**Sales Forecasting for Retail Stores Using LightGBM**

**Project Overview**

This project aims to predict future sales for various items across multiple retail stores. Accurate sales forecasting helps businesses manage inventory effectively, optimize marketing strategies, and enhance overall operational efficiency.

**Key Features**

* **Sales Forecasting**: Built a model to predict future store sales, aiding inventory management and marketing planning.
* **Tools Used**: Utilized Python and LightGBM for model implementation, ensuring accurate predictions.
* **Feature Engineering**: Developed time-based and lag features to enhance model accuracy.

**Data Description**

The project uses the following datasets:

* **train.csv**: Historical sales data for training the model.
* **test.csv**: Data for which sales predictions need to be made.
* **sample\_submission.csv**: Sample submission file for the predictions.

**Implementation Details**

1. **Data Loading and Preprocessing**
   * Loaded training, testing, and sample submission datasets.
   * Combined training and testing datasets for unified processing.
   * Handled missing values and ensured date columns were correctly parsed.
2. **Exploratory Data Analysis (EDA)**
   * Conducted initial analysis to understand the data range, unique stores, and items.
   * Aggregated sales data to calculate summary statistics such as sum, mean, median, and standard deviation.
3. **Feature Engineering**
   * Created new date-related features (month, day, week, etc.).
   * Developed lag features to incorporate previous sales information.
   * Computed rolling mean and exponentially weighted mean features to capture moving averages and trends.
4. **One-Hot Encoding**
   * Applied one-hot encoding to categorical variables like store, item, day of the week, and month.
5. **Modeling**
   * Transformed the sales data using a log transformation for normalization.
   * Defined custom SMAPE function for model evaluation.
   * Split the dataset into training and validation sets based on time.
   * Trained a LightGBM model with early stopping to prevent overfitting.
   * Evaluated the model using the SMAPE metric to measure prediction accuracy.
6. **Feature Importance**
   * Analyzed and visualized feature importance to identify key predictors in the sales forecasting model.

**Results**

The final model achieved high accuracy in predicting future sales, validated using the Symmetric Mean Absolute Percentage Error (SMAPE) metric. The feature importance analysis highlighted the most influential features, providing valuable insights into sales dynamics.

**How to Run the Project**

1. Clone the repository to your local machine.
2. Ensure you have the required libraries installed: **numpy**, **pandas**, **datetime**, **matplotlib**, **seaborn**, **lightgbm**, **warnings**.
3. Place the datasets (**train.csv**, **test.csv**, **sample\_submission.csv**) in the project directory.
4. Run the **sales\_forecasting.py** script to train the model and make predictions.

**Conclusion**

This project demonstrates the application of machine learning techniques to real-world sales forecasting, showcasing skills in data preprocessing, feature engineering, model training, and evaluation using Python and LightGBM.