

|  |  |
| --- | --- |
| Date: 14/06/2024 | FSD(Functional Specification Documentation) |

**FUNCTIONAL SPECIFICATION DOCUMENT**

**Document Review**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Department / Program | Stakeholder Name | Designation | Date | Signature |
| Market Analysis Project Mentor | Abhinav Kumar | Digital Engineer (Data Scientist) |  |  |
| Market Analysis Project Team | Team name | Intern (AI ML Developer)  Intern (Frontend)  Intern (Backend) |  |  |
|  |  |  |  |  |

**Disclaimer**: - Document sign-off indicates business user acknowledgement to the use case design. Future enhancements will be enabled post go-live (if needed) to ensure continuous improvement within the scope of this document.

**INTRODUCTION**

**Purpose of the document**

This document outlines the use case requirements, data entities, visualization mock-ups, and data sources necessary to achieve the objectives of the market sentiment analysis project. It describes how various analyses will be conducted on data collected from operational use cases to understand and interpret market sentiments. The primary focus is on analysing financial news articles and other public data sources to gauge the sentiment toward specific companies or sectors. This includes assessing sentiment from news, blogs, and social media about clients and partners to inform strategic business decisions.

The document is designed for business stakeholders, providing a clear understanding of the functional workflows being implemented to meet the project's requirements. These workflows encompass sentiment classification and the correlation of these sentiments with market movements. The insights from this document will serve as a foundational input to the technical specification document, which will detail the technical aspects necessary for implementing the sentiment analysis use cases.

Additionally, this document identifies all assumptions and dependencies related to the analytical processes involved in market sentiment analysis. This includes data gathering methodologies, the types of data to be analysed, and the integration of sentiment analysis findings with existing business strategies and decision-making processes.

**SCOPE**

* In Scope

The scope of the functional specification document includes:

* Objective and requirements of the use case with mock-ups.
* The business logics, rules and flowcharts involved in the use case implementation
* Visualizations with details about the objective of the visualization, target stakeholders and control actions
* The Data Model that is required to be stored in data lake or extracted from the source systems
* Data Requirements for the use case implementation
* Data latency details
* Out of Scope

This functional specification document does not cover:

* The technical implementation details including the details around the interface and integration specifications
* Non-functional aspects of the use case implementation
* Data Quality aspects of the data related to the use case implementation.

**FUNCTIONAL SPECIFICATION**

Use Case Summary

The primary objective of these analytical use cases is to monitor and forecast market sentiments based on financial news and public data sources. This involves:

1. Monitoring Market Sentiments: Continuously track and analyse the sentiment expressed in financial news articles and public data sources, providing a real-time understanding of how companies and sectors are perceived in the market.

2. Forecasting Market Movements: Utilize sentiment data to predict potential market movements, identifying trends and signals that can inform investment strategies and decision-making processes.

3. Recognizing Market Anomalies: Keep an eye out for odd sentiment trends or notable changes in public opinion that may point to hidden market possibilities or problems, such new trends or hazards.

4. Analysing Sentiment Impact: Link sentiment scores with financial measures and indicators to assess how sentiment affects market performance and corporate stock prices.

5. Segmenting Market Data: To offer focused insights for strategic planning and company optimization, analyse sentiments across several market segments, including firms, industries, or geographic areas.

This detailed focus will guide the development and implementation of the market sentiment analysis project, ensuring that the analytical use cases align with the business objectives and deliver actionable insights into market dynamics.

This section covers functional details relating to this use case scenario.

**Market Analysis**

The primary objective of these analytical use cases is to monitor and forecast market sentiments based on financial news and public data sources.

Use Case Description

In the market sentiment analysis project, we aim to analyse and interpret sentiments from financial news, social media, and public forums to gain insights into market perceptions of companies and sectors. Using advanced sentiment analysis, we will classify these texts as positive, negative, or neutral and correlate them with stock prices and market trends. The data will be visualized on dynamic dashboards, allowing stakeholders to track real-time sentiment changes and forecast future trends across different time horizons and regions. Additionally, integrating weather data will provide context for sectors influenced by weather patterns, helping to refine the analysis. This project will deliver actionable insights, enabling better-informed decision-making in market strategy and investment.

Stakeholders

* Market Analysis Team
* Data Scientists
* Private VCC Firm
* Finance Department

Notification

* Currently, no notification is created. Depending on requirements, further changes can be made.

**Data and System Parameters**

|  |  |  |
| --- | --- | --- |
| S.No | Variable | Description |
| 1 | Stock Name and Domain | Specific stocks and their sectors targeted for analysis. |
| 2 | Sentiment Analysis Model | LLM parameters and classification thresholds. |
| 3 | Historical Data | Historical stock prices and volume data from financial databases. |
| 4 | Model Selection Criteria | Accuracy, speed, and suitability of different machine learning models. |
| 5 | Prediction Horizon | Time frames for which predictions are made (e.g., daily, weekly, monthly). |
| 6 | Visualization Parameters | Dashboard settings, types of charts, and user interaction elements. |

Feature Descriptions for Market Sentiment Analysis

**Stock Price Volatility**

Description: Stock Price Volatility measures how much a stock's price fluctuates over a specific period. It is typically expressed as the standard deviation of returns. Higher volatility indicates greater price changes and uncertainty, suggesting potential for higher risk and reward.

Usage: In sentiment analysis, volatility is crucial for assessing the stability of a stock and predicting its future movements. High volatility may correlate with significant news or sentiment changes, impacting investor behaviour and market trends.

**Trading Volume Spike**

Description: A Trading Volume Spike refers to a sudden, significant increase in the number of shares traded compared to the stock’s average trading volume. This can signal increased interest or activity in the stock, often due to major news, earnings reports, or market events.

Usage: Detecting volume spikes helps identify periods of high investor interest or potential price changes. It can also highlight unusual trading activity that may be driven by sentiment shifts, providing insights into market reactions and potential future movements.

**Insider Trading Activity**

Description: Insider Trading Activity tracks the buying or selling of a company's shares by its executives, directors, or other insiders who have access to non-public information. Significant insider transactions are often seen as a signal of the insiders’ confidence (or lack thereof) in the company’s future performance.

Usage: Monitoring insider trading can offer valuable clues about a company’s health and prospects. For example, substantial insider buying may indicate positive sentiment among those most familiar with the company, potentially influencing broader market sentiment and investor decisions.

**News Sentiment Score**

Description: The News Sentiment Score quantifies the overall sentiment (positive, negative, or neutral) expressed in news articles about a company or sector. Using natural language processing and sentiment analysis algorithms, this score provides a numerical representation of how favourable or unfavourable recent news is.

Usage: This score helps investors understand the media's portrayal of a company and anticipate its potential impact on stock prices. Consistently positive news sentiment may boost investor confidence and stock performance, while negative sentiment can trigger selloffs and price declines.

**Number of News Mentions**

Description: The Number of News Mentions counts how frequently a company or topic appears in news articles over a given period. An increase in mentions typically correlates with heightened attention or significant developments related to the company.

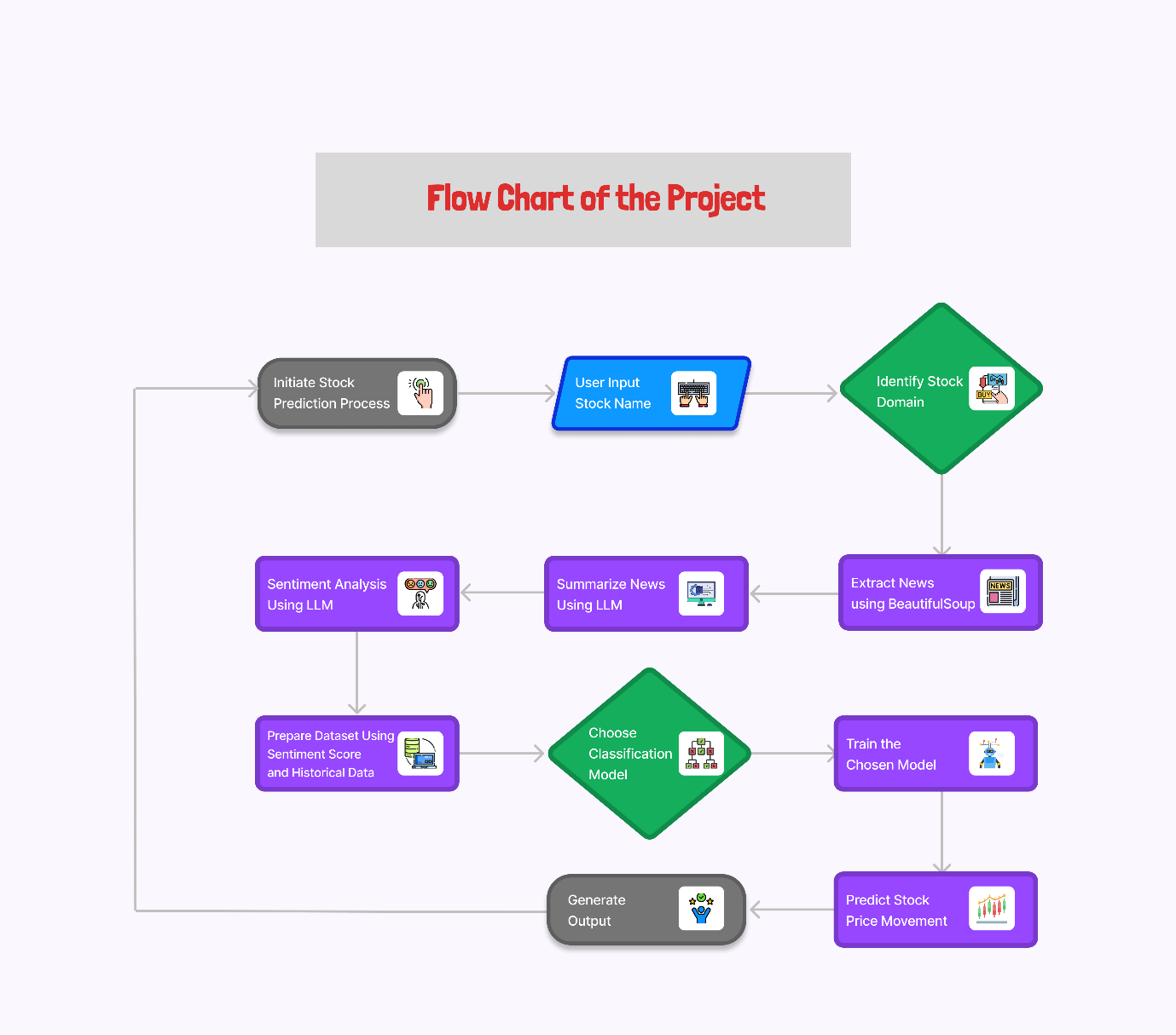
Usage: Tracking the frequency of news mentions helps gauge the level of public and media attention. A surge in mentions often aligns with major events or announcements that could drive sentiment and influence market behaviour. High media visibility can also amplify the impact of sentiment on stock prices.

**Data Integration**

This section provides the data integration details for the Market Analysis

|  |  |
| --- | --- |
| Stock Market Data | * Historical Prices: Data on past stock prices for analysis. * Real-time Prices: Up-to-date stock prices for live predictions. * Trading Volumes: Information on share volumes traded, crucial for detecting spikes. |
| News and Media | * News Articles: Continuous stream of news for sentiment analysis. * Press Releases: Company announcements impacting sentiment. * Economic News: Broader market and economic news influencing stock movements. |
| Sentiment Analysis | * Sentiment Scores: Quantitative sentiment from news and social media. * Sentiment Volatility: Metrics on how sentiment scores fluctuate over time. |
| Web Scraping | * Beautiful Soup: Used for extracting news articles and data from websites. |
| Large Language Models  (LLM) Tools | * Summarization Tools (LLMs): Large Language Models summarizing news content. * Sentiment Analysis Tools: LLMs tools assessing the sentiment of text data. |

**Flow Chart**



**Business Rules**

* Real-time Stock Market Data Monitoring
* Financial News Summarization and Classification
* Stock Price Movement Prediction
* Sentiment Analysis from News and Social Media
* Interactive Data Visualization on Dashboards
* Sentiment Score Calculation and Monitoring

**Market Sentiment Analysis - Summary**

In the Market Sentiment Analysis project, data from stock markets, financial news, and social media is collected and analyzed to predict stock price movements.

**Data Collection:**

* Stock Market Data: Real-time and historical stock prices and trading volumes are gathered from financial data providers.
* News and social media: Articles and posts related to stocks are collected using web scraping.

**Data Processing:**

* Sentiment Analysis: LLMs assess the sentiment of collected news and social media content, classifying them as positive or negative.
* Integration: Sentiment scores and market data are merged into a unified dataset and stored in a data lake.

**Prediction:**

* Model Training: Machine learning models are trained on the integrated dataset to predict future stock price movements.
* Optimization: Models are fine-tuned for better accuracy.

**Visualization:**

* Dashboards: Insights from the data and predictions are displayed on interactive dashboards, allowing stakeholders to monitor stock trends and sentiment changes in real-time.

**Data Requirements**

In this integrated approach, historical stock price data is retrieved from Yahoo Finance, serving as the foundation for training machine learning models to discern patterns and trends in stock movements over time. Real-time stock prices, obtained from the same or similar sources, are then used to predict future stock movements. Concurrently, news articles are scraped from various platforms using Beautiful Soup, extracting key information like article text. These articles are subjected to sentiment analysis via LLMs, assigning sentiment scores (positive, negative) to each article. Through integration and analysis of historical data, real-time prices, and sentiment scores, investors gain insights into the impact of news sentiment on stock prices, enabling them to make more informed trading decisions and refine investment strategies accordingly.

**Web Design**

This section entails details about visualization for all three analytical use cases and covers following points.

* Analysis of user’s input.
* Processing will be done.
* Output will be showed to user on the screen.

**Note**: - Once UI development begins, a virtual walk through of test dashboards shall be provided and further recommendations will be covered in revised FSD’s.

**Filters, Search selection:**

* **Extraction of News Data via Beautiful Soup:** Extracting news data through Beautiful Soup involves using the requests library to make HTTP requests to web pages, and Beautiful Soup to parse and extract relevant data such as headlines, article text, publication dates, and authors. This process typically involves scripting in Python to interact with web content, locate HTML elements on the page using techniques like CSS selectors, and extract the desired information for further analysis or use.
* **Summarisation:** Summarization of news data will be done by pre-trained language models like BART. This process utilizes the extracted text as input for the model, which then generates a shorter version of the content while retaining the key information and main points. The summarization output provides a condensed overview of the news articles, enabling users to quickly grasp the main ideas without having to read the entire text, which will be used for further processing.
* **Sentiment Analysis via LLM:** Sentiment analysis using LLM involves applying pre-trained models like RoBERTa to classify the sentiment of the news articles as positive, negative, or neutral. The process includes encoding the text, predicting sentiment scores, and calculating a compound sentiment score based on the probabilities of each sentiment class.

**ABBREVIATIONS**

The following table contains the list of abbreviations used throughout the document

|  |  |
| --- | --- |
| Abbreviation | Full Form |
| AI | Artificial Intelligence |
| ML | Machine Learning |
| LLM | Large Language Model |
| NLP | Natural Language Processing |

**Glossary**

The following sections describes terminologies often used in the document.

**System Alert** – An event triggered by the system to notify users of significant occurrences or changes in data or system status that may require attention or action.

**Notification** – A workflow which informs users about specific events or updates related to the system, such as changes in stock prices, sentiment analysis results, or other relevant data points, often delivered through emails, messages, or in-app alerts.

**Beautiful Soup** – A Python library used for web scraping purposes to pull the data out of HTML and XML files. It creates parse trees from page source code that can be used to extract data easily.

**Large Language Model (LLM)** – A type of artificial intelligence model trained on vast amounts of text data to understand and generate human-like language. Examples include BART and GPT.

**Sentiment Analysis** – The process of using natural language processing (NLP) to determine the sentiment expressed in a piece of text, typically classified as positive, negative, or neutral.

**Selenium WebDriver** – A tool used for automating web application testing to verify that it performs expectedly. It can also be used to automate web tasks such as scraping dynamic content.

**BART** – A denoising autoencoder for pretraining sequence-to-sequence models, used for tasks like text summarization and translation. The acronym stands for Bidirectional and Auto-Regressive Transformers.

**RoBERTa** – A robustly optimized version of BERT (Bidirectional Encoder Representations from Transformers) pretraining approach, used for tasks such as sentiment analysis.

**APPENDIX**

* **Google Scraping Methodology**: The Google Scraping Methodology involves the following steps:

1. **Finding URLs**: Construct a Google search query URL based on a given keyword and use the requests library to fetch the search results. Beautiful Soup is used to parse the HTML content and extract the URLs of news articles.
2. **Scraping Articles**: Extract news data through Beautiful Soup by navigating web pages, locating specific elements containing news content, and extracting relevant data such as headlines, article text, publication dates, and authors.
3. **Summarization**: Summarize the extracted text using a pre-trained BART model, which generates a shorter version of the content while retaining key information and main points.

* **Sentiment Analysis Using LLM**: The sentiment analysis process involves using a pre-trained sentiment analysis model to classify the sentiment of summarized news articles. The steps include:

1. **Loading the Sentiment Model**: Load a pre-trained RoBERTa model and tokenizer for sentiment analysis.
2. **Finding Sentiment**: Encode the text and predict sentiment scores using the loaded model. Calculate a compound sentiment score based on the probabilities of each sentiment class.

**Document References**

The following table contains the use case, use case workshop summary, clarifications sheet, data mapping sheet references.

|  |  |
| --- | --- |
| Document Title | Location |
| Use Case | |  | | --- | | Fsd-draft.docx |  |  | | --- | |  | |
| |  | | --- | | Sentiment Analysis |  |  | | --- | |  | | See Appendix for detailed methodology |
| Google Scraping Methodology | |  | | --- | | See Appendix for detailed methodology |  |  | | --- | |  | |