Tasks for AI/ML Engineering Interns

Objective: Work on diverse AI/ML projects to gain hands-on experience in anomaly detection, multi-label classification, satellite image analysis, and credit risk assessment.

Project Details

Task 1: Financial Time-Series Anomaly Detection

Objective: Build a tool to identify anomalies in stock price trends to detect unusual activities or market manipulations.

- **Dataset**: Yahoo Finance Stock Market Dataset
  - **Steps**:
    1. Download and preprocess historical stock price data for a few chosen companies. 2. Calculate financial indicators (e.g., SMA, EMA, RSI, Bollinger Bands).
    3. Use **Isolation Forest** or **DBSCAN** for unsupervised anomaly detection.
    4. Build a time-series forecasting model using **LSTM** or **Prophet** to identify deviations.
    5. Visualize detected anomalies on stock price trends.
- **Outcome**: A tool or report identifying anomalies in stock price trends and possible market manipulations.

Task 2: Multi-Label Emotion Recognition from Text

Objective: Develop a system to classify multiple emotions (e.g., joy, sadness, anger) present in textual data.

- **Dataset**: GoEmotions Dataset by Google
  - **Steps**:
    1. Preprocess the dataset, handling imbalanced data.
    2. Fine-tune a transformer model such as **BERT** for multi-label classification.
    3. Evaluate the model using metrics like **Hamming loss** and **F1 score**.
    4. Test the system on real-world textual data such as customer feedback or social media posts.
- **Outcome**: A system capable of analyzing emotional tones with

multiple labels.

Task 3: Satellite Image Analysis for Deforestation Monitoring

Objective: Use satellite imagery to detect and monitor deforestation areas over time.

- **Dataset**: Planet: Understanding the Amazon from Space
  - **Steps**:
    1. Preprocess multi-band satellite images for land cover analysis.
    2. Train **Convolutional Neural Networks (CNNs)** for land classification.
    3. Perform change detection using sequential images to identify deforestation areas.
    4. Visualize environmental changes over time.
- **Outcome**: A system that detects and monitors deforestation.

Task 4: Credit Risk Analysis

Objective: Build a model to assess the creditworthiness of customers and flag high-risk customers for financial institutions.

- **Dataset**: Give Me Some Credit Dataset
  - **Steps**:
    1. Preprocess the dataset, handling missing data and imbalances using techniques like **SMOTE**.
    2. Engineer features related to income, debt, and repayment history.
    3. Train models such as **Random Forest**, **Gradient Boosting**, or **XGBoost**.
    4. Evaluate the model using appropriate metrics.
- **Outcome**: A system that reduces default rates by flagging high risk customers.

Submission Requirements

For each task:

1. **Python Code**: Well-documented and commented scripts for the implemented models. 2. **Report**:
   - ○ Dataset preprocessing steps.
   - ○ Model selection and rationale.
   - ○ Challenges faced and solutions.

    - Results with visualizations and interpretations.
3. **Video Demonstration**: A 5-7 minute video for each task, showcasing the implementation and results.

Submission Deadline

16th May , 2025

Late submissions will only be considered with prior approval.