A Survey of Machine Learning Techniques For Product Sales Prediction

S. Usha[1], S.Petchimuthu[2]
[1]Dept of MCA
[1,]SRM Institute of Science and Technology
[2]Technical Trainer, SAI Ltd.

Abstract- *Product Sales Prediction is a vital a part of modern business intelligence. Customers became more fascinated by the standard of service (QoS) that organizations can provide them. Services provided by different vendors are not highly distinguished which increases competition between organizations to keep up and increase their QoS. Customer Relationship Management systems are accustomed enable organizations to amass new customers, establish a continual relationship with them and increase customer retention for more profitability. CRM systems use machine-learning models to investigate customers' personal and behavioral data to convey organization a competitive advantage by increasing customer retention rate. Predictions are accustomed design targeted marketing plans and repair offers. This paper tries to match and analyze the performance of various machine-learning techniques that are used for churn prediction problem.*

Keywords- Machine Learning Models, Sales Dataset, Sales forecasting, Sales Prediction

I. INTRODUCTION

One of the most important objectives of this research work is to search out the reliable sales trend prediction mechanism which is implemented by using data processing techniques to realize the simplest possible revenue. Today's business handles huge repository of information. the quantity of information is anticipated to grow further in an exponential manner. The measures are mandatory so as to accommodate process speed of transaction and to boost the expected growth in data volume and customer behavior. The E-commerce industry is badly in need of latest data processing techniques and intelligent prediction model of sales trends with highest possible level of accuracy and reliability. Sales forecasting gives insight into how an organization should manage its workforce, income and resources. It is a vital prerequisite for enterprise planning and higher cognitive process. It allows companies to plan their business strategies effectively.

II. MACHINE LEARNING MODELS

We can identify the three types of Machine learning methods [1]:

a) Supervised Learning

This kind of learning is characterized by using data structures that have a set of features pointing to a result, because the desired output is already known, supervised models learn progressively to imitate the desired output. There in purpose, the training system creates its own logic that will produce outputs to queries on new features. Supervised Learning is commonly used for classification.

b) Unsupervised Learning

In contrast to the aforementioned learning type, an unsupervised learning system isn't fed with an expected output or explicit feedback to form its rules; instead, the system is meant to uncover patterns by exploring a knowledge structure and extracting meaningful information. this kind of learning is for example used for data clustering, which could be a way of organizing a knowledge set into subgroups so as to acknowledge patterns or hidden structures. it's also utilized in dimensionality reduction, which consists in choosing the foremost meaningful variables for the model.

c) Reinforcement Learning

Reinforcement learning is related to Supervised Learning, It is an area of machine learning concerned with how software agents ought to take actions in an environment in order to maximize the notion of cumulative reward. Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

Time series forecasting is one in all the foremost building blocks of Machine Learning. There are many methods within the literature to attain this like Autoregressive Integrated Moving Average (ARIMA), Seasonal

Autoregressive Integrated Moving-Average (SARIMA), Vector Auto-regression (VAR), and so on.

III. APPLYING MACHINE LEARNING TO SALES PREDICTION

To demonstrate how previous sales data may be accustomed predict future sales [3], we will apply neural network library in R language to a dataset from UCI machine learning depository- Dresses_Attribute_Sales Data Set.

The idea is to potentially apply the identical method to enhance other areas of sales. Say, we are able to come up with a group of attributes that describe qualities of a product page in an e-Commerce store. Then, if these qualities do affect sales, we must always be ready to predict possible results of the merchandise page changes.

for example some textile attributes for data sets are,

- Style
- Price
- Rating
- Size, color
- Season etc..,

IV. SALES FORECASTING

Online and offline retailers understand: the external environment has become overly complex and unpredictable, because the number of products gets overlarge to manage by hand. Therefore, retail uses algorithms to line the correct prices, predict stocks and optimize other business processes. Before the start of predicting, the algorithm has to be trained. Often, the training is conducted on historical data—supervised learning [3]—where there's a target function sales, revenue, profit or market share. Through the educational process, the model analyzes all the variables that affect sales (prices, traffic, etc.) and outputs a function describing sales. After the algorithm has completed the training and showed sufficient accuracy of the forecast, it may be accustomed recommend which values should be taken within the future to maximize sales. If the retailer incorporates a significant amount of knowledge, it can use neural networks to recommend stock-ups or prices to maximize sales.

If there's not enough data to form and train a neural network, there are other algorithms that require fewer data. If the retailer’s portfolio incorporates a sufficient sales history just for about 30% of the products, its traffic is tiny, and sales are scarce, there's no space for the neural network to figure.

During this case, tree algorithms are ready to forecast sales at the extent of individual products. An example of a tree algorithm is XGboost or its younger and bold pursuers, LightGBM and CatBoost. For this kind, an energetic sales history of 150+ days is sufficient to predict optimal prices. This algorithm incorporates a disadvantage: it's typically poor at taking into consideration the interdependence of costs for various products. It may be used for KVI-products, while the remaining products may be managed with simple pricing scenarios (e.g., rule-based pricing).

On a little number of products (20-30 positions) a retailer can use regression to calculate the value elasticity of those goods by adding of 3-4 variables. it may be used for high-level decision making: is there room for a increase, or is it better to refrain from it.

Example: linear or polynomial regression (SVM – support vector machines)

This algorithm doesn't say what the value should be to maximize sales and margins but shows the trend. Another method that's used when the retailer doesn't have enough data is A/B testing, which uses mathematics and statistics. Beginner retailers can use it to assess the impact of ads and costs on sales. So, we can find sales forecasting must include the following fields,

1) Accelerates processing speed
2) Provides a additional correct forecast
3) Automates forecast updates supported the recent knowledge
4) Analyzes additional knowledge
5) Identifies hidden patterns in knowledge
6) Creates a strong system
7) will increase ability to changes

V. SALES PREDICTION

Sales prediction is quite a regression problem than a statistic problem. Practice shows that the use of regression approaches can often give us better results compared to statistic methods. Machine-learning algorithms make it possible to seek out patterns within the statistic. we are able to find complicated patterns within the sales dynamics, using supervised machine-learning methods[4]. a number of the foremost popular are tree-based machine-learning algorithms,

e.g. Random Forest, Gradient Boosting Machine.

One in all the most assumptions of regression methods is that the patterns within the past data are repeated in future. We studied linear models, machine learning, and probabilistic models for statistic modeling. For probabilistic modeling, we considered the utilization of copulas and Bayesian inference approaches. We studied the logistic regression within the problem of detecting manufacturing failures. For logistic regression, we considered a generalized linear model, machine learning and Bayesian models. We studied stacking approaches for statistic forecasting and logistic regression with highly imbalanced data. within the sales data, we are able to observe several kinds of patterns and effects. They are: trend, seasonality, autocorrelation, patterns caused by the impact of such external factors as promo, pricing, competitors' behavior. We also observe noise within the sales. Noise is caused by the factors which don't seem to be included into our consideration. Within the sales data, we are able to also observe extreme values outliers.

VI. CONCLUSION

In our research, we considered different machine-learning approaches for statistic forecasting. Sales prediction is very a regression problem than a statistic problem. The utilization of regression approaches for sales forecasting can often give us better results compared to statistic methods. One in all the most assumptions of regression methods is that the patterns within the historical data are repeated in future. The accuracy on the validation set is a very important indicator for selecting an optimal number of iterations of machine-learning algorithms.

VII. ACKNOWLEDGMENT

This work is a part of a Research Project and the authors would like to thank the anonymous reviewers for their thorough reviews, and constructive suggestions which significantly enhance the presentation of the paper.

REFERENCES

[1] Juan Pablo Usuga Cadavid, Samir Lamouri, Bernard Grabot. Trends in Machine Learning Applied to Demand & Sales Forecasting: A Review. International Conference on Information Systems, Logistics and Supply Chain, Jul 2018, Lyon, France. .

[2] Alex Gaggin https://rstudio-pubs-static.s3.amazonaws.com/105869_f6e7f8d4e0434c40bd939a3d1e792af9.html

[3] Nikolay Savin "Machine Learning for Sales Forecasting: Tackling insufficient data issue" , sep-7,2018

[4] Bohdan M. Pavlyshenko, "Machine-Learning Models for Sales Time Series Forecasting", MDPI open access Journals, 18 January 2019

[5] Pei-Yuan Zhou ,Keith C.C. Chan, Member, IEEE, and Carol XiaojuanOu, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE 2018.