EXPANDED STORE SALES PREDICTION

Maliha Aafrin Mujahid Sayyad

[1] Maliha Aafrin Mujahid Sayyad
Student, M.Tech CSE, Shreeyash College of Engineering and Tech Aurangabad, Maharashtra, India 2 Dr B.M Patil

Principal, Shreeyash College of Engg and Tech Aurangabad,Maharashtra, India

ABSTRACT

Sales forecasting is the process of predicting future sales. It is the vital part of the financial planning of the business. Most of the companies heavily depend on the future prediction of the sales. Accurate sales forecasting empower the organizations to make informed business decisions and it will help to predict the short-term and long-term performances. A precise forecasting can avoid overestimating or underestimating of the future sales, which may leads to great loss to companies. The past and current sales statistics is used to estimate the future performance. But it is difficult to deal with accuracy of sales forecasting by traditional forecasting. For this purpose, various machine learning techniques have been discovered. In this work, we have taken Black Friday dataset and made a detailed analysis over the dataset. Here, we have implemented the different machine learning techniques with different metrics. By analyzing the performance, we have trying to suggest the suitable predictive algorithm to our problem statement.

Keyword: - *1ML,2 Strategies,3 store 4 eco-friendly,5 workforce 6 enormous, 7 techniques, 8 enhances,9 resources ,10 historical ,*

Retail businesses face challenges in accurately forecasting sales due to dynamic market conditions and reliance on traditional methods. This project aims to predict daily sales for Crossman drug stores using machine learning techniques. By leveraging historical data and external factors, the project seeks to enhance operational efficiency, inventory management, workforce planning, and marketing strategies. Accurate sales predictions will help store managers make informed decisions and improve resource allocation.

Sales play a key role in the business. At the company level, sales forecasting is the major part of the business plan and significant inputs for decision-making activities. It is essential for organizations to produce the required quantity at the 6 7 specified time. For that, sales forecasting will gives the idea about how an organization should manage its budgeting, workforce and resources. This forecasting helps the business management to determine how much products should be manufacture, how much revenue can be expected and what could be the requirement of employees, investment and equipment. By analyzing the future trends and needs, Sales forecasting helps to improve the business growth. The traditional forecasting systems have some drawbacks related to accuracy of the forecasting

and handling enormous amount of data. To overcome this problem, Machine-Learning (ML) techniques have been discovered. These techniques help to analyses the big data and play a important role in sales forecasting. Here we have used supervised machine learning techniques for the sales forecasting.

2. PREVIOUS WORK

A lot of work has been done related to sales prediction, as it's one of the most important concern by the retailer. Sales prediction could be done by customer related features, store related features, and item related feature. For example, (Chen, Lee, Kuo, Chen, & Chen, 2010) has been working on forecasting sales model on fresh food. The prediction is based on both item and customer, as when consumers are making purchases of food products, they would first consider if the foods are fresh and if they are expired. But the methodology can't feed into our problem, as only store related features are provided and customer-item prediction could not be predicted at all. Another retail sale prediction problem has been described in (Giering, 2008). It's also a sales prediction problem based on customer related feature and item related feature where SVD and recommendation system is applied. Although the methodology can't be well applied in our problem, there is still inspiration from their work: using log(Sales) as the target in prediction as it might normalized the distribution. (Chang, Liu, & Lai, 2008) has described a way to make sales prediction only based on item related feature which is more similar to our problem where only store related feature is provided. In (Chang et

al., 2008), it obtained casebased reasoning model and k-nearest neighbours' algorithm to find the most similar item with sale history, given an item without sale history. We tried k-nearest neighbours' algorithm in our dataset to find the most similar store and time information. We got some result; however, its performance is not as good as expected. The model and the result could be found in the following section. (Thiesing, Middelberg, & Vornberger, 1995) has adapted Back-Propagation as a neural network method to make sales prediction on Transputer system. The article has also described how they applied parallel computing into the model to improve the efficiency of computing

1. Importance of Sales Prediction in Retail

Accurate sales forecasting has far-reaching implications for retail businesses. Below are some of the key benefits.

1.1 Inventory Optimization

One of the biggest challenges in retail is managing inventory effectively. Overstocking can lead to increased storage costs, product obsolescence, and wastage, particularly in industries dealing with perishable goods. On the other hand, under stocking can result in lost sales and dissatisfied customers. A well-structured sales prediction model allows retailers to adjust their stock levels based on anticipated demand, reducing costs and ensuring product availability.

1.2 Workforce Planning

Retailers often struggle with workforce management, particularly during peak seasons or promotional events. Predicting sales trends helps managers optimize employee schedules by ensuring that sufficient staff is available during high-traffic periods while reducing unnecessary labor costs during slower days. This improves overall operational efficiency and enhances customer service.

1.3 Financial Planning and Budgeting

Revenue prediction plays a crucial role in financial decision-making. By forecasting sales accurately, businesses can set realistic revenue targets, allocate budgets more effectively, and make informed decisions regarding expansion, marketing campaigns, and operational investments.

1.4 Personalized Marketing Strategies

Understanding future sales trends enables businesses to create targeted marketing strategies. By identifying which products are likely to perform well, retailers can implement personalized promotions and discounts that increase customer engagement and drive revenue growth.

1.5 Competitor Analysis

Retail businesses operate in a competitive landscape where external factors, such as new competitor stores, pricing strategies, and promotional campaigns, can significantly impact sales. A strong sales prediction model helps retailers understand these influences and develop strategic responses to maintain or enhance market share.

1.6 Seasonal Demand Adjustments

Retail sales exhibit strong seasonality, with higher demand observed during specific periods such as holidays, festivals, and major shopping events like Black Friday. Forecasting these fluctuations allows businesses to prepare by adjusting inventory, marketing efforts, and staffing levels accordingly.

1.7 Supply Chain Management

A precise sales prediction model helps suppliers manage production and logistics more efficiently. Knowing which products will be in demand allows suppliers to optimize manufacturing schedules, reduce supply chain delays, and improve distribution efficiency

2. Initial Data Preprocessing

Before building a predictive model, it is essential to clean and preprocess the dataset. Below are the key preprocessing steps.

2.1 Handling Missing Values

● Some records in the dataset may have missing values for attributes such as competition distance, promotional participation, or store opening dates.

● Imputation techniques, such as replacing missing values with the median or mean, can be used.

● Alternatively, missing values may indicate meaningful patterns (e.g., some stores never had competitors nearby), which should be analyzed.

Encoding Categorical Variables

● Machine learning models require numerical inputs, so categorical features (e.g., store type, holiday type, assortment level) need to be converted into numerical form using techniques like one-hot encoding or label encoding. 4.3 Feature Scaling

● Some features, such as competition distance, have large numerical ranges compared to others.

● Scaling methods like Min-Max scaling or Standardization help normalize feature distributions, improving model performance.

3.1 Feature Engineering

To improve prediction accuracy, new features can be derived from existing ones:

● Date-Based Features: Extracting additional time-related features such as day of the week, month, year, and whether the date falls on a weekend or holiday.

● Lag Features: Using past sales data as features to capture historical trends.

● Interaction Features: Creating new features that capture interactions between variables, such as the effect of promotions on different store types.

4. Advanced Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) provides critical insights into the dataset before model development. By understanding sales trends, seasonality, and feature relationships, we can extract meaningful patterns that help improve prediction accuracy.

4.1 Time Series

Analysis of Sales data is inherently time-dependent, meaning past sales trends influence future sales. A time-series analysis helps identify the following components:

● Trend: The overall direction in which sales are moving (upward, downward, or stable).

● Seasonality: Recurring sales patterns observed at specific times of the year (e.g., holidays, weekends, or promotional seasons).

● Cyclic Variations: Sales fluctuations due to economic cycles or industry trends.

● Irregular Variations: Unpredictable fluctuations caused by external events such as natural disasters, pandemics, or sudden market changes. 6.1.1 Sales Trend Analysis

● A rolling mean can smooth daily fluctuations and reveal long-term trends.

● Sales trends may show an increasing or decreasing pattern depending on external factors such as competitor presence or changes in store policies.

4.1.1 Sales Trend Analysis

● A rolling mean can smooth daily fluctuations and reveal long-term trends.

● Sales trends may show an increasing or decreasing pattern depending on external factors such as competitor presence or changes in store policies.

4.1.2 Seasonal Sales Pattern

● Sales may increase during the holiday season, Black Friday, or back-to-school periods.

● Visualizing sales data using line plots and box plots can help detect seasonal peaks. 4.1.3 Day of the Week Effects

● Sales often exhibit differences based on the day of the week (e.g., higher sales on weekends).

● grouping sales data by weekdays can highlight patterns such as increased customer traffic on Saturdays.

4.3 Impact of Promotions on Sales

Promotions play a critical role in driving customer engagement and increasing revenue.

4.3.1 Short-Term Promotions (Promo Feature)

● Temporary discounts and sales events can boost daily revenue.

● Analyzing the effect of "Promo" feature on sales reveals the percentage increase in revenue during promotional periods.

● Continuous promotional campaigns (e.g., loyalty programs) aim to sustain long-term customer retention.

● A comparative analysis of stores running "Promo2" versus those that do not can help evaluate the effectiveness of sustained promotions. 6.3.3 Promotional Intervals

● Stores often run promotions in specific months (e.g., summer sales, winter discounts).

● Examining the "Promo Interval" feature can highlight seasonal patterns in promotional effectiveness.

Conclusion and Future Scope

● Findings: The XGBoost model provides the most accurate sales forecasts, leveraging historical trends, promotions, and seasonal effects.

● Business Impact: Improved inventory management, better staffing decisions, and enhanced marketing strategies.

● Future Improvements: ○ Incorporating deep learning models (LSTMs). ○ Enhancing data preprocessing with external factors like weather.

hyper parameter tuning for further optimization.

5. REFERENCES

[1] Ching Wu Chu and Guoqiang Peter Zhang, “A comparative study of linear and nonlinear models for aggregate retails sales forecasting”, Int. Journal Production Economics, vol. 86, pp. 217-231, 2003.

[2] Giuseppe Nunnari, Valeria Nunnari, “Forecasting Monthly Sales Retail Time Series: A Case Study”, Proc. of

IEEE Conf. on Business Informatics (CBI), July 2017.

[3] https://halobi.com/blog/sales-forecasting-five-uses/. [Accessed: Oct. 3, 2018]

[4] Zone-Ching Lin, Wen-Jang Wu, "Multiple Linear Regression Analysis of the Overlay Accuracy Model Zone", IEEE Trans. on Semiconductor Manufacturing, vol. 12, no. 2, pp. 229 – 237, May 1999.

[5] O. Ajao Isaac, A. Abdullahi Adedeji, I. Raji Ismail, "Polynomial Regression Model of Making Cost Prediction In Mixed Cost Analysis", Int. Journal on Mathematical Theory and Modeling, vol. 2, no. 2, pp. 14 – 23, 2012.

[6] C. Saunders, A. Gammerman and V. Vovk, "Ridge Regression Learning Algorithm in Dual Variables", Proc. of Int. Conf. on Machine Learning, pp. 515 – 521, July 1998.IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 56, NO. 7, JULY 2010 3561.

[7] A. Koutsoukas, K. J. Monaghan, X. Li and J. Huan, Deep-learning: investigating deep neural networks hyperparameters and comparison of performance to shallow methods for modelling bioactivity data," Journal of Cheminformatics, 2017.

[8] Xinqing Shu, Pan Wang, "An Improved Adaboost Algorithm based on Uncertain Functions", Proc. of Int. Conf. on Industrial Informatics – Computing Technology, Intelligent Technology, Industrial Information Integration, Dec. 2015.

[9] A. S. Weigend and N. A. Gershenfeld, "Time series prediction: Forecasting the future and understanding the past", Addison-Wesley, 1994.

[10] N. S. Arunraj, D. Ahrens, A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting, Int. J. Production Economics 170 (2015) 321-335.