Analysis Report: Comparison of Different Models for Dispense Amount Prediction

Introduction:
In this analysis, we compared the performance of different models for predicting the dispense amount in an ATM system. The objective was to identify the most accurate and reliable model that can effectively forecast dispense amounts, aiding in optimizing ATM operations and ensuring sufficient cash availability.

Data:
The analysis was conducted using a dataset containing historical dispense amounts along with various features such as date, ATM ID, and transaction details. The dataset was split into training and test sets.

Models Explored:

ARIMA Model:

The ARIMA model was used to capture the temporal patterns and seasonality in the dispense amounts.
The model was trained on the training set and evaluated using the test set.
The model achieved a training score of 0.81 and a test score of 0.79, indicating good performance in capturing the underlying patterns in the data.
Hyperparameter tuning was performed, resulting in a slight improvement in the scores to 0.80 for training and 0.81 for the test set.

SARIMAX Model:

The SARIMAX model, an extension of ARIMA, incorporated exogenous variables in addition to the time series data to capture more complex relationships.
The model was trained on the training set and evaluated using the test set.
The model achieved a training score of 0.81 and a test score of 0.79, similar to the ARIMA model.
Hyperparameter tuning was performed, resulting in a marginal improvement with a training score of 0.80 and a test score of 0.81.

XGBoost Model:

The XGBoost model, a gradient boosting algorithm, was employed to capture complex relationships and non-linear patterns in the dispense amounts.
The model was trained on the training set and evaluated using the test set.
The model achieved a training score of 0.23 and a test score of 0.21, indicating lower performance compared to the ARIMA and SARIMAX models.
Hyperparameter tuning was performed, resulting in similar scores with a training score of 0.23 and a test score of 0.21.

Model Selection:

Based on the performance evaluation, the ARIMA model is recommended for dispense amount prediction. It demonstrated higher accuracy and consistency compared to the XGBoost model. Although hyperparameter tuning resulted in slight improvements in the XGBoost model's performance, it still lagged behind the ARIMA model.

Conclusion:

The analysis of different models for dispense amount prediction revealed that the ARIMA model outperformed the SARIMAX and XGBoost models in terms of accuracy. The ARIMA model effectively captured the underlying patterns and seasonality in the dispense amounts, making it a suitable choice for forecasting future dispense amounts. Further optimization and fine-tuning of the ARIMA model can be explored to enhance its performance and make it more robust for real-world applications.

Forecast Results:

Based on our model, the forecasted dispense amounts for each ATM for the next 7 days are as follows:

Day 11299: 0.0039223570783631
Day 11300: 0.0039223570783631
Day 11301: 0.0039223570783631
Day 11302: 0.0039223570783631
Day 11303: 0.0039223570783631
Day 11304: 0.0039223570783631
Day 11305: 0.0039223570783631

Validation Performance:
To evaluate the performance of our model, we calculated the Mean Absolute Error (MAE) on the validation set. The MAE measures the average absolute difference between the predicted and actual dispense amounts. In our case, the validation MAE is 0.8111986989147306, indicating the average error in dispense amount prediction.

Conclusion:
Our model provides forecasts for the dispense amount for each ATM for the next 7 days. These forecasts can be valuable for ATM operations and cash management teams to plan and optimize cash availability. However, it is important to regularly monitor the model's performance and update it with new data to ensure accurate and reliable predictions.