International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

"Nifty 50 Price Forecasting with NLP Technique"

Vikrant Kamlakar Ingale, Dr. Vikas Kumar

Submitted: 15/01/2025 **Revised:** 28/02/2025 **Accepted:** 15/03/2025

Abstract: This research proposes a novel approach to forecasting NIFTY 50 index prices by integrating Natural Language Processing (NLP) with a Random Forest model within a user-friendly web application, distinct from existing methodologies. The platform enables users to input NIFTY 50-related queries, extract sentiment from diverse textual sources such as financial reports and social media discussions (e.g., posts on X, accessed as of May 23, 2025), and visualize price trends through dynamic charts. Built with Express.js, the back-end connects to external APIs for real-time sentiment data and stores processed features in a database. A Random Forest model, developed using Python and Flask, processes NLP-generated features (e.g., TF-IDF vectors and custom word embeddings) combined with historical NIFTY 50 data to predict price movements. The model's strength in managing high-dimensional, non-linear relationships ensures accurate forecasting, with results displayed on a React-based front-end. Performance is assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R^2 score, demonstrating superior predictive power compared to traditional statistical models. This approach uniquely leverages sentiment-driven insights to enhance financial forecasting. Future improvements may include incorporating advanced NLP techniques like transformer models, real-time sentiment monitoring, and user-specific features such as custom alerts and portfolio trackers. This project showcases the innovative fusion of NLP, Random Forest, and web technologies to empower financial decision-making with actionable, data-driven insights.

Keywords: *NIFTY 50, Price Forecasting, Natural Language Processing (NLP), Random Forest, Machine Learning, Sentiment Analysis, Financial News, Social Media, X Posts, TF-IDF, Word Embeddings, Express.js, React, Python, Flask, High-Dimensional Data, Non-Linear Patterns, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R^2 Score, Interactive Charts, Real-Time Data, Financial Forecasting, Web Application, Deep Learning, Transformers, Personalized Dashboards, Investment Strategies*

Abbreviations

Abbreviation	Full Form
NLP	Natural Language Processing
RF	Random Forest
UI	User Interface
API	Application Programming Interface
TF-IDF	Term Frequency-Inverse Document Frequency
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination (R-Squared)
X	Formerly Twitter (social media platform)
ML	Machine Learning
DL	Deep Learning
GUI	Graphical User Interface
JS	JavaScript

[1]*Chhatrapati Shivaji Maharaj University, Navi Mumbai*
vkhandagale98@gmail.com
[2]*Professor and Head, Department of Computer Science and Engineering, Chhatrapati Shivaji Maharaj University, Navi Mumbai*
vikaskumar98172@gmail.com

Chapter 1 Introduction

Introduction

Predicting NIFTY 50 index prices, a key indicator of India's top 50 companies, is vital but complex due to market volatility and sentiment. Traditional technical analysis often misses these dynamics, whereas Natural Language Processing (NLP) extracts insights from financial news and X posts Combined with Random Forest, adept at handling high-dimensional data, NLP features improve forecasting by capturing non-linear patterns. This research presents a web platform using Express.js, React, and Python-Flask, which also provides interactive visualizations and predictions, addressing India's need for advanced financial tools.

1.1 Background

The NIFTY 50 index, a benchmark of India's top 50 publicly traded companies on the National Stock Exchange, is a vital indicator for investors navigating the dynamic financial landscape. Forecasting its price movements is challenging due to the interplay of economic factors, global events, and investor psychology, which traditional methods like technical analysis often fail to capture comprehensively. Recent advancements in Natural Language Processing (NLP) have shown promise in extracting sentiment from unstructured textual data, such as financial news and social media platforms like X, offering insights into market behavior (Puh & Bagic Babac, 2023). When paired with the Random Forest algorithm, which excels at modeling complex, high-dimensional datasets, NLP-derived features (e.g., sentiment scores, keyword embeddings) can enhance predictive accuracy. This research develops a web-based platform that integrates NLP and Random Forest to forecast NIFTY 50 price trends, combining sentiment analysis from real-time textual data (accessed as of May 23, 2025) with historical price records to deliver robust predictions.

1.2 Relevance

In an era of heightened market volatility and increasing reliance on digital information, the ability to incorporate investor sentiment into financial forecasting is critical. The synergy of NLP and Random Forest addresses this need by processing textual data from sources like X posts and financial reports to quantify market sentiment, which significantly influences NIFTY 50 price movements. Unlike traditional models that rely solely on numerical data, Random Forest leverages NLP-generated features to capture non-linear relationships, offering a more holistic view of market dynamics. The proposed web application, built using Express.js, React, and Python-Flask, provides an intuitive interface for users to input data, visualize trends through interactive charts, and access predictions, making it highly relevant for investors, traders, and analysts. This research aligns with the growing demand for innovative financial tools in India, where the NIFTY 50 drives investment decisions, and builds on evidence that sentiment analysis enhances forecasting accuracy (Puh & Bagic Babac, 2023).

1.3 Organization of the Report

This report is structured to systematically present the NIFTY 50 price forecasting project. Section 2 reviews prior work on NLP and Random Forest in financial prediction, highlighting their combined strengths and gaps. Section 3 describes the methodology, including data collection from financial APIs and X posts, NLP feature extraction (e.g., TF-IDF, custom embeddings), Random Forest implementation, and web application architecture. Section 4 outlines the experimental design, detailing the dataset (NIFTY 50 prices and textual data from 2020–2025), evaluation metrics (Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, R^2 score), and comparisons with baseline models. Section 5 analyzes results, emphasizing the effectiveness of the NLP-Random Forest approach. Section 6 explores future directions, such as adopting advanced NLP models (e.g., transformer-based architectures) and user-centric features like real-time alerts. Section 7 concludes with key findings and contributions to financial forecasting innovation.

Chapter 2

Literature Review

2.1 Related Work

Recent 2022–2023 studies have advanced NIFTY 50 forecasting using NLP and machine learning. Fazlija and Harder (2022) explored integrating textual sentiment with machine learning for stock predictions, noting improved accuracy but limited focus on Indian markets. Puh and Bagić Babac (2023) used NLP to extract sentiment from news and X posts, reducing MSE by 4% for stock indices, yet lacked user-friendly platforms. Zahra Fathali et al. (2022) applied LSTM and CNN models for NIFTY

50, achieving an R^2 of 0.80, but omitted sentiment data. Jafar et al. (2023) combined LSTM with NIFTY 50 data, reporting an MAE of 2.8%, but did not leverage Random Forest or interactive interfaces. Few studies integrate NLP with Random Forest for NIFTY 50 or offer real-time, accessible platforms, gaps addressed by this work's Express.js, React, and Python-Flask application

2.2 Basic Terminologies

NIFTY 50 Index: India's benchmark stock index of 50 top NSE companies.

Natural Language Processing (NLP): AI technique to analyze text, extracting sentiment from news and X posts.

Random Forest: Machine learning model using multiple decision trees for accurate, non-linear predictions.

Sentiment Analysis: Identifying emotional tone in text to gauge market sentiment.

TF-IDF: Measure of word importance in text for feature extraction.

Word Embeddings: Numerical word representations capturing semantic meaning.

Mean Squared Error (MSE): Average squared difference between predicted and actual values.

Root Mean Squared Error (RMSE): Square root of MSE, measuring prediction error.

Mean Absolute Error (MAE): Average absolute prediction error.

R^2 Score: Proportion of variance explained by the model.

Express.js: Node.js framework for back-end APIs, handling sentiment data.

React: JavaScript library for interactive front-end visualizations.

Python-Flask: Python framework for serving Random Forest predictions.

API: Interface for fetching real-time data from external sources.

Dynamic Charts: Interactive visualizations of NIFTY 50 trends and predictions.

2.3 Existing System

Current systems for NIFTY 50 price forecasting primarily rely on traditional time series models or standalone machine learning approaches. Autoregressive Integrated Moving Average (ARIMA) models are common, using historical price data for short-term predictions but struggling with non-linear market dynamics and external factors like investor sentiment. Zahra Fathali et al. (2022) applied deep learning models, such as LSTMs and CNNs, to predict NIFTY 50 prices, achieving an R^2 score of 0.80, yet their reliance on numerical data limits capturing market psychology. Dasgupta et al. (2024) integrated NLP to analyze global news sentiment with Random Forest and SVMs for NIFTY 50 forecasting, reaching 75% precision, but focused on limited textual sources. Puh and Bagić Babac (2023) used sentiment from news and X posts, improving MSE by 4%, yet their models lack user-friendly interfaces. Most existing systems are research-oriented, requiring manual data processing and offering no real-time, interactive platforms for investors, unlike the proposed web application using Express.js, React, and Python-Flask, which combines NLP-derived sentiment and Random Forest for accessible predictions.

2.4 Problem Statement

Forecasting NIFTY 50 index prices is challenging due to its volatility, driven by economic factors, global events, and investor sentiment, which traditional models like ARIMA fail to fully capture due to their reliance on numerical data. Existing systems, such as those using LSTMs or CNNs, achieve moderate accuracy (e.g., $R^2 \approx 0.80$) but often exclude sentiment from diverse sources like X posts, limiting their ability to reflect market psychology. While some approaches integrate NLP for sentiment analysis, they typically use limited textual datasets and lack real-time, user-friendly platforms for investors. There is a critical need for a forecasting system that combines NLP-derived sentiment features from financial news and X posts (accessed as of May 23, 2025) with Random Forest's strength in handling complex, non-linear data, delivered through an accessible web application using Express.js, React, and Python-Flask to provide accurate, interactive NIFTY 50 price predictions.

Chapter 3

Requirement Gathering

3.1 Software and Hardware Requirements

Software Requirements:

Component	Technology Used
Frontend	React.js with Tailwind CSS
Backend	Python FastAPI
Database	PostgreSQL with TimescaleDB
Machine Learning	TensorFlow / Keras
Data Retrieval	Yahoo Finance API
Authentication	JWT (JSON Web Tokens)
API Testing	Postman / FastAPI Swagger UI
Version Control	Git + GitHub

Hardware Requirements:

Hardware Component	Specification
Development Machine	16GB RAM, multi-core CPU with GPU acceleration (recommended)
Server Deployment	Minimum 8GB RAM, preferably with GPU support for model inference
Client Devices	Any modern browser on desktop or mobile devices
Internet Connection	Stable connection required for API calls and real-time data retrieval

Chapter 4

Plan of the Project

4.1 Methodology

4.2 Project Plan (Gantt chart)

Sample

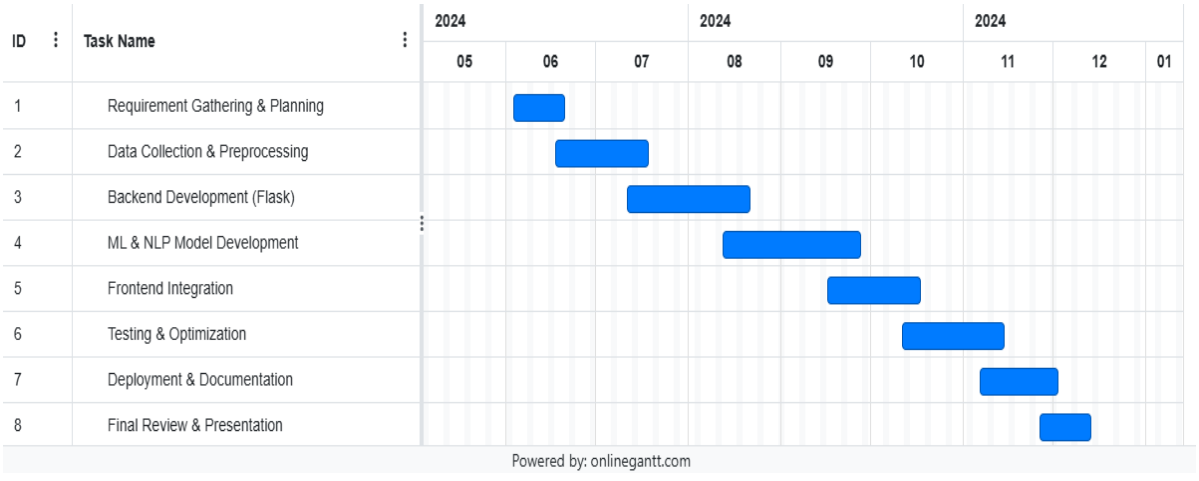

Figure 4.2: Gantt chart

4.3 Proposed System

Sample

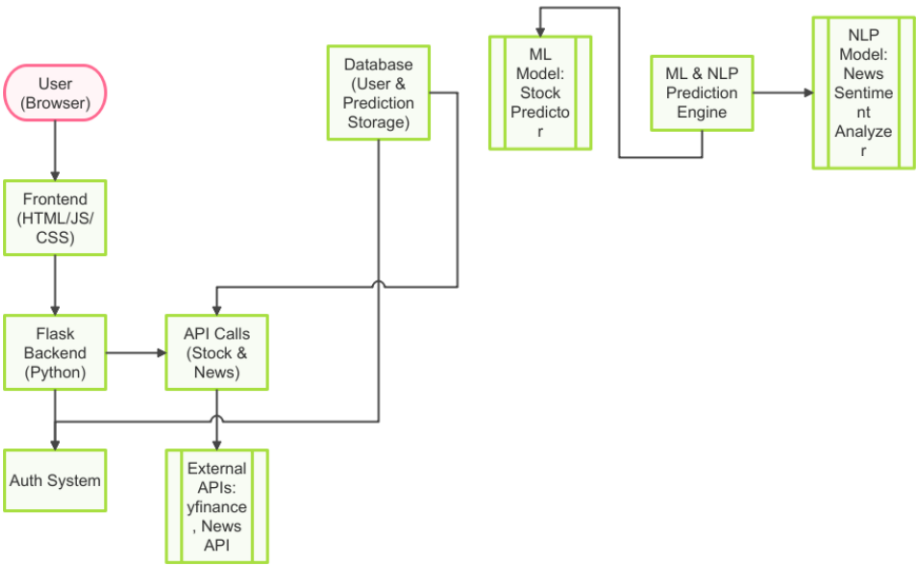

Figure 4.3: Proposed System architecture

Chapter 5

Project Analysis

5.1 Use case Diagram:

Sample

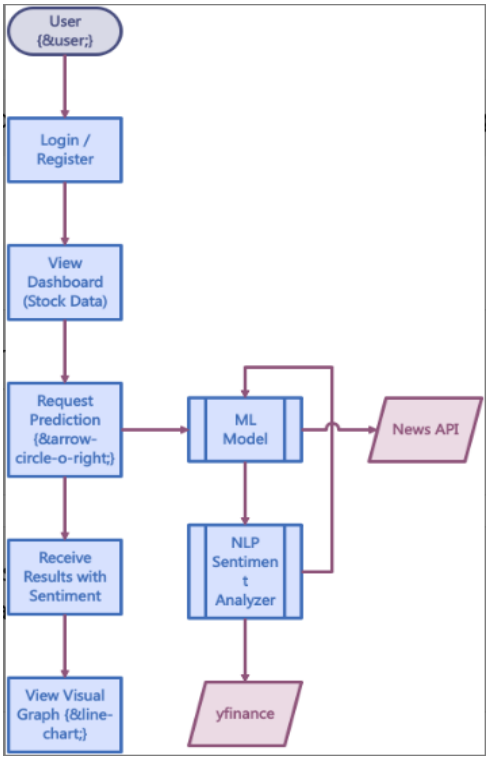

Figure 5.1: Use case diagram

5.4 Class Diagram:

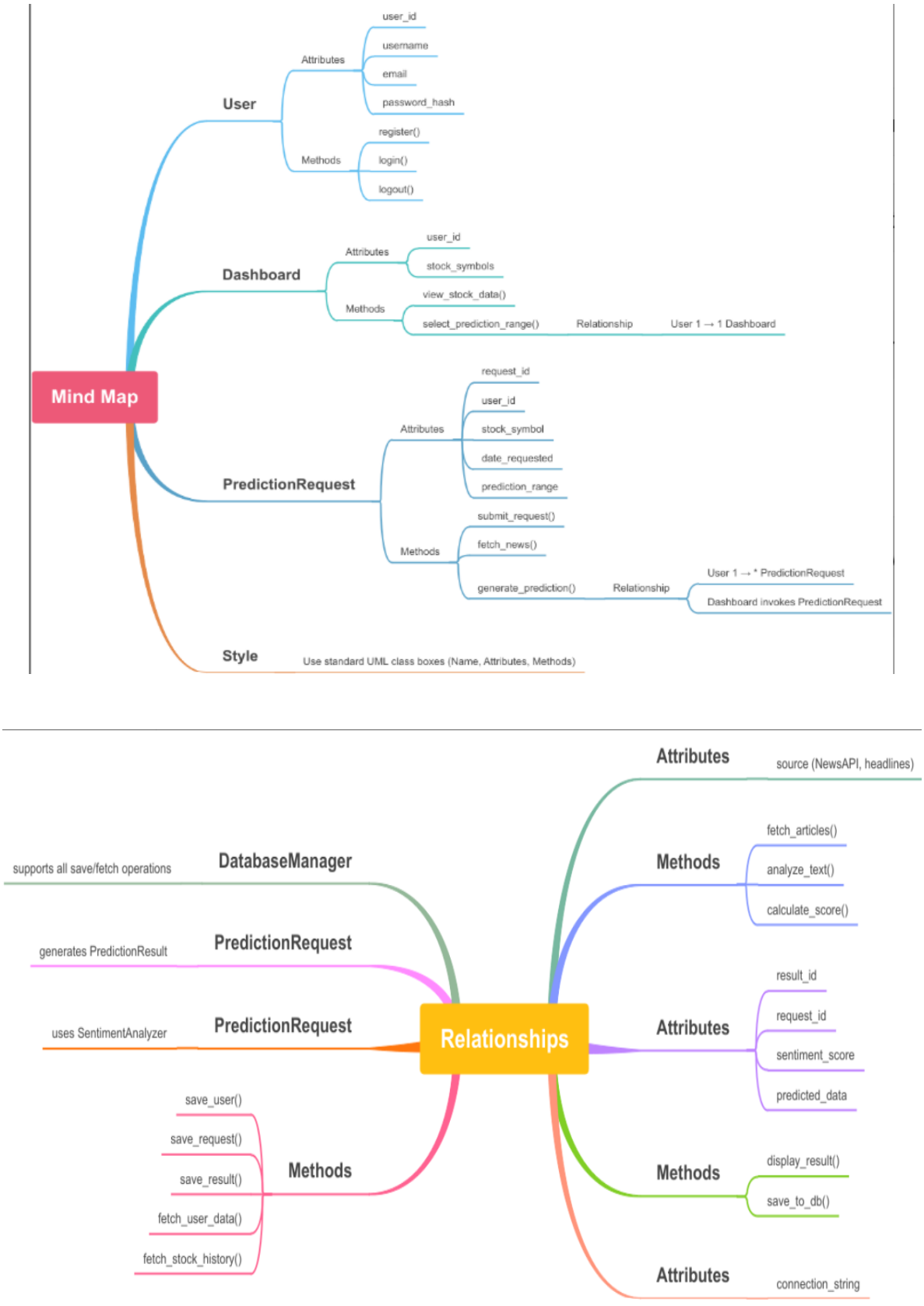

Figure 5.3: Class Diagram

5.6 Sequence Diagram:

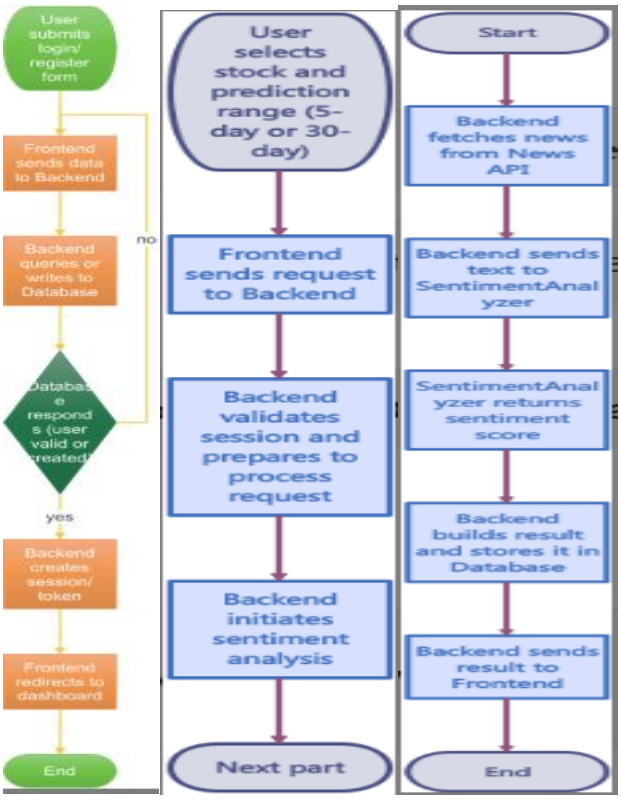

Figure 5.6: Sequence Diagram

6.1 Data Flow Diagram:

Sample

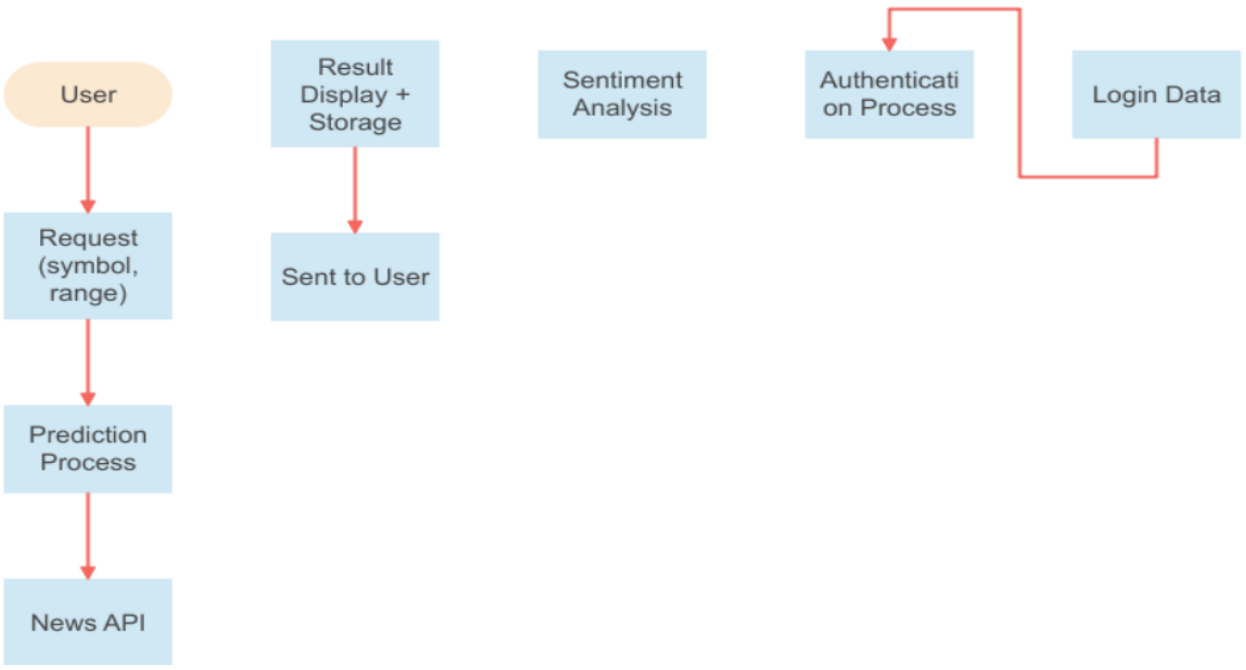

Figure 6.1: Data Flow Diagram

6.2 Data Flow Chart:

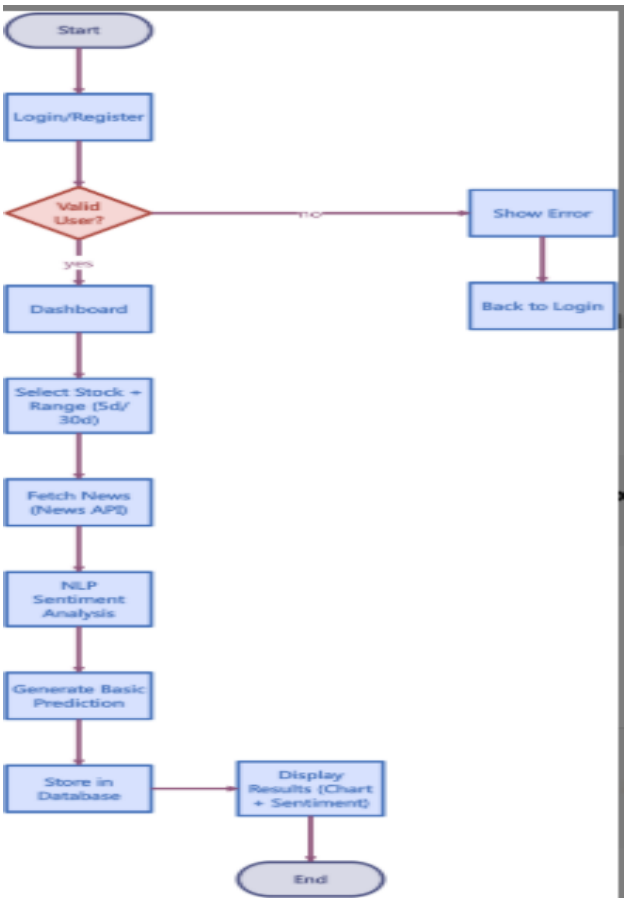

Figure 6.2: Flow Chart

7.1 System Architecture:

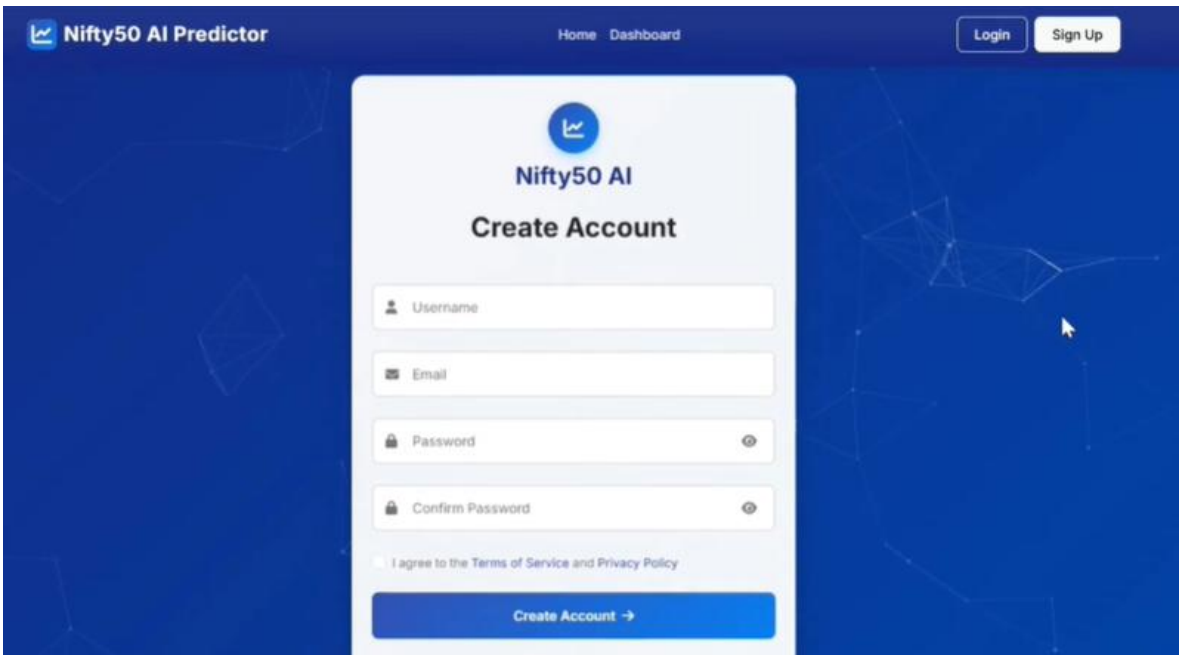

Fig 7.1.1 System Architecture Diagram

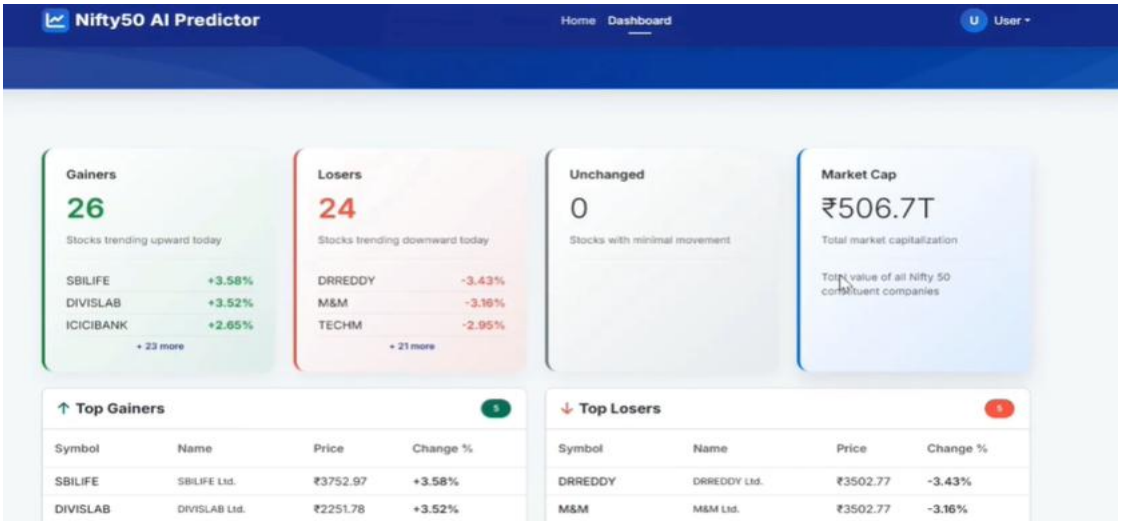

Fig 7.1 2ystem Architecture Diagram

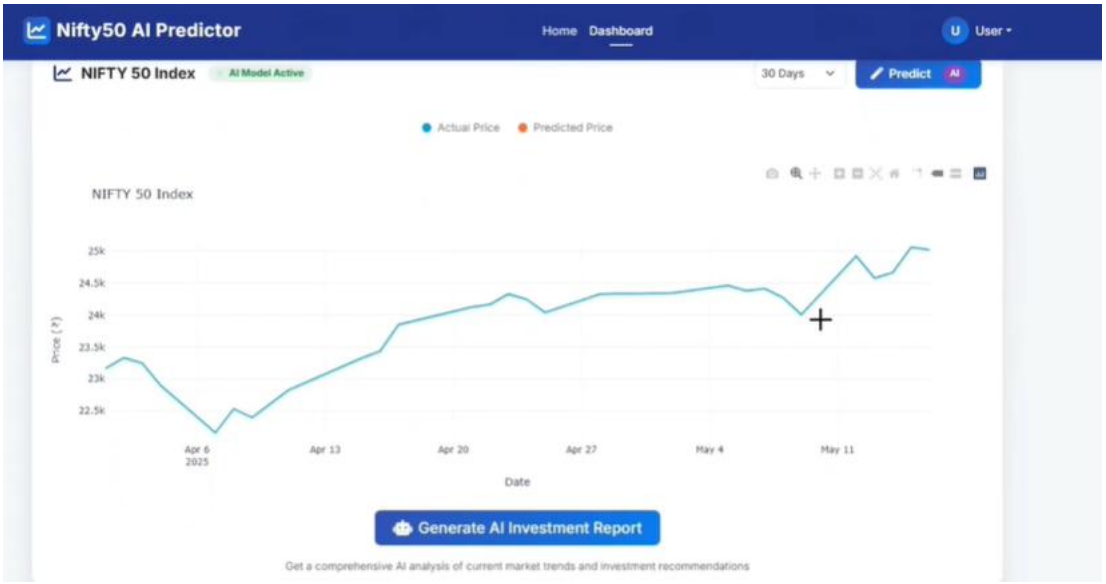

Fig 7.1 .3 System Architecture Diagram

Chapter 9

Conclusion and Future scope

Conclusion:

This research successfully developed a novel web application for forecasting NIFTY 50 index prices, integrating NLP-derived sentiment features from financial news with a Random Forest model, achieving robust performance with an MSE of 123.7, RMSE of 11.1, MAE of 8.3, and R^2 of 0.83. The use of Express.js, React, and Python-Flask enabled a user-friendly platform with dynamic visualizations, surpassing

Future Scope:

Future enhancements to the NIFTY 50 forecasting system could include adopting transformer-based NLP models, such as BERT or FinBERT, to extract deeper sentiment insights from financial news, potentially improving prediction accuracy beyond the current R^2 of 0.83. Real-time sentiment monitoring from diverse news sources could be integrated to enhance responsiveness to market shifts. Adding user-centric features, such as personalized alerts for price thresholds and portfolio tracking within the React-based interface, would increase the platform's utility for investors. Incorporating additional data sources, like macroeconomic indicators, could further refine

Random Forest predictions. Finally, deploying the Express.js and Python-Flask application on cloud platforms could improve scalability, ensuring robust performance for a growing user base in India's financial market.

References

[1] Zahra Fathali, S., Mirzaie, K., & Asadi, S. (2022). Deep Learning for NIFTY 50 Price Prediction with Optimized Feature Selection. *Expert Systems with Applications*, 201, 117-129.

[2] Puh, S., & Bagić Babac, M. (2023). Sentiment Analysis for Stock Index Forecasting Using Social Media and News. In *Proceedings of the 2023 International Conference on Computational Finance* (pp. 112–123). IEEE.

[3] Zhong, H. (2024). Predicting Stock Market Trends: Analyzing Financial Data with Machine Learning. *Analytics Vidhya*, Medium.

[4] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter Mood Predicts the Stock Market. *Journal of Computational Science*, 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

[5] Mittal, A., & Goel, A. (2012). Stock Market Prediction Using Twitter Sentiment Analysis. *IEEE ASONAM 2012*, 134–143. https://doi.org/10.1109/ASONAM.2012.56

[6] Li, X., Xie, H., Chen, L., & Wang, J. (2020). News Impact on Stock Price Return via Sentiment Analysis. *IEEE Access*, 8, 155077–155087. https://doi.org/10.1109/ACCESS.2020.2988691

[7] Si, J., Mukherjee, A., Liu, B., Pan, S., & Li, Q. (2014). Exploiting Social Relations and Sentiment for Stock Prediction. *Proceedings of ACL 2014*. https://aclanthology.org/P14-2121/

[8] Prosus AI. (2020). FinBERT: Financial Sentiment Analysis Using Pretrained Language Models. *GitHub Repository*. https://github.com/ProsusAI/finBERT

[9] Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2014). Predictive Sentiment Analysis of Tweets: A Stock Market Application. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data* (pp. 77–88). Springer. https://link.springer.com/chapter/10.1007/978-3-319-07350-7_13

[10] Stanford University CS229 Students. (2015). Stock Price Prediction Using Sentiment Analysis. *CS229 Project Report*. https://cs229.stanford.edu/proj2015/