**MLOps Report: Sales Forecasting and Demand Prediction Project**

**1. MLOps Implementation**

To ensure reproducibility and traceability, the following practices and tools were applied:

* **Experiment Tracking:**  
  MLflow was used to track model experiments. For each run, metrics such as MAE, MSE, and R², as well as model parameters, were logged to evaluate model performance and facilitate model selection.
* **Version Control:**  
  The codebase, including the trained model (model.pkl), encoders (encoders.pkl, onehotencoder.pkl), and preprocessing scripts (preprocessing.py), was version-controlled using Git and hosted on GitHub.
* **Artifacts Management:**  
  All trained models and preprocessing artifacts were saved and documented for reproducibility and reuse in deployment.

**2. Model Deployment**

* **Deployment Approach:**  
  The final model was deployed using **Streamlit** to provide an interactive user interface for batch and real-time forecasting.
* **Repository:**  
  The entire project, including the app and model files, is hosted on a public GitHub repository for open access:  
  👉 https://deplo-depi-mfep3nxjr3eyw5ftwgbxxc.streamlit.app/
* **Files Included:**
  + app.py: Streamlit app to serve the model.
  + model.pkl: The selected regression model trained on historical sales data.
  + encoders.pkl, onehotencoder.pkl: Encoders used during data preprocessing.
  + preprocessing.py: Contains data cleaning, encoding, and feature engineering steps.
* **Deployment Platform:**  
  The app is deployed using **Streamlit Sharing** or **Streamlit Community Cloud** (if applicable), making it accessible via a public web link.

**3. Model Monitoring**

* **Performance Monitoring:**  
  While active retraining is not implemented (due to no new data), the model's last performance was logged using:
  + **MAE**
  + **MSE**
  + **R² Score**
* **Model Drift:**  
  Since the data is static and no new inputs are expected, drift monitoring has not been implemented but the pipeline is modular to support future integration.

**4. Model Retraining Strategy**

* **Current Status:**  
  As of deployment, no new data has been added. Therefore, retraining is not scheduled.
* **Planned Strategy (If Needed):**
  + Periodic data checks (monthly/quarterly)
  + Trigger retraining based on:
    - Drop in forecast accuracy
    - Introduction of new seasonal trends or products

**5. Model Selection Justification**

The selected model was chosen based on its superior performance on the validation set.

|  |  |  |  |
| --- | --- | --- | --- |
| Model | MSE | Mae | R2 |
| GradientBoosting\_all\_features | 112.61 | 80.85 | 0.653 |
| RandomForest\_all\_features | 134.74 | 98.07 | 0.503 |
| ETS\_date\_only | 134.68 | 89.09 | 0.498 |
| LSTM\_all\_features | 187.75 | 147.17 | 0.044 |
| Prophet\_Date\_only | 1306098683.07 | 28828.36 | 0.81 |