**Title: Proposal for a Trading Bot Using TPOT, Ensemble Models, and Robinhood Data**

**Problem Statement**

The stock market is highly dynamic, with price movements driven by multiple factors, including economic indicators, news sentiment, and historical trends. Traditional trading strategies, such as moving average crossovers and mean reversion, have shown varying levels of success but lack adaptability to changing market conditions. The objective of this project is to develop a machine learning-based trading bot that integrates TPOT (Tree-based Pipeline Optimization Tool), an ensemble of Linear Regression, LSTM (Long Short-Term Memory networks), and XGBoost to optimize trading strategies. The bot will leverage stock data collected from Robinhood to generate trading signals with the goal of maximizing returns while managing risk effectively.

**Data Selection and Justification**

Robinhood provides a rich dataset that includes real-time and historical stock prices, trading volume, and additional market indicators. The dataset is relatively clean and structured, making it ideal for applying machine learning models. The data will be preprocessed and standardized to ensure consistency and improve model performance. Given the scope of this course, the selected dataset is manageable and aligns with best practices for time-series forecasting and predictive modeling.

**Methodology**

The trading bot will be developed using the following approach:

1. **Data Collection**: Retrieve stock market data from Robinhood’s API, including Open, High, Low, Close (OHLC) prices, volume, and additional technical indicators such as SMA, RSI, and MACD.
2. **Feature Engineering**: Extract and engineer features such as volatility, momentum indicators, and lag variables to enhance predictive power.
3. **Model Selection**:
   * **TPOT**: A genetic algorithm-based AutoML tool to identify optimal machine learning pipelines.
   * **Ensemble Model**: A combination of:
     + **Linear Regression**: For baseline trend analysis.
     + **LSTM**: To capture sequential dependencies in stock prices.
     + **XGBoost**: To handle non-linear relationships and feature interactions.
4. **Backtesting and Evaluation**: Utilize historical data to simulate trading performance, using metrics such as Sharpe Ratio, Profit Factor, and Drawdown.
5. **Deployment**: Implement the model to generate real-time trading signals and execute trades via Robinhood’s API.

**Significance and Justification**

The need for an intelligent trading system arises from the inefficiencies of traditional trading models that fail to adapt to complex market conditions. Machine learning-based trading bots offer several advantages, including:

* **Data-driven Decision Making**: Eliminates human biases and emotions in trading.
* **Adaptive Learning**: Ability to update trading strategies dynamically based on market conditions.
* **Automation**: Reduces the need for manual intervention, allowing for real-time execution of trades.

This project is significant as it explores the combination of traditional machine learning (Linear Regression, XGBoost) with deep learning (LSTM) and AutoML (TPOT) to develop an efficient trading strategy. By leveraging Robinhood data, this bot can be used for both paper trading and live execution.

**Computational Resource Estimation**

The computational requirements for this project will vary based on the dataset size and model complexity. An estimate developed in collaboration with the mentor includes:

* **Data Storage**: ~1GB for historical stock data (dependent on the number of tickers and time horizon).
* **Processing Power**: A GPU is preferred for training LSTM models efficiently, but a high-performance CPU can also suffice.
* **Training Time**: TPOT’s AutoML approach may require a few hours per iteration, while XGBoost and LSTM training can range from minutes to a few hours depending on hyperparameter tuning.
* **Memory**: At least 16GB RAM is recommended to handle large datasets and multiple model training processes.

**Presentation and Community Engagement**

The proposal follows a logical structure, articulating the problem, methodology, and computational requirements. The project aims to be novel by leveraging multiple models in an ensemble approach while utilizing Robinhood data.

By integrating TPOT, LSTM, and XGBoost into a cohesive trading strategy, this project aims to build a robust, adaptive, and data-driven trading bot capable of outperforming traditional strategies.