

Sales Prediction System using Machine Learning

MR. AHER SHUBHAM DNYANDEV, MS. AHIRE GAYATRI DILIP, MR. KUTE SUMEET VILAS,
MS. MANDALE HARSHADA SANJAY

Dept. of Computer Engineering,
Pune Vidyarthi Griha's College of Engineering & S. S. Dhamankar Institute of Management,
206, Dindori Rd, behind Reliance Petrol Pump, Gorksha Nagar, Mhasrul, Nashik, Maharashtra 422004.

Abstract: In recent years, with the continuous popularity of the Internet, the number of online shopping users in my country has reached 639 million, which contains huge commercial value. In order to maintain the prosperity, diversity and order of merchants, and fully meet consumers' one-stop shopping needs, it is necessary to analyze and predict user purchase behaviors more accurately. The traditional approach of sales and marketing goals no longer help the companies, to cope with the pace of competitive market, as they are carried out with no insights to customers' purchasing patterns. Major transformations can be seen in the domain of sales and marketing as a result of Machine Learning advancements. Owing to such advancements, various critical aspects such as consumers' purchase patterns, target audience, and predicting sales for the recent years to come can be easily determined, thus helping the sales team in formulating plans for a boost in their business

Keywords: Support Vector Machine (SVM), K-Neighbors Regressor, XGBoost Regressor & LightGBM

INTRODUCTION

Sales prediction is a very important area that focuses on. Efficient and optimal prediction methods have become essential for all vendors to keep their marketing organizations efficient. Performing this task manually can result in a large amount of error, poor organizational control, and most importantly, undesirably time-consuming in this fast-changing world. Much of the global economy depends on a corporate sector that is literally expected to produce enough products to meet aggregated demand. Market Audience alignment is the primary focus of the business sector. Therefore, it is significant that the company was able to achieve this target using a prediction system.

The prediction process involves analyzing his data from various sources such as market trends, consumer behavior and other factors. This analysis also helps manage its financial resources effectively. The prediction process can be used for many purposes, such as prediction future demand for a product or service, prediction the quantity of a product that will be sold in a given period of time, and so on. This is where machine learning can come in very handy. Machine learning is where machines acquire the ability to outperform humans on specific tasks. They are used to perform professional tasks in a logical manner and to achieve better results for the advancement of modern society.

The foundation of machine learning is the art of mathematics, which can be used to formulate different paradigms to get closer to the optimal output. For sales prediction, also found machine learning to be a boon. Helps predict future sales of more accurately.

1. PURPOSE

A sales forecast is an expression of expected sales revenue. A sales forecast estimates how much your company plans to sell within a certain time period (like quarter or year). The best sales forecasts do this with a high degree of accuracy. Sales forecasts help businesses make better decisions based on future revenue, which will help them to: Forecast likely profit (or loss) in a designated period. Organize staffing levels and create HR plans. Plan the required level of production needed to meet demand.

EXISTING SYSTEM

In order to improve the precision marketing ability of merchants and reduce the decision-making cost before users place an order, the e-commerce platform needs to dig into the logic behind users' behaviors so as to push personalized marketing information to users to improve the conversion rate. we aim at the problem that the original features are not strongly related, and use feature fusion to construct strong features.

In recent years, with the continuous popularity of the Internet, the number of online shopping users in my country has reached 639 million, which contains huge commercial value. In order to maintain the prosperity, diversity and order of merchants, and fully meet consumers' one-stop shopping needs, it is necessary to analyze and predict user purchase behaviors more accurately.

OBJECTIVE OF SYSTEM

1. We are making the system to analysis the user shopping behavior and provide best product to them.
2. Avoid the Time-consuming task of feature extraction.
3. Provide a effective way with 80-85 % accuracy.
4. To provide a cost-effective solution in market

LITERATURE SURVEY

Every newspaper publisher faced with the problem of determining the number of copies of newspaper and distributing them to the retail traders. Two aspects need to be balanced out in order to optimize the economical success which is the number of unsold copies should be minimal to reduce the cost of production, and the sell-out rate should be minimal to maximize the number of sold copies. Thus, a good sales rate prediction is necessary to optimize both antagonistic aspects. This paper utilized artificial neural network to

predict newspaper sales for one vendor in the area of Sungai Petani, Malaysia. The predicted sales value can help the company to optimize their sales. The main objective is to develop a prototype that apply artificial neural network so that it can predict the future trend as well as the future daily sale. The network will consist of three layer which is input layer, one hidden layer and output layer. The input layer will have six input node where this will be the factor that will affect the output which is the number of copies that sold. The network will be trained with history data of a one year records of data. The output produced has the error value as low as 1.24% while the correlation coefficient between prediction and actual value is 0.1197.

Sales forecasting is an important aspect of different companies engaged in retailing, logistics, manufacturing, marketing and wholesaling. It allows companies to efficiently allocate resources, to estimate achievable sales revenue and to plan a better strategy for future growth of the company. In this paper, prediction of sales of a product from a particular outlet is performed via a two-level approach that produces better predictive performance compared to any of the popular single model predictive learning algorithms. The approach is performed on Big Mart Sales data of the year 2013. Data exploration, data transformation and feature engineering play a vital role in predicting accurate results. The result demonstrated that the two-level statistical approach performed better than a single model approach as the former provided more information that leads to better prediction.

Baseline prediction is an important to devise marketing strategy for a consumer goods product. Simulation techniques, time series algorithms are often used to generate baseline for the future. However the algorithm that fits a particular point of sales (POS) data varies according to the datasets. Sample set of point of sales data were simulated under different conditions and constraints incorporating seasonal and non seasonal trends. This study has compared the performance of two time series models namely Winters model and linear exponential smoothening on the simulated datasets. Winters model was found to be a better fit for the point of sales data that were used for testing.

Services organization manage a pipeline of sales opportunities with variable enterprise sales engagement lifespan, maturity levels (belonging to progressive sales stages), and contract values at any given point in time. Accurate forecasting of contract signings by the end of a time period (e.g., a quarter) is a desire for many services organizations in order to get an accurate projection of incoming revenues, and to provide support for delivery planning, resource allocation, budgeting, and effective sales opportunity management. While the problem of sales forecasting has been investigated in its generic context, sales forecasting for services organizations entails the consideration of additional complexities, which has not been thoroughly investigated: (i) considering opportunities in multi-staged sales pipeline, which means providing stage-specific treatment of sales opportunities in each group, and (ii) using the information of the current pipeline build-up, as well as the projection of the pipeline growth over the remaining time period before the end of the target time period in order to make predictions. In this paper, we formulate this problem, considering the service-specific context, as a machine learning problem over the set of historical services sales data. We introduce a novel optimization approach for finding the optimized weights of a sales forecasting function. The objective value of our optimization model minimizes the average error rates for predicting sales based on two factors of conversion rates and growth factors for any given point in time in a sales period over historical data. Our model also optimally determines the number of historical periods that should be used in the machine learning framework to predict the future revenue. We have evaluated the presented method, and the results demonstrate superior performance (in terms of absolute and relative errors) compared to a baseline state of the art method.

The retail sector has widely adapted different inventory management applications and some retail chains even employ prediction software to analyze future sales. However, a lot of day-to-day shopping in India happens through local shops. The owners of such mom-and-pop shops do not necessarily have the capital to invest in proprietary applications for setting up an inventory management system. Needless to say that same is the case for any sales prediction software. As a result, many of the shopkeepers end up hoarding a lot of irrelevant and nonprofitable products that lead to financial losses. A very cost-effective and accessible solution for this problem is a mobile application that provides all the features of a point-of-sale system as well as gives future sales insights. It will enable shopkeepers to manage their current product purchases and invoicing. The predictive sales analysis will help them to modify their investments on products and supplies thereby ensuring maximum profits. If a shop houses relevant products that cater to customer needs, its customer reach will increase. The

Economic Times published an article in the May of 2019, which stated that the number of smartphone users in India is expected to rise by 84% to 859 million by 2022 from 468 million in 2017. It is safe to assume that a large population of shop owners will have smartphones in the following years. Hence, equipping the local shopkeepers with a mobile application will prove instrumental since it will give them exposure to all the aforementioned benefits.

PROPOSED SYSTEM

- First of all, we aim at the problem that the original features are not strongly related, and use feature fusion to construct strong features. For high-dimensional systems that are slow to calculate and prone to overfitting, we use SVM to filter features, reducing the complexity of the model.
- Finally, a fusion algorithm based on XGBoost and LightGBM is proposed to predict the purchase behavior of users, and the prediction effects of random forest and GBDT algorithm are compared at the same time.
- The results show that the fusion model has the best prediction effect, and can help merchants more accurately conduct marketing activities to potential customers based on the prediction results.

SYSTEM ARCHITECTURE

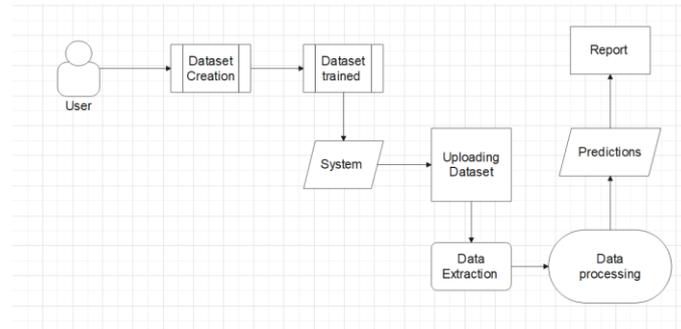


Fig -1: System Architecture Diagram

IMPLEMENTATION DETAILS (Modules)

1. Register and Login: Here we are allowing user to register first to our system which will be a security protocol used by us.
2. Dataset Creation and Trained: This module is based on machine learning where we create the dataset for system and trained the system and create model for analyzing the requirement of user.
3. Dataset Upload: Here we are uploading the excel data for analysis.
4. Processing: Here we will do the extraction of data and matching it with the dataset trained model.
5. Output: System will then provide Sales prediction report based on user behavior

ADVANTAGES

1. To gain valuable insight.
2. Sales prediction gets you into the habit of looking at past and real-time data to predict future demand.
3. To learn from past mistakes.
4. It can decrease costs

APPLICATIONS

1. Personal
2. Organization
3. Education

ALGORITHM/TECHNOLOGY

• SVM

Support Vector Machine (SVM) is a supervised machine learning algorithm used for both classification and regression. Though we say regression problems as well its best suited for classification.

The objective of SVM algorithm is to find a hyperplane in an N-dimensional space that distinctly classifies the data points. An SVM training algorithm builds a model that assigns new examples to one category or the other, making it (although methods such as exist to use SVM in a probabilistic classification setting). SVM maps training examples to points in space so as to maximize the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

• XGBoost

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now.

• Light GBM

LightGBM creates decision trees that grow leaf wise, which means that given a condition, only a single leaf is split, depending on the gain. Leaf-wise trees can sometimes over fit especially with smaller datasets.

Light GBM uses a histogram-based method in which data is bucketed into bins using a histogram of the distribution. The bins, instead of each data point, are used to iterate, calculate the gain, and split the data. This method can be optimized for a sparse dataset as well. Another characteristic of LightGBM is exclusive feature bundling in which the algorithm combines exclusive features to reduce dimensionality, making it faster and more efficient.

SYSTEM REQUIREMENTS

- **Software Used :**
 1. Python 4.4 or above
 2. Anaconda 2
 3. Windows 8 or above
 4. Pycharm
- **Hardware Used :**
 1. CPU: i3 or above
 2. RAM : 4GB or above
 3. Hard Disk : 80 GB or above

CONCLUSION

We propose a hybrid model based on Light GBM and XGBoost to predict user purchase behavior. It models from the perspective of common machine learning processes such as data processing, feature engineering, model establishment, model tuning, and result prediction, and compares it with traditional algorithms. A more accurate prediction result was obtained. Each process is crucial to the prediction of the final result, such as designing effective features, handling missing or outliers, and so on. By analyzing the importance of variables, we have identified the variables that contribute higher to the model. Through the above analysis, it can be found that the purchase prediction model widely used by most e-commerce platforms needs to be improved. The LightGBM algorithm will be one of the candidates for future improvement methods, rather than being stuck on the optimization of the old algorithm.

REFERENCES

- [1] X.Zhu,X.M.Liu,“Research on Network Purchase Behavior Prediction Based on Machine Learning Fusion Algorithm 焱”Statistics&Information Forum,vol.32,no.12,December 2017
- [2] Martínez,C.Schmuck,“A machine learning framework for customer purchase prediction in the non-contractual setting” , European Journal of Operational Research,pp.588-596,2018
- [3] C.Chen,C.Hou,“Purchase Behavior Prediction in E-Commerce with Factorization Machines ”, IEICE TRANSACTIONS on Information and Systems ,Vol.E99-D,no.1,January 2016
- [4] Y.Zhao,“Purchase prediction using Tmall-specific features”, CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE,vol.28,pp3879-3894,2016
- [5] T.Q.Chen,“XGBoost:A Scalable Tree Boosting System Tianqi Chen”, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,pp13-17,August 2016
- [6] C.Y.Hou,C.Chen,“Tree-Based Feature Transformation for Purchase Behavior Prediction”, IEICE TRANSACTIONS on Information and Systems,Vol.E101-D,no.5,pp1441-1444,May2018
- [7] R Genuer,“Variable Selection using Random Forests”, Pattern recognition letters,vol.31,no.14,pp.2225-2236,2010
- [8] H.Zhang,“The Application of XGBoost Algorithm Based on Bagging Strategy in the Prediction of Commodity Purchase”,intelligent processing and applicationVol.2,2017
- [9] S.peker,“A hybrid approach for predicting customers’ individual purchase behavior”, Kybernetes,2017
- [10] Kagan.Saar,Bekkerman.Ron,“Predicting Purchase Behavior of Website Audiences”, International Journal of Electronic Commerce,Vol.22,no.2,pp.510-539,2018