Predicting Automobile Risk Ratings based on three classifier models

Abdrabo M.Shahin , Mazen Mohamed Gamal .
AI Research Lab (AIRL)
Arab Academy for Science, Technology, and Maritime Transport (AASTMT)
Alexandria, Egypt
abdrabomohammed25@gmail.com

***Abstract*—This paper is predicting the risk rating of automobiles that is using three classifier models: "Decision Tree", "Support Vector Machine (SVM)", and "Neural Network" to get the higher accuracy. Our dataset includes 205 samples with 26 features such as: hourse power, drive-wheels, stroke, compression-ratio, and etc…. . The target variable, `symboling`, categorizes cars as unsafe (-3) or safe (3). Preprocessing steps like handling nulls by replacing them with mode and mean, encoding the features, and dropping the unneeded columns. The highest accuracy was 91 percent using decision tree. This paper demonstrates the potential of machine learning in automotive safety analysis, with future efforts focused on enhancing performance.**

INTRODUCTION

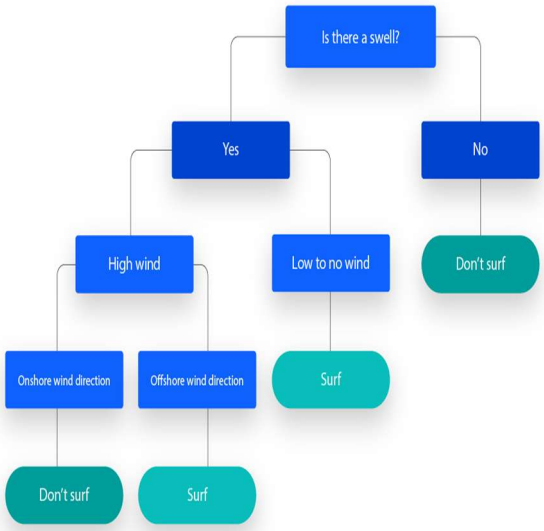
The automotive industry has come a long way since the early days of motor vehicles. Today, it stands as one of the most heavily regulated industries in the world. The safety of drivers and passengers is paramount, and the consequences of failing to meet safety standards can be dire. According to the World Health Organization (WHO), road traffic accidents rank as the eighth most common cause of mortality worldwide.

This paper explores the application to predict the car risk rating which is symbolling column in our dataset. Our data set has 26 columns which are the features like: the car maker ,wheel-base , num-of-cylinders,…etc we used three machine learning algorithms which are Decision Trees, Support Vector Machines (SVMs), and Neural Networks. A decision tree works by first asking and answering one main question called the root node and which should be a feature in the dataset for it to make sense and be useful at all.

At and from the root node, the tree has a series of yes and no questions; each designed to split the data into subsets with some specific attributes. For instance, the first question may be "Is it raining?" If yes, this defines which branch of the tree you will follow. With each question there is a different branch you follow, each of which leads to a different answer. If you answer "Yes," you follow one branch, if your answer is "No," then you will take a different branch.

This is a fine level of branching that is maintained throughout every decision. Each branch taken brings forth more questions and segments the data further. This segregational process remains until the questions posed have no more substantial weight.

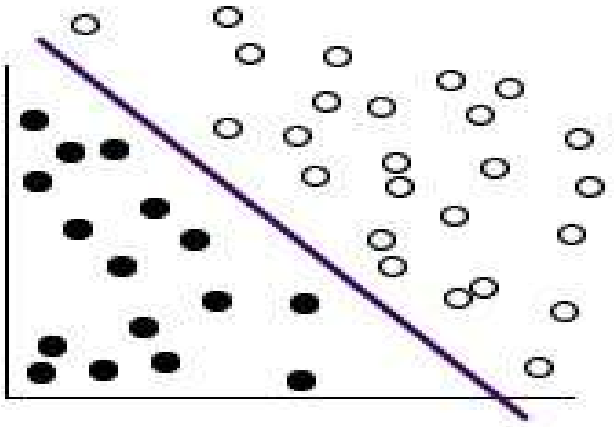
When you reach the end of a branch, you are given a final outcome which is the classification and prediction or an outcome such as determining if something is spam or not or establishing an estimated price.



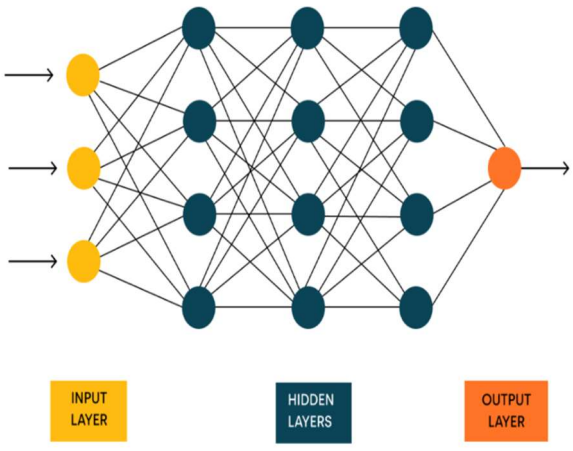
Support Vector Machine is a supervised learning algorithm designed for classification and regression tasks. To separate multiple classes of data points in an N-dimensional space, the algorithm calculates the best-suited hyperplane. The algorithm identifies the support vectors, the nearest data points to the decision boundary, and seeks to maximize the distance between them and the boundary. This usually results in better generalization to new data. Additionally, If the data is not linearly separable, SVM maps the data into a higher dimension using kernel functions such as polynomial or radial basis function. For more complex classification

978-1-4799-7165-7/15/$31.00 ©2015 IEEE

problems, this means SVM's performance improves but the susceptibility to overfitting decreases.



A neural network is a type of machine learning system that imitates the human brain and is built from layers of neurons (or nodes). It works with three types of layers: the input layer accepts raw data, single or multiple hidden layers perform computations, and the final output layer produces results. Each connection has specific importance denoted by a weight and an adjustable bias to modify predictions. The process of adjusting these parameters is neural learning and is obtained through the use of a technique called backpropagation along with gradient descent to reduce the error of the model. Nowadays, neural networks find their application in image classification, speech recognition, and predictive modeling, among others.



Methodology

we cleaned the data by drop the unnecessary columns"features". We found that we have 6 features to drop them like :fuel-type , price, fuel-type, make, num-of-doors,and aspiration .then check the nulls and replace them with the mode and median after that we encode the string to numbers so the algorithm can handle it at the end we dropped the target column .

RESULTS

After we used the three models we found that the decision tree was the higher accuracy which is the 87.07%

```
accuracy=accuracy_score(y_test,tree_y_pre)
print("model accuracy score:", accuracy*100 ," %")
```

```
model accuracy score: 87.09677419354838  %
```

Neural networks accuracy

```
print("Accuracy:", accuracy*100)
```

```
Accuracy: 77.41935483870968
```

SVM accuracy

```
print("Accuracy:", accuracy*100)
```

```
Accuracy: 72.58064516129032
```

Conclusion

Based on these three machine learning models, which we applied—"Decision Tree," "Support Vector Machine (SVM)," and "Neural Network"—the Decision Tree classifier obtained the best performance with an accuracy of 87%. The preprocessing steps followed were filling missing values with mode and mean, encoding categorical features, and omitting unnecessary and unnamed columns. Our paper demonstrates machine learning has a strong role in analyzing automotive safety with future work directed towards enhancing model performance.

References

https://www.ibm.com/think/topics/decision-trees
https://www.ibm.com/docs/en/spss-modeler/saas?topic=models-how-svm-works
https://cloud.google.com/discover/what-is-a-neural-network
https://www.geeksforgeeks.org/decision-tree/
https://www.kaggle.com/datasets/toramky/automobile-dataset