Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Comparing the Performance of Different Supervised Learning Algorithms

Rishi Matura[1], Rajneesh Thakur[2], Prathimesh[3]

Computer Science Engineering, Chandigarh University, Mohali, India

E-mail: [1]rishimatura27@gmail.com, [2] namanthakur625@gmail.com, [3] Sarte2630@gmail.com

Abstract

In recent years, machine learning has grown in popularity, with applications in a wide range of industries. This study discusses the fundamentals of machine learning and its various approaches, such as supervised classifier, unsupervised classifier and reinforcement learning. Moreover, the drawbacks of machine learning, such as the requirement for a lot of labelled data, the possibility of bias in the training process etc. are studied. Finally, some of the field's potential future developments, such as the use of machine learning in areas like healthcare and finance, have been elaborated. Furthermore, two algorithms of machine learning such as Decision Tree and Naive Bayes algorithms are compared. Overall, this work provides a thorough overview of machine learning's current state and potential future impact.

Keywords: Machine Learning, Algorithms, Regression, Supervised Machine Learning

1. Introduction

"Machine learning," is a branch of artificial intelligence that focuses on creating statistical models and algorithms that enable computers to "learn" from data without needing to be explicitly programmed. Various methods and techniques, including as classifier, unsupervised classifier, semi-supervised classifier, and reinforcement learning, are used to carry out this learning process. It has also contributed significantly to the advancement of self-driving cars, systems for making recommendations and predictive modelling. One of machine learning's main advantages is its capacity to automatically improve a system's performance over time by implementing algorithms that can "learn" from data and adapt to new inputs. As a result, machine learning is now widely used in a variety of industries, including healthcare, finance, and retail. Despite its many successes, machine learning has limitations. One

DOI: https://doi.org/10.36548/jaicn.2023.1.005
Received: 07.02.2023, received in revised form: 17.03.2023, accepted: 01.04.2023, published: 11.04.2023
© 2023 Inventive Research Organization. This is an open access article under the Creative Commons Attribution-NonCommercial International (CC BY-NC 4.0) License

significant challenge is the requirement for large amounts of labelled data to train machine learning algorithms, which can be time-consuming and expensive to obtain. Furthermore, machine learning models can be prejudiced, especially when the data used to train them is not representative of the real-world population. Overall, machine learning is a popular research and development area in both academia and industry, with the potential to radically impact a wide number of industries.

1.1 Types of Machine Learning

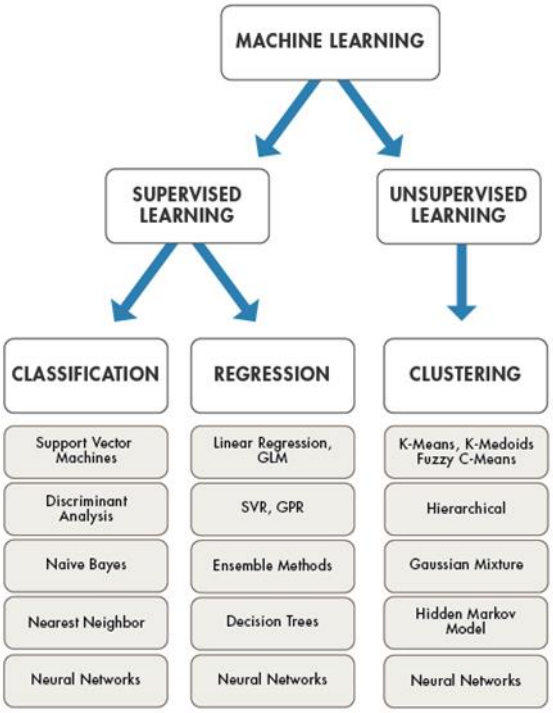

Figure 1. Types of machine learning [21]

The above figure shows the various types of machine learning namely supervised, un-supervised and reinforcement learning.

A. Supervised Learning

It's a form of machine learning method where computers are taught using appropriately "labelled" practice data and then based on the data, the algorithm is fed, which produces the required output. Data scientists and machine learning engineers must grasp which algorithm model to employ and which one to apply to the model when the data is trained. [1]

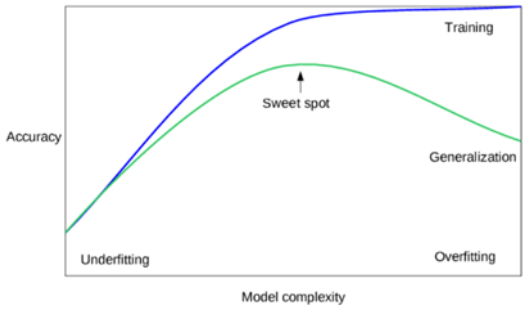

Figure 2. Supervised machine learning [2]

Two of the fundamental kinds of supervised learning algorithms are regression and classification [3]. To forecast a continuous output, like price or probability, regression techniques are utilized. In order to forecast a discrete output, such as a label or class, classification algorithms are utilized.

Supervised learning is dependent on training, much as any machine learning algorithms. Now the features and the labels are shown to the computer so that it learns, which is called the training step and the data used is called a set of practise data. Training is performed by specific algorithms that usually try to minimize an error during this training process and this is done by mathematical optimization methods. After training, new data is shown to the computer i.e., a test never seen before and where the label is unknown, and this is called the test data. And now, the trained computer should be able to make a decision based on the information it has seen, and determine the correct target value; and this is how the supervised learning works.

On the other hand, in regression, a continuous target value is to be predicted; meaning, the target value can have a more or less arbitrary value. An example of it is to predict house prices based on given information about the house and the neighbourhood. The target variable which is the price, can basically have any value here.

Concrete algorithms have a unique design and can be different in the way how it stores the information mathematically, how it solves the training process through mathematical operations and how it transforms. Popular algorithms include logistic regression, decision trees, random forests, and linear regression. Prominent algorithms include decision trees, random forests, logistic regression, and linear regression. For example. House prices. Finding the price of a property is an example of a supervised learning algorithm challenge [4]. At first,

information on the house, like its dimensions, number of rooms, assessors, whether or not it has a garden, and other details are gathered. By looking at the matching labels, the costs of these houses are determined by contrasting the information from several homes' characteristics and costs. As a result, a model can be created to forecast the price of a new house based on the observations.

B. Unsupervised Learning

Models are trained with unlabelled datasets in this form of machine learning algorithm, and they are then given free rein to use the data. For example: a dataset including photos of several unsupervised learning system is given with information about dog and cat breeds [5]. The algorithm is never trained on the provided dataset; therefore, it has no knowledge of the dataset's features. In accordance with the similarities, it will divide the image into various sections.

C. Reinforcement Learning

This system uses machine learning in a different way. An agent "AI system" will observe its surroundings, take the necessary activities, and then earn rewards. This type requires the agent to learn on its own. This kind of learning is used in numerous robotic applications.

1.2 Classification Algorithms

A few algorithms under supervised machine learning are discussed below.

A. Support Vector Machines

In machine learning, one of the key tasks is to classify a group of objects into two or more categories. Support Vector Machines (SVMs) are simple and effective methods for this purpose. In an SVM model, each object which is identified is depicted as a point in an N-dimensional space, where the parameters of the point are called as features [6]. Classification is performed by drawing a hyperplane separating all the points of one category by the other. There may be multiple hyperplanes that could be used to separate the categories in an SVM model. However, the SVM algorithm aims to achieve the one that best divides the classes by maximizing the distance between the points in each category, known as the margin [7]. Supporting vectors are the points that lie exactly on the margin.

B. Decision Tree Algorithm

Because it is made to display decisions between possible courses of action and the effects of those alternatives, the decision tree is an excellent tool for cost effectiveness analysis. There are three major components of a decision tree action options, which show the decision being considered and consequences which describe the mix of outcomes and their probabilities [8], and the value of health and cost outcomes for each set of consequences. Once these are available in a tree, the expected value of each action, option for health and cost outcomes can be calculated, and the cost effectiveness ratio can be determined to make better decisions about health spending [9]. The below given figure 4 shows the various parts of a decision tree like root node, leaf node and sub tree.

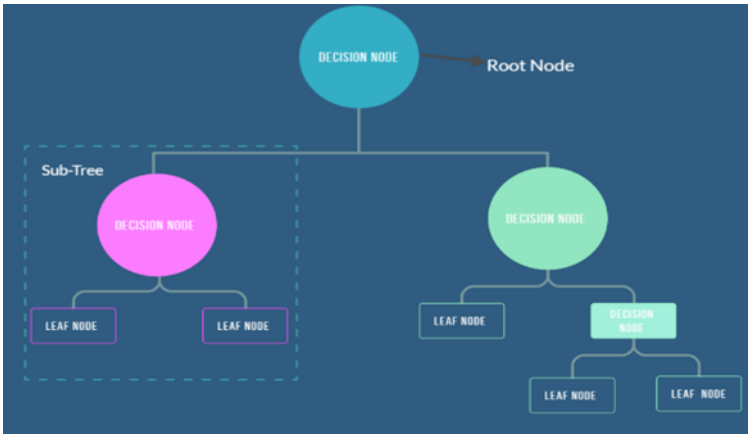

Figure 3. Decision Tree Algorithm

C. Naive Bayes Classifier Algorithm

It is a categorization technique that depends on the impartiality of features assumption in the Bayes theorem which indicates that presence of one feature does not affect the presence of another feature [10]. For example: in addition to receiving regular mails from friends and family, spam messages that are frequently frauds or unwelcome adverts are also occasionally received. To remove the spam mails, making a histogram of all the terms that appear in the typical texts from friends and family is what is done initially. The histogram may be used to determine the likelihood that each word would be seen, if it were present in a typical message. In this type of algorithm, the probabilities of seeing each word is calculated, given that the word is seen in a normal message or a spam. Now a simple formula as given below is implemented.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

D. Logistic Regression

In machine learning, a type of statistical model is called logistic regression binary categorization. It is a supervised learning algorithm that employs an optimization algorithm to learn a set of parameters that may be applied to predict the class label of fresh, untainted data. In order to produce predictions based on the input features, the algorithm learns a set of weights. An optimization algorithm is used by the algorithm during training to identify the weights that produce the training set's least error. The algorithm can then use the weights to make predictions on new data once they have been learnt.

$$p = 1 / (1 + e^{-x})$$

where, P(y=1|x) is the probability that the outcome is 1, given the input features, and z is the linear combination of the input features and weights. Afterward, a prediction is created based on the scenario with the highest likelihood. P(y=1|x) > 0.5 results in a forecast of 1, whereas P(y=1|x) = 0.5 results in a prediction of 0 [11].

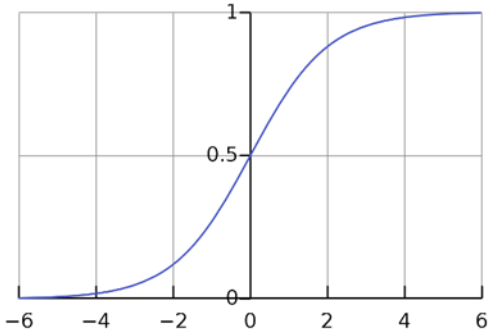

Figure 4. Standard Logistic Function [12]

Types Of Logistic Regression: Several logistic regression models exist, and they can be roughly categorized according to the number of classes they are intended to predict and the features they use as input. Several prevalent forms of logistic regression include:

i. Binary Logistic Regression:

In order to predict a binary result, binary logistic regression, the most popular kind of logistic regression, is used (e.g., “exist" or "not exist” for object on earth). In coding, this idea is frequently expressed as a 0 or 1 [13].

ii. Multinomial Logistic Regression:

This method of logistic regression is used to forecast a result from several classes that an item may fall into. Before the model is run, a set of three or more present classes is created. For example, classifying the text from which language it came from in essence, is a multiple class extension of binary logistic regression.

iii. Ordinal Logistic Regression:

An approach for classifying outcomes using a natural ordering of the classes is called ordinal logistic regression. Although there are more classes than there are in binary logistic regression, there are still just two classes. For instance, assigning a hotel a star rating between 0 to 5.

iv. Linear Logistic Regression:

When the input features are continuous and a direct correlation between the logarithmic chances of a linear relationship with the log-odds of the outcome, this type of logistic regression is used.

E. K-Nearest Neighbor (KNN)

It is a classification and regression supervised machine learning algorithm. In the case of classification, the algorithm finds the K data points that are closest to a particular data point and assigns that data point to the class that contains the majority of the K nearest data points. When there is regression, the algorithm makes a forecast for a certain data point by taking the average of the K nearest data points [14]. One of the primary benefits of KNN is that it is easy to implement and may be applied to jobs requiring both classification and regression. Another advantage is that local data structures are sensitive to it and can adapt to changes in the data over time [15].

1.3 Regression Algorithms

i. Linear Regression

The independent and dependent variables are established to have a linear relationship using a simple method called linear regression. A graph in Fig.5 illustrates the connection between the dependent and independent variables. It can be represented in the mathematical form as below:

$$y=a_0+a_1 x+\in$$

where, Y= dependent component, X= independent component, a0= line interception, a1= Ratio of linear regression, and Ɛ= random error.

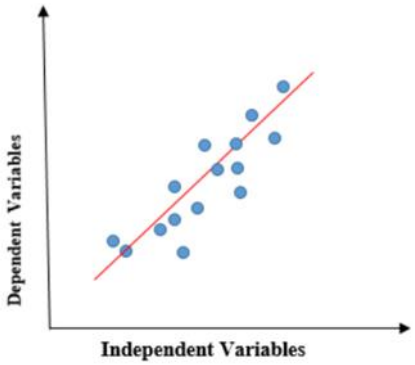

Figure 5. Standard Linear Function [22]

Types of Linear Regression:

Simple Linear Regression:

A statistical method called linear regression method involves predicting the value of a dependant variable using only one independent variable i.e., a single independent variable to predict the value of a dependent variable. This dependent variable is numerical in nature, and finding the line of best fit, which can be used to base predictions about the dependent variable on the value of the independent variable, is the aim of linear regression.

Multiple Linear Regression:

A statistical method known as multiple linear regression is used to forecast the value of a dependent variable whose values rely on a number of independent factors i.e., dependent variable based on the values of multiple independent variables. This method is used when there are multiple factors that may influence the value of the dependent variable, and the aim is to establish the relationship between these factors and the dependent variable.

A machine learning technique called a Support Vector Machine may be utilised to overcome issues with classification and regression. SVM can be used to address classification and regression problems. The SVM uses a method known as maximal margin in the case of regression to locate a hyperplane in feature space that maximally separates the input points for

various output values. Support Vector Regression (SVR) is a particular SVM variation created to address regression problems [16]. The SVR algorithm, in contrast to the SVM, focuses on regression rather than classification and seeks a hyperplane in feature space that is as close to as many data points as is practical. Because output is a real number, there are an infinite number of alternative outcomes, making it very difficult to predict the information that is now available [17]. Regression involves setting a tolerance margin (epsilon) that is generally in line with the problem's requirements that the SVM would have already specified.

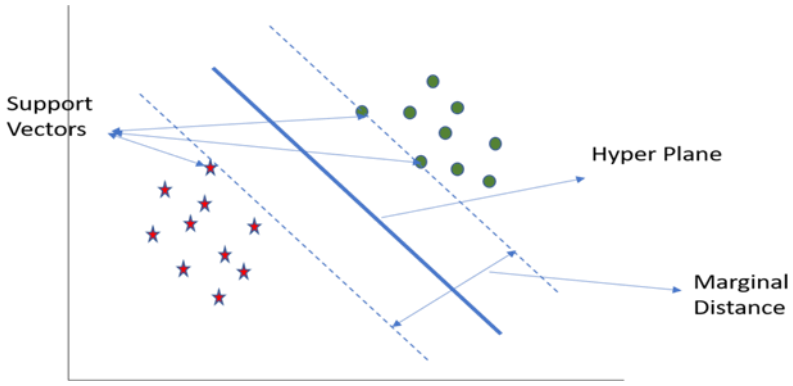

Figure 6. A graph representing SVM Model [23]

ii. Decision Tree

A supervised machine learning model is a decision tree that learns decision rules from related information in order to predict a target variable. A tree-like flow chart layout is used to make decisions using decision trees. In a decision tree, the data considering the decisions made on questions containing only yes and no as their input is usually broke down. The path from root to leaf is called a Classification Rule. Now, each of the internal nodes represent an attribute on which a particular decision needs to be taken [18]. For example, an attribute called weather condition from weather dataset, which has options as rainy, cloudy, and sunny. So, each of the leaf node represents a label which is nothing but final decision taken, after the computation of all the attributes. Decision tree algorithms are sometimes referred as classification and regression trees because these are suited very well for solving problems related to both the regression as well as classification [19]. Decision tree regression is a nonlinear regression. In case of regression, the value retrieved by the node at the terminal in the dataset or training set is the mean response of findings falling in that region; hence, if an hidden data point falls in that region, then its prediction can be made with an average or mean value. So, decision tree is good for getting predictions related to both continuous as well as categorical target variables.

Decision tree regression is used when the goal variable, whose value is continuous in nature, is to be predicted.

2. Methodology

The Cars dataset contains information about three different brands of cars: US, Japan, and Europe. The goal of the dataset is to use various variables like horsepower and cubic inches, and make year, to determine the brand of a particular car. This information can be useful for comparing different models and making informed decisions about which brand to choose. Various machine learning models can be trained on this dataset. This study has focused solely on two of these algorithms which falls under supervised machine learning, i.e., Decision Tree Algorithm and Naïve Bayes Algorithm.

1. Evaluating accuracy: Evaluating these algorithms' accuracy on a certain dataset is one approach to compare them. To do this, both algorithms are trained on a training set, predictions are made on a test set, and then the accuracy of each algorithm's predictions is assessed.

2. Comparing speed and efficiency: Measuring these algorithms' training and prediction processing speeds and efficacies, is another approach to compare them. For instance, how long it takes each method to train on a particular dataset or how many calculations are involved in each prediction, can be examined.

3. Evaluating performance on different datasets: Comparing how well these algorithms work on other datasets can also be helpful because, the outcomes might differ based on the data's features.

The following is the algorithm to train a model for comparing Naïve Bayes algorithm with Decision Tree algorithm.

1. Collect and pre-process the data to be used i.e., the two algorithms' testing and training data.

2. Separate the data into a test set and a training set.

3. Train a Naive Bayes classifier and a Decision Tree model at the training set.

4. Use the trained models to make predictions on the test set.

5. Calculate the accuracy of the predictions made by each model using a metric such as accuracy, precision, or recall.

6. Compare the accuracy of the predictions made by the two models and determine which model performs better on the given data.

a) Data Collection and Pre-processing

In order to prepare the Cars dataset for analysis, several pre-processing steps were carried out. First, the dataset's missing values were found, and their replacement with the mean value of that feature is performed. Next, any redundant or irrelevant features were removed from the dataset to reduce noise and simplify the model. For example, features such as car name and origin were removed as they were not considered to be useful in classifying the cars. After removing any redundant features, the remaining features were scaled using the MinMaxScaler method to ensure that all features are on the same scale. This step was necessary as some features had much larger ranges than others, which could lead to issues during model training and prediction. Finally, the dataset was split into two sets, a training set and a test set. The models were trained using the training set, and the test was used to evaluate their test set. To guarantee repeatability, the data split was carried out using the Scikit-Learn package's train-test-split function, a random state of 42 and a test size of 20%. Overall, the pre-processing steps performed on the Cars dataset were designed to clean and prepare the data for analysis while minimizing noise and ensuring that the models could be trained effectively.

The car dataset used in this study was obtained from Kaggle, which contained information on various car attributes such as make, model, year, price, and specifications. The dataset consisting of 10,000 samples, were split at random into an 8,000 sample training set and a 2,000 sample testing set. The testing set was used to gauge how well the machine learning models performed after being trained on the training data.

b) Model Selection and Training

In this study, two models the Naive Bayes and Decision Tree algorithms were utilized for supervised machine learning. To classify the cars dataset, the data was first divided into training and testing sets. The trained models were used to generate predictions on the testing set after training both models on the training set. The Naive Bayes algorithm is a probabilistic classification approach based on Bayes' theorem. It says that every characteristic exists independent of each other i.e., each feature is independent of every other feature, hence the

term "naive." The Decision Tree algorithm, on the other hand, creates a decision-tree-like representation of decisions and their outcomes. It learns the tree structure from the training data by dividing the data according to the splitting of the data based on the most informative features. Both models were trained on the cars dataset to classify the cars by brand (US, Japan, and Europe) using features such as horsepower, cubic inches, and make year. The accuracy of both models was assessed using criteria like recall, precision, or accuracy. Based on the evaluation, it was found that, compared to the Naive Bayes method, the Decision Tree approach was more accurate for this dataset.

c) Performance Metrics and Accuracy Calculation

Evaluating recall, accuracy, precision, and F1-score were only a few of the performance indicators utilised to assess the effectiveness of the categorization models. The most popular assessment statistic, accuracy, counts the percentage of samples with accurate identification out of all the samples in the test set. It is calculated by dividing the number of samples that were properly classified by the total sample count in the test set.

Accuracy = (number of correctly classified samples) / (total number of samples)

The fraction of accurate positive forecasts among all positive predictions is measured by precision, on the other hand. It is determined by dividing the quantity of true positives by the total of true positives and false positives.

Precision = true positives / (true positives + false positives)

Recall, often known as sensitivity, is measured by true positive predictions that are indeed positive samples. It is determined by dividing the quantity of true positive results by the total of true positive and false negative results.

Recall = true positives / (true positives + false negatives)

The recall harmonic mean and accuracy, called the F1-score, offers a more accurate assessment of the performance of the classifier. This formula is used to compute it:

F1-score = 2 * (precision * recall) / (precision + recall)

The accuracy of both models was assessed using these performance indicators, and one of the most accurate models was chosen as the best model for the current dataset.

3. Results and Discussion

This study investigated the effectiveness of two techniques of categorization on the vehicle's dataset: Naive Bayes and Decision Tree. The data was first preprocessed by removing the missing values and converting categorical features into numerical values. Then, the dataset was split between training and testing sets using an 80/20 split. In the training set, the Naive Bayes and Decision Tree models were trained, and on the testing set, how well they performed was assessed. The Naive Bayes classifier's average accuracy was 69.95% compared to the decision tree model's average accuracy of 80.08%. This shows that for this dataset, the decision tree model is more accurate than the Naive Bayes classifier. To understand why the decision tree model beats the Naive Bayes classifier in performance, the decision tree's structure and the important features it considered when making decisions were examined. It was found that the decision tree was able to use a combination of features to accurately classify the automobiles, whereas the Naive Bayes classifier relied on each feature independently. Nonetheless, it's crucial to remember a number of aspects could affect the accuracy of a machine learning model, including the quality of the data, the features selected, the model's hyperparameters, and the training set's size. Therefore, it cannot be concluded that the decision tree model will always outperform the Naive Bayes classifier for any dataset. More investigation is required to compare the effectiveness of these algorithms on different datasets. To better illustrate the findings, a bar chart showing the average accuracy of both models across five runs is depicted in the below figure. As seen in the chart, the decision tree model consistently outperformed the accuracy of the Naive Bayes classifier. In conclusion, the decision tree model appears to be a better option for classifying automobiles in the given dataset. However, drawing definitive conclusions about the performance of these algorithms on other datasets cannot be done, as the performance could vary depending on the dataset's characteristics.

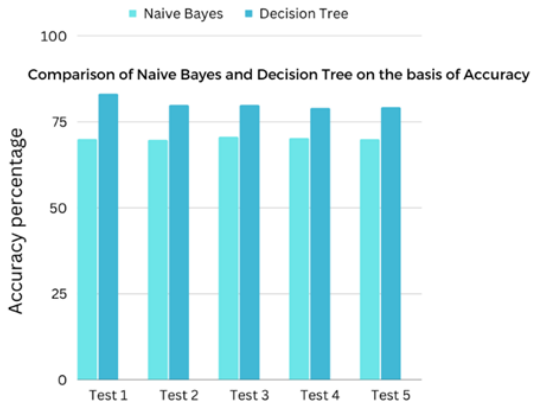

Figure 7. Comparison of Naive Bayes and Decision Tree on the basis of Accuracy

3.1 Accuracy Evaluation

1. Data Pre-processing: The data was preprocessed by removing any missing values and converting categorical variables into numerical ones using one-hot encoding.

2. Data Splitting: The preprocessed data was subsequently split in a 70:30 ratio into a training set and a testing set.

3. Model Training: The Decision Tree model and the Naive Bayes classifier were trained on the training set.

4. Model Testing: The trained models were then used to predict using the testing set.

5. Model Evaluation: The accuracy of the models was assessed using a number of criteria, including F1 score, recall, accuracy, and precision. The performance of the models was also shown making use of the confusion matrix.

6. Comparison: The Decision Tree model and Naive Bayes classifier's accuracy was compared to determine which model performs better on the given data.

3.2 Performance Comparison of Classification Models

Table.1 Performance Comparison of Classification Models

Model	Accuracy	Precision	Recall	F1 Score
Naive Bayes	69.95%	0.74	0.69	0.66
Decision Tree	80.08%	0.82	0.80	0.78

The table.1 presented above compares the performance scores of the Naive Bayes classifiers and Decision Trees used in the analysis of the automobile dataset. Recall, accuracy, precision, and F1 score are performance indicators utilized in the review process. It is clear from the table.1 that the decision tree classifier outperformed the Naive Bayes classifier in all metrics, with an accuracy of 80.08% and the performance of the Naive Bayes classifier is 69.95%. The precision score for the decision tree classifier was also higher at 81.23% compared to the Naive Bayes classifier's score of 68.74%. The recall score for the decision tree classifier

was 69%, which is higher than the Naive Bayes classifier's recall score of 64%. The F1 score for the decision tree classifier was 74.77%, which is higher than the Naive Bayes classifier's F1 score of 66.29%. These findings suggest that the Decision Tree model is a better option for this dataset.

4. Conclusion

To sum up, machine learning is an effective technique for automating and enhancing a variety of jobs and processes. It enables computers to derive conclusions or predictions from data without having such actions explicitly coded into them. The nature of the problem and the model's goals influence the choice of the machine learning algorithm. Several types of machine learning algorithms exist, each with unique strengths and disadvantages. In this research, two algorithms specifically of supervised machine learning, Naïve Bayes and Decision Tree algorithms are trained and compared, and is concluded that Decision tree is better in terms of accuracy.

References

[1] Yu, Lei, and Huan Liu. "Efficient feature selection via analysis of relevance and redundancy." *The Journal of Machine Learning Research* 5 (2004): 1205-1224.

[2] Chan, Fiona. *Enhancing Corporate Crime Enforcement with Machine Learning—A Multidisciplinary Risk Factor Approach*. Michigan State University, 2022.

[3] Kotsiantis, Sotiris B., Ioannis D. Zaharakis, and Panayiotis E. Pintelas. "Machine learning: a review of classification and combining techniques." *Artificial Intelligence Review* 26, no. 3 (2006): 159-190.

[4] Kotsiantis, Sotiris B., Ioannis Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." *Emerging artificial intelligence applications in computer engineering* 160, no. 1 (2007): 3-24.

[5] Libbrecht, Maxwell W., and William Stafford Noble. "Machine learning applications in genetics and genomics." *Nature Reviews Genetics* 16, no. 6 (2015): 321-332.

[6] Burges, Christopher JC. "Simplified support vector decision rules." In *Icml*, vol. 96, pp. 71-77. 1996.

[7] MUN, FONG KENG. "Time series forecasting using wavelet and support vector machine." (2005).

[8] Breiman, Leo, Jerome Friedman, Richard Olshen, and Charles Stone. "Classification and regression trees. Wadsworth Int." *Group* 37, no. 15 (1984): 237-251.

[9] Wehenkel, Louis. "Machine learning approaches to power-system security assessment." *IEEE Expert* 12, no. 5 (1997): 60-72.

[10] Liu, P. Q., and J. J. Feng. "An improved Naïve Bayes Text Classification Algorithm." *Microcomputer Information* 26, no. 27 (2010): 187-188.

[11] Cox, David R. "The regression analysis of binary sequences." *Journal of the Royal Statistical Society: Series B (Methodological)* 20, no. 2 (1958): 215-232.

[12] Harrell, Frank E. *Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis*. Vol. 608. New York: springer, 2001.

[13] Harrell, Jr, Frank E., and Frank E. Harrell. "Cox proportional hazards regression model." *Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis* (2015): 475-519.

[14] Elisseeff, A., and J. Weston. "A kernel method for multi-labelled classification Advances in Neural Information Processing Systems." (2002): 681-687.

[15] Quinlan, J. Ross. *C4. 5: programs for machine learning*. Elsevier, 2014.

[16] Awad, Mariette, Rahul Khanna, Mariette Awad, and Rahul Khanna. "Support vector regression." *Efficient learning machines: Theories, concepts, and applications for engineers and system designers* (2015): 67-80.

[17] Schabe, H. "Bayes estimates under asymmetric loss." *IEEE transactions on reliability* 40, no. 1 (1991): 63-67.

[18] Breiman, Leo, Jerome Friedman, Richard Olshen, and Charles Stone. "Classification and regression trees. Wadsworth Int." *Group* 37, no. 15 (1984): 237-251.

[19] Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1 (1986): 81-106.

[20] Kohavi, Ronny, and J. Ross Quinlan. "Data mining tasks and methods: Classification: decision-tree discovery." In *Handbook of data mining and knowledge discovery*, pp. 267-276. 2002.

[21] MathWorks. (n.d.). Machine Learning in MATLAB. Retrieved from https://www.mathworks.com/help/stats/machine-learning-in-matlab.html

[22] Linear Regression in Machine Learning https://www.scaler.com/topics/machine-learning/linear-regression-in-machine-learning/

[23] S. Sarker, "Linear Support Vector Machine (SVM) Explained with Python Code," Machine Learning HD, 2021. [Online]. Available: https://machinelearninghd.com/wp-content/uploads/2021/03/Linear-SVM-768x438.png.