Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Prediction and Forecasting of Financial Sales Analysis

Payal Praful Muni*, Shubhangi Kamble**
*(Computer Science, CKT ACS College, New Panvel (Autonomous))
** (Computer Science, CKT ACS College, New Panvel (Autonomous))
Email: shubhangi.ak4@gmail.com

----------**********----------

Abstract:

Prediction and forecasting of financial sales analysis is a critical aspect of modern business strategy, enabling companies to make data-driven decisions and optimize resource allocation. This study explores various predictive models and techniques, such as time series analysis, machine learning algorithms, and statistical methods, to forecast future sales trends based on historical financial data. The aim is to enhance the accuracy of sales predictions, which can guide inventory management, budgeting, marketing strategies, and investment decisions. Our study showcases how historical demand data can be harnessed to predict future demand and how these forecasts impact the supply chain. By employing the Box–Jenkins time series procedure, we developed several autoregressive integrated moving average (ARIMA) models using the historical demand information. The most suitable model was then selected based on its adequacy. By leveraging data from past sales performance, economic indicators, and market conditions, this research highlights how businesses can anticipate changes in demand and adjust their strategies accordingly to stay competitive and financially sustainable.

Keywords — **SQL, Power BI, Power Query sales prediction autoregressive integrated moving average (ARIMA).**

----------**********----------

I. INTRODUCTION

In order to enhance the logistics service experience of customers in the e-commerce industry chain, supply chain collaboration [1] requires that commodities are stocked in advance in local warehouses of various markets around the world, which can effectively reduce logistics time. However, for cross-border e-commerce enterprises, the production and sales areas of e-commerce products are globalized, which takes them longer to make preparations from the procurement of commodities, transportation, to customs quality inspection, etc. Therefore, algorithms and technologies of big data analysis are widely applied to predict sales of e-commerce commodities, which provide the data basis for the supply chain management and will provide key technical support for the global supply chain scheme of cross-border e-commerce enterprises.

In recent years, advancements in data analytics and computational technologies have transformed the way businesses approach financial sales analysis. Traditional methods of sales forecasting, which relied heavily on basic statistical techniques, are being replaced or supplemented by sophisticated predictive models. These models incorporate large volumes of structured and unstructured data, including historical sales records, market trends, customer demographics, and external economic factors. Techniques such as machine learning, time series analysis, and regression modelling have proven to be highly effective in improving the accuracy and reliability of sales forecasts.

The integration of artificial intelligence (AI) and machine learning (ML) into financial sales analysis offers significant advantages. These technologies enable the identification of hidden patterns and relationships within datasets that would be difficult to uncover using conventional approaches. Additionally, they allow for real-time forecasting and scenario analysis, providing businesses with the agility to adapt to rapidly

changing market conditions. For instance, predictive algorithms can forecast seasonal sales fluctuations, evaluate the impact of promotional campaigns, and anticipate shifts in consumer demand due to economic changes or global events. Despite these advancements, several challenges persist in the field of financial sales forecasting. Data quality and availability remain key concerns, as inaccurate or incomplete data can lead to flawed predictions. Moreover, the interpretability of complex machine learning models is often limited, making it difficult for decision-makers to fully trust or understand the outputs. Addressing these challenges requires a combination of rigorous data preprocessing, model validation, and the use of explainable AI techniques.[1]

This research paper aims to explore the methodologies and technologies employed in the prediction and forecasting of financial sales. By analysing the effectiveness of various predictive models and their applications across industries, the study seeks to provide insights into best practices and emerging trends in the field. Furthermore, the paper highlights the potential of integrating advanced analytics with domain expertise to enhance decision-making and drive business success.

A) SQL Database for Data Collection and Integration and SQL Server Analysis Services (SSAS) for Data Modelling

SQL databases were employed to collect, store, and integrate data from multiple sources. By using SQL queries, structured data from sales records, market trends, and customer demographics were consolidated into a centralized repository, ensuring data accuracy and consistency for analysis.
ARIMA is an innovative statistical model for analyses of statistical data and helps forecast a series of data within a specific time period for taking appropriate decisions for the reinvention of sales management strategies in the organization. [2]

SQL Server Analysis Services (SSAS) provided a robust platform for building multidimensional and tabular data models. These models enabled efficient processing and aggregation of large datasets, laying the foundation for advanced analytics and enabling better insight into sales patterns and trends.[3]

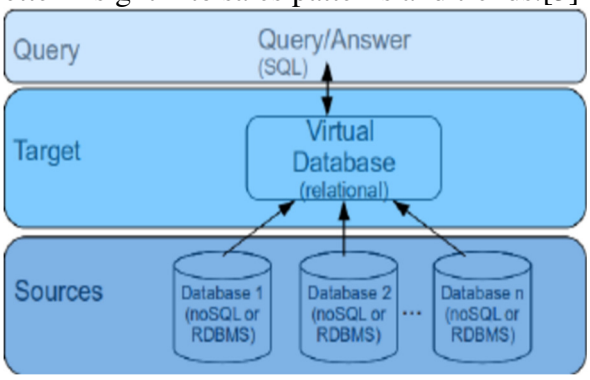

Fig 1. SQL Database for Data Collection and Integration

B) Power Query Editor for Data Analysis and Calculation

Power Query Editor was utilized to perform data cleansing, transformation, and calculations. Through its user-friendly interface, complex data preparation tasks such as merging datasets, handling missing values, and creating calculated columns were streamlined, ensuring high-quality inputs for predictive models.

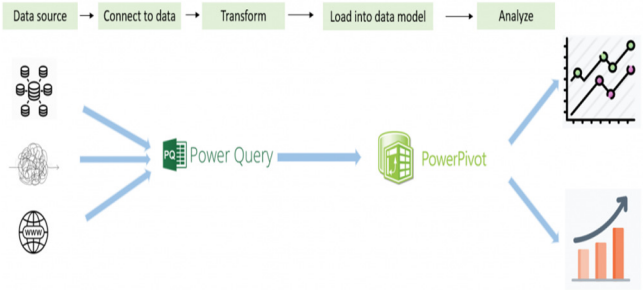

Fig. 2 Power Query Editor for Data Analysis and Calculation.

Apart from this, data pre-processing is essential for predictive models in the

organization. Evaluation of different data is essential for the maintenance of the sales growth rate in the organization. Python is used to calculate data to focus on different external factors, such as pricing, competitor's behaviour and promotion for enhancement of sales rate in the organization. Python is essential for Random Forest, linear regression, and design tree algorithms for deradicalization of complexity in the sales management rate.

C) Power BI for Data Visualization

Fig. 3 Power BI for Data Visualization for sales analysis

Power BI was used to create interactive dashboards and visualizations, allowing stakeholders to explore and interpret data effectively. Through dynamic charts, graphs, and KPIs, complex sales trends were represented visually, aiding in intuitive understanding and actionable decision-making [4].

II. MATERIALS AND METHODS

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

The methodology section guides the research toward reliable conclusions by adopting a positivist philosophy and leveraging Machine Learning (ML) techniques to improve financial sales forecasting. Secondary data from peer-reviewed journals via Google Scholar was used for analysis, focusing on quantitative data and the application of ML models to identify patterns and predict future sales trends. Descriptive statistics further enhance understanding of ML's role in sales prediction.

Power BI was employed for data visualization, enabling interactive dashboards to track real-time sales trends.[5] The ARIMA model was used for time-series forecasting, capturing data dependencies to predict future sales. Additionally, the Decision Tree Regressor was applied to model relationships between sales and key features, offering interpretable insights into sales performance.

A non-probability sampling method streamlined data collection, with Power Query used for processing and forecasting. The research considers inclusion/exclusion criteria to ensure authenticity, with qualitative data analysis to explore ML applications in sales prediction. The data spans nearly a decade, supporting accurate forecasting and improved business decision-making.

A) Power BI for data visualisation

Power BI was utilized as the primary tool for data visualization, playing a critical role in transforming raw financial sales data into meaningful insights. Power BI's interactive dashboards and reporting capabilities enabled the creation of dynamic visualizations that allowed for easy identification of trends, patterns, and anomalies in the sales data. By integrating diverse data sources, such as historical sales figures and external market indicators, Power BI provided a comprehensive view of the financial landscape.[6]

The tool's user-friendly interface allowed stakeholders to explore data intuitively, offering a flexible platform for visualizing key performance indicators (KPIs) and sales forecasts. This was crucial for enhancing decision-making by providing real-time insights into sales performance and enabling users to drill down into specific data points for a deeper understanding of underlying factors influencing sales trends. Moreover, Power BI's

ability to handle large volumes of data and present it in an interactive format supported a more efficient and accessible analysis process, which was essential for accurate sales forecasting. By leveraging Power BI, this study was able to present a clear, data-driven narrative, contributing to more informed and strategic business decisions in the context of financial sales analysis[7]

B) ARIMA (Auto Regressive Integrated Moving Average):

This model is a widely used statistical method for time series forecasting that combines three key components[8]The graph displays historical sales data with evident fluctuations and periodic spikes, which could suggest seasonal trends or specific events that affect sales volumes. Superimposed on this, it shows forecasted sales, which continue the pattern observed in the historical data, indicating expectations of similar future trends. The results seem to suggest that the sales pattern is expected to continue with similar variability and periodic spikes as seen in the past.[9]

C) Decision Tree Regressor

The model has provided in the script is a Decision Tree Classifier, which is used for binary classification tasks. In Script, the Decision Tree Classifier is being trained to predict whether sales will be above a certain threshold, making it a binary classification problem. This is indicated by the creation of a binary target variable named High Sales, which is set to 1 if sales exceed a 75th percentile threshold (considered high sales) and 0 otherwise (considered low sales).[10]

Decision Tree model is trained on historical sales data, which is then used to forecast sales for the following year. The graph presents historical sales data (blue dots) alongside the next year's sales forecasts (green dots) produced by a Decision Tree model. The x-axis indicates time, spanning from 2021 into 2024, and the y-axis represents sales values up to 1 million. The model's forecasts, represented by green dots, start after the last historical data point, indicating an attempt to project future sales based on past trends. The results demonstrate the model's capacity to anticipate future performance, which can be crucial for business planning and decision-making.[11]

III. RESULTS AND DISCUSSION

3.1 Data Visualization Results

For data visualization and reporting, Power BI was utilized to transform raw data into interactive visualizations. This tool enabled the creation of dashboards and reports that provided real-time insights into financial sales trends.

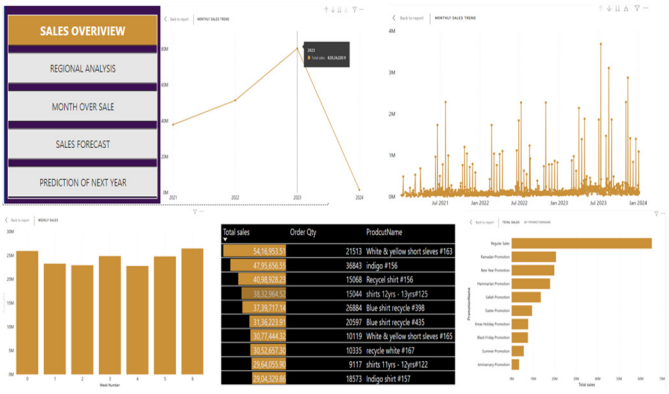

Fig. 4 Core data representation using Power BI

Sales Overview (Top Left): This section acts as a navigation or summary panel. It lists key aspects of the data: Regional Analysis, Month Over Sale, Sales Forecast, and Prediction of Next Year. Each item likely links to a more detailed view within the dashboard (though this isn't shown).

Yearly Sales Trend (Top Middle): A line graph showing sales figures over four years (likely 2021-2024). There's a sharp increase in sales in 2023, followed by a significant drop in 2024.

Daily Sales Trend (Top Right): A detailed bar chart showing daily sales fluctuations over roughly a two-year period. This gives a granular view of sales

performance over time, highlighting peaks and valleys.

Monthly Sales Performance (Bottom Left): A bar chart presenting monthly sales for a period (likely one year, 7 months shown). This provides a summarized view of sales performance across months.

Sales by Product (Bottom Middle): A table lists the top-selling products. It shows the total sales value for each product and the corresponding order quantity, sorted by sales volume in descending order.

3.2 ARIMA result

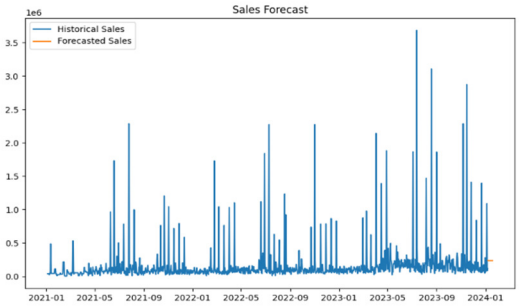

Fig. 5 ARIMA representation

The graph shows a comparison of historical sales and forecasted sales over time from January 2021 to early 2024. Sales values are represented on the Y-axis, which goes up to 3.5 million, and time is on the X-axis. Historical data is shown alongside the forecasted data, likely to project future trends based on past patterns.

3.3 Decision tree regressor results

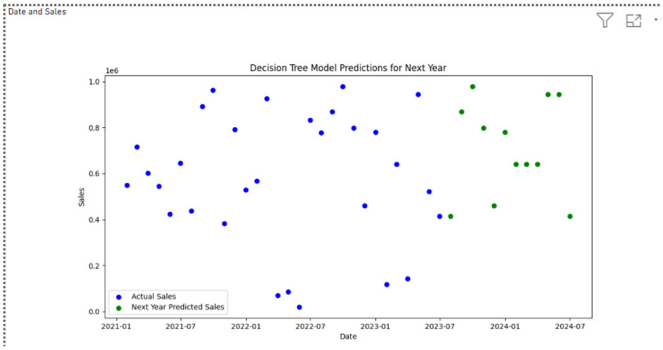

Fig. 6 Decision Tree regressor representation

The scatter plot shows the actual and predicted sales from 2021 to 2024. Blue dots are actual sales, and green dots are model predictions. In some years, the Green dots (predictions) are close to the blue dots (actual sales), suggesting accurate predictions for those periods. In other years, there's a noticeable gap between the predictions and actual sales, indicating less accurate predictions. The model has varying accuracy, with some predictions close to actual sales and others not. The sales values go up to 1 million. The plot indicates the model's predictive performance over time.

IV. CONCLUSIONS AND RECOMMENDATIONS

This study successfully demonstrated the application of advanced analytics tools and methodologies for predicting and forecasting financial sales. The integration of SQL databases, SSAS, Power Query Editor, and Power BI enhanced the accuracy, efficiency, and usability of the sales forecasting process. Key conclusions from the research are as follows:

Data integration and quality are critical consolidating data from diverse sources using SQL ensured consistency and reliability, underscoring the importance of robust data integration frameworks in financial analysis.[10]

Advanced data models enhance forecasting leveraging multidimensional models through SSAS significantly improved the precision of sales predictions, enabling businesses to address complex trends effectively.

User-Friendly Analytics Drive Decision-Making Power BI's interactive visualizations allowed stakeholders to gain actionable insights, highlighting the value of intuitive tools for decision-making.

Security Measures Ensure Trust in Analytics The implementation of role-based access controls in Power BI safeguarded sensitive data, emphasizing

the need for strong security protocols in financial analytics[11].

Recommendations:

Incorporate Real-Time Data Streams Future implementations should include real-time data to capture immediate market changes and enhance forecasting accuracy.[12]

Adopt Machine Learning Models Incorporating machine learning algorithms could further improve prediction precision by identifying complex, non-linear patterns in sales data.

Expand External Data Sources Integrating external factors such as economic indicators, weather data, and competitor activity could provide deeper insights into sales trends.

Train Stakeholders on Analytics Tools Providing training on tools like Power BI can ensure that all users can effectively interact with dashboards and make data-driven decisions.

ACKNOWLEDGMENT

The heading of the Acknowledgment section and the References section must not be numbered.

Causal Productions wishes to acknowledge Michael Shell and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template. To see the list of contributors, please refer to the top of file IEEETran.cls in the IEEE LaTeX distribution.

REFERENCES

[1] S. Ji, X. Wang, W. Zhao and D. Guo, "An Application of a Three-Stage XGBoost-Based Model to Sales Forecasting of a Cross-Border E-Commerce Enterprise", Mathematical Problems in Engineering, vol. 2019, pp. 1-15, 2019.

[2] S. Das, "Data Visualization and Stock Market and Prediction ", International Research Journal of Engineering and Technology (IRJET), vol. 06, issue 09, 2019.

[3] J. Heitz & W. Breymann, "Scalable architecture for Big Data financial analytics: user-defined functions vs. SQL", Journal of BigData, vol. 6, article number 46, 2019.

[4] S. Murugan, "Retail Store Sales Analysis: Unveiling Insights Through Power BI Business Analytics", International Conference on Innovative Computing & Communication (ICICC) 2024.

[5] M. Raje, P. Jain, V. Chole, "Sales Analysis And Prediction Dashboard Using Power BI", International Research Journal of Modernization in Engineering Technology and Science, Vol. 3, Issue 06, 2021.

[6] M. Belghith, S. Bouajaja, "Developing a Sales Dashboard with Power BI – A Case Study in a Pharmaceutical Company", Decision Making Advances VL - 2 DO - 10.31181/dma21202438.

[7] S. Shivankar, S. Mehetar, N. Darade, S. Bhimanpalli, D. Dafale, "Global Superstores Sales Prediction and Data Visualization Using Power BI", International Journal of Research in Engineering, Science and Management Vol. 6, Issue 4, 2023.

[8] H. Al-Chalabi, Y. K. Al-Douri, J. Lundberg, "Time Series Forecasting using ARIMA Model", The Twelfth International Conference on Advanced Engineering, Computing and Applications in Sciences, VL- ISBN: 978-1-61208-677-4.

[9] A. Fattah, L. Ezzine, Z. Aman, H. E. Moussami, A. Lachhab, "Forecasting of demand using ARIMA model", International Journal of Engineering Business Management, Vol. 10, 1–9, 2018.

[10] S. Ravi Teja Reddy, P. Malathi,,"Design and Implementation of Sales Prediction Model Using Decision Tree Regressor over Linear Regression Towards Increase in Accuracy of Prediction**",** Advances in Parallel Computing, Vol. 41: Advances in Parallel Computing Algorithms, Tools and Paradigms.

[11] S. Thomassey, A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees", Decision Support Systems, Vol. 42, Issue 1, 2006, Pages 408-421.

