Weekly Report: Rafid Ul Karim - Alpha AI

Week 4 (April 14 - April 19)

1. Tesla - S Stock Prediction Project

a. Progress & Learnings:

- Worked on the core phases of the Tesla - S stock forecasting project.
- Completed full data preprocessing workflow including:
    - Time-aware train/validation/test splitting.
    - Feature standardization.
    - Dataset versioning for reproducibility.
- Performed Exploratory Data Analysis (EDA), highlighting price trends, feature correlations, and temporal seasonality.
- Implemented and evaluated three predictive models:
    - **Linear Regression**: Used as a baseline; exposed high bias when excluding engineered features.
    - **XGBoost**: Tuned using cross-validation with RMSE minimization and regularization (L1/L2) to prevent overfitting.
    - **LSTM**: Used sequential time-series data with scaled inputs and MSE/MAE evaluation.

b. Key Concepts Covered:

- Model evaluation metrics: RMSE, MAE, R^2 across temporal validation folds.
- Use of train_test_split and K-fold Cross Validation while preventing data leakage in a time-series setting.
- Impact of data standardization and engineered features on model convergence and accuracy.

c. TODO:

- Finalize inference routines.
- Structured notebooks (currently under development) to be separated into:
    - EDA + Data Prep notebook: will save preprocessed datasets for reusability.
    - Model Training notebook: will include training scripts for all three models.
- Push all components to a version-controlled GitHub repository with README and usage docs.

2. Test-Driven Development in ML

a. Progress & Learnings:

- Received lecture on "Test Driven Development"
- Authored a technical documentation based off of lecture on Test-Driven Development (TDD) principles applied to machine learning workflows.
- Covered theoretical principles and practical strategies including:
    - Unit testing, Integration testing, System testing, Regression testing, Acceptance testing.
    - Use of mocking in testing dependencies.
    - Setting seed values for reproducibility in randomized ML pipelines.

b. Key Concepts Covered:

- Importance of deterministic testing in ML workflows with emphasis on reproducibility and test isolation.
- Best practices for structuring test suites for ML pipelines and data transformation functions.

c. Deliverables/Resources:

- Lecture notes and supplementary materials on Agile Development (no formal deliverables).
- Available in ‘Cloudly-Alpha-AI-Team-1” repository (branch: main):
    - Test Driven Development/ Documentation-Test-Driven-Development-Rafid-Ul-Karim.md