WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

E-COMMERCE PRODUCT SALES PREDICTION USING MACHINE LEARNING

Asik Akram .S ,Mr. Arun M,MCA., (P. hD)[2]

[1]UG Student, Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Coimbatore.
[2] Assistant Professor, Department of Computer Science ,Sri Krishna Adithya College of Arts and Science, Coimbatore

ABSTRACT :

The E-commerce Product Sales Dataset is a comprehensive collection of sales data for a wide range of products available on the E-commerce e-commerce platform. This dataset provides invaluable insights into customer behaviour, product performance, and market trends, making it an essential resource for data analysis, market research, and business strategy development. This dataset is indispensable for market research, allowing businesses to discern market trends, consumer preferences, and competitive landscapes. It supports competitive analysis by enabling sellers to benchmark their performance against others and refine their strategies. Customer profiling and targeted marketing are empowered by customer-specific data, while product development can be guided by customer feedback and demand analysis. Furthermore, this dataset can be applied for price optimization, sales forecasting, sentiment analysis of product reviews, marketing campaign strategies, and evaluating the influence of external factors on sales.

Keywords:Text-to-SQL, Tree-based Architecture, Large Language Models, Database Schema Understanding, Natural Language Processing, Query Generation, Schema-aware Processing.

Introduction :

An E-commerce product sales dataset is a valuable and comprehensive collection of data that provides insights into the performance and trends of products sold on the E-commerce e-commerce platform. This dataset encompasses a wide range of information, including product details, pricing, customer reviews, sales volume, and more. Analysing this dataset can be immensely beneficial for businesses, market researchers, and data scientists seeking to understand consumer preferences, optimize pricing strategies, and make informed decisions in the competitive world of e-commerce.

With the ever-expanding reach and influence of E-commerce in the global marketplace, access to such data has become increasingly essential for businesses looking to thrive in the online retail space. This dataset can shed light on

consumer behaviour, product popularity, and seasonal fluctuations, enabling companies to fine-tune their marketing efforts, identify niche markets, and tailortheir product offerings to meet customer demands effectively.

In this era of data-driven decision-making, an E-commerce product serves as a goldmine of information that can be harnessed to enhance market competitiveness, drive sales growth, and ultimately provide customers with theproducts they desire.

Problem Definition :

Existing System

Moving Averages and Seasonal Trends: Many businesses use moving averages or seasonal indices to predict demand, assuming that sales patterns will be similar to those in the past.

Linear Regression: Simple regression models are often used to understand the relationship between historical sales and time, though they lack the complexity needed to account for multiple factors affecting demand.

Problem Statement

Accurately predicting product sales is a significant challenge for e-commerce businesses. Existing sales forecasting methods, such as simple statistical models (e.g., moving averages or linear regression) and spreadsheet-based tracking, are inadequate in capturing the complex dynamics of consumer demand. These methods are typically based on historical sales data alone and fail to account for multiple influential factors such as seasonal trends, promotions, customer demographics, competitor actions, and market fluctuations.

The current approaches often result in inaccuracies, leading to inefficient inventory management, stockouts, overstocking, and suboptimal pricing strategies. These inefficiencies impact both operational costs and customer satisfaction. For instance, overstocking leads to higher holding costs and potential product markdowns, while stockouts can result in missed sales opportunities and customer frustration.

Furthermore, traditional forecasting systems struggle to scale with growing e-commerce platforms that offer a large variety of products with different sales patterns and demand drivers. As customer preferences change rapidly and market conditions fluctuate, these outdated methods are unable to adapt quickly enough to provide accurate predictions.

The problem, therefore, is to develop an advanced, machine learning-based solution that can predict future product sales with high accuracy by incorporating a wide range of factors, including historical sales data, customer behavior, external variables (such as economic indicators and competitor pricing), and product-specific characteristics. This predictive model should be scalable, adaptable, and capable of providing insights that will optimize inventory management, improve revenue, and enhance customer satisfaction.

By addressing this problem, e-commerce businesses can significantly improve their forecasting accuracy, leading to more efficient operations, reduced costs, and better alignment with customer demand.

Proposed System :

The proposed system for E-Commerce Product Sales Prediction aims to forecast future sales using machine learning techniques, helping businesses optimize inventory management, marketing, and pricing strategies. It starts with collecting comprehensive data such as historical sales, product details, customer behavior, external factors like seasonality, and marketing campaigns. This data undergoes preprocessing, including cleaning, feature engineering, and scaling, to make it suitable for machine learning models.

Various machine learning algorithms can be applied depending on the data and prediction goals, such as linear regression, decision trees, gradient boosting, LSTM networks for time-series forecasting, or ARIMA for trend and seasonality analysis. The model is trained on historical data, evaluated using metrics like RMSE and MAE, and then fine-tuned through techniques like cross-validation and hyperparameter optimization.

Literature Review :

prediction systems for e-commerce reveals a growing body of research and practical applications focusing on forecasting demand, optimizing inventory, and enhancing marketing strategies. The majority of studies have explored various machine learning models, including traditional regression techniques, time series models, and more advanced ensemble methods and deep learning algorithms.

Early works predominantly focused on linear regression and ARIMA models, which are effective for predicting sales based on historical data but often struggle to account for complex patterns such as non-linear relationships or sudden shifts in trends. As e-commerce data has grown in complexity, more sophisticated models like Random Forests and Gradient Boosting Machines (GBM), particularly XGBoost, have gained popularity for their ability to handle interactions between variables and capture non-linear relationships.

Deep learning approaches, especially Long Short-Term Memory (LSTM) networks, have been extensively explored for time-series forecasting due to their ability to model sequential dependencies over time, making them particularly effective for predicting product sales with seasonality and trends. Research has also investigated neural networks for capturing intricate, high-dimensional data patterns, including customer behavior and product attributes.

Studies often emphasize the importance of feature engineering—creating meaningful features like promotional activity, customer demographics, and website traffic—which significantly enhance model performance. Additionally, some research has integrated external data sources (e.g., weather, economic indicators) to improve accuracy, recognizing that sales are often influenced by factors beyond historical data.

Another common theme in the literature is the challenge of model interpretability. While more complex models like deep learning provide high accuracy, they are often criticized for being "black boxes." Many studies highlight the need for transparent models that stakeholders can understand and trust, particularly in business contexts where decisions must be justified.

In terms of deployment, research has shown the effectiveness of integrating machine learning models into real-time decision-making systems, linking sales predictions directly to inventory and marketing strategies. However, challenges remain in model drift, where the accuracy of predictions declines over time due to changing consumer behavior and market dynamics, highlighting the need for continuous model monitoring and retraining.

In conclusion, while machine learning has proven highly effective in sales prediction for e-commerce, the literature emphasizes the importance of selecting the right model, continuously updating the system, and addressing challenges related to interpretability and feature engineering to ensure long-term success in real-world applications.

5.Methodology :

The methodology for an E-Commerce Product Sales Prediction System using machine learning involves several key steps. First, data is collected from various sources such as historical sales, product details, customer behavior, and external factors like weather and promotions. This data is then cleaned, processed, and transformed through feature engineering and normalization to make it suitable for model training.

Next, machine learning models such as linear regression, decision trees, XGBoost, and LSTM networks are selected based on the nature of the data. The models are trained using historical data, with the dataset split into training and testing sets to ensure robust performance. Model evaluation is done using metrics like RMSE, MAE, and R-squared, and hyperparameters are tuned for optimal accuracy.

In conclusion, In conclusion, an E-Commerce Product Sales Prediction System using machine learning offers significant value by providing accurate forecasts that can drive better business decisions. By leveraging historical sales data, product details, customer behavior, and external factors, businesses can optimize inventory management, marketing strategies, and pricing. Machine learning models such as decision trees, XGBoost, and LSTM networks can effectively capture complex patterns and trends in sales data. The system's deployment allows for real-time predictions, enabling e-commerce platforms to proactively adjust to demand fluctuations. Continuous monitoring and model updates ensure sustained accuracy and alignment with changing market conditions, ultimately leading to improved operational efficiency and customer satisfaction.

REFERENCES :

[1] Dong, L., & Lapata, M., "Language to logical form with neural attention," 2016.
[2] Iyer, S., Gardent, C., & Clark, P., "Learning to map questions to SQL queries," 2017.
[3] Krishna, P., & Kollar, I., "Hierarchical modeling of SQL queries with large pre-trained language models," 2021.
[4] Yu, T., & Yao, H., "SQLNet: Generating structured queries from natural language using reinforcement learning," 2018.
[5] Wei, J., Schuster, M., & Pruksachatkun, Y., "Chain-of-thought prompting elicits reasoning in large language models," 2022.
[6] Zhang, Y., & Sun, Y., "T5 for SQL: A pretrained transformer-based model for Text-to-SQL tasks," 2023.
[7] Zhang, Y., & Wang, Q., "Productionizing neural models for Text-to-SQL in real-world applications," 2022.
[8] Gao, J., & Chang, M., "Efficient and scalable database query generation using neural networks in production environments," 2020.
[9] Herzig, J., & Shieber, S., "Database query generation using schema-aware neural architectures," 2019.
[10] Berg, S., & Finkel, H., "Neural approaches for schema understanding in text-to-SQL models," 2021.