**FUTURE SALES PREDICTION PHASE 1**

* Forecasting sales is a common and essential use of machine learning (ML). Sales forecasts can be used to identify benchmarks and determine incremental impacts of new initiatives, plan resources in response to expected demand, and project future budgets. In this article, I will show how to implement 5 different ML models to predict sales.
* **Data Collection** : Gather historical sales data, including timestamps, product information, quantities sold, prices, and any relevant external factors (e.g., promotions, holidays, economic indicators).
* **Data Preprocessing** :Clean the data by handling missing values and outliers.Convert timestamps into a consistent format and extract features like day of the week, month, and year.Explore and visualize the data to understand trends, seasonality, and any patterns.
* **Feature Engineering**:Create additional features that can impact sales predictions, such as lag features (previous sales), rolling statistics, and holiday indicators.
* **Data Splitting**:Split the dataset into training, validation, and testing sets. Typically, you use historical data for training and validation, reserving a more recent period for testing.
* **Model Selection**:Choose an appropriate predictive model based on the nature of your data. Common choices include:Time Series Models (e.g., ARIMA, SARIMA, Prophet): Suitable for capturing temporal patterns.
* **Machine Learning Models** (e.g., XGBoost, Random Forest, LSTM): Effective for capturing complex relationships and non-linear patterns.
* **Model Training**:Train the selected model on the training dataset, using hyperparameter tuning if necessary.Consider normalizing or scaling features to ensure model stability.
* **Validation and Model Evaluation**:Validate the model's performance on the validation dataset using appropriate evaluation metrics (e.g., Mean Absolute Error, Mean Squared Error, Root Mean Squared Error).Fine-tune the model based on validation results.
* **Hyperparameter Tuning**:Optimize hyperparameters through techniques like grid search or Bayesian optimization to improve model performance.
* **Testing and Deployment**:Evaluate the final model on the test dataset to simulate real-world performance.Once satisfied with the results, deploy the model in a production environment where it can generate future sales predictions.
* **Monitoring and Maintenance**:Implement monitoring systems to track model performance in real-time.Regularly retrain the model using fresh data to adapt to changing market conditions.Continuously update the model to incorporate new features or improve its architecture as needed.
* **Feedback Loop**:Gather feedback from end-users and stakeholders to identify areas for improvement.Iterate on the model and the prediction pipeline to enhance accuracy and reliability.
* **Documentation and Reporting**:Document the entire process, including data sources, preprocessing steps, model architecture, and evaluation metrics.Create regular reports summarizing sales predictions, model performance, and insights for decision-makers.
* My role in this project is to train the dataset using hyperparameter.
* Evaluate the final model with real time environment.