IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FORECASTING ACCOMPLISHMENTS OF TRADES OF AN ESTABLISHMENT USING ML TECHNIQUES

[1]Pydi Meghana, [2]P. Priyanka, [3]P. Appala Raju, [4]S. Rohith

[1]Student, [2] Student, [3]Student, [4]Student

[1]Department of CSE,

[1]Raghu Institute of Technology, Visakhapatnam, India

Abstract: In the Business commerce, companies always compete to grab high-valued sales opportunities to maximize their profitability. In this regard, a key factor for maintaining a successful business enterprise is the "task of forecasting the outcome of sales". Most of the Customer Relationship Management (CRM) systems allow salespersons to manually assign a probability of winning for a new sales opportunity. This directly affect the revenue cause, often each salesperson develops a non-systematic intuition to forecast the likelihood of winning a sales opportunity with little to no quantitative rationale, neglecting the complexities of business dynamics. In this project, we address the problem of forecasting/predicting the outcomes of business sales by a thorough data-driven Machine-Learning (ML) workflow with Python.

Index Terms **– Forecasting, machine-learning, sales, regression, random forest**

I. INTRODUCTION

Machine learning is a domain of computer science which evolved from the study of pattern recognition in data, and also from the computational theory in artificial intelligence [1]. In today's world, it is the best path to most interesting careers in data analytics. As data sources increase rapidly along with the computing power to process them, going directly into the data is one of the best ways to quickly gain insights and make predictions [1].

Machine learning is a sub-domain of computer science which emerged from the computational learning theory in artificial intelligence and from the study of pattern recognition in data [3]. As the data size is growing day by day, human made analysis is becoming difficult.

So, using the computing power to process the huge data, we can quickly gain insights and make predictions using Machine learning [5,6]. Using supervised learning, we can predict the sales outcome by using the past data which is in bulk in size.

II. LITERATURE SURVEY

2.1 IMPLICIT (SALES CLOUD BY SALESFORCE.COM)

Implicit helps sales teams focus on specific actions, proven to drive revenue by scanning signals and understanding text from your CRM, calendar and email data [8]. Implicit, sales reps get timely alerts on deals and drive better performance with powerful insights on team performance [4].

2.2 "Assimilation of Machine Learning upon portending sales pipeline forecast."

The ability of sales pipelines to win forecast serves as a basic point for effective management of sales. Apart from using ratings from humans naturally, an alternate way that we propose is to use a modern machine learning technique to calculate the winning probability of sales leads [2]. A model which is specific to related profiles which is a 2-D Hawkes processes model is also developed in order to apprehend the influence from activities of sellers on their leads to the outcome of winning [6,7,9]. It is directed by two considerations:

a) sellers mostly focus on their selling tasks along with efforts put on a few leads during a period of time. This is further supported by their interactions with the pipeline which includes logging in, browsing and also updating of sales leads which are logged by the system [12];

b) the pending opportunity is prone to reach its win outcome shortly after such temporally concentrated intercommunications. Since the model is very flexible, it has the potential ability to be applicable to any other real problems [10].

2.3 "On machine learning towards integration of its insights into organizational learning - A case of Business sales forecasting."

The forecasting of business sales is best described as a process which involves decision-making and is based on internal and external data of past, characterized rules, and implicit knowledge of organization [6]. The focus of research in this paper is designed to reduce the gap between real and theoretical performance by proposing a new approach on the basis of machine learning techniques [9]. Also, the fundamental outcomes of performance of machine learning model are presented focusing on clear visualizations that enhance powerful, yet human understandable points which enable the scope for contribution to continuous organizational learning [15].

2.4 "Explaining machine learning models in sales predictions."

Complications in a business system forces the people responsible for taking decisions in an organization to choose their opinions on the basis of subjective models by reflecting their insights [11]. But research shows that companies are tend to do better by applying data driven decision-making apart from subjective ones. This supports an additional plan to invent a creative, data-driven decision models which are both comprehensive and also support the evaluation of decision options which is essential for a business environment. In recent times, a new general explanation methodology has been proposed, which supports the definition of black-box models of prediction [14]. We present a novel use of this methodology inside an intelligent system in a real- world case of business sales forecasting, a complex task frequently done judgmentally.

III. PROBLEM STATEMENT

Inaccurate sales forecast is proving to be costly to abusiness organization as inventory purchases is tiedto forecasted sales. Low inventory levels have resulted in placing rush orders to the vendor. Likewise, over stocking i.e., low inventory turnover is costing the business as cash sits idle and held up in inventory [13].

IV. OBJECTIVE – FORECAST STRATEGY

Using Regression Analysis to forecast sales for the coming period will accurately determine the quantity to be ordered. Regression analysis is a mathematical way (statistical model) forecasting relationship among dependent variables and independent variables. The analysis assimilates the time series and apprehend the recent time scenarios using time period index. Sales data will be the dependent variable in the analysis. The time period along with some other variables come under the category of independent variables [14]. In addition, the regression analysis will provide the quarterly estimates of sales which management can confidently link with the number of sales they would need from the vendor beforehand. This further makes the company to get rid of any expensive rush orders and unbalanced level of inventory and at the same time meet normal production levels [8,10,14].

V. DATASET DESCRIPTION

Majority of the trades need correct predictions for the earnings of each of their stores. These predictions allow us to plan, customize, and also to make sure that every provision has required resources [14]. In the absence of these forecasts, organizations may lose money by overstocking a store, or may have shortage of supply which leads to loss of income.

In this project, we use historical data from the XYZ Inc. chain of department stores and grocery stores to build a predictive model to forecast the revenue of each of their stores. This model can be run monthly or quarterly or annually and provide business actors with accurate predictions about the revenue for coming months or years.

We start with 3 different data sources:

- three datasets, split between our historical data and our forecasting data
- a dataset with information about each store.

Similar to other projects related to data, we also go ahead with next steps:

1. **Data Cleaning** - Here, all the unnecessary variables are removed from data and the necessary features are added into it
2. **Predictive Modelling** - Here, a predictive model framework is created and deployed
3. **Visualization** - The predicte d data is then enhanced and represented by using visualization techniques

VI. IMPLEMENTATION

6.1 Linear Regression and Model Representation

Linear Regression model is widely used as its representation and concept is comparatively simpler [1]. The representation of a linear regression is a linear equation which combines a specific set of input values (x) and the solution for them is the estimated output for that set of input values (y). Also, both of the input values and the output values are of numeric type. The linear equation assigns factor to each input value also referred to as a column which is called a coefficient indicated by the capital Greek letter Beta (B). One additional coefficient is also added, giving the line an additional degree of freedom, which is referred as the intercept or the bias coefficient. Let us consider a simple regression problem which has only a single x and y, then the form of the model is –

$$y = B0 + B1*x \quad (1)$$

In complex and advanced situations when we comprise of more than one input (x), then that line is called a hyper-plane [9]. In this representation, we use specific values for the coefficients (B0 and B1 in the above example) and it is a form of the equation. Any regression model such as linear regression has the concept of complexity to be noted. This refers to the number of coefficients which are to be used in the model [14]. Whenever a coefficient becomes 0, it removes the influence of the input variable on the model and also from the prediction made from the model (0 * x = 0). This is almost similar to the cases of regularization methods which differ the learning algorithm in order to reduce the complexity of regression models by pressurizing the absolute size of the coefficients, equalizing some of them to zero as well.

6.2 Decision Tree Regression

Decision tree is one of the predictive modelling approaches used in statistics, data mining and machine learning. Decision trees are assembled with an algorithmic approach which analyzes options to divide a data set based on unique scenarios. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non- parametric supervised learning method used for both classification and regression tasks [5].

Classification trees are the tree models in which the target variable consists of a different set of values. Regression trees are those decision trees in which a target variable can take any continuous values which also includes real numbers. The generally used terminology for this is CART (Classification and Regression Tree) [8].

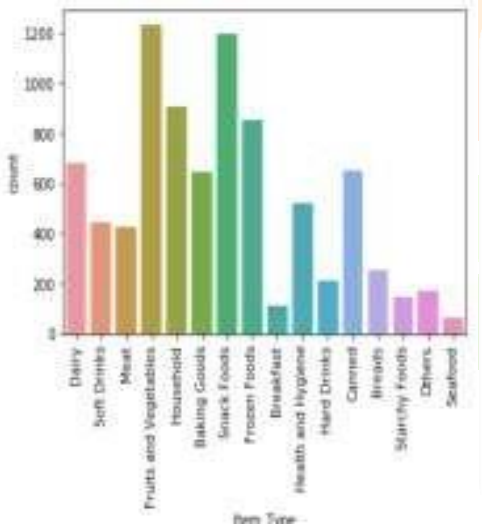
6.3 Random Forest Regression

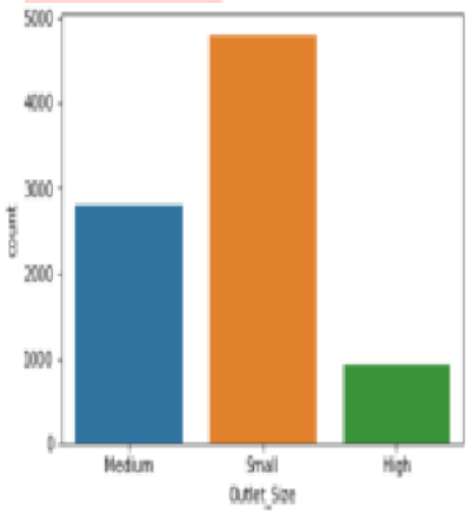
A Random Forest is an ensemble technique which can achieve classification and regression by using multiple decision trees as well as by techniques - Bootstrap and Aggregation [2]. The preliminary ideology of this is to associate more than one decision tree for determining the final outcome instead of relying on a single individual decision tree. The base learning models in a Random Forest are multiple decision trees. Also, we randomly execute feature sampling and row sampling from the dataset by forming sample datasets for every model. This process is known as Bootstrap.

For Data Pre-processing, we have used mean for each type of product rather than taking the average of all products as we got more accuracy on using product-wise means. We have normalized the data for even more best results from the input data [3].

6.4 Data Visualization

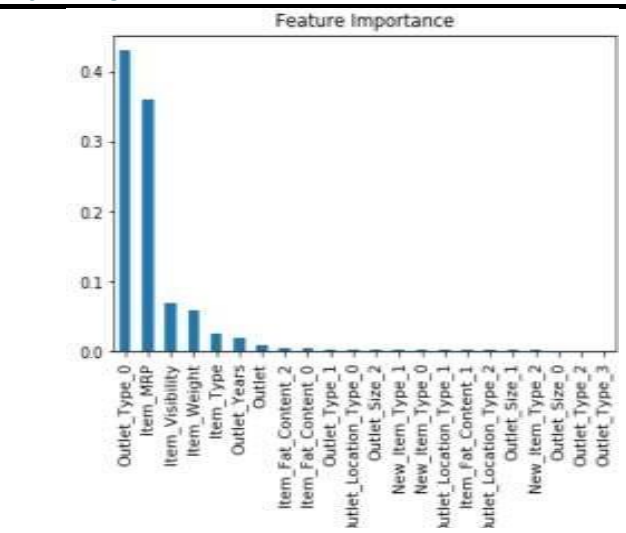
Data visualization refers to the process of presenting data in a meaningful way through the creation of charts, graphs, maps, and other visually appealing tools. Graph use in financial reports is especially advantageous as financial graphs can make complicated data easy to understand. Visuals give clear and faster ideas while doing the task of sales forecasting [6]. On the whole the strategies used in business can be formulated by geographic locations, time periods. Also choose only those variables that can highlight where it works, where it scores and join more than one variable which works in a particular geographical location [1]. When all of these scenarios are combined the result that is formulated is the data visualization which plays a vital role in forecasting the predictions for sales.



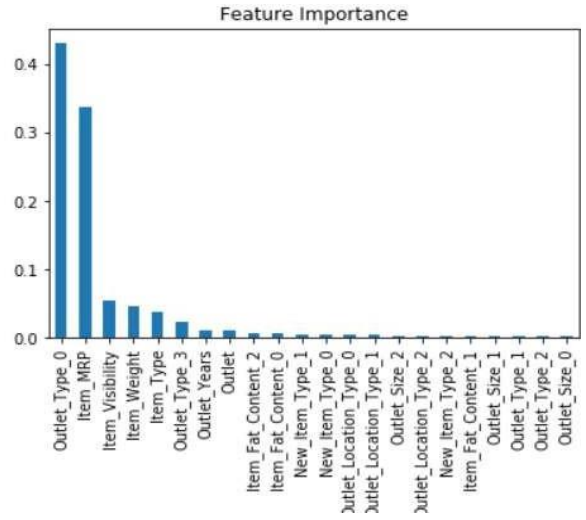


VII. RESULT ANALYSIS

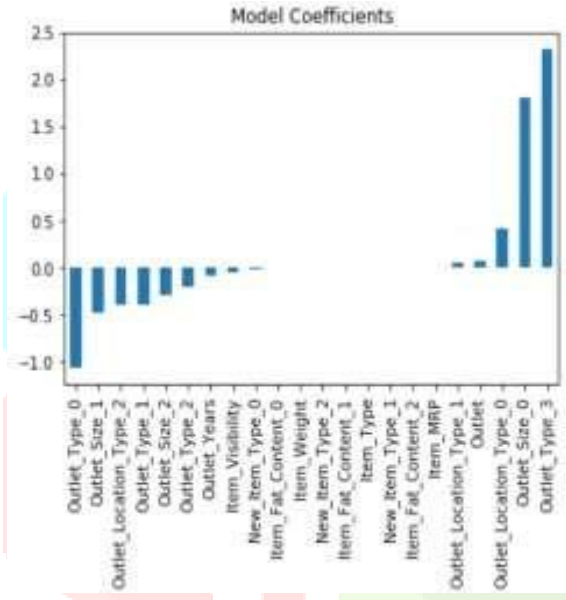
Having different predictive models with different sets of features, it is useful to consider all these results to make further decision. Also, any other linear model or any machine learning algorithm can be considered, for example, Random Forest. We can use a conventional cross validation approach; we have to split a historical data set on the training set. Predictions on the validation sets are treated with the linear regression model and other advanced regression models, which will further show the results obtained on the regression models. For the cases of sales datasets, the results i.e., cross validation and mean square value can be different and models can play more essential role in the forecasting.



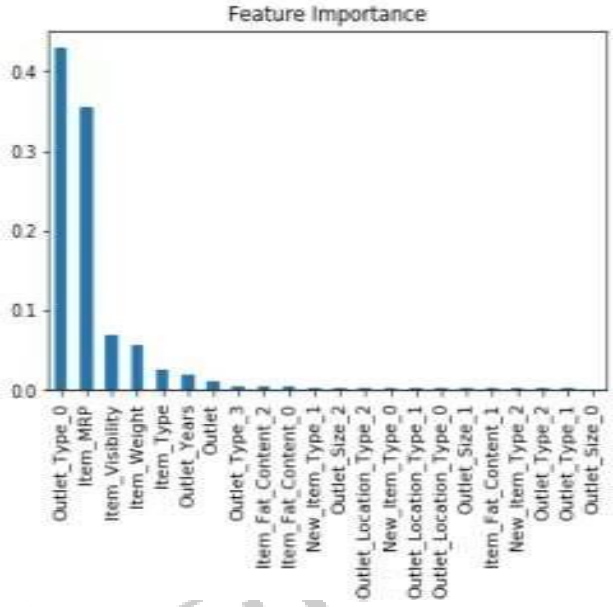
Decision Tree Regressor



Extra tree Regressor



Linear Regressor



Random Forest Tree Regression

VIII. CONLCUSION

In our project, we have considered many different machine-learning approaches for forecasting. Prediction of sales can be considered as a regression problem but not a time series problem. The usage of regression techniques for forecasting of sale can sometimes enhance good outcomes in comparison with time series models. One of the main presumptions of regression techniques is that whatever patterns are found in historical data would definitely be repeated in near future. The correctness on the validation part is an important sign for choosing an ideal number of iterations of machine learning algorithms. The effect of machine learning abstraction consists of the idea of capturing the patterns in the entire data. This effect is used in making trades forecasting whenever there exists a small number of historical data for specific sales time series in the case when a new product or store is started. In this approach, the results of multiple model predictions on the validation set are treated as input regressors for the next level models. As the next Data level model, Linear Regression, Decision Tree Regression, Extra Tree regression, Lasso regression can be used. Using it makes possible to take into account the differences in the results for multiple models with unique arrangements of domains and enhance efficiency on the validation and on the real-data.

IX. FUTURE ENHANCEMENT OF PROJECT WORK:

To better achieve the objective of predicting open opportunities, it would be prudent to capture and model how opportunity fields change over time, perhaps via periodic snapshots. This way, the company would be able to make predictions at different stages in the opportunity lifecycle. Another important application of these kinds of prediction models is to assist in determining where to invest sales time and resources for business planning. Predictions from accurate models are also worth rolling up into aggregate sales forecasts and adjusting existing “bottom-up” methods.

X. REFERENCES:

[1]. "Intro to Machine Learning | Udacity." Intro to Machine Learning | Udacity, 2021. https://www.udacity.com/course/intro-to- machine-learning ud120.

[2]. "Elements of Statistical Learning: DataMining, Inference, and Prediction. 2ndEdition. Datasets: Coronary Heart Disease Dataset." Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition, 2021.http://statweb.stanford.edu/~tibs/Elem StatLearn/.

[3]. "No Free Lunch Theorems." No Free Lunch Theorems,2021. http://www.no-free-lunch.org/.

[4]. Hastie, Trevor, Robert Tibshirani, and J. H. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-color Illustrations.New York: Springer, 2021.

[5]. Scholkopf, Bernhard, Christopher J.C. Burges, and Alexander J. Smola. Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: MIT Press, 2020.

[6]. Norming, Peter, and Stuart Russel. Artificial Intelligence: A Modern Approach. S.1.: Pearson Education Limited,2020.

[7]. Witten, I. H., and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques. Amsterdam: Morgan Kaufman, 2020.

[8]. https://zoo.cs.yale.edu/classes/cs470/ /materials/aima2020.pdf

[9]. https://www.pearson.com/us/higher-education/program/Russell-Artificial-Intelligence-A-Modern-Approach-4th-Edition/

[10]. Hastie, Trevor, Robert Tibshirani, and J. H. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-color Illustrations.New York: Springer, 2021.

[11]. Christopher J.C. Burges, and Alexander J. Smola. Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: MIT Press, 2020.

[12]. Peter, and Stuart Russel. Artificial Intelligence: A Modern Approach. S.1.: Pearson Education Limited,2020.

[13]. Witten, and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques. Amsterdam: Morgan Kaufman, 2020.

[14]. https://zoo.cs.yale.edu/classes/cs470/ /materials/aima2020.pdf

[15]. https://www.springer.com/us/higher-education/program/Artificial-Intelligence-A-Modern-Approach-/PGM1263338.html