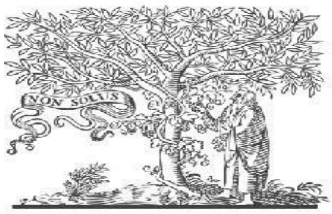
International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

COPY RIGHT



IJIEMR Transactions, online available on 29th Aug 2021.

Link :http://www.ijiemr.com/downloads.php?vol=Volume-10&issue=ISSUE-08

DOI: 10.48047/IJIEMR/V10/I08/19-20

Title:- **SALES PREDICTION USING MACHINE LEARNING TECHNIQUES**

Volume 10, Issue 08, Pages:257-262

Paper Authors
Mr.P.V.Kishore Kumar1, Bandi Sowjanya 2, Pinnamaneni Nandini Devi 3 , Jujjuvarapu Anusri 4 , B Sowrya Vara Prasad Rayulu 5

Editor IJIEMR

www.ijiemr.com

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

SALES PREDICTION USING MACHINE LEARNING TECHNIQUES

Mr.P.V.Kishore Kumar1, Bandi Sowjanya 2, Pinnamaneni Nandini Devi 3 , Jujjuvarapu Anusri 4 , B Sowrya Vara Prasad Rayulu 5
[1]Assistant Professor, Dept. of CSE, [2]17ME1A0566, [3]17ME1A05A8, [4]17ME1A0582, [5]17ME1A0569
Ramachandra College of Engineering, A.P., India

ABSTRACT:

In this project, we will be creating a model which will help us to predict the sales in future. We will be training the model with 5 different algorithms to see which algorithm will give us the best performance ,based on that we will be using the same model for predicting the sales in an organization. For this we would be using several classification algorithms like KNN, SVM, LDA, Logistic Regression, Decision Tree, Naive Bayes etc. model. The dataset will be divided into training set and test set. Using the training set the model will be built and using the test set we will determine the accuracy of the algorithm by applying the test set to the training set.
We will be generating a confusion matrix which will help to determine our results accurately. Using this project several reports will be generated to determine the accuracy of the model.

1.INTRODUCTION

Earlier companies used to produce goods without considering the number of sales and demand. For any manufacturer to determine whether to increase or decrease the production of several units, data regarding the demand for products on the market is required. Companies can face losses if they fail to consider these values while competing on the market. Different companies choose specific criteria to determine their demand and sales.
In today's highly competitive environment and ever-changing consumer landscape, accurate and timely forecasting of future revenue, also known as revenue forecasting, or sales forecasting, can offer valuable insight to companies engaged in the manufacture, distribution or retail of goods. Short-term forecasts primarily help with production planning and stock management, while long-term forecasts can deal with business growth and decision-making.
Sales forecasting is particularly important in the industries because of the limited shelf-life of many of the goods, which leads to a loss of income in both shortage and surplus situations. Too many orders lead to a shortage of products and still too few orders lead to a lack of opportunity.
In this project, we will be creating a model which will help us to predict the sales in future. We will be training the model with 5 different algorithms to see which algorithm will give us the best performance based on that we will be using the same model for predicting the sales in an organisation. For this we would be using several classification algorithms like KNN, SVM, LDA, Logistic Regression,

Decision Tree, Naive Bayes model. The term 'machine learning' is often, incorrectly, interchanged with Artificial Intelligence , but machine learning is actually a subfield/type of AI. Machine learning is also often referred to as predictive analytics, or predictive modelling.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

Coined by American computer scientist Arthur Samuel in 1959, the term 'machine learning' is defined as a "computer's ability to learn without being explicitly programmed". In supervised learning, the machine is taught by example. The operator provides the machine learning algorithm with a known dataset that includes desired inputs and outputs, and the algorithm must find a method to determine how to arrive at those inputs and outputs. While the operator knows the correct answers to the problem, the algorithm identifies patterns in data, learns from observations and makes predictions. The algorithm makes predictions and is corrected by the operator – and this process continues until the algorithm achieves a high level of accuracy/performance.

2.RELATED WORK

Existing System :

The existing system focusses on the algorithms such as Decision tree, LDA, Random Forest and XGBoost to determine the accuracy. These algorithms especially Random Forest works well when we are working with small samples of data. Random Forests tend to train each tree independently, using a random sample of the data. This would help to make the model more robust than a single decision tree, and less likely to overfit on the training data.
XGBoost build trees one at a time, where each new tree helps to correct errors made by previously trained tree. With each tree that is being added, the model becomes even more expressive.

Proposed System

In the proposed system we will be using the same old LDA and Decision tree. In addition to the we would using other algorithms like Support Vector Machine, K nearest Neighbours and Naïve Bayes to analyse the sales of an organisation. It helps us to work on the data which contains large samples, thereby helping us to determine the result accurately.These algorithms tend to be more reliable.

System Design

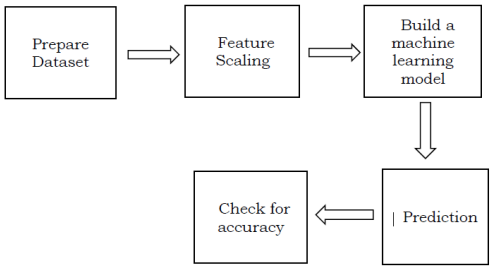


Figure 1:System Architecture

MODULES DESCRIPTION

Prepare Datasets

Use the credit score dataset and separate them into independent and dependantvariables . Convert the dependant variables to binary numbers by using one hot encoder.

Apply Feature Scaling

Divide the dataset into test set and training set. Apply feature scaling to the test set and training set. 9

Build A Machine Learning Model

Use different classification algorithms like KNN, Naïve Bayes, SVM,Decision trees etc., and test the data against these models.The model which gives lesser errors in the confusion matrix is then decided to be the best machine learning model.

Prediction

check whether the predicted result is appropriate or not.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

Check for accuracy

Validate the result against the graph and the confusion matrix and check which algorithm is giving the accurate result.

3. METHODOLOGY:

Feature Scaling

Feature Scaling or Standardization: It is a step of Data Pre Processing which is applied to independent variables or features of data. It basically helps to normalise the data within a particular range. Sometimes, it also helps in speeding up the calculations in an algorithm. Package Used.Real world dataset contains features that highly vary in magnitudes, units, and range. Normalisation should be performed when the scale of a feature is irrelevant or misleading and not should Normalise when the the scale is meaningful.

The algorithms which use Euclidean Distance measure are sensitive to Magnitudes. Here feature scaling helps to weigh all the features equally. Formally, If a feature in the dataset is big in scale compared to others then in algorithms where Euclidean distance is measured this big scaled feature becomes dominating and needs to be normalized.

K-Means uses the Euclidean distance measure here feature scaling matters. 2. K-Nearest-Neighbours also require feature scaling. 3. Principal Component Analysis (PCA): Tries to get the feature with maximum variance, here too feature scaling is required. 4. Gradient Descent: Calculation speed increase as Theta calculation becomes faster after feature scaling.

Naive Bayes, Linear Discriminant Analysis, and Tree-Based models are not affected by feature scaling. In Short, any Algorithm which is Not Distance based is Not affected by Feature Scaling.

In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. This data set may simply be bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class too. Some examples of classification problems are: speech recognition, handwriting recognition, bio metric identification, document classification etc.

4.STUDYOF RESULTS:

Figure 2:One Hot Encoder

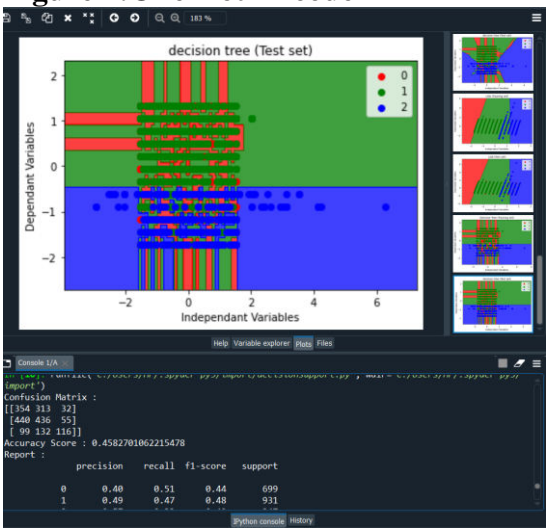


Figure 3:Output of Decision Tree algorithm

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com

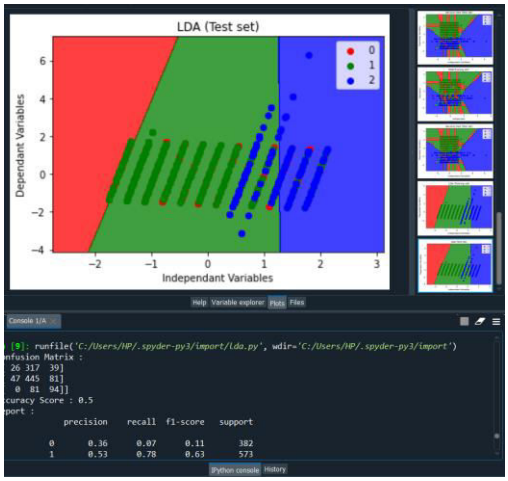


Figure 4:Output of Linear Discriminant Analysis

Figure 5:Output of Support Vector Machine

5.CONCLUSION

Sales forecasting plays a vital role in the business sector in every field. With the help of the sales forecasts, sales revenue analysis will help to get the details needed to estimate both the revenue and the income. Different types of Machine Learning techniques such as Support Vector Machine, Decision tree, Linear Discriminant Analysis , K- nearest neighbour , Naïve bayes have been evaluated on sales data to find the critical factors that influence sales to provide a solution for forecasting sales. In this project of building the best machine learning model using classifiers we would be training our model against several classifiers and thereby generate a confusion matrix and a graph where we can decide the best classifier that can be used to predict the sales of a particular organization .Based on the dataset that we have we have concluded that Naïve Bayes is the best algorithm for our sales data with 51% accuracy.

In future work one can attempt performance metrics such as time while predicting the sales. These metrics can play a crucial role in evaluating multiple Machine Learning algorithms. And also one can attempt to implement more accurate data in the continued study. Machine Learning has the advantage of analyzing data and key variables so that you can aim to develop a systematic approach using a variety of Machine Learning techniques.

6. REFERENCE

1.Machine learning approach for forecasting the sales by **Venishetty Sai Vineeth**

2.SALES PREDICTION USING MACHINE LEARNING ALGORITHMS Purvika Bajaj, Renesa Ray, ShivaniShedge, ShraveniVidhate, Prof. Dr. NikhilkumarShardoor.

3.INTELLIGENT SALES PREDICTION USING MACHINE LEARNING TECHNIQUES Mr. AkshayGodse, Miss. Poonam Pawar, Mr.SairajSawant, Miss.ShirinMujawar, Prof. Prashant Karkalle. Department of Computer

Engineering, SKN Sinhgad Institute of Technology and Science, Lonavala, SavitribaiPhule Pune University, Pune

4.General Sales Forecast Models for Automobile Markets and their Analysis Marco H¨ ulsmann, DetlefBorscheid, Christoph M. Fridrich, and Dirk Reith

5.Predicting sales using Machine Learning **Techniques Sai Nikhil Boyapati, Ramesh Mummidi**

6.Sales Prediction System using Machine Learning ArchishaChandel , Akanksha Dubey, SaurabhDhawale, MadhuriGhuge.

7.Intelligent Sales Prediction Using Machine Learning Technique

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.com