Hooghly Engineering & Technology College

Vivekananda Road, Pipulpati, Hooghly-712103

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Research Methodology

(PROJ-CS601)

3RD YEAR

6TH SEMESTER

TEAM MEMBERS

Arkodip Das (17600121022)

Subhadip Samanta (17600121065)

Sandeep Mondal (17600121113)

Kousik Biswas (17600121063)

PROJECT TITLE :- Stock Market Scraper : Harnessing Web Scraping techniques to research about market insights.

TOOLS AND TECHNOLOGIES

Programming Language: Python

Database(Relational): SQL

Web Scraping Library: yfinance

Data Analysis Libraries: pandas, matplotlib

Additional Libraries: Requests (for website interaction)

PROJECT DESCRIPTION :- The project revolves around the development of a web scraping tool capable of extracting textual data from online sources related to the stock market, investment forums, and open trading platforms. We will utilize popular web scraping libraries like **yfinance**, **Pandas and Matplotlib** in **Python** to gather data efficiently and robustly.

Once the data is collected, we will employ NLP techniques to preprocess and analyse the textual content. This involves tokenization, stemming, and sentiment analysis to extract meaningful insights from the gathered text. The processed data will be structured in the form of a tree, which will then be parsed to extract targeted data from a specific webpage.

OBJECTIVES

a) To develop a web scraper to extract relevant stock market data from financial websites.

b) Target specific data points such as stock prices, historical data, news articles, and analyst ratings.

c) Organize the scraped data into a structured format for further analysis.

d) Explore potential applications of the scraped data for market research.

EXPECTED OUTCOME :-

A functional stock market scraper that can extract and organize relevant data from financial websites.

The project will demonstrate the capabilities of web scraping for market research by:

a) Providing a readily available dataset for further analysis.

b) Highlighting potential applications like sentiment analysis of news articles or identifying undervalued stocks based on historical data.

c) Proper visualization of the structured raw data as given below of a famous company(Apple Inc.).