**Cracking market code withai driven stock price prediction using time series analysis**

**Student Name:** HARIHARAN.C

**RegisterNumber:**513523106012

**Institution:** ANNAI MIRA COLLEGE OF ENGINEERING AND TECHNOLOGY **Department:**ECE **Date of Submission:26/04/2025**

# ProblemStatement

Stock price prediction has always been a critical challenge in financial markets due to the volatile and non-linear nature of stock behavior. Traditional statistical models often fail to capture complex patterns in time series data. This project aims to leverage artificial intelligence, specifically deep learning models, to analyze stock market trends and predict future prices using historical time series data. The goal is to make informed predictions that can assist investors and traders in decision-making.

# Objectives of the Project

The main objective of this project is to build an intelligent stock price prediction system using AI-driven time series analysis techniques. The aim is to develop a model that can analyze historical stock market data and forecast future price trends with improved accuracy. Specifically, the project seeks to apply deep learning algorithms such as LSTM to understand the sequential nature of stock data.

These models are designed to handle temporal dependencies, which are essential when predicting time series data like stock prices. The project will also compare the performance of AI-based models with traditional methods like ARIMA to demonstrate the added value of deep learning in financial forecasting. Key outcomes include a well-trained and evaluated predictive model, insights into which features and algorithms provide the best predictive accuracy, and a clear interpretation of results through visualizations. The ultimate goal is to empower investors and traders with better forecasting tools that can support decision.

# Scope of the Project

The scope of this project involves the development of an AI-based system to predict stock prices using time series data. The project is centered around using deep learning models, particularly Long Short-Term Memory (LSTM) networks, to capture sequential patterns and dependencies in historical stock data. Key features to be analyzed include daily open, high, low, close (OHLC) prices, volume, and derived technical indicators like moving averages and RSI. Data will be sourced from publicly available APIs such as Yahoo Finance, ensuring accessibility and reproducibility.

The analysis will be confined to selected stocks over a fixed historical period. Traditional models like ARIMA will be implemented as benchmarks to evaluate the improvement brought by AI-driven methods. The models will be evaluated using RMSE, MAE, and other relevant metrics.

# Data Sources

For this project, the primary data source is the Yahoo Finance API, accessed programmatically using the Python library yfinance. This source provides comprehensive and up-to-date stock market data for a wide range of companies across various sectors. The dataset includes essential features such as the stock’s opening price, closing price, highest and lowest price of the day, traded volume, and the corresponding timestamps.

This data is crucial for time series modeling and provides the historical foundation upon which the prediction model will be trained. Since Yahoo Finance is a public platform, the data is freely accessible, making it ideal for academic and exploratory research. The dataset is considered dynamic, as it can be updated in real time, allowing the model to be retrained periodically to maintain accuracy with recent market conditions.

# High-Level Methodology

## The stock market is dynamic and influenced by numerous factors. The primary objective of this project is to leverage AI and time series analysis techniques to predict future stock prices.

* + **DataCollection**–Yahoo Finance: Free API for daily stock data.

Alpha Vantage/Quandl: APIs with additional economic and financial indicators.

* + **DataCleaning**–In the data cleaning phase, one of the key steps is identifying and handling missing values. This involves checking for null or NaN entries in important columns such as Open, High, Low, Close, and Volume. Techniques like .isnull().sum() in Python are used to quantify missing data.

Once identified, missing values can be addressed by forward-fill, backward-fill, interpolation, or dropping rows, depending on the extent and importance of the missing data

* + **Exploratory Data Analysis (EDA)** – EDA reveals patterns and guides feature selection

Trend and Seasonality: Use line plots, decomposition (STL).

Stationarity: ADF test and differencing to check if the data has constant mean and variance .Volatility Analysis-Use rolling standard deviation.

Correlation: Pearson/Spearman to detect relationships among indicators.

EDA helps understand whether the data needs transformation .

* + **FeatureEngineering**–Feature engineering is a crucial step in time series stock prediction. It involves transforming raw data into meaningful inputs that enhance a model’s ability to detect patterns and make accurate predictions
  + **Model Building** – In the model building phase, multiple experiments are conducted using a variety of models to predict stock prices. Baseline models like Naïve Forecast and Moving Averages provide reference points for performance.Statistical models such as ARIMA and SARIMA are explored for capturing linear temporal dependencies.

Machine learning models like Random Forest, XGBoost, and Support Vector Regression are employed to capture non-linear relationships in the data

* + **ModelEvaluation**–Model evaluation is conducted using metrics such as MAE, RMSE, and MAPE to quantify prediction errors, while R-squared (R²) assesses the model's explanatory power.

Additionally, directional accuracy is used to evaluate the model's ability to correctly predict the movement (up or down) of stock prices, which is particularly valuable for making trading decisions. These metrics ensure a balanced and comprehensive performance assessment.

* + **Visualization&Interpretation**–Key findings, insights, and predictions will be visualized using a combination of line plots, bar charts, and heatmaps. Line plots will compare actual vs. predicted stock prices over time to show model accuracy.

Bar charts will highlight error metrics across different models for easy comparison. Feature importance plots and correlation heatmaps will help interpret which indicators influenced predictions most, aiding in decision-making and model explainability.

* + **Deployment** – This project is primarily research-focused and aims to explore the effectiveness of AI-driven models for stock price prediction. While deployment is not the main objective, a basic prototype may be developed using Streamlit or Flask to demonstrate real-time predictions.

# Existing system

## Several existing systems leverage AI and time series analysis for stock price prediction, using a range of powerful tools and technologies. Platforms like QuantConnect and Alpaca enable algorithmic trading with support for Python and C#, utilizing libraries such as pandas, TensorFlow, Keras, and TA-Lib. Numerai crowdsources machine learning models using Python and R with tools like scikit-learn and xgboost.

## Research efforts like DeepMind explore advanced models using PyTorch and attention mechanisms. Many developers also build custom pipelines using Yahoo Finance APIs, combining ARIMA, Prophet, and deep learning frameworks .

# Proposed System

# **The proposed system for AI-driven stock price prediction using time series analysis will be developed using Python as the core programming language due to its versatility and strong ecosystem. Key tools and platforms include Jupyter Notebook for prototyping, GitHub for version control, and yFinance or Alpha Vantage APIs for historical stock data retrieval.**

# **The system will utilize pandas and numpy for data preprocessing, seaborn for visualization, and TA-Lib for technical indicator calculations. For modeling, TensorFlow, Keras, and PyTorch will be used to build LSTM and Transformer-based deep learning models.**

# ToolsandTechnologies

# **The proposed system leverages a powerful stack of tools and technologies tailored for AI-driven stock price prediction. Python serves as the primary programming language due to its strong support for data science and machine learning. Jupyter Notebook is used for model development and testing.**

* + **ProgrammingLanguage**–The main programming language used in the proposed system is Python. It is widely adopted in the fields of data science, machine learning, and financial analytics due to its simplicity, readability, and a rich ecosystem of libraries and tools specifically designed for data processing, modeling, and visualization.
  + **Notebook/IDE**–The proposed system is developed using Jupyter Notebook, a widely used interactive development environment ideal for data analysis and machine learning. It allows for real-time code execution, visualization, and documentation in a single interface.
  + **Libraries** –

pandas – Data manipulation and analysis

numpy – Numerical computing

matplotlib – Data visualization

seaborn – Statistical data visualization

TA-Lib – Technical analysis indicators

* + **Optional Tools for Deployment** – Flask – Lightweight web framework for creating APIs

FastAPI – Modern, high-performance framework for building RESTful APIs

Streamlit – Tool for building interactive data science web apps

Docker – Containerization tool for deploying the system in any environment

Heroku / Render / AWS EC2 – Cloud platforms for hosting the application

GitHub Actions – For CI/CD automation and deployment workflows

# TeamMembersandRoles

HEMAMALINI DEVEDIRAN –Team leader

HARIPRIYA JOYHILINGAM- Data collection and processing lead

HARIHARAN.C- visualization and reporting lead

IRSHAD AHMED.P- Model development and evaluation lead

GURUMOORTHY.G- Deployment