ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 19, Issue 2, 2025

Using Machine Learning Algorithms To Forecast Divorce Cases

[1] Ch. Sravan Kumar, [2] K. Lavanya,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract—

Worldwide, there has been a dramatic spike in the number of divorces. The divorce rate in India has increased from 1 in 1000 in the last several decades to 13 in 1000 now. For this reason, marital therapists and counselors find it to be a really pressing issue. Therefore, a marital counselor or therapist needs a reliable method to forecast divorce that can assess the severity of a situation. This paper reports the results of a research that used preexisting machine learning algorithms to forecast divorce cases. For the purpose of predicting divorce cases, the authors have used a number of classifiers and compared their accuracy. These include Perceptron, Decision Tree, Random Forest, Naive Bayes, K-Nearest Neighbor, and Support Vector Machine. The criteria used to create predictions in this research are based on the Gottman approach. Following training, the algorithms will be able to forecast the likelihood of a divorce. Psychologists may use this information to gauge the level of tension in a couple's relationship and provide appropriate treatment. The authors have used the Perceptron model to get the maximum accuracy of 98.5%. Machine learning, decision trees, the Gottman Method of Relationship Therapy, logistic regression, support vector machines, and divorce prediction are all terms that may be found in the index.

I. INTRODUCTION

A major problem in the modern society is the ever-increasing number of divorces. For all people, the family is the most important and fundamental social unit. Helping it remain together and preventing breakdowns caused by misunderstandings is, then, an issue that requires significant care. However, worldwide, the number of divorces has been on the rise lately. Divorce rates in India have jumped from 1 in 1000 in the 1980s to 13 in 1000 in the 2000s. [1]: Yes. In order to assist therapists or marital counselors identify the source of the couple's conflict, this research places a focus on obtaining a divorce prediction to some degree. Relationship therapy according to the Gottman model is the basis for this evaluation [2]. University of Washington psychology professor John Gottman created this technique. According to this approach, the so-called "Four horseman" elements are the root of all marital problems. Criticism, defensiveness, stonewalling, and scorn are the elements that contribute to this [3]. Having a common purpose in life and improving the ability to resolve conflicts constructively are two goals of the approach. Here are the seven pillars upon which this thesis rests: • Love Maps—This theory stresses the need of paying attention to one's partner's emotions, worldview, values, habits, and aspirations. • Learning to live with diversity is a crucial component of marriage, according to this idea, which is based on the sharing of admiration and fondness. • Coming together and talking it out — Couples who work through their differences rather than avoiding them tend to be more stable, according to the research. • Keeping a positive outlook—According to this concept, one should treat their spouse with kindness and understanding, rather than hostility, and be flexible and open to new ideas. • Work through conflicts as a couple — Couples who are able to resolve conflicts peacefully while also understanding one another's frailties and ways of talking usually end up with a good solution. • Conflict management — Happily married couples say that most arguments happen periodically and are handled as they arise. Rather of resolving conflicts, the Gottman approach teaches people how to handle them. the third A couple may do this by discussing and reaching an agreement on the relationship's objectives and the expectations each partner brings to the table. To achieve this goal, one must refrain from criticism and accept that some things will stay unchanged until both sides are satisfied. According to the shared meaning concept, a successful marriage is the result of the partners' ability to find common ground in their shared experiences of life's challenges and triumphs. Combining relationship theory with machine learning methods, this research proposes a way to anticipate divorce instances. In therapy or while trying to determine the root cause of a couple's divorce, the suggested strategy could be useful for therapists or marital counselors. A plethora of machine learning algorithms with varying degrees of accuracy have emerged in recent years. The aforementioned concepts provide the basis for the collection of characteristics that inform the algorithmic decisions made. The couples score these traits on a scale from 0 to 4, where 4 indicates significant agreement and 1 indicates no agreement at all. The likelihood of the couple's divorce may be predicted using this collection of characteristics. We may compare the various machine learning algorithms for a given dataset based on how well they perform in a certain application. In this research, we tested the accuracy

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

of many classifiers to identify the best fit for this case. These include Perceptron, Naive Bayes, Logistic Regression, K-nearest neighbors, and Support Vector Machine. The therapist may gauge the severity of the couple's situation and provide appropriate assistance based on this forecast. Here is how this document is structured: The dataset used for this investigation is detailed in Section II, while the relevant work is presented in Section III. The suggested method is detailed in Section IV, while the experimental work is detailed in Section V. The study is concluded and its potential extensions are stated in Section VI.

II. DATASET USED

We utilize the UCI Machine Learning Repository Dataset[4] for this investigation. For the purpose of making divorce case predictions, this dataset contains all the relevant information. A total of 170 couples have contributed their responses to the 54-question data collection. This survey follows the guidelines laid forth by the Gottman couple theory; responses range from 0 (strongly disagree) to 4 (completely agree). Following training is complete, the algorithm uses the answers to these questions as features to create predictions. A portion of the dataset that was used is seen in Figure 1. The result is given as a binary number between 0 and 1, with 1 representing a divorce and 0 representing no divorce. The data was collected using a 5-point scale, where 0 signifies never, 1 means seldom, 2 means averagely, 3 means often, and 4 means always. The following are examples of some of the traits: 1) Rather than being considered family, we are more like two strangers living in the same house. 2) Every time my wife and I go on vacation, I rejoice. 3. I love going on vacation with my wife. 4) My wife and I have a lot of same aims. We're too busy being parents to spend much time 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

Presented below is the body of work that pertains to our investigation. Academics have only offered a handful of machine learning-based methods in the literature [5, 6, 7]. For the purpose of predicting divorce situations, M. Irfan et al.[5] compared KNN and Naive Bayes classifiers. They were able to get a 72.5% success rate by using the dataset provided by the Cimahi Religious Court Office. For the purpose of predicting divorce cases, Yontem et al. [6] used artificial neural networks and correlation-based feature selection. Classification accuracy reached a peak of 98.23% when they utilized the same dataset as UCI Machine Learning. An algorithm called Augur Justice, developed by Somya Goel et al. [7], may determine the user's likelihood of winning or losing a lawsuit based on their faith. Accuracy rates of 84.09% on the Hindu dataset, 55.56% on the Christian dataset, and 100% on the Muslim dataset have been acquired. Despite the lack of research into utilizing ML for divorce case prediction, ML techniques like categorization and estimate are already used in several psychiatric and psychological studies. In 509 instances of adult suicide attempts, Baca-Garcia et al.[8] employed data mining methods to predict psychiatrists' decisions to hospitalize. Using five characteristics such as drug usage during the attempt and a family history of suicide attempts, etc., they were able to reach a 99% accuracy rate using forward selection. Using Machine Learning models such as KNN, Naive Bayes, and SVM, Song[9] studied data from college students' psychological evaluations. When they used SVM, they got an accuracy of 79.1 percent, which was their best result. In order to identify adverse medication responses in the inpatient psychiatric population, Erikson[10] used temporal data mining methods in their research. Minimizing the possibility of human error in reporting adverse medication reactions, it will aid in their detection.

IV. PROPOSED APPROACH

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

To predict divorce cases, this study employs a variety of machine learning models, each with its own set of advantages and disadvantages. These models include Perceptron Classifier, Random Forest Classifier, Naive Bayes Classifier, Logistic Regression, K Nearest Neighbor, Support Vector Machine, and Decision Tree. Next, we compared their predictions to see which model was the most accurate. Why? Because prior studies have compared only few machine learning models, and their results have been inconclusive. Therefore, we can learn which models are more accurate for this application by comparing all of them. The suggested methodology for this investigation is shown in Figure 2 by means of a flow diagram. Here we show the outcomes of many training and testing splits (80-20, 70-30, and 60-40) that we used for the benefit of our model.

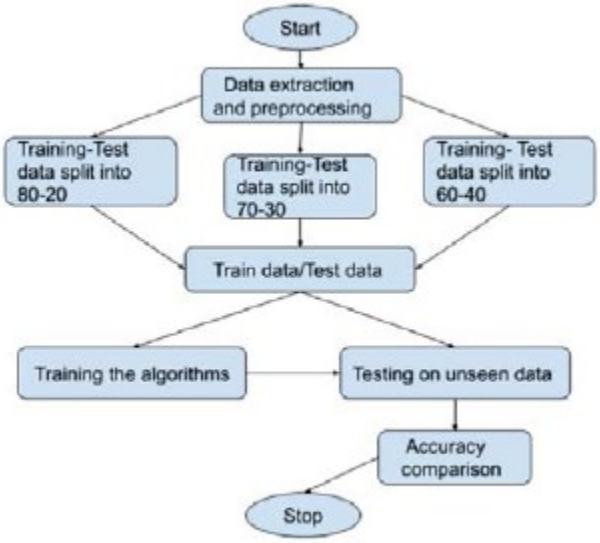


Fig. 2: Proposed methodology flowchart

The Perceptron

A perceptron is a linear binary classifier that uses a single layer of neural networks. [11] An input layer comes first, then a concealed layer, and finally an output layer. The input is multiplied by the weights assigned to it, and the resulting total is summed to get the weighted sum. Finally, the weighted sum is transformed into the target output by applying an activation function. In order to train the network, this makes use of supervised learning.

Part B: Naïve Bayes

One classifier that utilizes probability to generate predictions is Naives Bayes. Bayes Theorem is the foundation. [12] Each feature is thought to be completely autonomous and to have an equal impact. It is assigned a class depending on its predictions, which are based on whether the characteristics utilized favor a given event or not. It use the formula (1). In terms of speed, Naive Bayes outperforms competing categorization systems.

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

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

P(C)- Prediction prior probability

C. LOGISTIC REGRESSION

It is possible to classify data using Logistic Regression. It uses a sigmoid function to compute a cost function, which it then uses to classify data into appropriate categories. in [13] Equation (2) represents the sigmoid function. A

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

training-time output and a hypothesis are used by the cost function. The hypothesis gives a best guess as to how likely it is that something will happen. How much the hypothesis deviates from the outcome may be inferred from the cost function. Next, optimization methods like gradient descent, conjugate gradient, etc., are used to minimize the cost. When it comes to training, logistic regression is both efficient and straightforward to understand and use.

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

D. K-Nearest Neighbour

K-nearest neighbours is a classification technique that uses supervised learning. It uses the training to determine which category best fits the new data and then assigns it to that category. The data upon which it is based is not presumptuous. [14] It uses the distance between known data elements and other known data to assign a class. Because of this, the classes of nearby properties are always grouped together. This approach uses the new and old cases to do a proximity search.

Section E: Support vector.

Assigning new instances to each category is the job of Support Vector Machine (SVM), a supervised machine learning technique that uses a pre-existing collection of training examples that are all part of different categories. [15] A further way of looking at logistic regression is via a support vector machine. Also called a large margin classifier, it creates a decision boundary using margin, which means that there is enough distance between the objects. Different data points are divided into two categories by the decision boundary. The support vector machine (SVM) assists in locating a decision boundary where the data points are ideally separated from one other, based on the largest distance between them. SVM uses the "one-versus-one" method to classify many classes at once. Despite its rather lengthy training period, it excels on huge datasets. F. Decision Tree A decision tree is a categorization and prediction tool with a tree-like structure, as the name implies. A tree's interior nodes contain testing criteria, its branches have outcomes for those circumstances, and the leaf nodes or terminating nodes carry the class label. [16] A well-liked machine learning tool, a decision tree is composed of options and their likely results, such as many outcomes for a given circumstance, and it offers efficient cost management and resource usage. In cases when several options exist, decision trees may be an invaluable tool.

V. EXPERIMENTAL SETUP

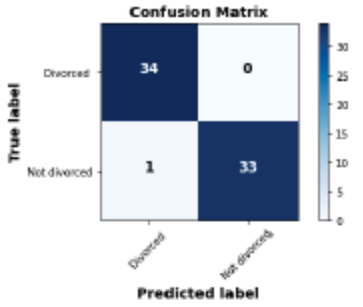
The model was trained using the Python programming language, and all experiments were conducted on Google Collaboratory. The default values for all parameters, including the constant by which updates are multiplied and the number of CPUs utilized, were left at their default values. The Perceptron learning rate was set to 0.0001 and the tolerance to 0.001. The default values of var smoothing and prior probability were set to 1 ∏ e−9 and none, respectively, for Naive Bayes, indicating that they were not altered based on the data. All of the settings, including the tolerance (0.001) and the maximum number of iterations required for solvers to converge, were left at their default values for Logistic Regression. All other parameters, including the degree of the polynomial kernel function and the regularization parameter, were left at their default values in the case of Support Vector Machine, with a tolerance of 0.001. In the instance of KNN, the default values for parameters such as leaf size and weight were maintained, and the number of neighbours was set to 15. Finally, for the Decision Tree, we left all the default values for the parameters, such as the maximum tree depth and the minimum amount of samples you need at the leaf node. A. Measures of performance When testing the accuracy of a classification model, a confusion matrix may help determine whether an item is true or false. The real (or "ground truth") data is compared with the projected data in this matrix prediction. There are two possible representations for a confusion matrix: positive and negative. Generally speaking, a positive class indicates aberrant conduct, whereas a negative class indicates normal behavior. In the confusion matrix, you'll find four sections: A true positive (TP) is the result when the model assigns the right positive label to a class. This is the result that is produced when the model successfully identifies a negative class: True Negative (TN). The result is known as a false positive (FP) when the model mistakenly classifies a positive class. When the model assigns the wrong negative label, the result is a False Negative (FN). The equations used for computation were:

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

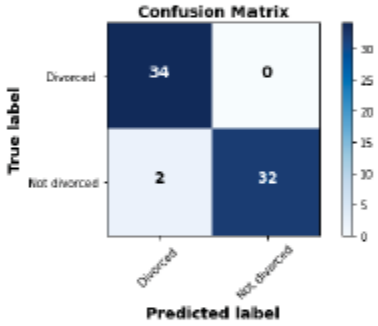
$$Accuracy = \frac{Number\ of\ correct\ values * 100\%}{total\ data} \quad (3)$$

B. Experimental Results

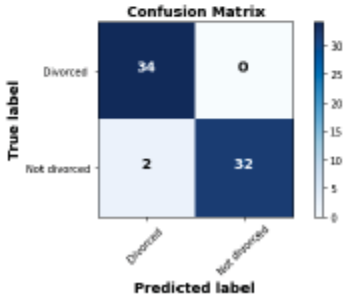
The study's stated divorce dataset was subjected to the following machine learning algorithms: Support Vector Machine, Decision Tree, Naive Bayes, K Nearest Neighbour, Logistic Regression, and Perceptron. An optimal configuration for a Perceptron would be a 60-40 split between training and testing. In the same vein as Logistic Regression and Support Vector Machine, Perceptron, Naive Bayes, K-Nearest Neighbor, and Support Vector Machine all achieved their best results when we divided the training and testing sessions evenly. Aside from that, the decision tree's accuracy was the lowest. With the exception of the decision tree model, we found that as training time decreased



(a) Perceptron



(b) Naive Bayes



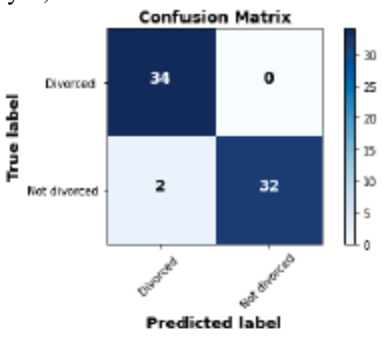
(c) Logistic Regression

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

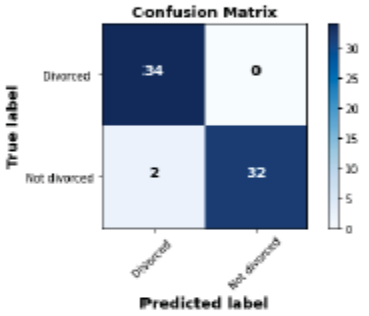
It led to an improvement in the model's accuracy. During different train and train splits, we've seen that a lot of our models exhibit a similar level of accuracy. Due to the small sample size (only 170 couples are included), this might be due to a lack of data. You can see the results and the confusion matrix of the machine learning models employed in Figures 3 and 4. The first thing that happened was that the Perceptron model accurately identified 34 couples as

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN 2454-9940
www.ijasem.org
Vol 19, Issue 2, 2025

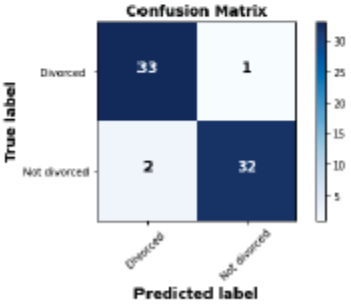
divorced and 33 couples as not divorced. Following this, 34 patients were accurately classified as divorced and 32 couples as not divorced using Logistic Regression, Naive Bayes, K Nearest Neighbor, and Support Vector Machine, respectively. Decision Tree properly identified 32 patients as not divorced and 33 as divorced in Last. Table 1 provides a comprehensive evaluation of the performance of the machine learning models used for various training and test splits. Perceptron earned the highest performance with a 98.5% accuracy rate. K-Nearest Neighbor, Support, Naive Bayes, and



(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.

Logistic Regression, Support Vector Machine, and each of them reached 97.1% accuracy. A decision tree with a 96.1% accuracy rate had the poorest performance. We may conclude that Perceptron is the best Machine Learning model because of this. Beyond these results, the most important takeaways from our work might be: 1) Unlike previous research, which only employed a few of machine learning models, we utilized six distinct models with distinct training and test splits. 2) Since our focus is on couples dealing with divorce. Datasets made available by courts are readily available. Thirdly, our model is incredibly user-friendly as we haven't utilized feature selection.

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

The therapist counseling the couple may greatly benefit from early divorce case prediction in order to undertake efforts to preserve the marriage, especially given the alarming increase in the number of divorce cases. Using a variety of training and test splits, this research evaluated our model's performance and achieved varying accuracy. With a 98.5% accuracy rate, Perceptron was the most effective machine learning model. Adding more couples to the dataset improves the categorization performance of various machine learning models, according to future research. We can improve our model's accuracy and save training time in the future by using feature selection. Marriage counselors and the court system may both benefit from a desktop-based application that can forecast 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/happilydivorced- indian-women-are-breaking-the-stigma-around-separation-likenever- 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-thegottman- method-of-relationship-therapy/, May 2013.
[3] G. Therapy, "Gottman method," https://www.goodtherapy.org/learnabout- 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.
[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," *Nev, sehir Hacı Bekta, s Veli Universitesi 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 naïve 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.