Gerontology: Series B, Volume 67, Issue 6, November 2012, Pages 731–741, https://doi.org/10.1093/geronb/gbs089.
14. Pal, S. S., Paul, S., Dey, R., Das, S., & Chaudhuri, A. K. Determining the probability of poverty levels of the Indigenous Americans and Black Americans in US using Multiple Regression.
15. Saha, S., Mondal, J., Arnam Ghosh, M., Das, S., & Chaudhuri, A. K. Prediction on the Combine Effect of Population, Education and Unemployment on Criminal Activity Using Machine Learning.
16. Hetherington, E. M. (1999). For better or for worse: Divorce reconsider. New York, NY: W.W. Norton.
17. Johnson, M. P., & Jacobson, N. S. (1994). The relationship between divorce and psychological distress: A longitudinal study. Journal of Social and Personal Relationships, 11(4), 481-498.
18. Beutel, A. M., & Cui, Y. (2017). Predictive modeling of marital dissolution using demographic, relationship quality, and individual characteristics. Journal of Family Issues, 38(5), 589-613.
19. Ahmadi, M. H., Ojaghi, Z., & Mohammad, R. (2017). Predictive modeling of divorce using machine learning techniques. Expert Systems with Applications, 79, 128-137.
20. Fagan, J., & Leon, K. R. (2015). Predictors of marital dissolution: Results from the national survey of families and households. Journal of Marriage and Family, 77(2), 361-377.
21. Ramya, S., Senthil Kumar, R. (2014). Survey of Accident Severity Estimation Using Data Mining Techniques, International Journal of Computer Science & Engineering Technology, 5, 1041-1044.
22. Karthick, R., and M. Sundararajan. “SPIDER-based out-of-order execution scheme for HtMPSOC.” International Journal of Advanced Intelligence paradigms 19.1 (2021): 28-41. https://doi.org/10.1504/IJAIP.2021.114581
23. Sabarish, P., et al. “An Energy Efficient Microwave Based Wireless Solar Power Transmission System.” IOP Conference Series: Materials Science and Engineering. Vol. 937.No. 1.IOP Publishing, 2020. doi:10.1088/1757-899X/937/1/012013
24. Tutul, Md Ashrafujjaman, Md Mehedi Hasan, Shashata Kumar Mondol, Moshaddek Hossain, and Ahmed Al Marouf. "Divorce Prediction using Machine Learning Methods-Bangladesh Perspective." In 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-5. IEEE, 2021.
25. Arpino, Bruno, Marco Le Moglie, and Letizia Mencarin. "Machine-learning techniques for family demography: an application of random forests to the analysis of divorce determinants in

Germany." University Pompeu Fabra Barcelona Reseaerch and Expertise Centre for Survey Methodology.

26. Nusrath Unnisa A; Manjula Yerva; Kurian M Z. "Review on Intrusion Detection System (IDS) for Network Security using Machine Learning Algorithms". International Research Journal on Advanced Science Hub, 4, 03, 2022, 67-74. doi: 10.47392/irjash.2022.014

27. Mohammadibrahim Korti; Basavaraj S. Malapur; Smita Gour; Rajesh M. Biradar. "Shuchi 1.0: Robotic System For Automatic Segregation of Waste & Floor Cleaning". International Research Journal on Advanced Science Hub, 4, 02, 2022, 31-37. doi: 10.47392/irjash.2022.007

28. Bhuneshwari Nayak; Rachana Choudhary; Roymon M. G.. "Isolation, Screening and Morphological characterization of Laccase producing fungi". International Research Journal on Advanced Science Hub, 4, 02, 2022, 38-43. doi: 10.47392/irjash.2022.008

29. Khaled Salem Ahmad Amayreh; Ahmad Taufik Hidayah Bin Abdullah. "Conjunction in Expository Essay Writing by Jordanian Undergraduate Students Studying English as a Foreign Language (EFL)". International Research Journal on Advanced Science Hub, 4, 02, 2022, 24-30. doi: 10.47392/irjash.2022.006

30. Pravin T, M. Subramanian, R. Ranjith,Clarifying the phenomenon of Ultrasonic Assisted Electric discharge machining, "Journal of the Indian Chemical Society", Volume 99, Issue 10, 2022, 100705, ISSN 0019-4522, Doi: 10.1016/j.jics.2022.100705

31. R. Sudhakaran, P.S. Sivasakthivel, M. Subramanian, Investigations on the effect of surface coatings on the weldment properties on chromium manganese stainless steel gas tungsten arc welded plates, Materials Today: Proceedings,Volume 46, Part 17,2021, Pages 8554-8560, ISSN 2214-7853, Doi: 10.1016/j.matpr.2021.03.541

32. Sivalakshmi B; Mahisha sri K.B; Swetha P. "Survey on Mammogram, Ultrasound, MRI, Spectroscopy, Biopsy for Detecting Tumor in Breast". International Research Journal on Advanced Science Hub, 3, 2, 2021, 30-37. doi: 10.47392/irjash.2021.027

33. Hari Prasada Raju Kunadharaju; Sandhya N.; Raghav Mehra. "Detection of Brain Tumor Using Unsupervised Enhanced K-Means, PCA and Supervised SVM Machine Learning Algorithms". International Research Journal on Advanced Science Hub, 2, Special Issue ICSTM 12S, 2020, 62-67. doi: 10.47392/irjash.2020.262

34. Suresh P; Justin Jayaraj K; Aravintha Prasad VC; Abishek Velavan; Mr Gokulnath. "Deep Learning for Covid-19 Identification: A Comparative Analysis". International Research Journal on Advanced Science Hub, 4, 11, 2022, 272-280. doi: 10.47392/irjash.2022.068

35. Rajesh P.; Vetrivel Govindarasu. "Analyzing and Predicting Covid-19 Dataset in India using Data Mining with Regression Analysis". International Research Journal on Advanced Science Hub, 3, Special Issue 7S, 2021, 91-95. doi: 10.47392/irjash.2021.216

36. Menonjyoti Kalita; Golam Imran Hussain. "Determining the Influencing Factors of COVID 19 on Mental Health Using Neural Network". International Research Journal on Advanced Science Hub, 3, Special Issue 6S, 2021, 126-129. doi: 10.47392/irjash.2021.177

37. Ajitha K; Samuel Joseph C; Mahila Vasanthi Thangam D. "Online marketing of agricultural products during COVID pandemic: Farmers and customers perspectives". International Research Journal on Advanced Science Hub, 3, Special Issue 6S, 2021, 94-101. doi: 10.47392/irjash.2021.173

38. Yeshi Ngima; Dorjee Tsering. "Impact of COVID-19 on Education". International Research Journal on Advanced Science Hub, 2, Special Issue ICAMET 10S, 2020, 34-39. doi: 10.47392/irjash.2020.196

39. Siddavatam rammohan reddy; Balaji krushna potnuru. "3D Printing Innovation during Covid-19 Pandemic". International Research Journal on Advanced Science Hub, 2, 8, 2020, 62-67. doi: 10.47392/irjash.2020.95

40. Pooja Dahiya; Roopsi Kaushik; Anil Sindhu. "Corona virus: an Overview Along with Its Alternative Diagnostic Measures". International Research Journal on Advanced Science Hub, 2, Special Issue ICARD 2020, 2020, 163-169. doi: 10.47392/irjash.2020.113

41. Remya S. "Covid19 and Environment-A Theoretical Review from Higher Education Students Perspective". International Research Journal on Advanced Science Hub, 2, Special Issue ICARD 2020, 2020, 227-230. doi: 10.47392/irjash.2020.124
42. Mohd. Akbar; Prasadu Peddi; Balachandrudu K E. "Inauguration in Development for Data Deduplication Under Neural Network Circumstances". International Research Journal on Advanced Science Hub, 2, 6, 2020, 154-156. doi: 10.47392/irjash.2020.55
43. Salini Suresh; Suneetha V; Niharika Sinha; Sabyasachi Prusty; Sriranga H.A. "Machine Learning: An Intuitive Approach In Healthcare". International Research Journal on Advanced Science Hub, 2, 7, 2020, 67-74. doi: 10.47392/irjash.2020.67
44. Trupti S. Gaikwad; Snehal A. Jadhav; Ruta R. Vaidya; Snehal H. Kulkarni. "Machine learning amalgamation of Mathematics, Statistics and Electronics". International Research Journal on Advanced Science Hub, 2, 7, 2020, 100-108. doi: 10.47392/irjash.2020.72
45. Salini Suresh; Suneetha V; Niharika Sinha; Sabyasachi Prusty. "Latent Approach in Entertainment Industry Using Machine Learning". International Research Journal on Advanced Science Hub, 2, Special Issue ICARD 2020, 2020, 304-307. doi: 10.47392/irjash.2020.106
46. Suneetha V; Salini Suresh; Niharika Sinha; Sabyasachi Prusty; Syed Jamal J. "Enhancement in the World of Artificial Intelligence". International Research Journal on Advanced Science Hub, 2, Special Issue ICARD 2020, 2020, 276-280. doi: 10.47392/irjash.2020.132
47. Logeswari T.. "Performance Analysis of Ml Techniques for Spam Filtering". International Research Journal on Advanced Science Hub, 2, Special Issue ICIES 9S, 2020, 64-69. doi: 10.47392/irjash.2020.161
48. Hari Prasada Raju Kunadharaju; Sandhya N.; Raghav Mehra. "Detection of Brain Tumor Using Unsupervised Enhanced K-Means, PCA and Supervised SVM Machine Learning Algorithms". International Research Journal on Advanced Science Hub, 2, Special Issue ICSTM 12S, 2020, 62-67. doi: 10.47392/irjash.2020.262Export Citation
49. Maneesha M; Savitha V; Jeevika S; Nithiskumar G; Sangeetha K. "Deep Learning Approach For Intelligent Intrusion Detection System". International Research Journal on Advanced Science Hub, 3, Special Issue ICARD-2021 3S, 2021, 45-48. doi: 10.47392/irjash.2021.061
50. Kalki N; Karthick M; Mr Kavin; Keerthana S; Sangeetha K. "Advanced Face Mask Detection System". International Research Journal on Advanced Science Hub, 3, Special Issue ICARD-2021 3S, 2021, 112-115. doi: 10.47392/irjash.2021.076
51. Naveenkumar S; Kirubhakaran R; Jeeva G; Shobana M; Sangeetha K. "Smart Health Prediction Using Machine Learning". International Research Journal on Advanced Science Hub, 3, Special Issue ICARD-2021 3S, 2021, 124-128. doi: 10.47392/irjash.2021.079
52. Kavya Shakthi R.P; Kavin Raja A.S; Janani S.R; Sangeetha K. "Industrial Machine Identification Using Augmented Reality". International Research Journal on Advanced Science Hub, 3, Special Issue ICARD-2021 3S, 2021, 68-71. doi: 10.47392/irjash.2021.066
53. Salma Begum; Sampurna P.. "A Study on growth in Technology and Innovation across the globe in the Field of Education and Business". International Research Journal on Advanced Science Hub, 3, Special Issue 6S, 2021, 148-156. doi: 10.47392/irjash.2021.181
54. Dhanya S Karanth; Kumaraswamy H.V; Rajesh Kumar. "Workaround prediction of cloud alarms using machine learning". International Research Journal on Advanced Science Hub, 3, 8, 2021, 164-168. doi: 10.47392/irjash.2021.231
55. Gyanendra Kumar Pal; Sanjeev Gangwar. "Discovery of Approaches by Various Machine learning Ensemble Model and Features Selection Method in Critical Heart Disease Diagnosis". International Research Journal on Advanced Science Hub, 5, 01, 2022, 15-21. doi: 10.47392/irjash.2023.003