International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 3827-3832 www.ijaem.net ISSN: 2395-5252

Predicting Business Sales Success Using Machine Learning

Mr. Ravi Teja (Asst.Professor)
Department of CSE, Raghu Institute of Technology, India.
1. A Sai kumar
2. G Noothaneswari
3. Abhimanyu JaythRai
4. B Pratyusha

Submitted: 15-07-2021 Revised: 29-07-2021 Accepted: 31-07-2021

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". Conventionally, forecasting the outcome of sales opportunities is carried out mostly relying on subjective human rating. Most of the Customer Relationship Management (CRM) systems allow salespersons to manually assign a probability of winning for a new sales opportunities. This directly effect 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.

I. INTRODUCTION

Machine learning is a sub-domain of computer science which evolved from the study of pattern recognition in data, and also from the computational learning theory in artificial intelligence. It is the first-class ticket to most interesting careers in data analytics today[1]. As data sources proliferate along with the computing power to process them, going straight to the data is one of the most straightforward ways to quickly gain insights and make predictions.

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. 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. Using supervised learning, we can predict the sales outcome by using the past data which is in bulk in size.

II. LITERATURE SURVEY

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. Implicit, sales reps get timely alerts on deals and drive better performance with powerful insights on team performance.

"On Machine Learning towards Predictive Sales Pipeline Analytics."

Sales pipeline win-propensity prediction is fundamental to effective sales management. In contrast to using subjective human rating, we propose a modern machine learning paradigm to estimate the win-propensity of sales leads over time. A profile-specific two-dimensional Hawkes processes model is developed to capture the influence from seller's activities on their leads to the win outcome, coupled with lead's personalized profiles. It is motivated by two observations: i) sellers tend to frequently focus their selling activities and efforts on a few leads during a relatively short time. This is evidenced and reflected by their concentrated interactions with the pipeline, including login, browsing and updating the sales leads which are logged by the system; ii) the pending opportunity is prone to reach its win outcome shortly after such temporally concentrated interactions. Our model is deployed and in continual use to a large, global, B2B multinational

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 3827-3832 www.ijaem.net ISSN: 2395-5252

technology enter-prize (Fortune 500) with a case study. Due to the generality and flexibility of the model, it also enjoys the potential applicability to other real-world problems.

"Integration of machine learning insights into organizational learning: A case of Business sales forecasting."

Business sales forecasting can be described as a decision-making process, which is based on past data (internal and external), formalized rules, subjective judgment, and tacit organizational knowledge. Its consequences are measured in profit and loss. The research focus of this paper is aimed to narrow the gap between planned and realized performance, introducing a novel approach based on machine learning techniques. Preliminary results of machine learning model performance are presented, with focus on distilled visualizations that create powerful, yet human comprehensible and actionable insights, enabling positive climate for reflection and contributing to continuous organizational learning.

"Explaining machine learning models in sales predictions."

A complexity of business dynamics often forces decision-makers to make decisions based on subjective mental models, reflecting their experience. However, research has shown that companies perform better when they apply data-driven decision-making. This creates an incentive to introduce intelligent, data-based decision models, which are comprehensive and support the interactive evaluation of decision options necessary for the business environment. Recently, a new general explanation methodology has been proposed, which supports the explanation of state-of-the-art black-box prediction models. Uniform explanations are generated on the level of model/individual instance and support what-if analysis. 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.

PROBLEM STATEMENT

Inaccurate sales forecast is proving to be costly to a business organization as inventory purchases is tied to 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.

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 will incorporate the time series in quarters and capture trend using time period index. Sales data will be the dependent variable in the analysis. The time period and dummy variables will form part of the independent variable. 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 in advance. This will allow the company to get rid of costly rush orders and uneven level of inventory and at the same time meet normal production levels.

We want to build a project to answer the following questions:

- What is the expected revenue of the store?
- What factors influence the revenue of a store most?

DATASET DESCRIPTION:

Many retail businesses need accurate forecasting of the revenue produced by each of their stores. These forecasts allow for planning, staffing optimization, as well as sure that each store has the necessary supply. Without these forecasts, businesses may waste money by overstocking a store, or worse yet, lose out on revenue because a store does not have enough supplies to handle predicted revenue.

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. This information can then be used to optimize business practices and streamline operations.

We start with 3 different data sources:

- three datasets, split between our historical data (used to train the model), and our forecasting data (used to deploy our model)
- a dataset with information about each store.

Like many data projects, we then proceed with three steps:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 3827-3832 www.ijaem.net ISSN: 2395-5252

1. **Data Cleaning:** we clean our data and build our features
2. **Predictive Modelling:** we build and deploy a predictive model
3. **Visualization:** we create a useful visualization of our predicted data

III. IMPLEMENTATION:

Linear Regression Model Representation [2]

Linear Regression is an attractive model because the representation is so simple.

The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output value are numeric.The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta (B). One additional coefficient is also added, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient.

For example, in a simple regression problem (a single x and a single y), the form of the model would be:

y = B0 + B1*x

In higher dimensions when we have more than one input (x), the line is called a plane or a hyper-plane. The representation therefore is the form of the equation and the specific values used for the coefficients (e.g. B0 and B1 in the above example).It is common to talk about the complexity of a regression model like linear regression. This refers to the number of coefficients used in the model.When a coefficient becomes zero, it effectively removes the influence of the input variable on the model and therefore from the prediction made from the model (0 * x = 0). This becomes relevant if you look at regularization methods that change the learning algorithm to reduce the complexity of regression models by putting pressure on the absolute size of the coefficients, driving some to zero.

Decision Tree Regression

Decision tree is one of the predictive modelling approaches used in statistics, data mining and machine learning.

Decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. 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.

Tree models where the target variable can take a discrete set of values are called **classification trees**. Decision trees where the target variable can take continuous values (typically real numbers) are called **regression trees**. Classification And Regression Tree (CART) is general term for this.

RandomForest Regression

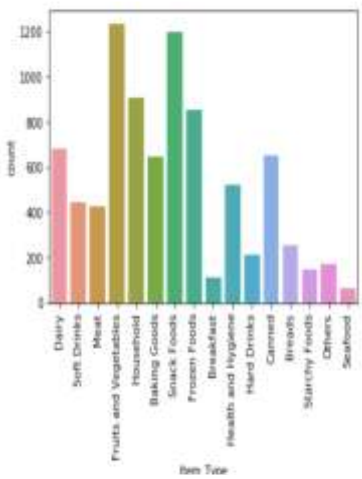
A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.Random Forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called Bootstrap.

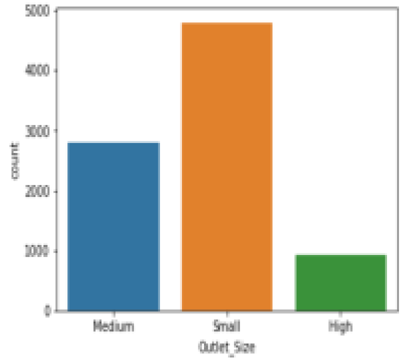
For Data Pre-processing, we have used mean for each type of product rather than taking the average of all products as we got more accuracy on using product-wise means. We have normalised the data for even more best results from the input data.

DATA VISUALISATION:

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 present better and quicker insights when forecasting sales. At a glance business strategies can be planned - time periods, geographic locations, pick variables that can highlight what works or doesn't, where it scores or doesn't, join two or more variables that work in specific geographical locations or don't, etc. All this put together makes data virtualization a very nifty tool to project what can make or break your predictions for sales!

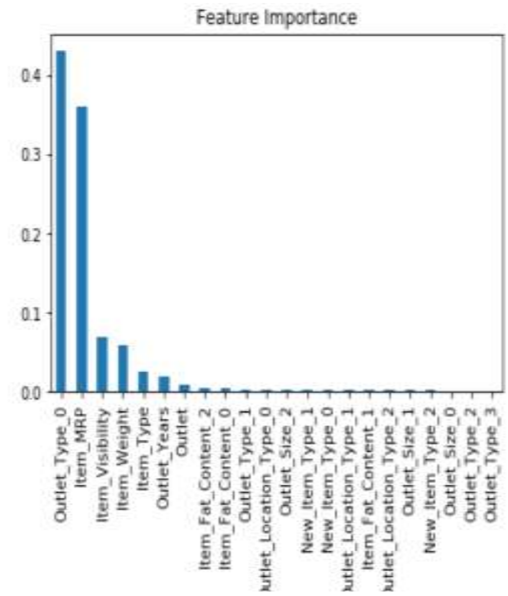
International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 3827-3832 www.ijaem.net ISSN: 2395-5252



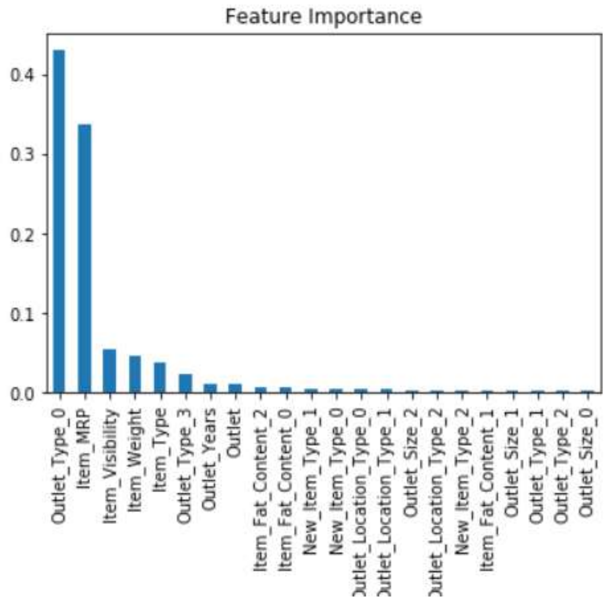


IV. RESULT ANALYSIS:

Having different predictive models with different sets of features, it is useful to consider all these results to make further decision. We can consider a linear model or anothertype of a machine-learning algorithm, e.g., 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 shows the results obtained on the regression models. For the cases of sales datasets, the resultsi.e cross validation and mean square value can be different and models can play more essential role in the forecasting.

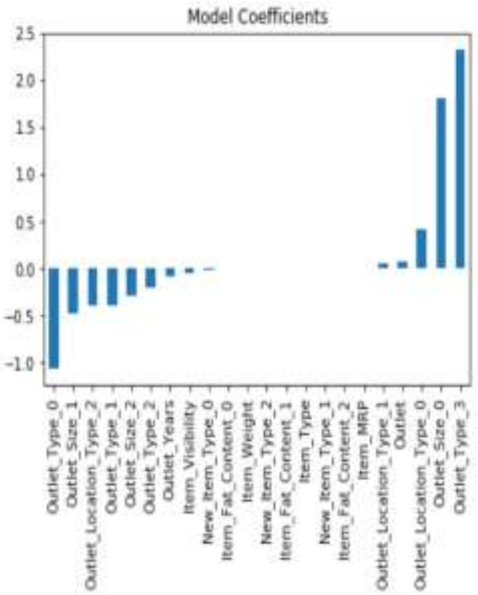


*DecisionTree Regressor

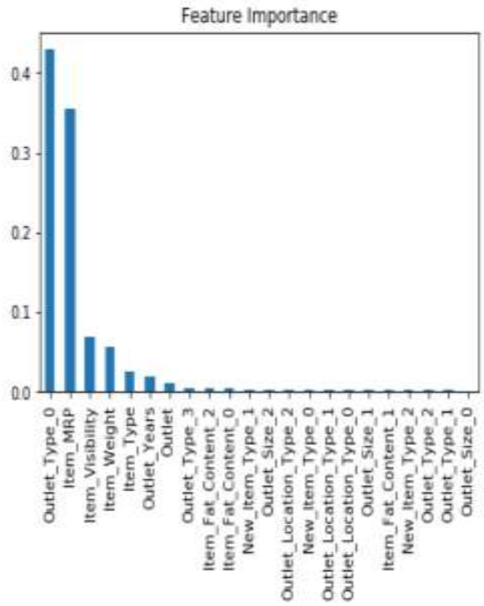


*Extra tree Regressor

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 3827-3832 www.ijaem.net ISSN: 2395-5252



***Linear Regressor**



***Random Forest Tree Regression**

V. CONCLUSION:

In our project, we have considered many different machine-learning approaches for forecasting. Sales prediction is rather a regression problem than a time series problem. The use of regression approaches for sales forecasting can often give us better results compared to time series methods. One of the main assumptions of regression methods is that the patterns in the historical data will be repeated in future. The accuracy on the validation set is an important indicator for choosing an optimal number of iterations of machine-learning algorithms. The effect of machine-learning generalization consists in the fact of capturing the patterns in the whole set of data. This effect can be used to make sales prediction when there is a small number of historical data for specific sales time series in the case when a new product or store is launched. 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 different sets of parameters and improve accuracy on the validation and on the out-of-sample data sets.

VI. 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 optimization. Predictions from accurate models are also worth rolling up into aggregate sales forecasts and adjusting existing “bottom-up” methods.

REFERENCES:

[1]. "Intro to Machine Learning | Udacity." Intro to Machine Learning | Udacity. Accessed April 27, 2016. https://www.udacity.com/course/intro-to-machine-learning--ud120.

[2]. "Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition.Datasets: Coronary Heart Disease Dataset." Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition. Accessed April 27, 2016.http://statweb.stanford.edu/~tibs/ElemStatLearn/.

[3]. "No Free Lunch Theorems." No Free Lunch Theorems. Accessed April 27, 2016.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, 2001.

[5]. Schölkopf, Bernhard, Christopher J. C. Burges, and Alexander J. Smola. Advances in Kernel Methods: Support Vector Learning. Cambridge, MA: MIT Press, 1999.
[6]. Norving, Peter, and Stuart Russel. Artificial Intelligence: A Modern Approach. S.l.:Pearson Education Limited, 2013.
[7]. Witten, I. H., and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques. Amsterdam: Morgan Kaufman, 2005.