<u>Stock Market Analysis and Predictive Modelling for AAPL Using Deep Learning</u>

<u>Abstract:</u> The project aims to create a hybrid Vector AutoRegression (VAR) and Long Short-Term Memory (LSTM) model for predicting stock prices that takes advantage of both linear and nonlinear trends in financial data. Our strategy begins with implementing VAR as a baseline model to establish a reference for linear stock price predictions. Given VAR's limitations in capturing long-term dependencies, **We use a Long Short-Term Memory (LSTM) model to predict future stock prices accurately.** The model enhances accuracy over standard forecasting approaches by using past stock prices and connected stock movements (for example, MSFT, GOOG, and AMZN). **The predicted outcome is a highly optimized stock price prediction model that helps investors make informed decisions.**

<u>Introduction</u>: Stock price prediction is difficult due to its high volatility and non-linearity. Traditional models, such as ARIMA and simple regression, struggle to capture long-term relationships. To solve this, we combine the strengths of VAR (which captures multivariate correlations and establishes a baseline) with LSTM (which models sequential dependencies and forecasts future prices).
This initiative is valuable for financial analysts, investors, and academics who aim to understand Apple's stock market behavior and forecast future trends. Including traditional statistical models (VAR) and advanced machine learning models (LSTM) enables a thorough evaluation of their efficacy in stock price prediction.

Objective and research Questions:

- How effectively can a hybrid VAR-LSTM model improve stock price forecasting?
- What impact does incorporating multiple stock trends have on accuracy?
- How does the VAR baseline compare to the LSTM model?

<u>Problem Statement:</u> Stock prices are influenced by a variety of external factors, making it difficult to forecast precisely. Most present models struggle to balance linear dependencies (VAR) with complicated sequential patterns (LSTM).

Challenges in Existing Solutions:

- Linear models (e.g., ARIMA, VAR) struggle with non-linearity but can serve as valuable baselines.
- Deep learning models, such as LSTMs, require extensive data for training.
- Traditional forecasting makes insufficient use of connected stocks.

<u>Project Architecture:</u>

- System Components:
    - i. Data Collection: Get historical Apple Inc. (AAPL) stock data from Yahoo Finance.
    - ii. Data Processing Module: cleans, normalizes, extracts features, and deals with missing values.
    - iii. Baseline VAR model: creates a benchmark for predicting stock prices.
    - iv. LSTM model: gains knowledge of sequential patterns for predicting future prices.
    - v. The hybrid Prediction module optimizes forecasting by comparing LSTM results with the VAR baseline.
- Data Flow and Processing Pipeline:
    - i) Fetch stock data
    - ii) Preprocess Data
    - iii) Train the VAR model to establish a baseline
    - iv) Train LSTM model to predict future prices
    - v) Compare VAR vs LSTM predictions and analyze performance improvements
    - vi) To find more predictive information, compare the prices of AAPL and other connected stocks.

Tools and Technologies:

- Software:
    1. Programming Languages: python
    2. Libraries: TensorFlow, Scikit-learn, Statsmodels, Pandas, NumPy, Matplotlib, Seaborn, streamlit
    3. Development Environment: VSCode (EDA tasks), Google Colab (Model Building)
- Hardware:
    1. Processing Unit: Using Google Colab T4-GPU for model building
    2. Storage: Local Hardware

Dataset:

- Source: Yahoo Finance API
- Preprocessing: missing value handling, normalization, EDA, feature engineering, correlation between stocks, rolling average, lag features
- Features: The dataset includes features such as the stock price of Apple Inc. (AAPL) as the target variable, along with date, close price, high price, low price, open price, and volume as features.

Methodology:

- Data Collection: Historical stock data for Apple Inc. (AAPL) and other companies is retrieved using the yfinance library.
- Preprocessing: cleans, normalizes, extracts features, and deals with missing values, EDA
- Feature Engineering: Create rolling averages and lag features and normalize data.
- Train VAR Model: Establish a linear baseline for stock price trends.
- Train LSTM Model: Learn sequential dependencies to improve future predictions.
- Compare VAR and LSTM Model: Analyse improvements from deep learning over the linear baseline.
- Performance Evaluation: Compare accuracy with existing models.

Evaluation Metrics:

- Accuracy: MSE, RMSE (generally used for stock price predictions), MAPE
- Precision, Recall, and F1 Score are not applicable for regression tasks
- Computational Performance Metrics: Training time and inference time for both VAR and LSTM models.

Expected Outcomes:

- Illustrations of returns, trends, and moving averages of Apple's stock.
- The baseline for predicting stock prices is a trained VAR model.
- An LSTM model trained for sophisticated stock price forecasting.
- An analysis of how well the VAR and LSTM models forecast the prices of Apple's shares.

Challenges and Risks:

- Possible Obstacles:
    1. Market Volatility: Accuracy may be affected by unexpected fluctuations in stock prices.
    2. LSTM model overfitting as a result of insufficient data.
    3. Ensuring the VAR model is stationary.
- Mitigation Strategies:
    1. Employ regularisation strategies to keep the LSTM model from overfitting.
    2. Conduct comprehensive differencing and stationarity checks for the VAR model.

References:

- Yahoo Finance: https://finance.yahoo.com/quote/AAPL/
- Keras Documentation: https://keras.io/
- [1] Sepp Hochreiter, Jurgen Schmidhuber, "LONG SHORT-TERM MEMORY", Neural Computation, 1997
- [2] Chatfield, Time-series Forecasting, Chapman & Hall, 2016.
- [3] Qi Li, Norshaliza Kamaruddin, Stock prices changes using long-short term memory neural network with symbolic genetic programming
- [4] Tanvi, Rishabh, Sheetal, Stock Price Prediction using LSTM
- [5] Y.S. Abu-Mostafa and A.F. Atiya, Introduction to financial forecasting, Applied Intelligence 6.3(1996), pp. 205-213.