**Interactive Dashboard for Data Visualization and Prediction**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Anant Kaulage | Soureesh Dalal | Rahul Nagpure | Deepak Gavit | Ashutosh Gore |
| *Department of Computer Engineering* | *Department of Computer Engineering* | *Department of Computer Engineering* | *Department of Computer Engineering* | *Department of Computer Engineering* |
| *Vishwakarma Institute of Technology* | *Vishwakarma Institute of Technology* | *Vishwakarma Institute of Technology* | *Vishwakarma Institute of Technology* | *Vishwakarma Institute of Technology* |
| *Pune, Maharashtra, India* | *Pune, Maharashtra, India* | *Pune, Maharashtra, India* | *Pune, Maharashtra, India* | *Pune, Maharashtra, India* |
| [anant.kaulage@vit.edu](mailto:anant.kaulage@vit.edu) | [soureesh.dalal20@vit.edu](mailto:soureesh.dalal20@vit.edu) | [rahul.nagpure20@vit.edu](mailto:rahul.nagpure20@vit.edu) | [deepak.gavit20@vit.edu](mailto:deepak.gavit20@vit.edu) | [ashutosh.gore20@vit.edu](mailto:ashutosh.gore20@vit.edu) |

***Abstract:* This paper presents the development and implementation of an interactive dashboard for analyzing and predicting sales data from a Superstore dataset. Utilizing Streamlit for the web interface and incorporating various data visualization libraries such as Plotly and Matplotlib, the dashboard enables users to filter, visualize, and analyze sales data efficiently. Additionally, a linear regression model is employed to predict future sales trends. The dashboard's design, functionalities, and predictive capabilities are discussed in detail, highlighting its utility for business intelligence and decision-making.**

***Keywords —Superstore, Sales Analysis, Interactive Dashboard, Streamlit, Data Visualization, Linear Regression, Time Series Analysis.***

1. INTRODUCTION

In today’s data-driven business environment, the ability to analyze and visualize large datasets is crucial for making informed decisions. The retail industry generates extensive data from sales transactions, customer interactions, and inventory management. Harnessing this data effectively can lead to significant operational improvements and strategic advantages. This paper introduces an interactive dashboard designed for the comprehensive analysis and prediction of sales data from a Superstore dataset, leveraging Streamlit, a powerful Python library for creating web applications.

The dashboard provides a user-friendly interface that allows users to upload datasets, apply various filters, and visualize sales data through interactive charts and graphs. Key functionalities include filtering by region, state, city, and date range, enabling users to drill down into specific data subsets. By utilizing Plotly for data visualization, the dashboard offers an engaging way to explore sales performance across different dimensions, such as categories, regions, and time periods. This interactivity helps users gain deeper insights and make data-driven decisions more efficiently.

Beyond visualization, the dashboard incorporates a linear regression model for predicting future sales trends. Predictive analytics plays a vital role in business intelligence, offering foresight into future performance. The model is trained on historical sales data and provides sales forecasts up to 2025, aiding in inventory management and strategic planning. By combining data visualization and predictive analytics, the dashboard serves as a comprehensive tool for business analysts and decision-makers, demonstrating the practical application of data science in the retail industry.

1. LITERATURE REVIEW

The development and utilization of interactive dashboards for data visualization has seen significant advancements, driven by the need for effective data analysis and decision-making tools. Recent research, such as DashBot's application of deep reinforcement learning, highlights the complexity and demand for creating effective dashboards, often requiring extensive data analysis expertise and familiarity with professional tools like Power BI [1]. The use of machine learning and data visualization has been crucial in various fields, including criminology, where data visualization aids in crime analysis and prevention strategies [2]. MultiVision's approach to designing analytical dashboards with deep learning-based recommendations further simplifies the creation of multi-view visualizations by selecting meaningful data combinations automatically [3]. Additionally, a quantitative approach in human resource management underscores the increasing research interest and the use of bibliometric analysis to identify influential trends and relationships in data visualization [4]. Learning analytics dashboards utilizing predictive modeling provide valuable insights for early warning systems, demonstrating the application of data visualization in educational contexts [5]. The integration of predictive analytics with data visualization tools underscores the importance of these technologies in decision-making across various sectors, emphasizing the need for comprehensive analytics and business intelligence tools [6]. Utilizing tools like Power BI for business strategy recommendations highlights the practical applications of data visualization in financial and operational analysis [7]. The review of data visualization tools across diverse industries further elaborates on their utility in analyzing historical data, patient information, IT operations, and market research, while also addressing challenges like data quality and user adoption [8]. Specific implementations, such as the interactive dashboard for ICU COVID-19 data, demonstrate the technical methodologies involved in creating effective visualization tools for clinical data analysis [9]. The design and development of real-time data visualization dashboards emphasize the importance of an intuitive interface and the integration of reliable technologies to facilitate data exploration and decision-making [10]

1. METHODOLOGY

The methodology for developing the interactive dashboard encompasses several critical steps, from data collection and preparation to the implementation of various functionalities within the dashboard, including data visualization and predictive analytics. Each step is designed to ensure the effective transformation of raw sales data into actionable insights.

1. Data Collection and Preparation
2. Data Source and Loading:

The primary data source is the Superstore data set, which contains detailed sales records. The dashboard allows users to upload this dataset through a file uploader component that supports CSV, TXT, XLSX, and XLS formats. If no file is uploaded, the application defaults to a pre-existing dataset stored locally.

1. Data Preprocessing:

Once the dataset is loaded, it undergoes preprocessing to manage common data issues:

* Encoding: The dataset is read using "ISO-8859-1" encoding to manage special characters.
* Missing Values: The dataset is checked for missing values, and appropriate measures (such as filling, interpolation, or exclusion) are applied.
* Date Conversion: The "Order Date" column is converted to datetime format, with invalid dates coerced into NaT (Not a Time) for later exclusion.
* Column Verification: The presence of key columns like "Category" or "Item Type" is verified to ensure that essential data fields are available for analysis. If these columns are missing, an error message is displayed, and the application stops.

1. Interactive Dashboard Design
2. User Interface:

The dashboard is designed using Streamlit, providing an intuitive and interactive user experience:

* Sidebar Filters: Users can filter data based on Region, State, and City using multi-select dropdowns.
* Date Range Selector: A date picker allows users to define a start and end date for filtering the sales data.
* Main Display Area: The central part of the dashboard displays various visualizations and data tables.

1. Data Filtering:

The data is dynamically filtered based on user selections:

* Region, State, and City Filters: Data is subset according to the regions, states, and cities selected by the user.
* Date Filtering: The dataset is further filtered by the selected date range, ensuring that only relevant records are analyzed.

1. Visualization Components:

The dashboard leverages Plotly to create interactive and visually appealing charts:

1. Bar Charts: Display sales by category, showing the total sales for each category with detailed labels.
2. Pie Charts: Illustrate the distribution of sales by region and segment, providing a clear visual comparison.
3. Line Charts: Used for time series analysis, showing monthly sales trends over the selected period.
4. Treemaps: Provide a hierarchical view of sales by region, category, and sub-category, allowing for detailed exploration of sales data.
5. Scatter Plots: Highlight the relationship between sales and profit, offering insights into profitability patterns.
6. Data Export:

Users can download filtered datasets and visualizations:

* CSV Export: Filtered data and summary tables can be exported as CSV files, enabling further offline analysis.

1. Predictive Analytics
2. Model Training:

A linear regression model is employed to predict future sales:

* Feature Engineering: The "Order Date" column is converted to ordinal values, which are numerical representations of dates, facilitating the training process.
* Training and Testing: The dataset is split into training and testing sets, typically with an 80-20 split. The training set is used to fit the linear regression model, while the testing set evaluates the model's performance.

1. Future Sales Prediction:

The trained model is used to forecast future sales:

* Date Generation: Future dates are generated from the last date in the dataset to January 1, 2025.
* Ordinal Conversion: These future dates are converted to their ordinal equivalents.
* Sales Prediction: The model predicts sales for these future dates, providing a forecast of sales trends.

1. Visualization of Predictions:

Predicted sales are visualized alongside actual historical sales:

* Line Plot: A line plot shows both actual and predicted sales over time, with actual sales represented by a solid line and predicted sales by a dashed line. This visualization helps in comparing the predicted trends with historical data.

1. Error Handling:

Robust error handling mechanisms are implemented to manage potential issues:

* Date Conversion Errors: Any errors during date conversion are caught and displayed to the user, with the application stopping if critical issues are encountered.
* Data Quality Issues: Checks are in place to ensure that the dataset meets the necessary criteria for analysis and prediction.

1. SYSTEM DESIGN AND ARCHITECTURE

The design and architecture of the interactive dashboard for Superstore sales analysis and prediction are meticulously crafted to ensure robustness, scalability, and user-friendliness. The system leverages a combination of Python libraries and follows a modular approach, integrating various components to provide comprehensive data analytics and visualization capabilities.

1. Data Ingestion Layer

The data ingestion layer is responsible for acquiring and loading the dataset into the application. This layer supports multiple file formats, including CSV, TXT, XLSX, and XLS. Key functionalities include:

* File Upload Component: A Streamlit file uploader widget allows users to upload their datasets. This component supports various file formats and ensures that the uploaded data is accessible for further processing.
* Default Data Loading: If no file is uploaded, the system defaults to a pre-existing dataset stored locally. This ensures that the application remains functional even in the absence of user-provided data.

1. Data Processing Layer

The data processing layer manages the preprocessing and filtering of the dataset to ensure it is suitable for analysis and visualization:

* Data Cleaning: This involves managing missing values, converting date columns to appropriate formats, and verifying the presence of essential columns like "Category" and "Item Type".
* Filtering Mechanisms: The system provides multi-level filtering options based on Region, State, City, and Date Range. These filters allow users to subset the data dynamically, focusing on specific areas of interest.
* Date Handling: The "Order Date" column is processed to manage various date formats and to filter data based on user-selected date ranges.

1. User Interface Layer

The user interface layer is designed using Streamlit, providing an interactive and intuitive experience for users:

* Sidebar Filters: The sidebar contains multi-select dropdowns and date pickers, enabling users to filter the data by Region, State, City, and Date Range.
* Main Display Area: The central part of the dashboard displays visualizations, summary tables, and other key information. This area is dynamically updated based on user inputs from the sidebar filters.

1. Visualization Layer

The visualization layer leverages Plotly to create interactive and visually appealing charts and graphs:

* Bar Charts: These charts display sales data categorized by different dimensions such as Category and Region, providing clear visual comparisons.
* Pie Charts: Used to show the distribution of sales across different segments and regions, offering a quick overview of proportional data.
* Line Charts: Essential for time series analysis, these charts illustrate sales trends over selected time periods.
* Treemaps: Provide a hierarchical view of sales data, enabling users to explore sales across multiple levels of granularity.
* Scatter Plots: Visualize the relationship between sales and profit, helping users identify patterns and correlations.

1. Predictive Analytics Layer

The predictive analytics layer incorporates a linear regression model to forecast future sales trends:

* Model Training: The system trains a linear regression model on historical sales data, converting dates to ordinal values for accurate modeling.
* Future Predictions: The trained model predicts future sales up to January 2025. Future dates are generated and converted to ordinal values for prediction purposes.
* Visualization of Predictions: Predicted sales are visualized alongside actual historical sales, providing a comprehensive view of both past performance and future trends.

1. Error Handling and Logging

Robust error handling and logging mechanisms are integrated throughout the system to ensure reliability and user-friendliness:

* Error Messages: Informative error messages are displayed to users in case of issues such as missing essential columns or date conversion errors.
* Logging: Critical operations and errors are logged for debugging and auditing purposes. This helps in maintaining the system and addressing issues promptly.

1. IMPLEMENTATION AND RESULT

The development of the interactive dashboard for Superstore sales analysis and prediction involved several key steps:

1. Development Environment Setup:

The application is built using Python, leveraging libraries such as Streamlit for the UI, Plotly for data visualization, Pandas for data manipulation, and Scikit-learn for machine learning.

Necessary libraries are installed via pip.

1. Data Ingestion and Preprocessing:

Users can upload datasets in various formats (CSV, TXT, XLSX, XLS) via a Streamlit file uploader. If no file is uploaded, a default dataset ("Superstore.csv") is loaded.

The dataset is cleaned by managing missing values and converting the "Order Date" column to datetime format. Essential columns are verified for their presence.

1. User Interface and Filtering:

Sidebar filters include multi-select dropdowns and date pickers for filtering data by Region, State, City, and Date Range.

Data is dynamically filtered based on user selections, creating subsets for analysis and visualization.

1. Data Visualization:

Bar charts show sales by category, pie charts visualize sales distribution by region and segment, line charts display monthly sales trends, treemaps provide a hierarchical view of sales, and scatter plots illustrate the relationship between sales and profit.

1. Predictive Analytics:

A linear regression model is trained on historical sales data, with dates converted to ordinal values.

The model predicts future sales up to January 2025, visualized alongside historical data for comparison.

1. Error Handling and Logging:

Informative error messages are displayed for issues like missing columns or date conversion errors. Logging is implemented for debugging and auditing.

1. Deployment:

The application is deployed on a web server using platforms like Streamlit's sharing service or Heroku.

A screenshot of a computer

Description automatically generated

Figure 1: Category and Region

A screenshot of a computer

Description automatically generated

Figure 2: Time Series

A screenshot of a computer

Description automatically generated

Figure 3: Segment wise and Category wise

A screenshot of a computer

Description automatically generated

Figure 4: Hierarchical

A screenshot of a computer

Description automatically generated

Figure 5: Summary

A screenshot of a computer

Description automatically generated

Figure 6: Prediction

1. CONCLUSION

In conclusion, the development of the interactive dashboard marks a significant step towards enhancing data-driven decision-making in the retail sector. By harnessing the capabilities of modern data science and machine learning techniques, businesses can gain deeper insights into their sales performance and anticipate future trends with greater accuracy. The dashboard's user-friendly interface, coupled with its robust analytics capabilities, empowers stakeholders across the organization to make informed decisions that drive growth and profitability.

Looking ahead, the ongoing refinement and expansion of the dashboard holds promise for further enhancing its value proposition. Future iterations could explore the integration of advanced predictive models, real-time data streaming, and enhanced visualization techniques to provide even richer insights and actionable intelligence. By embracing innovation and leveraging the power of data analytics, businesses can stay agile and competitive in an ever-evolving marketplace, positioning themselves for success in the digital age.

1. REFERENCES

[1]DashBot: Insight-Driven Dashboard Generation Based on Deep Reinforcement Learning by Dazhen Deng, Aoyu Wu, Huamin Qu, Yingcai Wu

[2]Application of Data Visualization and Machine Learning Algorithms for Better Decision Making by  Olta Llaha 1, Azir Aliu 2

[3]MultiVision: Designing Analytical Dashboards with Deep Learning Based Recommendation By Aoyu Wu, Yun Wang, Mengyu Zhou, Xinyi He, Haidong Zhang, Huamin Qu, Dongmei Zhang

[4]A Quantitative approach for Data visualization in Human resource Management By Bandar Abdullah AlMobark

[5]Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies By Gomathy Ramaswami, Teo Susnjak, Anuradha Mathrani, Rahila Umer

[6]Predictive Analytics With Data Visualization By Selvaraj Boopathy, Palanisamy Satheesh Kumar,  Muharrem Karaaslan

[7]Data Visualization approach for business strategy recommendation using Power BI Dashboard By Siddhartha Das1, Dr. Kallal Banerjee2, Soumen Nath2, Sourav Chatterjee2

[8]DATA VISUALIZATION TOOLS USED FOR DECISION  MAKING: A REVIEW by Abhishek Chavan

[9]An Interactive Dashboard for Statistical Analysis of Intensive Care Unit COVID-19 Data By Rúben Dias 1, Artur Ferreira 1,2,\* , Iola Pinto 1,3 , Carlos Geraldes 1,4 , Cristiana Von Rekowski 1 and Luís Bento

[10]DESIGN AND DEVELOPMENT OF AN INTERACTIVE DASHBOARD FOR REAL-TIME DATA VISUALIZATION AND ANALYSIS