DETAILED PROJECT REPORT (DPR)

Thomas Antony

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Back Order Prediction project aims to enhance inventory management, product merchandising, and customer satisfaction by predicting which products may go into backorder. This predictive system leverages historical data from ERP systems, employing supervised machine learning models for binary classification.

Objectives:

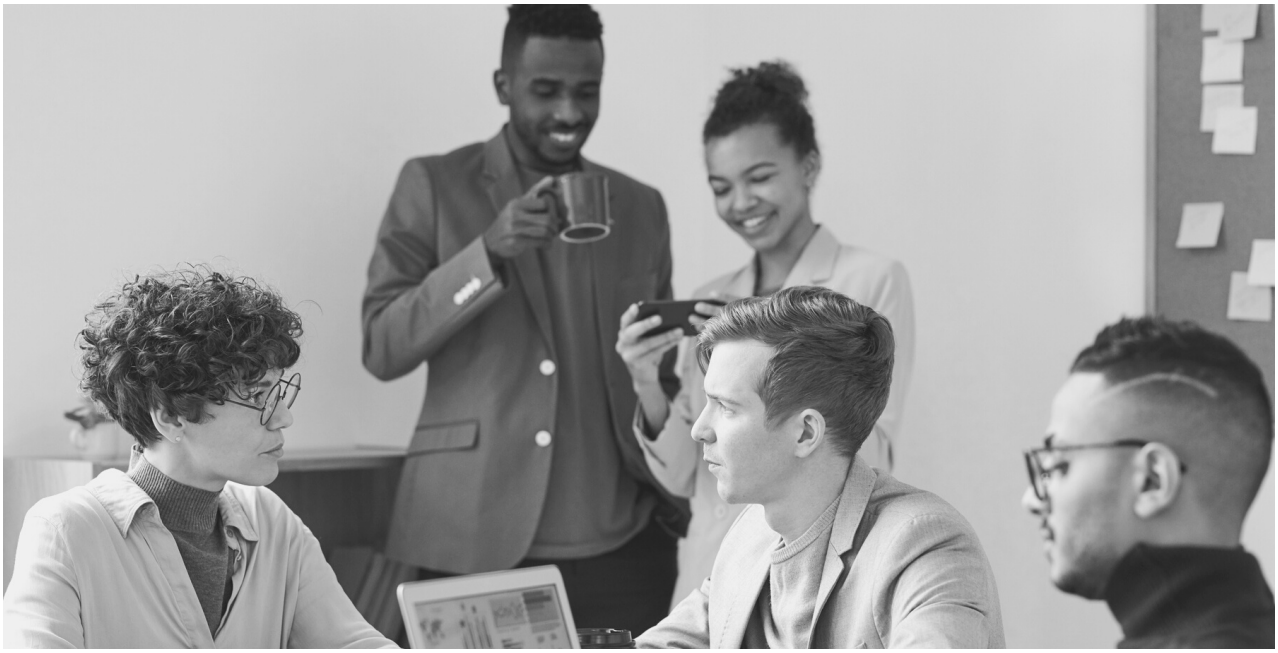
1. Predict backorders accurately to streamline inventory management.
2. Improve product merchandising strategies based on forecasted demand.
3. Enhance customer satisfaction by minimizing instances of product unavailability.

Achievements:

1. Developed a robust system using technologies such as Python, Flask, MLflow, and DVC.
2. Implemented a Model Training Module with Data Preprocessing, Feature Selection, and Model Evaluation components.
3. Ensured version control with DVC and MLflow, tracking data versions and model metrics for effective model comparison.
4. Established an Event Logging Module to monitor and debug system activities.
5. Created an intuitive user interface with wireframes for easy interaction.

The project exhibits promising results with an accuracy of approximately 91.25% and a recall of 72.47%. Despite challenges related to class imbalance, the system demonstrates its potential in strategic decision-making processes for inventory, merchandising, and customer satisfaction. Moving forward, opportunities for refinement include addressing class imbalance issues and exploring model parameter adjustments for a more balanced and effective Back Order Prediction system. The choice between precision and recall should align with specific business or operational use case priorities. Overall, the project provides valuable insights and recommendations for future enhancements.

PROJECT BACKGROUND



In the dynamic landscape of business and commerce, effective inventory management is crucial for ensuring operational efficiency and customer satisfaction. Traditional methods often fall short in predicting and preventing backorders, leading to challenges in product availability, customer dissatisfaction, and potential revenue loss.

The Back Order Prediction project arises from the necessity to address these issues by leveraging the wealth of historical data generated by ERP systems. These systems, while producing significant amounts of structured data, often lack the ability to proactively forecast backorders. The aim is to capitalize on this historical data and construct predictive models that can classify products as likely to go into backorder or not.

By harnessing the power of predictive analytics, businesses can anticipate and plan for potential backorders, enabling more proactive inventory management. This, in turn, facilitates improved product merchandising strategies and contributes to higher levels of customer satisfaction.

OBJECTIVES

1. **Accurate Backorder Prediction**:
   - Develop models capable of accurately predicting which products are likely to go into backorder based on historical data.
2. **Streamlined Inventory Management**:
   - Provide businesses with a tool to streamline inventory management processes by forecasting and preparing for potential backorders.
3. **Enhanced Product Merchandising**:
   - Improve product merchandising strategies by aligning them with forecasted demand, reducing instances of product unavailability.
4. **Customer Satisfaction**:
   - Contribute to higher levels of customer satisfaction by minimizing disruptions caused by product unavailability.
5. **Utilization of ERP Data**:
   - Capitalize on the vast amount of historical data generated by ERP systems, transforming it into actionable insights for better decision-making.

METHODOLOGY

The Back Order Prediction project adopts a systematic approach to project development, centering around supervised machine learning models for binary classification. The methodology includes the following key steps:

1. Data Collection:
    - Gather historical data from ERP systems, focusing on past instances of backorders.
2. Data Preprocessing:
    - Conduct data cleaning, handle missing values, and transform the data for model compatibility.
    - Implement feature selection to include only the most relevant attributes for prediction.
    - Address class imbalance through synthetic oversampling techniques.
3. Model Selection:
    - Choose a suitable machine learning model for binary classification.
    - Experiment with various models and parameters using Jupyter notebooks.
    - Opt for RandomForestClassifier due to its performance in experiments.
4. Model Training:
    - Train the selected model on the preprocessed data using the chosen algorithm.
5. Model Evaluation:
    - Assess model performance using metrics such as accuracy, precision, recall, and F1 score.
    - Utilize these metrics to understand the model's ability to correctly predict backorders.
6. Deployment:
    - Create an HTML webpage to serve as an API for predicting new data based on the trained model.

CHALLENGES FACED

Despite the systematic methodology, the project encountered challenges, notably related to class imbalance in the dataset. The metrics, such as precision and recall, were disproportionately affected by the imbalance, leading to a lower F1 score. Class imbalance is a common issue in binary classification tasks, where the dataset is skewed towards one class.

CONCLUSION

The Back Order Prediction project has provided valuable insights into the effectiveness of predictive models in addressing inventory management challenges. The model demonstrated commendable accuracy, reaching approximately 91.25%. However, challenges such as class imbalance impacted precision and recall, highlighting the need for further refinement.

Key Takeaways:

- The model's recall of approximately 72.47% indicates its effectiveness in capturing actual backorders.
- Class imbalance impact on precision emphasizes the trade-off between precision and recall in binary classification.
- Recommendations for Improvement: Address class imbalance using techniques like resampling, adjust class weights, fine-tune model parameters, or explore different algorithms for better precision-recall balance.

In conclusion, while the Back Order Prediction application holds promise for inventory management, product merchandising, and customer satisfaction, further optimization and refinement are recommended. The choice between precision and recall should align with specific business priorities, and ongoing efforts to enhance the model's performance will contribute to a more balanced and effective system.