DONE BY

DILLI GANESH

B TECH IT

Sales Prediction System Using Machine

Abstract:

The objective is to perform techniques like Clustering Model and measures for sales predictions. On the basis of a performance evaluation, a best suited predictive model is suggested for the company sales trend forecast. The results are summarized in terms of accuracy of machine learning techniques taken for prediction . The main objective of the system is to analyze the future sales of a particular company and to predict whether a particular sales will increase or decrease by using different machine learning algorithms like – Random Forest just one commodity is meaningless to sellers. A general prediction for all commodities is needed.Data mining is a discipline that can be used to gather information by data classifier, Decision Tree, and Linear Regression.

Introduction

Sales prediction plays an important role in many fields and helps to improve the sales of a company by making future plans by predicting the sales of a company.Sales prediction is an important prerequisite for enterprise planning

and correct decision making, allowing companies to better plan their business activities.Previous research on sales prediction has always used a single prediction model. However, no single model can perform the best for all kinds of merchandise.Accurate prediction results for classification in data mining, this research By using the method of decision tree, sellers can boost sales.Several prediction models are included as basic models for classification. We compared the results of the prediction model with other single models.The results show that the accuracy of the prediction model is better than that of a single model. Sales forecasting plays an important role in guiding the sales and marketing of warehousing department planning warehouse location. At an equivalent time, sales data can better reflect future sales trends.

Bigmart is a big supermarket chain, with stores all round the country and its current board began a challenge to all or any Data Scientists out there to assist them create a model which will predict the sales, per product, for every store, period of time. The forecasts are generated using the flow of demands from the past as well as by considering other known factors in future. Various machine learning models are developed for the same. The purpose of forecasting sales is mainly to help the organization predict their targets and modify their strategy to improve their productivity in the coming future.

Literature Survey

[1],[9],[15],[16] In order to improve the predictive algorithm , these paper introduces observable data and qualitative analysis of previous data prediction models by using Markov model to predict the hidden values, to further improve the predictability of the model.The experimental results of this

papers indicate that the grey DNN model, an emerging and pioneering artificial intelligence technology, predict he accurate sales volume.

[2],[12],[19],[20] Sales prediction is playing a important role in many fields, such as economic

describes patterns of the market and predict the potential region of the national market commodities.However

research which is concerned on small medium enterprise with data mining. Using classification processing the forecasting, electric power forecasting, resource prediction, etc. Sales prediction is an important pre requirement for enterprise planning and correct decision making, allowing companies to predict the sales and plan accordingly . Sales prediction is important for offline businesses, The predictions are generally done by applying statistical methods, such as regression or many other models to predict the future sales and plan accordingly for the company sales

[5],[18] Supply and demand are the two main fundamental concepts of sellers and consumers. Predicting demand accurately is crucial for organizations in order to be able to set future plans for the sales. Sales Prediction is based on predicting the sales for different outlets of Big Mart companies so that they can change their future plans of the company based on the predicted sales. They propose an approach for demand prediction for Big Mart companies. The business model used by the Big Mart companies, for which the model is

implemented, that contains many outlets that sell the same product throughout the country.

[4],[8],[10] The techniques are typically different methods, which produce forecasts considering only the historical sales data of the products. However, in a situation where large quantities were applicable to only short-term sales so the related data are available, conditioning the forecast of an individual time series on past sales of the company .

[3],[7],[14],[17] The growth of internet users in the world have given impact on many aspects of daily life, including commerce. World's small medium enterprises took this advantage of new media to derive their activity by the meaning of online commerce.There were not known practical implementations of how to predict their sales using their historical transaction, therefore creating a prediction model of e – commerce sales based on data.They have built a sales prediction model on small and medium industries using previous data to predict the future sales bt due to lack of previous data and predicting model, The sales predicted were not accurate.

[6],[11],[14] Sales prediction is concerned with estimating future sales of companies such as supermarkets, groceries, restaurants, bakeries and patisseries. Sales prediction helps the company to minimize the goods whose sales is to decrease and increase the stock of the goods that are going to increase which will lead to increase in sales of the

company and the representation of the sales output variable. But the company can't find the predicted sales for a long-term.

Proposed System

A.Architecture overview

The proposed system is used to predict the sales using the machine learning models.
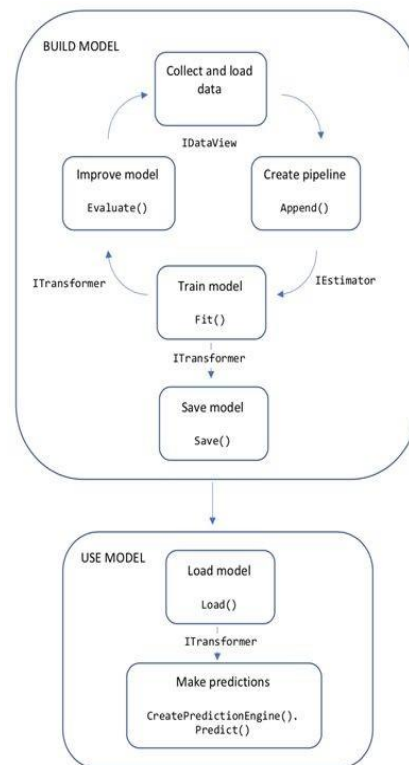

Figure 1

The fig 1 shows the system architecture in which the model trains the previous data of a mart provided using the csv file and create a prediction model using various different Machine learning algorithms like Linear Regression, Random Forest and Decision

Tree and the model is been saved, after creating the prediction model the very next moment the current data is tested using csv file. The prediction provides the predicted value (RMSE , Mean , Standard , Minimum, Maximum).

As the data source is been provided, the data is been pre-processed i.e filling all the zeros and empty spaces with the mean value of the column and clustering is done and now the testing and training is done sing various prediction models like Linear Regression , Random Forest and Decision Tree and the output of the model i.e the predicted sales is provided to the user.

B. Algorithms Used

The algorithms we used were Linear Regression,Random Forest and Decision Tree.

Random Forest: Random forests is an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and giving the output of the class. Random forests correct for decision trees' habit of overfitting to their training set.. Also random forests work well for a vast range of data items than a one decision tree does and the scaling of data does not require a random forest classifier as well. This actually forces more diversification and more variation amongst the trees in the model and finally results in lower correlation across trees. So in the random forest, we are left with the trees which are not only trained on various different sets of data but also which use different features which helps them to make decisions. In our project the parameter selected to predict the future sales is the Item_outlet_sales .

Decision Tree:A choice tree may be a simple representation for classifying examples. For this section, assume that every one among the input features have finite discrete domains, and there is one target feature called the "classification". Each element of the domain of the classification is named a category .A decision tree or a classification tree may be a tree during which each internal (non-leaf) node is labeled with an input feature.. Each leaf of the tree is labeled with a category or a probability distribution over the classes, signifying that the data set has been classified by the tree into either a specific class, or into a particular probability distribution. In our project the parameter selected to predict the future sales is the Item_outlet_sales.

Model	70% Data split	50% Data split
Random Forest Classifier	83.96%	58.52%
Decision Tree Classifier	81.21%	57.87%

From all three used algorithms of which we tested all three algorithms on 50%-50% train set-test set split and on 70%-30% train set-test set split, in which we got the highest accuracy of random forest classifier with 83.86% accuracy followed by linear regression with 82.01% and finally followed by decision tree classifier with 81.21%

Evaluation Measures

A. Home Screen

Fig 1 Home Screen

Fig. 2 Selecting Sales

On the Home Screen, the user gets 5 different options to select out of which the first one is viewing the graph of the past sales of the mart in graphical format . The second option for a user is to select Linear Regression as the model to run predict the sales, the third option for a user is to select Decision Tree as the model to run predict the sales, the fourth option for a user is to select Random Forest as the model to run predict the sales and the final option for the user is to directly provide the item_type and predict the sales and finally gives the result.
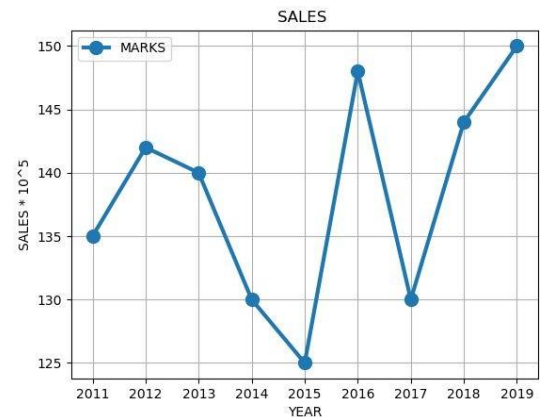
C. Selecting any Model .
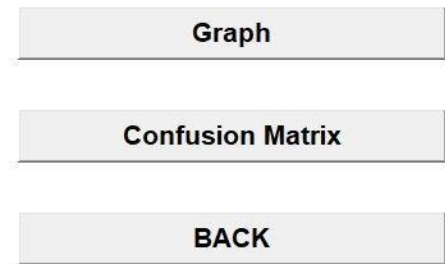
Fig 3 Selecting Model

B. Selecting Sales
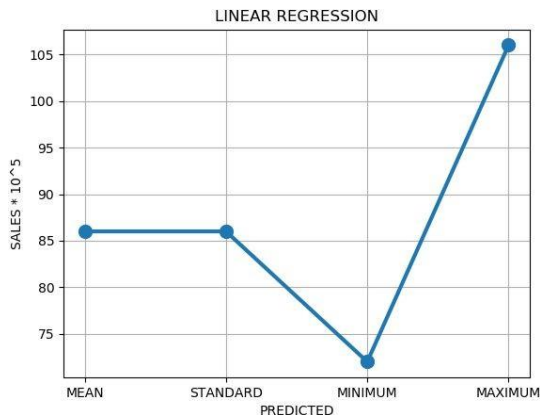
D. Selecting Graph.

Fig 4 Predicted Graph

(eg Linear Regression)

E.Selecting Confusion Matrix and Accuracy.

Fig 5 Confusion Matrix and Accuracy

(eg Random Forest)

The above figure shows the Training Accuracy and the Confusion Matrix of the selected algorithm, on choosing Random Forest Classifier, we get 83.33% accuracy.
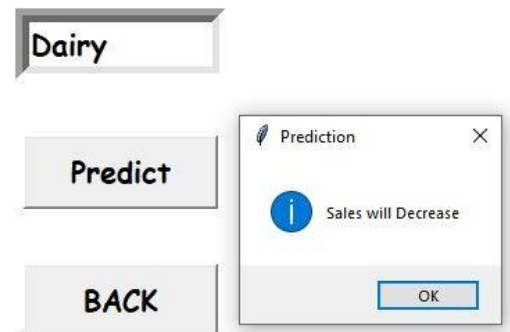
F.Selecting Predict at the Home page.
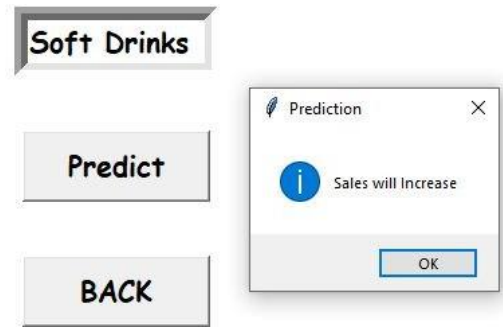
Fig 6 Predicting sales

(eg Increasing)

Fig 7 Predicting sales

(eg Decreasing)

CONCLUSION

In this paper, Making a prediction model for predicting sales and to select the best model that can be reliable i.e it's accuracy should be highest . We applied three typical forecasting models and several attributes to the trigger model through training and testing the classification model with present and past sales data. Finally we obtained more accurate forecasting results than could be obtained by comparing the models. We have also tested results of single classifiers separately with the general model.

REFERENCES:

[1]Research on sales information prediction system of e-commerce enterprises based on time series model
Authors: Jian Liu, Chunlin Liu, Lanping Zhang & Yi Xu

[2] a completely unique Trigger Model for Sales Prediction with data processing Techniques
Authors: Wenjie Huang , Qing Zhang, Wei Xu, Hongjiao Fu, Mingming Wang, Xun Liang

[3] Sales Prediction Model Using Classification Decision Tree Approach For Small Medium Enterprise Based on Indonesian E – Commerce Data
Authors: Raden Johannes H. P., Andry Alamsyah

[4] Sales Demand Forecast in E-commerce Using a Long Short-Term Memory Neural Network Methodology
Author:KasunBandara,Peibei Shi,Christoph,Bergmeir,Hansika,Hewamanage, Quoc Tran,Brian Seaman.

[5] Sales Prediction System using Machine Learning
Authors:Archisha Chandel, Akanksha Dubey, Saurabh Dhawale, Madhuri Ghuge

[6] A survey of machine learning techniques for food sales prediction.
Author:Grigorios Tsoumakas

[7] Schroeder, G., Klim, A., Heinz, G., et al. (2010) System for predicting sales lift and profit of a product supported historical sales information: U.S. Patent 7,689,456.

[8] Yuan, H., Xu, W & Wang, M. (2014) Can online user behavior improve the performance of sales prediction in E-commerce? IEEE International Conference on Systems, Man, and Cybernetics, Patent 2377–2382.

[9] Pavlyuchenko, B.M. Linear, machine learning and probabilistic approaches for statistical analysis. In Proceedings of the IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, 23–27 August 2016; IEEE: Piscataway, NJ,USA, 2016; Patent . 377–381

[10] Gao, M., Xu, W., Fu, H., Wang, M. & Liang, X. (2014). A novel forecasting method for large-scale sales prediction using an extreme learning machine. The Seventh International Joint Conference on Computational Sciences and Optimization, Patent 602–606
Processing Systems 25. Curran Associates, Inc., Patent. 2951–2959

[11] Graves, A., r. Mohamed, A. & Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Ieeexplore.ieee.org.

[12] Box, G.E.P. & Cox, D.R., 1964. An Analysis of Transformations. Journal of the Royal Statistical Society. Series B, Statistical methodology, 26(2) Patent.211–252

[13] Kulkarni, G., Kannan, P.K. & Moe, W., 2012. Using online search data to forecast new product sales. Decision support systems, 52(3), Patent.604–611.

[14] Trapero, J.R., Kourentzes, N. & Fildes, R., 2015. On the identi□cation of sales forecasting models within the presence of promotions. The Journal of the Operational Research Society, 66(2), Patent.299–307

[15] Rokach, L. Ensemble methods for classifiers. Data Mining and Knowledge Discovery Handbook; Springer: Cham, Switzerland, 2005; Patent. 957–980.

[16] Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; Patent 785–794.

[17] Michis AA (2015) A wavelet smoothing method to improve conditional sales forecasting.

[18] Seeger, M.W., Salinas, D. & Flunkert, V., 2016. Bayesian Intermittent Demand Forecasting for Large Inventories. In D. D. Lee et al., eds. Advances in Neural Information Processing Systems 29. Curran Associates,Inc., Patent. 4646–4654

[19] Snoek, J., Larochelle, H. & Adams, R.P., 2012. Practical Bayesian Optimization of Machine Learning Algorithms. In F. Pereira et al.,eds. Advances in Neural Information

[20] Elman, J.L., 1990. Finding Structure in Time. Cognitive science, 14(2), Patent.179–211