

## Survey on Car Sales Prediction System

Priya Gosavi, Rutuja Kamble, Ashwini Miraje, Saurabh Magdum , Prof.Ganesh I. Rathod

Computer Science and Engineering, Dr. J. J. Magdum College of Engineering,

Maharashtra India

### ABSTRACT

Sales Prediction is the process of estimating future revenue by predicting the amount of product or services a sales unit (which can be an individual salesperson, a sales team, or a company) will sell in the next week, month, quarter, or year. Now a days there lies a great interest in automobile industry. Thus, we get familiar with the thinking of the other people. Now the automobile industry creates, packs, runs, moves the world on its wheels and also allows other fields to flourish parallel along with them. This paper describe the serves an guiding guardian by providing accurate results for purchasing a car. This survey would cover various machine learning algorithms and methodology used in car sales prediction system.

**Key words:** Sales prediction, model, accurate, price, year, month.

### I.INTRODUCTION

Estimation investigation is a data gathering assignment to accomplish client's sentiments. Utilizing opinion investigation researchers can be examining huge quantities of archives these sentiments can be communicated into various way positive, negative and neutral routes as remarks, inquiries and solicitations. Generally, forecasts are about the future. It's hard to overstate how important it is for a company to produce an accurate sales forecast. Privately-held companies gain confidence in their business when leaders are able to trust forecasts. For publicly-traded companies, accurate forecasts confer credibility in the market. Sales forecasting adds value across an organization. Finance, for example, relies on forecasts to develop budgets for

capacity plans and hiring. Production uses sales forecasts to plan their cycles. Forecasts help sales ops with territory and quota planning, supply chain with material purchases and production capacity, and sales strategy with channel and partner strategies. Selling is the art of matching product benefits with customer needs or desires. Sell your businesses offerings by communicating the value of your product or service to your potential customers. Lead the customer through the buying decision and facilitate a satisfying transaction. Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days. Forecasting these sales in automobile industries can be performed with various and variety of technologies and one among them is the machine learning technique. That may help in classifying the prediction of an automobile for a say let us take it as car the yearly sales of it if know before then it will provide the manufacturer the huge boost in designing it, getting spare parts, getting key parts and reducing the waste products and tracking its revenue model its generation and various other activity. The classifiers used such as logistic regression, decision tree and random forest provides us with accurate prediction results.

### II. LITERATURE SURVEY

[1] The first paper is machine learning models and bankruptcy prediction. In this paper, the improvement that takes place in academics industry with the aid of machine learning algorithms in predicting bankruptcy. The data is derived from integrated resource of Salomon center database

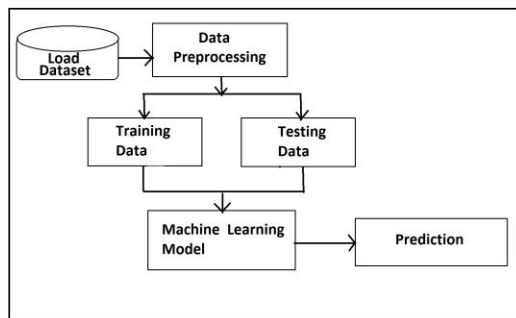
which contains the details about the North American firms from the period between 1985 to 2013.

[2] The Second paper is explaining machine learning models in sales predictions. Original Altman's Z-score variables are used as predictive variables with addition of extra variables such as the operating margin, sales, growth measures related to assets, change in return-on-equity, change in price-to-book, and number of employees based on carton and Hofer(2006). And a comparison is made between the models and these variables, the machine learning techniques and the algorithm with most accuracy is determined.

[3] The Third paper is using machine learning algorithms for housing price prediction. In this paper, Explaining machine learning models in sales prediction is a generic manuscript that discusses about the recent trends of predictive models, real time scenarios in order to gain a deep insight about buyers and sellers interaction and the forecasting of house sales.

### III. METHODOLOGY

The prediction system is divided in 3 modules.



#### Module I:

**Load Dataset:**

A data set is a collection of data. In the case of tabular data, a data set corresponds to one or more database tables, where every column of a table represents a particular variable, and each row corresponds to a given record of the data set in question.

**Data Pre-Processing:**

Data processing is, generally, "the collection and manipulation of items of data to produce meaningful information. Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

**Cleaned Dataset:**

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining multiple data sources, there are many opportunities for data to be duplicated or mislabeled.

#### Module II:

**Training and Testing Module:**

Train/Test is a method to measure the accuracy of your model. It is called Train/Test because you split the data set into two sets: a training set and a testing set. 80% for training, and 20% for testing. You train the model using the training set.

**Training Dataset:**

The sample of data used to fit the model. The actual dataset that we use to train the model. The model sees and learns from this data.

#### Module III:

**Machine Learning :**

Analyze the data set and process data set according to Machine Learning Algorithms such as Linear Regression, Random Forest, etc

**1: Linear Regression:**

Linear regression is one of the most commonly used predictive modelling techniques. It is represented by an equation  $Y=a+bX+c$ , where  $a$  is the intercept,  $b$  is the slope of the line and  $e$  is the error term. This equation can be used to predict the value of a target variable based on given predictor variable(s).

**2. Random Forest:**

Random forest is a supervised learning algorithm. The "forest" it builds, is an ensemble of decision trees, usually trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result

Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

Prediction:

By using above algorithms we are predicting how much cars are going to be sold. The output of the prediction system is the sales of car that is number of cars sold as per car attributes. The output is in the form of integer value.

Result:

Results will display in form of integers.

#### IV. REQUIREMENTS

Hardware Requirements:

- RAM – 8GB
- Processor – Intel core i5
- Hard Disk – 500GB
- Operating system – Windows 10

Software Requirements:

- Python Programming
- Jupyter Notebook
- Oracle
- Chrome

#### V. FUTURE SCOPE

In future this machine learning model may bind with various website which can provide real time data for sales prediction. Also we may add large historical data of car sales which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with user. For better performance, we plan to judiciously design deep learning network structures, use adaptive

learning rates and train on clusters of data rather than the whole dataset.

#### VI. CONCLUSION

The increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase. Therefore, there is an urgent need for a Car Sales Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help showrooms or customers, how many cars are going to be sold in next year. So that showrooms can decide how many cars they have to keep in showroom for sale. This paper tells us about algorithms for machine learning : Linear Regression and Random Forest.

#### VII. REFERENCE

- [1] J. Burez, D. Van den Poel \*, 2008, Handling class imbalance in customer churn prediction, journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa).
- [2] Byeonghwa Park a, Jae Kwon Bae b,, 2014, Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data, journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa)
- [3] Srikumar Krishnamoorthy, 2015, Linguistic features for review helpfulness prediction, journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa).
- [4] Yaya Xie a, Xiu Li a,\*, E.W.T. Ngai b, Weiyun Ying c, 2008, Customer churn prediction using improved balanced random forests, journal homepage: [www.elsevier.com/locate/eswa](http://www.elsevier.com/locate/eswa).
- [5] Zhen-Yu Chen , Zhi-Ping Fan, 2012, Distributed customer behavior prediction using multiplex data: A collaborative MK-SVM approach, journal homepage: [www.elsevier.com/locate/knossys](http://www.elsevier.com/locate/knossys).