EXPERIMENT REPORT

Student Name	Asmita Sheshrao Kamble
Project Name	ML as a service
Date	10/10/2023
Deliverables	Model-2 ARIMA as SARIMA **Git link:** https://github.com/asmitaskamble/AT2_ML-as-a-Service

---

1. EXPERIMENT BACKGROUND	
Create a time-series forecasting model to predict total sales revenue for all stores and items over the next seven days. In order to improve inventory management and financial planning, it is necessary to improve sales revenue predictions.	
1.a. Business Objective	The goal of the proposal is to provide the business with accurate forecasts of sales revenue, as well as optimize inventory and resource allocation. By achieving real results, profitability can be enhanced, stockouts can be reduced, and financial planning can be enhanced. This may result in inefficiency in stock management and loss of revenue due to incorrect results.
1.b. Hypothesis	Hypothesis: When using a SARIMAX time-series model, sales revenue forecasting can be significantly improved. The improvement of forecasting accuracy may result in the reduction of stock inefficiencies, the elimination of revenue losses caused by stockouts, and the optimization of resource allocation, ultimately resulting in increased profitability and business efficiency.
1.c. Experiment Objective	The purpose of this experiment is to identify a more accurate way of forecasting sales revenues for the next 7 days by using the SARIMAX model. Our objective is to reduce forecasting error significantly by achieving a Mean Squared Error (MSE) below a predetermined threshold so that we can minimize forecasting error.

2. EXPERIMENT DETAILS	
An objective of the proposed approach is to preprocess data first (standardization and label encoding), detect outliers using Z-scores, model time-series forecasting using SARIMAX, and assess the prediction accuracy using Mean Squared Error. These techniques were selected to ensure data quality, model suitability, and accurate forecasts based on the information available.	
2.a. Data Preparation	The 'date' column was converted to a datetime index to prepare the data, continuous features were standardized for model compatibility and categorical variables were labelled for numerical representation. Furthermore, I identified and removed outliers using Z-scores to enhance the accuracy of my model. No feature engineering was performed in this experiment, or external variables were incorporated. Nevertheless, these steps could be crucial for future experiments to improve forecast accuracy and uncover hidden patterns in the data.
2.b. Feature Engineering	Data preprocessing and categorical variable encoding were the focus of feature generation in this experiment. SARIMAX uses only univariate time series and therefore does not require explicit feature engineering. As a result, there were no additional features that needed to be generated or removed.

2.c. Modelling	Install libraries # pip install pmdarima # pip install statsmodels In this study, SARIMAX was the primary model trained. The main reason I chose SARIMAX was its suitability for time-series forecasting tasks as well as the ability to consider seasonal factors. A number of hyperparameters were tuned, including the order of the ARIMA components (p, d, q), and different combinations of these were tested in order to optimize the model's fit. As part of future experiments, it might be worthwhile to consider incorporating exogenous variables, such as promotional data or economic indicators, into the SARIMAX model, which could increase its forecasting accuracy. It may also be helpful to explore different seasonality's and window sizes for the model in order to capture complex patterns in the data.

3. EXPERIMENT RESULTS	
Analyse in detail the results achieved from this experiment from a technical and business perspective. Not only report performance metrics results but also any interpretation on model features, incorrect results, risks identified.	

3.a. Technical Performance	Score of the relevant performance metric(s). Provide analysis on the main underperforming cases/observations and potential root causes.
	
	Performance of ARIMA
	`Mean Squared Error (MSE): 463632191.69441235`
	`Root Mean Squared Error (RMSE): 21532.120000000286`
	
	Observation: It's common for forecasted sales revenue to differ from actual sales, and t hings like unexpected events or changes in customer behavior cause this.

3.b. Business Impact	The model does a reasonably good job regarding forecasting accuracy, but it is not perfect. Inaccurate results may affect the allocation of resources, the amount of stock in the warehouse, and the level of revenue, resulting in future operational efficiency issues. Underestimating sales can result in missed opportunities, potentially affecting customer satisfaction and profitability. Overestimating sales could result in excess inventory costs while underestimating sales could result in excess inventory costs.
3.c. Encountered Issues	List all the issues you faced during the experiments (solved and unsolved). Present solutions or workarounds for overcoming them. Highlight also the issues that may have to be dealt with in future experiments. Outliers: Outliers were present in the model, so it was detected using the Z-score technique. Performance Threshold: In this study, the MSE and RMSE were used as evaluation metrics; however, the ideal threshold may vary from case to case.

4. FUTURE EXPERIMENT	
In this experiment, we highlight the potential of SARIMAX modeling for forecasting sales revenue in order to improve customer satisfaction.	
4.a. Key Learning	• In terms of sales revenue forecasts, the SARIMAX model is reasonably accurate. • Preprocessing of the data, detecting outliers, and tuning the model are all essential for making better predictions based on the data. • Forecasting accuracy may be affected by external factors as well as abrupt changes in demand. • In the future, experiments should consider external variables in order to be able to improve forecasting and address potential irregularities in data.
4.b. Suggestions / Recommendations	Given the results achieved and the project's overall objective, list the potential next steps and experiments. For each of them assess the expected uplift or gains and rank them accordingly. If the experiment achieved the required outcome for the business, recommend the steps to deploy this solution into production. • To handle patterns that are more complex, it may be necessary to explore more sophisticated models, such as neural networks and ensemble methods. • Ensure that the model is continually monitored and updated as new data is available to ensure its relevance at all times.