Anomaly detection in financial market news sentiment analysis

# MICRO PROJECT REPORT

**Submitted by**

**DIVI MANOJ KUMAR**

**9921004859**

**in partial fulfilment for the award of the degree of**

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

****

**SCHOOL OF COMPUTING**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING KALASALINGAM ACADEMY OF RESEARCH**

**AND EDUCATION KRISHNANKOIL 626 126**

APRIL 2025

# DECLARATION

We affirm that the micro project work titled **“Anomaly detection in financial market news sentiment analysis”** being submitted in partial fulfilment for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** is the original work carried out by us. It has not formed part of any other project work submitted for the award of any degree or diploma, either in this or any other University.

Divi Manoj Kumar

9921004859

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Date:

Signature of the Mentor **Mrs. J Benita Associate/Assistant Professor**

**Department of Computer Science and Engineering**



# BONAFIDE CERTIFICATE

Certified that this project report “Anomaly detection in financial market news sentiment analysis” is the Bonafide work of “DIVI MANOJ KUMAR (9921004859)” who carried out the Micro project work under my supervision.

**Mrs. J Benita Dr. N. Suresh Kumar**

**SUPERVISOR HEAD OF THE DEPARTMENT**

**Associate/Assistant Professor Professor and Head**

Computer Science and Engineering Computer Science and Engineering Kalasalingam Academy of Research Kalasalingam Academy of Research and Education and Education

Krishnankoil 626126 Krishnankoil 626126

Virudhunagar District. Virudhunagar District.

Submitted for the Micro Project Viva-voice examination held on

**Internal Examiner External Examiner**

# ACKNOWLEDGEMENT

First and foremost, we thank the ‘Supreme Power’ for the immense grace showered on us which enabled us to do this project. We take this opportunity to express sincere thanks to the late, **“Kalvivallal” Thiru T. KALASALINGAM, Chairman, Kalasalingam Group of Institutions, “Illayavallal” Dr. K. SRIDHARAN, Ph.D., Chancellor, Dr. S. SHASI ANAND, Ph.D., Vice President,** who is the guiding light for all the activities in our university.

We thank our Vice Chancellor **Dr. S. NARAYANAN, Ph.D.,** for guiding every one of us and infusing us with the strength and enthusiasm to work successfully.

We wish to express our sincere thanks to our respected Head of the Department **Dr. N. SURESH KUMAR,** whose moral support encouraged us to process through our project work successfully.

We offer our sincerest gratitude to our Project Supervisor, **Mrs. J Benita,** for his/her patience, motivation, enthusiasm, and immense knowledge.

We are extremely grateful to our Micro Project Coordinator **Dr. P. Anitha**, Faculty In Charges **Dr. M. Rajasekaran, Mrs. B. Lavanya, Ms. P. J. Kiruthiga** for their constant encouragement in the completion of the Project.

Finally, we thank all, our Parents, Faculty, Non-Teaching Faculty, and our friends for their moral support.



**SCHOOL OF COMPUTING COMPUTER SCIENCE AND ENGINEERING**

**MICRO PROJECT SUMMARY**

|  |  |  |
| --- | --- | --- |
| Micro Project Title | “Anomaly detection in financial market news sentiment analysis**”** | |
| Micro Project Team Members (Name with Register No) | DIVI MANOJ KUMAR (9921004859) | |
| Guide Name/Designation | Mrs. J Benita  Assistant Professor  Computer Science and Engineering | |
| Program Concentration Area | Anomaly detection | |
| Technical Requirements | TextBlob for Sentiment Analysis , VADER , Linear Discriminant Analysis | |
| Engineering standards and realistic constraints in these areas | | |
| **Area** | **Codes & Standards / Realistic Constraints** | **Tick** ✓ |
| Economic |  | ✓ |
| Environmental |  |  |
| Social |  | ✓ |
| Ethical |  |  |
| Health and Safety |  |  |
| Manufacturability |  | ✓ |
| Sustainability |  |  |

**ABSTRACT**

Knowing how news items' emotion affects them is essential for making wise decisions in the ever-changing world of financial markets. The importance of anomaly detection in financial market news sentiment research is clarified in this abstract, emphasizing its value in exposing underlying market   
dynamics.   
  
A number of factors have a significant impact on financial markets, and sentiment is a key one.   
The wide range of emotions expressed in news items, from optimism to pessimism, can have a big impact on market trends. But among all of this data, anomalies—differences from the norm or anticipated patterns—frequently surface. In order to understand the actual attitude of the market and predict future changes, analysts and investors must be able to identify these anomalies.   
  
In order to find anomalies in the enormous corpus of sentiment data from financial news, anomaly detection techniques are essential. Anomaly detection models search through the noise using sophisticated machine learning algorithms to identify odd trends, attitudes, or occurrences that might portend future market moves. Investors may obtain important insights into market dynamics and successfully reduce risk by identifying anomalies, which will allow them to modify their strategy.

In order to improve the precision and dependability of sentiment research in financial markets, anomaly identification is crucial, as this abstract highlights. Stakeholders can make better decisions and maximize returns by navigating the financial landscape's intricacies with more confidence when anomalies are routinely detected.   
  
Ultimately, anomaly detection in financial market news sentiment analysis is a potent instrument for interpreting market dynamics and predicting changes in investor mood. Keeping ahead of the curve and making wise investment decisions will require incorporating anomaly detection techniques into sentiment research frameworks as financial markets continue to change.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **TITLE** | | **PAGE NO.** |
| **ABSTRACT** | |  |
| **LIST OF TABLES** | |  |
| **LIST OF FIGURES** | |  |
| **LIST OF NPTEL/COURSE ERA/ UDEMY COURSES** | |  |
| **CHAPTER I** | **INTRODUCTION** |  |
| **CHAPTER II** | **LITERATURE REVIEW** |  |
| **CHAPTER III** | **MICRO PROJECT IMPLENTAION** |  |
| **CHAPTER IV** | **RESULTS & DISCUSSION** |  |
| **CHAPTER V** | **CONCLUSION & FUTURE SCOPE** |  |
| **COURSE CERTIFICATION** | |  |
| **PUBLICATION** | | |
| **REFERENCES** | | |
| **PLAGIARISM REPORT** | | |
| **APPENDIX** | | |

|  |  |  |
| --- | --- | --- |
| **TABLES** | **DETAILS** | **PAGE NO.** |
| **Table 1** | Table of contents |  |
| **Table 2** | List of tables |  |
| **Table 3** | List of Figures |  |
| **Table 4** | List of courses |  |

|  |  |  |
| --- | --- | --- |
| **FIGURES** | **DETAILS** | **PAGE NO.** |
| Figure 1 | System Design |  |
| Figure 2 | Results |  |
| Figure 3 | Coursera Certificate 1 |  |
| Figure 4 | Coursera Certificate 2 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **S. NO.** | **COURSE NAME** | **COURSE DURATION** | **COURSE PLATFORM** |
| 01 | AI and Machine Learning Algorithms and Techniques | 45 hours | Coursera |
| 02 | Machine Learning for All | 33 hours | Coursera |

Coursera's "AI and Machine Learning Algorithms and Techniques" course provides a thorough examination of the fundamental algorithms and approaches used in ML and AI. Deep learning techniques, with a special emphasis on pre-trained large language models (LLMs), are included, along with other learning paradigms like supervised, unsupervised, and reinforcement learning. The focus of the training is on applying these strategies practically, stressing both their advantages and disadvantages while dealing with various business problems.  
  
**Course Organization:**

The five courses that make up the curriculum are all intended to build on one another and provide students a thorough understanding of AI and ML methods. These courses address the following important topics, even though the information provided does not specify their exact titles:  
  
  
  
**Learning Outcomes:**

After finishing this course, students will:  
  
Acquire a strong theoretical grasp of several AI and ML algorithms. Gain hands-on experience utilizing pertinent programming frameworks and tools to build these algorithms. Gain the skills to evaluate the benefits and drawbacks of various AI and ML approaches for resolving actual business issues.  
Anyone looking to learn more about AI and ML, especially those who are curious about how these technologies are used in real-world business settings, should take this course.

"Machine Learning for All" is an introductory course offered by the University of London on Coursera. Designed for learners without a background in programming or mathematics, the course aims to demystify machine learning concepts and their societal implications. citeturn0search0

**Course Details:**

* **Modules:** The course consists of four modules:
  1. **Machine Learning:** Introduces artificial intelligence and machine learning techniques, including practical experience in training learning models.
  2. **Data Features:** Explores how data representation affects machine learning, focusing on data features and their role in simplifying learning processes.
  3. **Machine Learning in Practice:** Prepares learners to undertake their own machine learning projects, emphasizing testing, applications, and understanding potential challenges.
  4. **Your Machine Learning Project:** Guides learners through collecting datasets, training models, and testing them, culminating in a hands-on project.
* **Assessments:** The course includes seven assignments, one per module, designed to reinforce learning and provide practical experience.
* **Language Support:** Taught in English, the course offers subtitles in 24 languages, making it accessible to a global audience.
* **Certificate:** Upon completion, learners receive a shareable certificate that can be added to their LinkedIn profile or resume.

**Course Outcomes:**

By the end of the course, learners will:

* Understand the basics of modern machine learning technologies.
* Be able to explain and predict how data affects machine learning outcomes.
* Use non-programming-based platforms to train machine learning models using datasets.
* Form informed opinions on the benefits and potential dangers of machine learning in society.

This course is ideal for beginners interested in grasping the fundamentals of machine learning without delving into complex programming or mathematical concepts. citeturn0search0

**INTRODUCTION**

Finding odd or surprising patterns in sentiment data taken from news stories, social media posts, or other textual sources pertaining to financial markets is known as anomaly detection in financial market news sentiment research. In order to identify anomalous shifts in sentiment that can potentially indicate important events, market moves, or new trends, this field makes use of statistical methods, machine learning, and natural language processing (NLP).   
  
The purpose of financial market sentiment analysis is to determine how traders and investors feel generally about certain assets, businesses, or market circumstances. In order to get insight into the dynamics of market sentiment, which can affect trading choices and market outcomes, analysts examine the sentiment expressed in news stories, blog posts, tweets, and other textual sources.   
  
Analysts and traders can learn a great deal about new market trends, foresee possible hazards, and make well-informed judgments to minimize negative effects or seize opportunities in the financial markets by successfully identifying abnormalities in sentiment data from the financial markets.

**LITERATURE REVIEW**

**1.** Pradeep K. Atrey, Mohsen Jamali, and Ehsan Fazlollahtabar's paper "Anomaly Detection for Financial Market Surveillance" (2018):   
  
The several anomaly detection approaches, such as statistical approaches and machine learning algorithms, that are used to monitor financial market data are covered in this study. Anomaly detection in trade and news sentiment data is covered.   
  
**2.** The second is "Detecting Anomalies in Financial News Streams" by Puneet Agarwal, Gautam Shroff, Lovekesh Vig, and Pankaj Malhotra (2016):   
  
The authors suggest use time-series analysis in conjunction with textual attributes to identify irregularities in financial news streams. Their main goal is to spot notable variations in sentiment patterns that can point to unusual market activity.

**3.** The study "Anomaly Detection in Text for the Financial Domain" by Sara Shahzad, Syed Asad Hussain, Kamran Ahsan, and Muhammad Nihal Hussain (2018):   
  
The methods for detecting anomalies in financial text data are examined in this research. The difficulties of identifying irregularities in the sentiment of financial news are covered, and a framework fusing machine learning algorithms with natural language processing approaches is shown.   
  
**4.** In 2019, David W. Guerra and Mykel J. Kochenderfer published "Detecting Anomalies in Financial Time Series Data Using Deep Learning":   
  
An strategy based on deep learning is suggested by the authors to identify irregularities in financial time series data. Using convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to detect anomalies in financial markets, including sentiment analysis data, is covered.

**5.** Asuman Dogac, Samet Kadioglu, and Hakan Ferhatosmanoglu's "Anomaly Detection in Financial News Using Autoencoders" (2020):   
  
In this work, the application of autoencoder neural networks to anomaly identification in emotion data from financial news is examined. Experimental results showing how well autoencoders detect anomalous patterns in sentiment time series data are presented.   
  
**6.** Gidofalvi, S. (2001) "Financial News Sentiment Analysis: A Survey and Taxonomy":   
  
Although it is not exclusively centered on anomaly identification, this survey article offers insightful information about sentiment analysis methods employed with financial news data. It discusses several sentiment analysis techniques, such as lexicon-based approaches, machine learning models, and their uses in the research of financial markets.

**MICRO PROJECT IMPLEMENTATION**

1. **Data Collection:**

Obtain market news information from a variety of sources, including social media platforms, financial news websites, RSS feeds, and more.   
Make sure the data includes a variety of financial products, such as commodities, stocks, and currencies.

1. **Preprocessing:**

Clear the gathered data by eliminating formatting errors, noise, and unnecessary information.   
Tokenize the text by turning it into words or sentences.   
Normalize the text by managing abbreviations, deleting punctuation, and changing it to lowercase.

1. **Sentiment Analysis:**

To examine the sentiment of the market news stories, apply Natural Language Processing (NLP) techniques.  
Use supervised learning techniques to train a sentiment analysis model utilizing labeled data or pre-trained models like BERT or GPT. Sort each news story according to the model's sentiment score into categories with positive, negative, or neutral sentiment.

1. **Anomaly Detection:**

The fourth step is anomaly identification, which involves using algorithms to find outliers or odd patterns in market news data.   
Employ statistical techniques such as Z-score, Grubbs' test, or machine learning algorithms like autoencoders, isolation forest, or one-class SVM.   
Using past data or subject-matter expertise, establish cutoff points or standards for spotting irregularities.

1. **Integration:**

To find important market trends or occurrences, combine the findings of anomaly detection and sentiment analysis.   
Examine news stories' attitude in relation to anomalies found to determine any possible effects on financial markets.   
Provide traders, investors, and financial analysts with insights by visualizing the sentiment trends and anomalies that have been identified over time.

1. **Feedback Loop:**

By using user input and performance data, continuously assess and improve the sentiment analysis and anomaly detection algorithms.   
"-" Adjust the models to reflect changing market conditions and trends by incorporating new data sources.

**7. Deployment:**

Either install the anomaly detection and sentiment analysis system as a stand-alone program or incorporate it into already-existing financial analysis platforms in a production setting.   
"-" Give users access to up-to-date information and alerts so they can make wise choices when trading and investing.

Financial institutions and traders can use sentiment analysis and anomaly identification to improve decision-making in the financial markets and obtain actionable insights from market news by using this methodology.

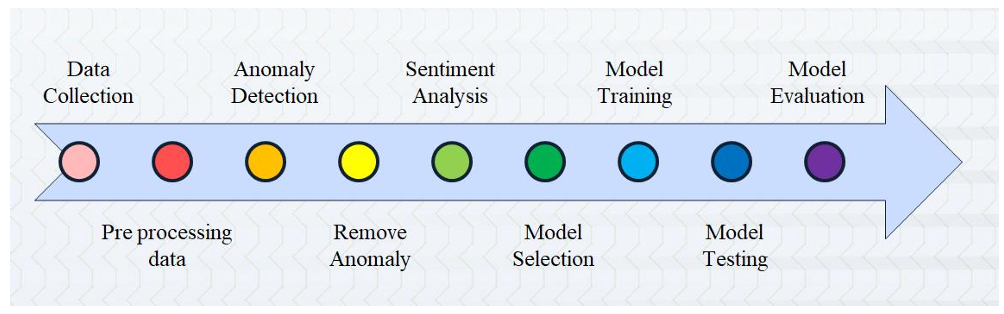
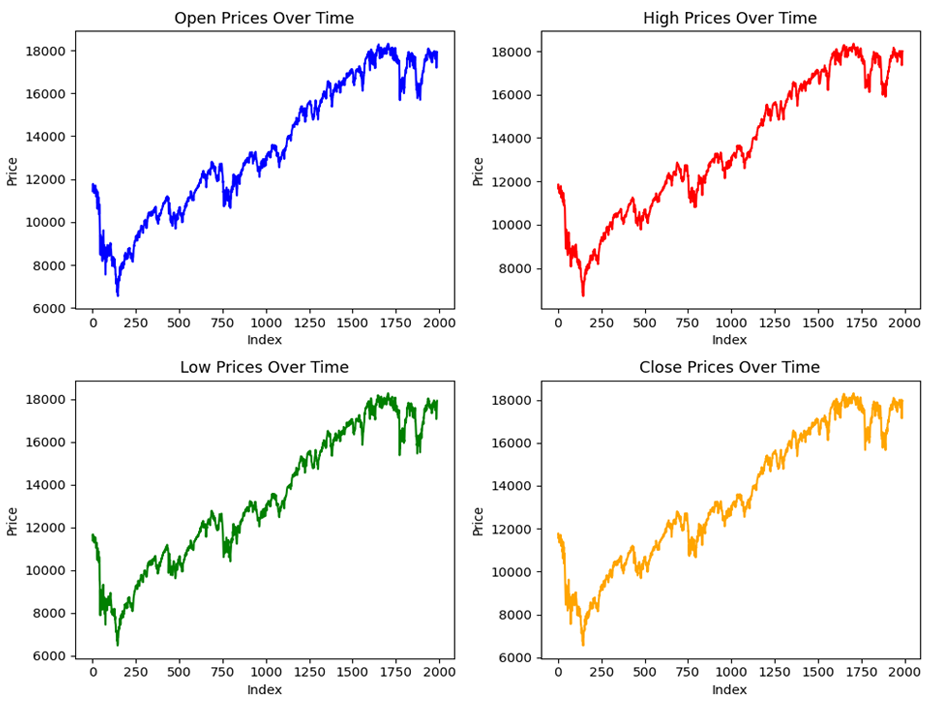


Fig.1 – System Design

**RESULTS & DISCUSSION**



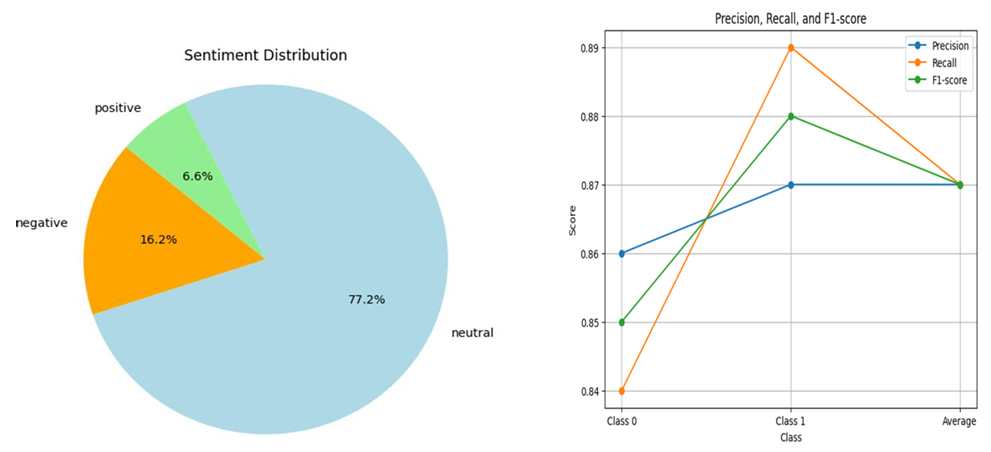


Fig.2 - Results

**CONCLUSION AND FUTURE SCOPE**

**Conclusion:**  
In financial market news sentiment research, anomaly identification is essential for spotting odd trends or departures in sentiment data taken from different textual sources. Using statistical, machine learning, and natural language processing (NLP) methods, analysts can learn a great deal about the dynamics of market sentiment and possible market-moving events.   
  
  
Overall, due to developments in data analytics, technology, and market dynamics, anomaly identification in financial market news sentiment analysis is still a dynamic and developing subject. Analysts may make better selections in the quickly evolving financial landscape of today, identify anomalies more accurately, and comprehend market sentiment trends by utilizing cutting-edge approaches and methodologies.

**Future Development:**

Research on anomaly detection in financial market news sentiment analysis may concentrate on creating increasingly sophisticated methods that can manage the intricacies of financial language in the future.   
  
data, combining data from other sources including news sentiment APIs and social media, and investigating the connections between anomaly detection and other financial analytics fields like event detection and predictive modeling.

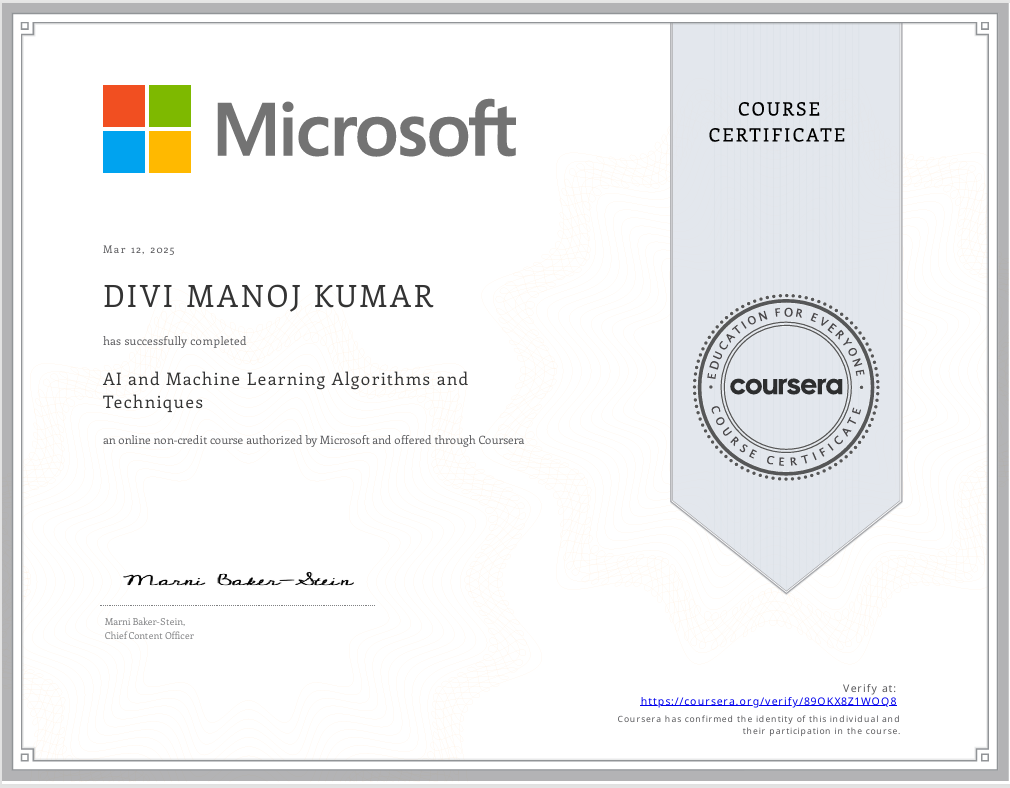


Fig.3 – Coursera Certificate.1

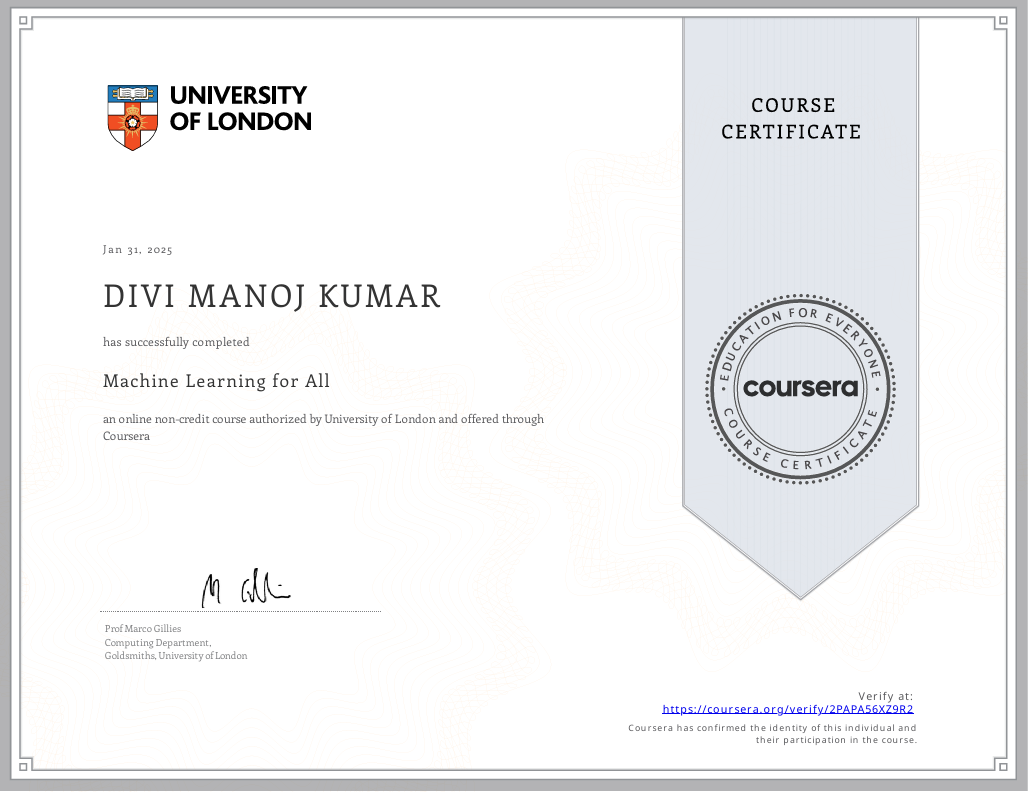


Fig.4 – Coursera Certificate.2

* [Real Time Text Analytics Software – Medallia – Medallia](https://www.medallia.com/platform/text-analytics/?utm_campaign=monkeylearnmigration)
* <https://monkeylearn.com/sentiment-analysis/>
* [https://towardsdatascience.com/understanding-outliers-in-text-data-with-transformers-cleanlab-and topic-modeling-db3585415a19](https://towardsdatascience.com/understanding-outliers-in-text-data-with-transformers-cleanlab-and%20%20%20%20%20%20topic-modeling-db3585415a19)
* <https://www.kaggle.com/datasets/aaron7sun/stocknews>
* <https://www.kaggle.com/datasets/mnassrib/dow-jones-industrial-average>

**(Project Report & Paper )**



# INTERNAL QUALITY ASSURANCE CELL MICRO PROJECT AUDIT REPORT

This is to certify that the micro project work entitled **“Anomaly detection in financial market news sentiment analysis”** categorized as an internal project done by **DIVI MANOJ KUMAR** of the Department of Computer Science and Engineering, under the guidance of **Mrs. J Benita** during the Even semester of the academic year 2024 - 2025 are as per the quality guidelines specified by IQAC.

**Quality Grade**

**Deputy Dean (IQAC)**

**Administrative Quality Assurance Dean (IQAC)**

**APPENEDIX**

**(Project Code)**

pip install VaderSentiment

import pandas as pd

import numpy as np

from textblob import TextBlob

import re

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

from sklearn.model\_selection import train\_test\_split

from sklearn.metrics import accuracy\_score, classification\_report

from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis

df1 = pd.read\_csv('dow\_jones\_news.csv')

df2 = pd.read\_csv('dow\_jones\_stock.csv')

df1.head(5)

df1.shape

df2.head(5)

df2.shape

merge = df1.merge(df2, how='inner', left\_index=True, right\_index=True)

merge

merge.drop(columns=['Date\_y'], inplace=True)

merge

headlines = []

for row in range(0, len(merge.index)):

  headlines.append(' '.join(str(x) for x in merge.iloc[row, 2:27]))

clean\_headlines = []

for i in range(0, len(headlines)):

  clean\_headlines.append(re.sub("b[(')]",'', headlines[i]))

  clean\_headlines[i] = re.sub('b[(")]', '', clean\_headlines[i])

  clean\_headlines[i] = re.sub("\'", '', clean\_headlines[i])

clean\_headlines[20]

merge['combined\_news'] = clean\_headlines

merge['combined\_news'][0]

merge.head(5)

def getsubjectivity(text):

  return TextBlob(text).sentiment.subjectivity

def getPolarity(text):

  return TextBlob(text).sentiment.polarity

merge['subjective'] = merge['combined\_news'].apply(getsubjectivity)

merge['Polarity'] = merge['combined\_news'].apply(getPolarity)

merge.head(5)

def getSIA(text):

  sia = SentimentIntensityAnalyzer()

  sentiment = sia.polarity\_scores(text)

  return sentiment

compound = []

neg = []

pos = []

neu = []

SIA = 0

for i in range(0, len(merge['combined\_news'])):

  SIA = getSIA(merge['combined\_news'][i])

  compound.append(SIA['compound'])

  neg.append(SIA['neg'])

  neu.append(SIA['neu'])

  pos.append(SIA['pos'])

merge['compound'] = compound

merge['negative'] = neg

merge['neutral'] = neu

merge['positive'] = pos

merge.head(5)

keep\_columns = [ 'Open', 'High', 'Low', 'Volume', 'subjective', 'Polarity', 'compound', 'negative', 'neutral' ,'positive',  'Label' ]

df = merge[keep\_columns]

df

X = df

X = np.array(X.drop(['Label'], 1))

y = np.array(df['Label'])

x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.2, random\_state = 32)

model = LinearDiscriminantAnalysis().fit(x\_train,y\_train)

pred = model.predict(x\_test)

pred

y\_test

print(classification\_report(y\_test, pred))

from flask import Flask, render\_template, request

from nltk.sentiment.vader import SentimentIntensityAnalyzer

from sklearn.model\_selection import train\_test\_split

from sklearn.discriminant\_analysis import LinearDiscriminantAnalysis

from sklearn.metrics import classification\_report

import numpy as np

import pandas as pd

app = Flask(\_\_name\_\_)

def getSIA(text):

    sia = SentimentIntensityAnalyzer()

    sentiment = sia.polarity\_scores(text)

    return sentiment

@app.route('/')

def home():

    return render\_template('index.html')

@app.route('/analyze', methods=['POST'])

def analyze():

    if request.method == 'POST':

        merge = pd.read\_csv('path\_to\_your\_csv\_file.csv')  # Adjust the path to your CSV file

        compound = []

        neg = []

        pos = []

        neu = []

        SIA = 0

        for i in range(0, len(merge['combined\_news'])):

            SIA = getSIA(merge['combined\_news'][i])

            compound.append(SIA['compound'])

            neg.append(SIA['neg'])

            neu.append(SIA['neu'])

            pos.append(SIA['pos'])

        merge['compound'] = compounda

        merge['negative'] = neg

        merge['neutral'] = neu

        merge['positive'] = pos

        x\_train, x\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=32)

        model = LinearDiscriminantAnalysis().fit(x\_train, y\_train)

        pred = model.predict(x\_test)

        classification\_result = classification\_report(y\_test, pred)

        return render\_template('result.html', classification\_result=classification\_result)

if \_\_name\_\_ == '\_\_main\_\_':

    app.run(debug=True)