INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

Sales Prediction using Machine Learning

Selvi P, G. Palanisamy

Assistant Professor, Department of CSE, Jayam College of Engineering and Technology, Dharmapuri,

Tamil Nadu, India

Student, Department of Master of Computer Application, Jayam College of Engineering and Technology, Dharmapuri,

Tamil Nadu, India

ABSTARCT: Sales prediction is a critical task for businesses aiming to optimize inventory management, resource allocation, and overall profitability. Machine learning (ML) techniques have emerged as powerful tools for making accurate sales forecasts by analyzing historical data and identifying patterns. This paper explores various ML algorithms and approaches commonly used for sales prediction, such as linear regression, decision trees, and neural networks. It discusses the preprocessing steps involved, including data cleaning, feature selection, and normalization, which are crucial for improving prediction accuracy. Additionally, the paper examines the challenges associated with sales prediction, such as seasonality, market dynamics, and outlier detection, and proposes strategies to address these challenges effectively. Finally, it highlights the importance of evaluating model performance and selecting appropriate metrics to assess the accuracy and reliability of sales predictions. Through this exploration, the paper aims to provide a comprehensive overview of the application of machine learning in sales prediction and its potential impact on business decision-making processes.

KEYWORDS: Sales prediction, machine learning, data pre-processing, mean squared error (MSE), Market basket analysis

I. INTRODUCTION

Sales prediction using machine learning is a critical component of modern business strategy, leveraging historical data and advanced algorithms to forecast future sales accurately. This predictive capability is essential for businesses across various industries to optimize inventory levels, allocate resources effectively, and formulate proactive marketing strategies.

Machine learning algorithms such as linear regression, decision trees, and neural networks play pivotal roles in this process. They analyze past sales data, identify underlying patterns, and learn from historical trends to make predictions. However, before applying these algorithms, thorough data pre-processing is crucial. This involves cleaning the data, handling missing values, normalizing variables, and selecting relevant features that contribute most to the predictive accuracy.

Challenges in sales prediction include managing seasonality effects, responding to market dynamics, and detecting outliers that may skew predictions. Addressing these challenges requires robust pre-processing techniques and selecting algorithms that can handle complex and noisy data environments effectively.

Evaluation metrics such as mean squared error (MSE) and root mean squared error (RMSE) are commonly used to assess the performance of predictive models. By evaluating these metrics, businesses can gauge the accuracy of their sales forecasts and refine their models accordingly.

Overall, sales prediction using machine learning empowers businesses to make data-driven decisions, enhance operational efficiency, and gain a competitive edge in the market by anticipating future demand and trends more accurately.

1.1. SALES PREDICTION USING MACHINE LEARNING

Sales prediction using machine learning is a process of using historical sales data to train a model that can forecast future sales. The abstract for sales prediction in machine learning describes the application of different machine learning algorithms for predicting sales. Collecting and preprocessing the data, choosing pertinent features, and training the model are all steps in the procedure. The model is evaluated using different metrics to determine its accuracy and effectiveness. The goal of the study is to provide information on the best machine learning algorithms for predicting

sales and their application to various enterprises. The findings of this study can aid companies in making better business decisions, optimizing inventory management, and increasing the accuracy of their sales forecasts. This research paper evaluates the effectiveness of machine learning models for sales prediction in various industries. The study analyzes historical sales data and compares the performance of different machine learning algorithms such as linear regression, random forest, K-Neighbors Regressor in predicting future sales. Results show that machine learning techniques can accurately predict sales with high accuracy, outperforming traditional statistical methods. Companies can leverage this approach to optimize their resources and increase profitability.

II. LITERATURE SURVEY

1.Sales Prediction of Market using Machine Learning
Author - JiantingNing, Xinyi Huang
Year-2022
Connected devices, sensors, and mobile apps make the retail sector a relevant testbed for big data tools and applications. We investigate how big data is, and can be used in retail operations. Based on our state-of-the-art literature review, we identify four themes for big data applications in retail logistics: availability, assortment, pricing, and layout planning. Our semi-structured interviews with retailers and academics suggest that historical sales data and loyalty schemes can be used to obtain customer insights for operational planning, but granular sales data can also benefit availability and assortment decisions. External data such as competitors' prices and weather conditions can be used for demand forecasting and pricing. However, the path to exploiting big data is not a bed of roses. Challenges include shortages of people with the right set of skills, the lack of support from suppliers, issues in IT integration, managerial concerns including information sharing and process integration, and physical capability of the supply chain to respond to real-time changes captured by big data.

2. Intelligent Sales Prediction Using Machine Learning Techniques
Author - Mrs. T. Ratnamala 2 Syed Ameer Sohail
Year-2023
Intelligent Decision Analytical System requires integration of decision analysis and predictions. Most of the business organizations heavily depend on a knowledge base and demand prediction of sales trends. The accuracy in sales forecast provides a big impact in business. Data mining techniques are very effective tools in extracting hidden knowledge from an enormous dataset to enhance accuracy and efficiency of forecasting. The detailed study and analysis of comprehensible predictive models to improve future sales predictions are carried out in this research. Traditional forecast systems are difficult to deal with the big data and accuracy of sales forecasting. These issues could be overcome by using various data mining techniques.

3. Sales Prediction Using Machine Learning
Author - Narayan Hampiholi
Year-2023
The fundamental concepts of sellers and customers are supply, demand. It is quite critical to predict demand accurately sales prediction is predicting the sales for Bigmart companies through which the performance is predicted by changing the business model. In this paper, a new approach for demand prediction for Big Mart companies is proposed. The business model used by the Big Mart companies includes many outlets that sell the product which are same throughout the country at the same time where the company operates a market place model. This actually helps in business to predict realistic sales which helps in increasing the business .In this study linear regression, gradient boosting and random forest are used to predict sales . This method is carried out on data from various companies where those are processed according the algorithms in machine learning and final data is extracted which helps in sales prediction.

III. EXISTING SYSTEM

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 analysing the performance, we have trying to suggest the suitable predictive algorithm to our problem statement.

DISADVANTAGES

➢ 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.

➢ 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.

IV. PROPOSED SYSTEM

Most of the business organizations heavily depend on a knowledge base and demand prediction of sales trends. Sales forecasting is the process of estimating future sales. Accurate sales forecasts enable companies to make informed business decisions and predict short-term and long-term performance. Companies can base their forecasts on past sales data, industry wide comparisons, and economic trends. Sales forecasts help sales teams achieve their goals by identifying early warning signals in their sales pipeline and course correct before it's too late. The goal is to improve the accuracy from the existing project. So that the sales and profit could be increased for the companies. Choosing an efficient algorithm from comparing different algorithms to improve the prediction further more.

ADVANTAGES

➢ Intelligent Decision Analytical System requires integration of decision analysis and predictions. Most of the business organizations heavily depend on a knowledge base and demand prediction of sales trends.

➢ The accuracy in sales forecast provides a big impact in business. Data mining techniques are very effective tools in extracting hidden knowledge from an enormous dataset to enhance accuracy and efficiency of forecasting.

➢ The detailed study and analysis of comprehensible predictive models to improve future sales predictions are carried out in this research. Traditional forecast systems are difficult to deal with the big data and accuracy of sales forecasting.

V. SYSTEM IMPLEMENTATION

MODULES DESCRIPTION
BIG DATA

Big data is an "imprecise description of a rich and complicated set of characteristics, practices, techniques, ethical issues, and outcomes all associated with data". A more technical and complementary definition is "datasets that could not be perceived, acquired, managed, and processed by traditional [information technology] and software/hard-ware tools within a tolerable time" For companies, big data is a cornucopia of digitalized content about consumers' cognitions, emotions, behaviours, and reactions critical to the ongoing data-driven industrial revolution. 'Big data' has different definitions that focus on size, range, and speed. While some incorporate a dynamic definition as beyond the ability of typical software, some use a static number as 100 terabytes or have a growth rate that is higher than 60% annually.

2. SALES PREDICTION FOR BIG MART

A retailer company wants a model that can predict accurate sales so that it can keep track of customer's future demand and update in advance the sale inventory. In this work, we propose a technique to optimize the parameters and select the best tuning hyper parameters, further ensemble with Xgboost techniques for forecasting the future sales of a retailer company such as Big Mart and we found our model produces the better result.

3. Intelligent Decision Analytical

Intelligent Decision Analytical System requires integration of decision analysis and predictions. Most of the business organizations heavily depend on a knowledge base and demand prediction of sales trends. The accuracy in sales forecast provides a big impact in business. Data mining techniques are very effective tools in extracting hidden knowledge from an enormous dataset to enhance accuracy and efficiency of forecasting. The detailed study and analysis of comprehensible predictive models to improve future sales predictions are carried out in this research. Traditional forecast systems are difficult to deal with the big data and accuracy of sales forecasting.
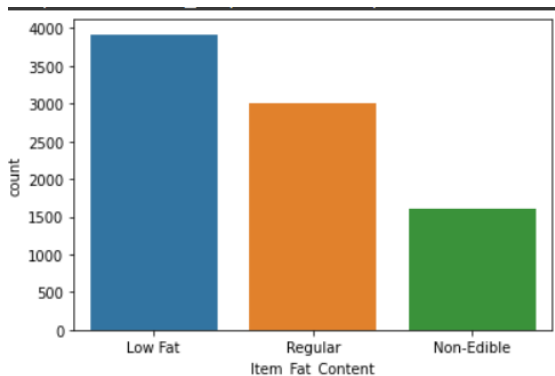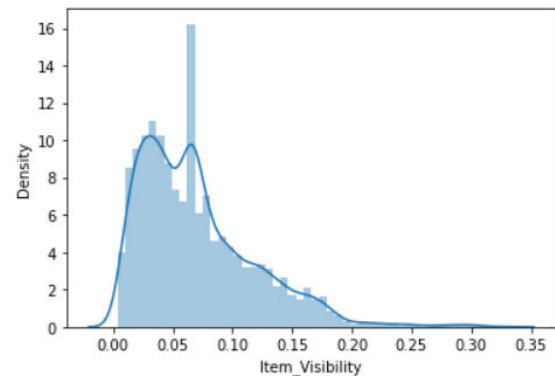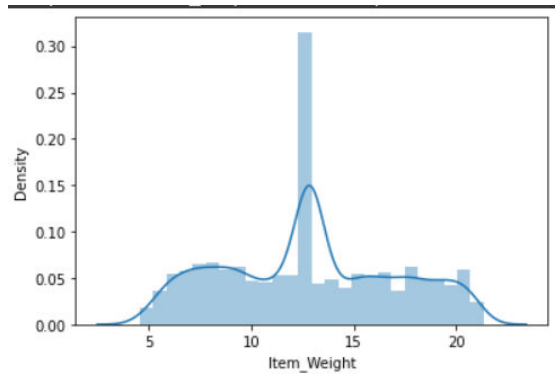
4. Data Pre-processing

➢ Pre-processing the dataset and formatting it into a form that is relevant to us

➢ Next is plotting our time series and coming the conclusion that we have a clear decreasing trend component and a seasonal component with an interval of approximately 12 months

➢ Then gaining insights from our dataset by aggregating it to monthly level to form. Our time series and plotting several graphs like rolling mean and rolling standard deviation and other datavisualization visualization analyze our dataset.

➢ Next we decompose time series into its constituent component which are trend , seasonality and residue.

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013	1987

	Item_Weight	Item_Visibility	Item_MRP	Outlet_Establishment_Year	Item_Outlet_Sales
count	7060.000000	8523.000000	8523.000000	8523.000000	8523.000000
mean	12.857645	0.066132	140.992782	1997.831867	2181.288914
std	4.643456	0.051598	62.275067	8.371760	1706.499616
min	4.555000	0.000000	31.290000	1985.000000	33.290000
25%	8.773750	0.026989	93.826500	1987.000000	834.247400
50%	12.600000	0.053931	143.012800	1999.000000	1794.331000
75%	16.850000	0.094585	185.643700	2004.000000	3101.296400
max	21.350000	0.328391	266.888400	2009.000000	13086.964800

VI. CONCLUSION

In conclusion, developing a sales prediction system using machine learning involves meticulous planning, from data collection and preprocessing to model selection and deployment. Technical feasibility is supported by advanced algorithms and scalable infrastructure, ensuring accurate predictions. Economic feasibility is justified by potential cost savings and revenue gains through optimized inventory management and targeted marketing strategies. Social feasibility requires transparent communication and addressing privacy concerns. Rigorous testing ensures reliability and performance. Ultimately, such a system enhances decision-making, improves customer satisfaction, and drives business growth by leveraging predictive analytics to anticipate market trends and meet customer demands effectively in dynamic business environments.

REFERENCES

[1] Trnka., "Market Basket Analysis with Data Mining Methods", International Conference on Networking and Information Technology (ICNIT) ,2010.

[2] W Yanthy, T. Sekiya, K. Yamaguchi , "Mining Interesting Rules by association and Classification Algorithms", FCST 09.

[3] Chiu, K.S.Y., Luk, R.W.P, Chan, K.C.C., and Chung, K.F.L, "Market-basket Analysis with Principal Component Analysis: An Exploration", IEEE International Conference on Systems, Man and Cybernetics, Vol.3, 2002.

[4] Cunningham , S.J. and Frank, E., "Market Basket Analysis of Library Circulation Data", International Conference on Neural Information Processing, Vol.2. 1999.

[5] Vo,B.and Le,B.,"Mining traditional association rules using frequent itemsets lattice",International Conference on Computers & Industrial Engineering. 2009.

[6] Rastogi, R.. and kyuseok Shim, "Mining optimised association rules with Categorical and numerical attributes", IEEE transactions on Knowledge and Data Engineering, vol.14, 2002.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462   6381 907 438   ijircce@gmail.com

Scan to save the contact details