

Advanced Stock Price Prediction forecasting using a hybrid model of numerical and textual analysis

Minor Project Report



Submitted To
Chhattisgarh Swami Vivekanand Technical University, Bhilai
For
Bachelor of Technology (Hons.)
in
Computer Science & Engineering
By

Name: Sneha Jha
Roll No: 300012821019
En. No. CB4668
Semester: 7th
Branch: C.S.E.(D.S.)

Name: Madhurima Rawat
Roll No: 300012821042
En. No. CB4689
Semester: 7th
Branch: C.S.E.(D.S.)

Name: Geetanshu Dev
Roll No: 300012821043
En. No. CB4691
Semester: 7th
Branch: C.S.E.(D.S.)

Under the Guidance of
Mr. Ramakant Ganjeshwar
Assistant Professor
Department of Computer Science & Engineering
UTD, CSVTU, Bhilai (C.G.)



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

Session: 2024 -25



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

DECLARATION BY THE CANDIDATE

We the undersigned solemnly declare that the Minor project report entitled “**Advanced Stock Price Prediction forecasting using a hybrid model of numerical and textual analysis**” is based our own work carried out during the course of our study under the supervision of **Mr. Ramakant Ganjeshwar**.

We assert that the statements made and conclusions drawn are an outcome of the project work. We further declare that to the best of our knowledge and belief that the report does not contain any part of any work which has been submitted for the award of any other degree/diploma/certificate in this University/Deemed university of India or any other country.

Sneha Jha
300012821019
CB4668

Madhurima Rawat
300012821042
CB4689

Geetanshu Dev Meshram
300012821043
CB4691



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

CERTIFICATE BY SUPERVISOR

This is to certify that the Minor project report entitled **“Advanced Stock Price Prediction forecasting using a hybrid model of numerical and textual analysis”** is a record of project work carried out under my guidance and supervision for the fulfillment of the award of degree of Bachelor of Technology (Hons.) in the faculty of Computer Science & Engineering of Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.) India.

To the best of my knowledge and belief the report

- I Embodies the work of the candidate himself
- II Has duly been completed
- III Fulfills the partial requirement of the ordinance relating to the B.Tech(Hons.) degree of the University
- IV Is up to the desired standard both in respect of contents and language for being referred to the examiners.

Mr. Ramakant Ganjeshwar
Assistant Professor
Department of CSE

Forwarded to
Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G.)

Dr. Toran Verma
HOD
Department of CSE
UTD, CSVTU, Bhilai (C.G.)

Dr. P.K. GHOSH
Director
UTD, CSVTU, BHILAI (C.G.)



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

CERTIFICATE BY EXAMINER

The project report entitled “**Advanced Stock Price Prediction forecasting using a hybrid model of numerical and textual analysis**” has been examined by the undersigned as a part of the examination of Bachelor of Technology (Hons.) in the faculty of Computer Science & Engineering of Chhattisgarh Swami Vivekanand Technical University, Bhilai.

Internal Examiner

Date:

External Examiner

Date:



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

ACKNOWLEDGEMENT

The real spirit of achieving a goal is through excellence and serious discipline. We would like to express our profound gratitude to **Mr. Ramakant Ganjeshwar**, Assistant Professor , and **Dr. P K Ghosh**, Director of university for their contributions to the completion of our project titled “Advanced Stock Price Prediction forecasting using a hybrid model of numerical and textual analysis”.

We would like to express our special thanks to our mentor **Mr Ramakant Ganjeshwar** for his time and efforts she provided throughout the semester. Your useful advice and suggestions were really helpful to us during the project’s completion. In this aspect, We are eternally grateful to you.

We would like to acknowledge that this project was completed entirely by us and not by someone else.

Sneha Jha

Roll No: 300012821019

Enroll No: CB468

Semester: 7th

(CSE (DS))

Madhurima Rawat

Roll No: 300012821042

Enroll No: CB4689

Semester: 7th

(CSE (DS))

Geetanshu Dev Meshram

Roll No: 300012821043

Enroll No: CB4691

Semester: 7th

(CSE (DS))



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

LIST OF ABBREVIATIONS

1	GOOG	Google - Alphabet Inc.
2	AMZN	Amazon.com
3	APPL	Apple Inc
4	META	Meta Platforms
5	MSFT	Microsoft Corporation
6	NFLX	Netflix Inc
7	NVDA	Nvidia Corporation
8	TCS	Tata Consultancy Services



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

LIST OF FIGURES

S.No.	Title	Page No.
1	Figure 4.1 DFD level 0 diagram	14
2	Figure 4.2 DFD level 1 diagram	15
3	Figure 4.3 Entity Relationship diagram	16
4	Figure 5.1 Stock Price Comparison	24
5	Figure 5.2 Close Price Moving Average Comparison of all 8 Companies	25
6	Figure 6.1 Target Variable Distribution	31
7	Figure 6.2 Word Cloud of News Articles	31
8	Figure 6.3 Top 20 Words of News Articles	32
9	Figure 7.1 InfluxDB database showing all measurements	38
10	Figure 7.2 Overview of Grafana Dashboard Showing all Rows	39
11	Figure 7.3 Overview of Grafana Dashboard Showing First row	39
12	Figure 7.4 Dashboard Panel Overview: Description and Dataset Information	40
13	Figure 7.5 Numerical Analysis Panel Google Close data	40
14	Figure 7.6 EDA of 'close' data of all company	41
15	Figure 7.7 Numerical Model Evaluation	41
16	Figure 7.8 Numerical Model Visualization	42
17	Figure 7.9 Textual Model Visualization	42
18	Figure 7.10 Textual Model Output Categories	43
19	Figure 7.11 Hybrid Model Visualization	43
20	Figure 7.12 Hybrid Visualization with Numerical Features	43
21	Figure 7.13 Hybrid Model Visualization with Numerical Features	43
22	Figure 7.13 Hybrid Model Visualization with Text Features	44
23	Figure 7.14 GitHub Repository	47

24	Figure 7.15 Deployed Streamlit App Overview	48
25	Figure 7.16 Deployed Streamlit App Data Exploration	48
26	Figure 7.17 Flask App showing User Input	49
27	Figure 7.18 Flask App showing Output Prediction	50



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

LIST OF TABLES

S. No.	Title	Page no.
1	Table 5.1 Companies Datasets summary	19
2	Table 5.2 Overview of key in dataset	20
3	Table 7.1 R2 score of all model	44
4	Table 7.2 Accuracy of model use in textual analytics	45



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

ABSTRACT

The prediction of stock prices has always been a major focus in financial sciences, driven by the need to anticipate variable movements in an unpredictable market. Traditional stock prediction models often rely solely on numerical data, overlooking critical external factors like news and social media trends. This project introduces an advanced stock price prediction model that leverages both numerical and textual data analysis for enhanced accuracy. The model forecasts stock prices for eight major tech companies by integrating historical stock data and sentiment analysis of financial news and social media. Machine learning algorithms in Scikit-learn process the numerical data, while natural language processing (NLP) techniques are used to assign sentiment scores to textual data. InfluxDB and Grafana support interactive monitoring and visualization of time-series data, facilitating effective data management. This hybrid approach incorporates both numerical data and unstructured sentiment to provide a comprehensive understanding of nonlinear stock market dynamics. The project code is available on GitHub, ensuring open-source accessibility and collaboration, and includes detailed documentation for ease of use. A Streamlit app has been developed and deployed, allowing users, including those new to the stock market, to interact with and understand the prediction model through an intuitive interface. This app features various sections, such as text-based prediction models and real-time analysis, making it an excellent resource for beginners and experienced users alike. Additionally, a Flask app has been created as a backend API for handling model predictions and future integration needs, providing a robust foundation for extending the application with additional features or other systems. The combination of these tools ensures that the project is both user-friendly and scalable, opening the door for further innovation in stock market forecasting.



Department of Computer Science & Engineering
University Teaching Department
Chhattisgarh Swami Vivekanand Technical University
Bhilai (C.G.) 491107

Table of content

Sno.	Title	Page No.
1	1 Introduction	2
2	1.1 Objective	3
3	2 System analysis	5
4	2.1 Identification of Need	5
5	2.2 Preliminary Investigation	5
6	3 Feasibility Study	10
7	3.1 Technically Feasibility	10
8	3.2 Operational Feasibility	10
9	4 Analysis	13
10	4.1 DFD	13
11	4.2 ER Diagram	15
12	5 Methodology	18
13	5.1 Project Overview	18
14	5.2 Data Collection	18
15	5.3 Dataset Description	20
16	5.4 Data analysis	22
17	5.5 For Imputing missing Values	23
18	5.6 Exploratory Data Analysis (EDA)	24
19	6 Implementation	28
20	6.1 Development Environment	28
21	6.2 Project Execution	28
22	6.3 Challenges Faced	33
23	7 Result and Discussion	37
24	7.1 Exploratory Data analysis	37
25	7.2 Prediction and actual closing price from google	37
26	7.3 Regression Results of Ensemble Model	44
27	7.4 Textual data Classified Result	45
28	7.5 Hybrid Model Performance	45
29	7.6 GitHub Repository and Deployment	46
29	8 Conclusion and Discussion	53
30	8.1 Summary	53
31	8.2 Achievement	53
32	8.3 Implication and Recommendation	54

33	8.4 Future Scope	55
34	Reference	57