posable.ai

Assessment Test: Forecasting Sales with Machine Learning

Objective:

Develop a predictive model that forecasts **future sales quantity** (ACTUAL) for **each store and each product** using historical data. The goal is to achieve **high accuracy** and **low error rates**, measured using **Mean Absolute Error (MAE)**, **Mean Absolute Percentage Error (MAPE)**, and **R-squared (R^2 Score)**.

Dataset Overview:

You are provided with **weekly sales data** for a single store (`STORE_CODE: 89888`). The dataset contains the following columns:

- `STORE_CODE` – Store identifier
- `PRODUCT_CODE` – Unique product identifier
- `CATEGORY_CODE` – Product category
- `SALES_WEEK` – The start date of the sales week
- `ACTUAL` – Actual quantity sold (Target Variable)
- `AVG_UNIT_PRICE` – Average selling price per unit

Task Instructions:

1. Data Preparation (25%)

- Load and explore the dataset.
- Handle missing values and outliers appropriately.

2. Feature Engineering (25%)

- Create relevant features that can improve the model's predictive power

3. Model Development (30%)

- Select an appropriate regression-based machine learning model (e.g., Linear Regression, Random Forest, XGBoost, LSTM).
- Train the model on historical sales data.
- Tune hyperparameters to improve performance.

4. Model Evaluation & Visualization (20%)

- Evaluate model performance using the following metrics:
  - **Mean Absolute Error (MAE)**
  - **Mean Absolute Percentage Error (MAPE)**
  - **R-squared (R^2 Score)**
- Compare results against a simple baseline (e.g., moving average).
- **Plot a chart comparing forecasted sales vs. actual sales over time** to visualize performance.
- Provide an analysis of the model's accuracy and any limitations.

Bonus Challenge:

If time permits, implement a **time-series forecasting approach** (e.g., ARIMA, Prophet, or LSTM) and compare performance with your initial model.

Expected Deliverables:

- A **Python script or Jupyter Notebook** containing the full implementation.
- A short **write-up (1-2 pages)** summarizing the methodology, model selection, and key findings.
- **A chart comparing forecasted vs. actual sales over time** to illustrate model accuracy.

Evaluation Criteria:

Criteria	Weight	Description
Data Preparation	25%	Handling missing values, outliers, and feature extraction
Feature Engineering	25%	Meaningful features to improve prediction accuracy
Model Development	30%	Correct model selection, training, and tuning
Evaluation & Visualization	20%	Clear explanation of error metrics and performance, including forecast vs. actual chart

Performance Benchmarks (Market Norm):

- **MAE**: < 5-10% of average sales per week
- **MAPE**: < 15%
- **R^2 Score**: > 0.75

Submission Deadline:

Candidates must **submit their completed assessment within 3 to 5 days** from the date they receive the test.

- **3 days**: Recommended for experienced candidates who can implement a model efficiently.
- **5 days**: Allows additional time for fine-tuning and documentation.

Late submissions may still be accepted but will be evaluated with consideration of timeliness.