Logo

Description automatically generated

**BE-CSE (Artificial Intelligence)**

**PROJECT SYNOPIS**

**Utilizing Natural Language Processing for Stock Market Prediction**

**Yuvraj (2110993863)**

**Aniket Thakur (2110993872)**

**CSE(AI) Group-1**

**CHITKARA UNIVERSITY INSITUTE OF ENGINEERING AND TECHNOLOGY**

**CHITKARA UNIVERSITY, RAJPURA**

**Introduction:** In the dynamic landscape of financial markets, accurate prediction of stock price movements has long been a goal pursued by investors, analysts, and researchers alike. The challenge lies in the complexity of factors influencing these price fluctuations, ranging from macroeconomic indicators to company-specific news and sentiments. Traditional quantitative models and technical analysis have been extensively used for stock prediction, but the advent of Natural Language Processing (NLP) techniques has introduced a novel approach that harnesses the power of textual data to enhance predictive capabilities.

**Business problem:** This project aims to predict stock market movements using Natural Language Processing (NLP) techniques. By extracting insights from financial news articles and social media posts, the project seeks to enhance the accuracy of stock price movement forecasts. The project involves data collection, text preprocessing, sentiment analysis, feature engineering, and predictive modelling.

**Literature Review:** This literature review aims to provide a comprehensive analysis of the existing research at the confluence of NLP and stock market prediction. We delve into the methodologies, models, datasets, and evaluation metrics employed by researchers to explore the predictive power of textual data. By synthesizing insights from a range of studies, this review aims to highlight the advancements made, the challenges faced, and the potential future directions in this evolving field.

**Objectives:**

1. **Data Collection**: Gather a comprehensive dataset of financial news articles and relevant social media posts related to target stocks.

2. **Text Preprocessing**: Clean and preprocess the text data, including tasks like tokenization, stop-word removal, and stemming/lemmatization.

3. **Sentiment Analysis**: Apply sentiment analysis techniques to gauge the sentiment expressed in the collected text data. Determine whether the sentiment is positive, negative, or neutral.

4. **Feature Engineering**: Extract relevant features from the text data, such as sentiment scores, keyword frequencies, and temporal trends.

5. **Stock Price Movement Labels**: Define labels for stock price movements over specific time intervals, considering thresholds like percentage changes.

6. **Model Selection**: Choose appropriate machine learning algorithms (e.g., LSTM, GRU, Transformer-based models) to predict stock price movements using the engineered features.

7. **Model Training**: Train the selected model using historical data, incorporating the engineered features and sentiment analysis results.

8. **Evaluation Metrics**: Assess the model's performance using metrics like accuracy, precision, recall, F1-score, and mean squared error (MSE).

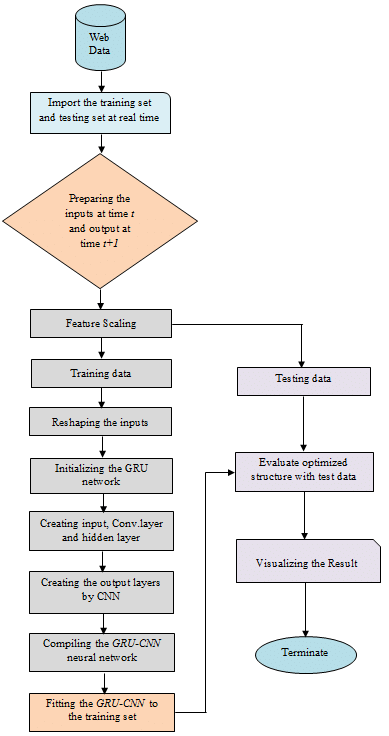
9. **Hyperparameter Tuning**: Fine-tune the model's hyperparameters to achieve optimal performance on the validation set.

10. **Prediction and Visualization**: Implement the trained model to make predictions on new data and visualize the predicted vs. actual stock price movements.

11. **Risk Analysis**: Evaluate the potential risks and limitations of the developed model, considering factors like data quality, generalizability, and market unpredictability.

12. **Conclusion and Future Work**: Summarize the project's findings, discuss its implications, and propose potential avenues for future enhancements.

**Internal Working:**



**Technologies to be used:**

* TensorFlow
* Keras
* NLP
* HTML, CSS ,Javascript

**Result**: By combining NLP techniques with stock market data, this project aims to offer a valuable tool for investors and financial analysts seeking more accurate stock price movement predictions. It also acknowledges the challenges inherent in predicting market fluctuations and highlights the potential impact of sentiment analysis on enhancing predictive models.