|  |  |
| --- | --- |
| Internship Project Title | TCS ION RIO 125-  Forecasting System - Project Demand of Products at a Retail Outlet Based on Historical Data |
| Name of the Company | TCS ION |
| Name of the Industry Mentor | Mr. Debashis Roy |
| Name of the Institute | Vishwakarma University,Pune |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Start Date | End Date | Total Effort (hrs.) | Project Environment | Tools used |
| 19/06/2024 | 18/07/2024 | 125 | Python | Google Collab |

**TABLE OF CONTENTS**

* Acknowledgements
* Objective
* Introduction/Description of Internship
* Approach/Methodology
* Assumptions
* Table
* Challenges & Opportunities
* Recommendations
* Outcome/Conclusion
* Enhancement Scope

**ACKNOWLEDGEMENT**

I would like to extend my heartfelt gratitude to our esteemed academic mentor, Prof. Shriprada Chaturbhuj of Vishwakarma University, and our industry mentor, Mr. Debashis Roy of TCS ION, for their invaluable contributions to the successful completion of my project, "Forecasting System - Project Demand of Products at a Retail Outlet Based on Historical Data". I am profoundly grateful for the time and effort they have dedicated throughout the internship. Your insightful guidance and unwavering support have been instrumental to my progress.

I also wish to express my sincere thanks to Vishwakarma University and TCS ION for granting me this exceptional opportunity.

I affirm that this project was undertaken and completed solely by me. It has been submitted to the Department of Computer Science of Vishwakarma University, under the TCS ION Industry Honour Program, during the academic year 2023-2024.

Furthermore, I am deeply appreciative of the support provided by the Head of the Department, Dr. Rajkumar Jagadale .

**Objective**

The primary objective of the forecasting system is to accurately predict the future demand for products at a retail outlet using historical sales data. This prediction will help in optimizing inventory management, reducing stockouts and overstock situations, and improving overall customer satisfaction. By leveraging advanced machine learning techniques and data analysis, the system aims to provide actionable insights that will enable the retail outlet to:

1. **Enhance Inventory Management:**
   * Minimize stockouts by predicting high-demand periods.
   * Reduce overstock by forecasting lower demand periods, leading to reduced storage costs and spoilage.
2. **Improve Customer Satisfaction:**
   * Ensure product availability by aligning stock levels with predicted demand, enhancing the shopping experience.
3. **Optimize Supply Chain Operations:**
   * Streamline ordering processes by providing accurate demand forecasts to suppliers, reducing lead times and ensuring timely restocking.
4. **Increase Sales and Revenue:**
   * Leverage demand predictions to design effective sales strategies and promotions, targeting products likely to be in high demand.
5. **Data-Driven Decision Making:**
   * Empower managers with data-driven insights to make informed decisions regarding product assortment, pricing strategies, and marketing campaigns.

**Introduction/Description of Internship**

The objective of this project is to build a forecasting system to predict demand of products at a retail outlet based on historical data.

**Approach/Methodology**

### Approach

1. **Data Acquisition and Preparation**
2. **Feature Engineering**
3. **Model Training and Validation**
4. **Prediction**
5. **Evaluation and Visualization**

### Methodology

#### 1. Data Acquisition and Preparation

**Objective:** Load and preprocess the dataset to make it suitable for modeling.

**Steps:**

* **Load Data**: Import the dataset from a CSV file or other data sources.
* **Convert Data Types**: Ensure date columns are in the datetime format for time series analysis.
* **Set Index**: Set the date column as the index to facilitate time-based operations.
* **Aggregate Sales**: Aggregate data on a monthly basis (or other desired frequency) to standardize the time intervals.
* **Feature Extraction**: Extract relevant features such as year and month from the date column.

#### 2. Feature Engineering

**Objective:** Create features and target variables that the model will use to learn patterns.

**Steps:**

* **Feature Matrix (X)**: Create a feature matrix with variables such as 'Year' and 'Month'.
* **Target Variable (y)**: Define the target variable as the sales units.

#### 3.Model Training and Validation

**Objective:** Train a predictive model using historical data and validate its performance.

**Steps:**

* **Select Model**: Choose an appropriate machine learning model, such as RandomForestRegressor.
* **Train Model**: Fit the model on the historical data.
* **Validate Model**: Optionally, use cross-validation or a validation set to evaluate the model’s performance.

#### 4. Prediction

**Objective:** Use the trained model to forecast future sales.

**Steps:**

* **Create Future Data**: Generate future time periods (e.g., each month of a specific year).
* **Predict Sales**: Use the model to forecast sales for these future periods.

#### 5. Evaluation and Visualization

**Objective:** Assess the model's performance and visualize the results.

**Steps:**

* **Compare Predictions**: Plot actual versus predicted sales to visually inspect the model’s accuracy.
* **Adjustments**: If necessary, refine the model based on evaluation results.

**ASSUMPTIONS**

**1. Data Quality and Availability**

* Historical Data Completeness: The historical sales data provided is complete, accurate, and representative of past sales patterns.
* Consistent Data Format: The data is formatted consistently across different time periods and sources, ensuring compatibility for analysis.
* Data Granularity: The level of detail in the data (e.g., daily, weekly) is sufficient to capture important patterns and trends for forecasting.

**2. Stationarity of Time Series**

* Stationarity: The time series data is assumed to be stationary or can be made stationary through transformations. This means that the statistical properties (e.g., mean, variance) do not change over time.
* Seasonality and Trends: Any inherent seasonality or trends in the data can be identified and modeled appropriately.

**3. Model Validity and Assumptions**

* Historical Patterns Predict Future Trends: The assumption is that historical patterns and trends in the data will continue into the future, and past behavior can be used to predict future demand.
* Model Assumptions: Forecasting models such as ARIMA, SARIMA, or machine learning algorithms have underlying assumptions that are satisfied (e.g., linearity in ARIMA models, independence of residuals).

**4. External Factors**

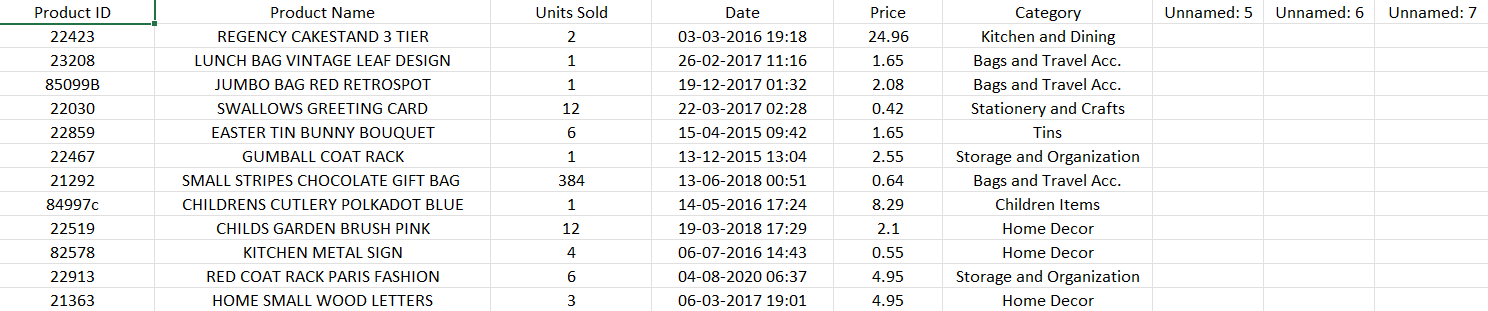
* Stable External Conditions: External factors influencing demand (e.g., economic conditions, market trends) are assumed to be stable or accounted for in the model. Significant changes in these factors could impact forecast accuracy.
* Promotions and Special Events: Assumptions are made about the predictability of promotional activities and special events, which may need to be incorporated into the model for accurate forecasting.

**5. Accuracy of Categorization**

* Product Categorization: The categorization of products into relevant groups or categories is accurate and consistent, ensuring that similar products are grouped together for forecasting purposes.
* Keyword Matching: The process of categorizing products based on keywords is assumed to be effective, and the keywords used are comprehensive and relevant.

**Table**

* Dataset Attributes:

****

**Challenges & Opportunities**

Data Quality and Quantity

Missing Data: Historical data might have missing values which can lead to inaccuracies in forecasting.

Inaccurate Data: Data might be incorrect due to entry errors or inconsistencies.

Volume: Large volumes of data can be challenging to process and analyze effectively.

Seasonality and Trends

Seasonal Variations: Retail demand often fluctuates with seasons, holidays, and events. Accurately capturing these patterns is crucial.

Trends: Long-term changes in consumer behavior or market conditions can affect demand, requiring models that can adapt to such shifts.

External Factors

Economic Conditions: Economic downturns or booms can impact consumer spending and demand.

Competitor Actions: New promotions or changes in competitors’ strategies can influence demand.

Supply Chain Issues: Disruptions in the supply chain can affect product availability and demand forecasts.

Model Complexity

Choosing the Right Model: Selecting and tuning forecasting models (e.g., ARIMA, Exponential Smoothing, Machine Learning) can be complex.

Overfitting: Models might perform well on historical data but fail to generalize to future data.

Integration with Business Processes

Actionable Insights: Forecasts need to be actionable and integrated with inventory management and supply chain planning.

Real-Time Updates: Incorporating real-time data can be challenging but is necessary for accurate forecasting.

**Data Collection**

Gather historical sales data, including relevant features (e.g., product type, sales volume, date).

Collect additional data sources if needed (e.g., customer behavior, economic indicators).

Data Preparation

Clean the data, handle missing values, and preprocess it for analysis.

Feature engineering to create meaningful predictors (e.g., day of week, promotions).

Model Selection and Training

Choose appropriate forecasting models based on the data characteristics.

Train and validate models using historical data, and tune hyperparameters for better performance.

Evaluation and Validation

Evaluate model performance using metrics such as RMSE, MAE, and MAPE.

Test models on unseen data to ensure generalization.

Deployment and Integration

Deploy the model into a production environment.

Integrate the forecasting system with inventory and supply chain management systems.

Monitoring and Maintenance

Continuously monitor model performance and update it as needed.

Adapt to changes in data patterns or external factors by retraining the model.

**Recommendations**

Enhance Data Collection and Quality

Recommendations:

Expand Data Sources: Incorporate additional data sources such as customer demographics, competitive pricing, and market trends to enrich the forecasting models.

Improve Data Accuracy: Implement regular data audits and validation checks to maintain high data quality and address inconsistencies promptly.

Automate Data Collection: Use automated data pipelines to streamline data ingestion and reduce manual errors.

2. Refine Forecasting Models

Recommendations:

Model Diversity: Explore and integrate advanced forecasting models, such as deep learning techniques (LSTM, GRU) or hybrid models that combine multiple approaches.

Continuous Learning: Implement model retraining strategies to update forecasts based on recent data and changing market conditions.

Error Analysis: Regularly perform error analysis to understand and address the causes of forecast inaccuracies.

3. Optimize Feature Engineering

Recommendations:

Dynamic Features: Incorporate more dynamic features that reflect real-time changes in demand, such as weather conditions or social media trends.

Feature Selection: Continuously evaluate and refine feature selection to include only the most relevant variables for improving model performance.

Experiment with Interactions: Explore interactions between features to capture complex relationships that may impact demand.

4. Improve System Integration and Usability

Recommendations:

User Feedback: Collect regular feedback from users to identify pain points and areas for improvement in the forecasting system.

Training and Support: Provide comprehensive training and support materials to ensure users can effectively utilize the system.

User Interface: Enhance the system’s user interface to make it more intuitive and accessible for users with varying levels of technical expertise.

5. Monitor and Evaluate Performance

Recommendations:

Performance Metrics: Establish a set of key performance indicators (KPIs) to continuously monitor the system’s effectiveness, such as forecast accuracy, inventory turnover, and stockouts.

Benchmarking: Compare the performance of the forecasting system against industry standards and best practices to identify areas for improvement.

Regular Reviews: Schedule periodic reviews of the system’s performance to ensure it aligns with business goals and adapts to changing conditions.

6. Enhance Collaboration and Communication

Recommendations:

Cross-Functional Teams: Foster collaboration between data scientists, IT professionals, and business stakeholders to ensure alignment on objectives and requirements.

Transparent Reporting: Implement transparent reporting practices to keep stakeholders informed about forecast accuracy, system performance, and any issues encountered.

Knowledge Sharing: Encourage knowledge sharing and best practice discussions within the team to drive continuous improvement.

**Outcome/Conclusion**

In this analysis, we leveraged a RandomForestRegressor model to forecast monthly sales for the year 2022 based on historical data. Here’s a summary of the approach and key findings:

1. **Data Preparation**:
   * We began by loading and preparing the dataset, including converting the 'Date' column to a datetime format and aggregating sales data on a monthly basis.
   * We extracted relevant features such as 'Year' and 'Month' to serve as inputs for our predictive model.
2. **Model Training**:
   * A RandomForestRegressor was trained on historical monthly sales data. This model learned from past patterns and relationships between input features (year and month) and the target variable (units sold).
3. **Prediction for 2022**:
   * Using the trained model, we predicted sales for each month of the year 2022.
   * The predicted sales values were generated for every month, allowing us to forecast future sales based on the historical trends observed.
4. **Visualization**:
   * We plotted both actual and predicted sales to visually compare the historical sales data with the forecasted values for 2022.
   * The plot highlighted the model’s ability to capture trends and patterns in the data, showing how well the predictions align with historical sales figures.
5. **Insights**:
   * By forecasting sales for the entire year of 2022, we gain insights into expected sales performance, which can be valuable for inventory management, budgeting, and strategic planning.
   * The model’s predictions provide a data-driven approach to anticipate future sales, helping businesses make informed decisions.

**Enhancement Scope**

Data Management and Integration

Scope:

Expand Data Sources: Integrate additional data sources such as external market trends, economic indicators, and customer feedback.

Data Quality Improvement: Implement automated data cleaning and validation processes to maintain high-quality data.

Real-Time Data: Develop mechanisms for real-time data ingestion to ensure up-to-date information is used in forecasting.

2. Advanced Forecasting Techniques

Scope:

Model Diversity: Explore and implement advanced forecasting models such as:

Time Series Analysis: ARIMA, SARIMA, Prophet.

Machine Learning: Random Forest, Gradient Boosting, XGBoost.

Deep Learning: LSTM, GRU, Transformer models.

Hybrid Models: Combine different forecasting techniques to leverage the strengths of each.

Ensemble Learning: Use ensemble methods to improve forecast accuracy by combining predictions from multiple models.

3. Feature Engineering and Selection

Scope:

Dynamic Features: Create and test dynamic features such as weather data, promotional activities, and social media sentiment.

Feature Selection: Utilize techniques like recursive feature elimination, feature importance from models, and dimensionality reduction (e.g., PCA) to select the most relevant features.

Interaction Features: Investigate and incorporate interaction features to capture complex relationships between variables.

4. System Architecture and Integration

Scope:

Scalable Architecture: Design a scalable system architecture that can handle increased data volumes and user load.

Cloud Solutions: Use cloud-based platforms for flexible data storage, processing, and model deployment.

APIs and Integration: Develop APIs for seamless integration with other business systems such as ERP and CRM.

5. User Experience and Interface

Scope:

User Interface Design: Improve the user interface to make it more intuitive and user-friendly, catering to different user roles and technical expertise levels.

Interactive Dashboards: Implement interactive dashboards and visualizations to help users understand forecasts and insights more effectively.

Feedback Mechanism: Create channels for users to provide feedback and suggest improvements.