**Phase-1 Submission**

**Student Name:** DHARSHINI N

**Register Number:** 712523205019

**Institution:** PPG INSTITUTE OF TECHNOLOGY

**Department:** BTECH INFORMATION TECHNOLOGY

**Date of Submission:**04/05/2025

**1.Problem Statement**

* *Stock market price prediction is one of the most challenging tasks in financial analytics due to the market's volatile and non-linear behavior. Investors and analysts seek more accurate forecasting models to make informed decisions. This project addresses the need for reliable stock price prediction using AI-driven methods and time series analysis to assist traders, investors, and financial institutions.*

**2.Objectives of the Project**

* *To predict the future stock prices of a publicly traded company using historical data.*
* *To develop an LSTM-based deep learning model for time series forecasting.*
* *To analyze historical trends and evaluate model accuracy using appropriate metrics.*
* *To visualize predictions against actual data to assess model performance.*

**3.Scope of the Project**

* ***Features Analyzed****: Historical stock prices (close price), time-based sequences.*
* ***Model Scope****: Focused on LSTM neural networks for sequential prediction.*
* ***Constraints****: Uses only historical price data (no sentiment/news analysis).Model not deployed to a live system (not integrated with a brokerage API). Data source limited to Yahoo Finance via yfinance API.*

**4.Data Sources**

* ***Source****: Yahoo Finance (via yfinance Python API)*
* ***Dataset Type****: Public and dynamically updated in real-time.*
* ***Stock Example****: Apple Inc. (AAPL), with data from 2015 to 2024*.

**5.High-Level Methodology**

* ***Data Collection:*** *Collected historical stock price data using yfinance API for selected stock symbols.*
* ***Data Cleaning:*** *Checked for and handled missing/null values. Converted date columns and ensured proper time series ordering.*
* ***Exploratory Data Analysis (EDA):*** *Used line plots to visualize stock price trends. Analyzed trends, seasonality, and volatility.*
* **Feature Engineering:** *Created time-based sequences (60-day window) for LSTM input. Normalized data using Min Max Scaler for better model performance*.
* **Model Building:** *Built an LSTM model using Keras with two stacked LSTM layers and one dense output layer. Chosen due to LSTM's effectiveness in time-dependent pattern learning.*
* **Model Evaluation:** *Used Mean Squared Error (MSE) to evaluate model loss. Compared predicted prices to actual prices visually.*
* **Visualization & Interpretation:** *Plotted predicted vs. actual stock prices using matplotlib. Interpreted trends and checked prediction alignment*.
* **Deployment:** *For this phase, deployment is not included .Future deployment could be in Streamlit or a web dashboard*.

**6.Tools and Technologies**

* ***Programming Language****: Python*
* ***Notebook/IDE****: Google Colab / Jupyter Notebook*
* ***Libraries:*** *Data Processing: pandas numpy*
* ***Visualization****: matplotlib, seaborn*
* ***Modeling****: keras, tensorflow, scikit-learn*
* ***Data Source****: yfinance*
* ***Optional Tools for Deployment****: Streamlit (planned for later**phase)*

**7.Team Members and Roles**

|  |  |  |
| --- | --- | --- |
| **NAME** | **ROLL** | **ROLL DESCRIPTION** |
| **DHARSHINI N** | **TEAM LEADER** | **Oversees project milestones, manages timelines and coordinates communication.** |
| **DHARANI A** | **TEAM MEMBER** | **Leads model development using time series analysis and machine learning techniques** |
| **KISHORE P** | **TEAM MEMBER** | **Responsible pipeline setup data cleaning and transformation** |
| **LAKSHIN S** | **TEAM MEMBER** | **Provides domain expertise stock trends and validates predictions** |
| **ASHOK KUMAR R** | **TEAM MEMBER** | **Builds the server-side logic and integrates the ML model into a scalable API** |