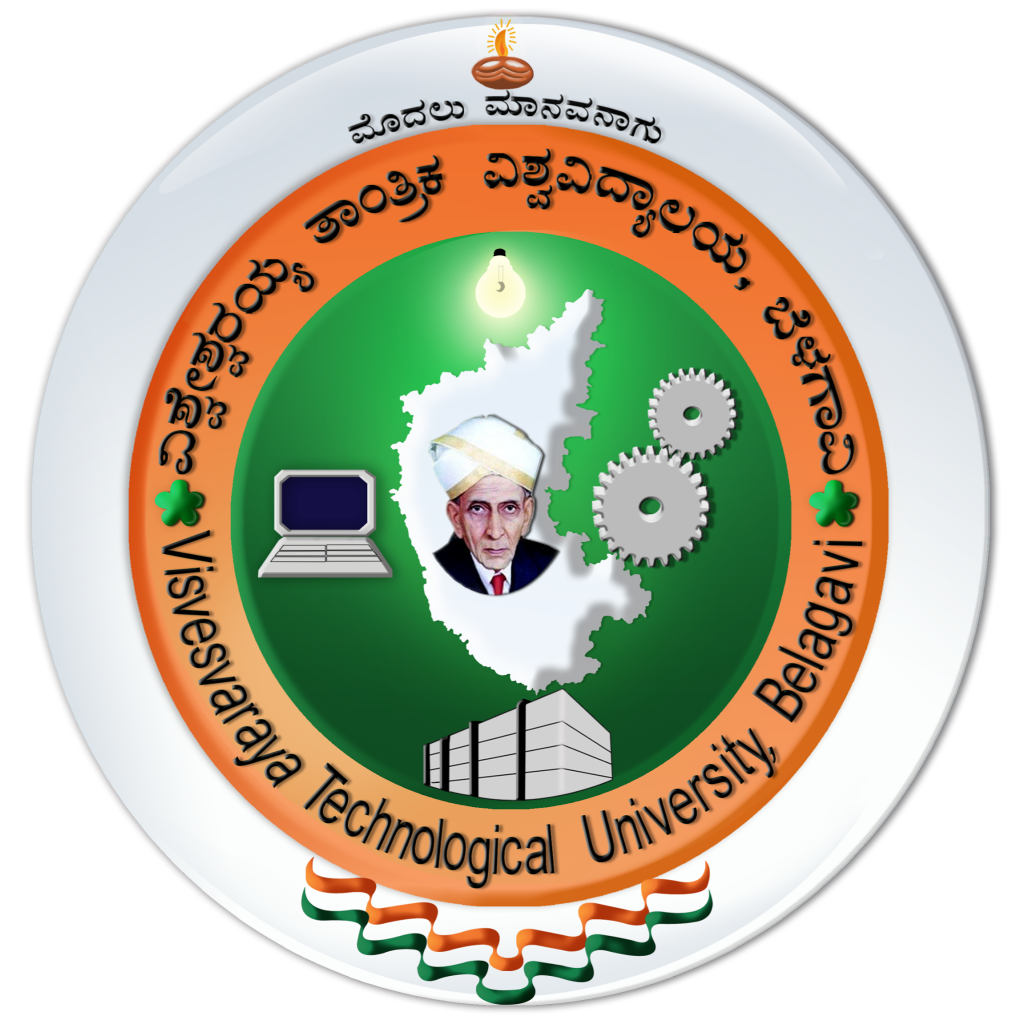
**VISVESVARAYA TECHNOLOGICAL UNIVERSITY**

**BELAGAVI – 590018, Karnataka INTERNSHIP REPORT ON**

**“Stockport | Predictive Sentiment Analysis**

***Submitted in partial fulfillment for the award of degree (18CSI85)***

**BACHELOR OF ENGINEERING IN INFORMATION SCIENCE AND ENGINEERING**

***Submitted by:***

**VAISHNAVI R**

**1MV19IS062**



Conducted at

**Varcons Technologies Pvt.Lt**



**SIR M VISVESVARAYA INSTITUTE OF TECHNOLOGY**

**Department of Information Science**

**Accredited by AICTE, New Delhi**

**International Airport Road, Hunasamaranahalli, Yelahanka,**

**Sir M Visvesvaraya Inst Rd, Yelahanka, Bengaluru, Karnataka 562157**

**CERTIFICATE**

This is to certify that the Internship titled **“Stockport | Predictive Sentiment Analysis”** carried out by **VAISHNAVI R (1MV19IS062),** a bonafide student of Sir M Visvesvaraya Institute of Technology, in partial fulfillment for the award of **Bachelor of Engineering**, in **Information Science and Engineering** under Visvesvaraya Technological University, Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respect of Internship prescribed for the course Internship / Professional Practice (18CSI85)

**Signature of Guide Signature of HOD Signature of Principal**

**External Viva:**

Name of the Examiner Signature with Date

1)

2)

**DECLARATION**

I, VAISHNAVI R, final year student of Information Science, Sir M Visvesvaraya Institute of Technology - 562 157, declare that the Internship has been successfully completed, in **Varcons technologies Pvt.Ltd**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Branch name, during the academic year 2022-2023.

Date :23/09/2022 :

Place : Bangalore

USN : 1VM19IS062

NAME: VAISHNAVI R

**CERTIFICATE OF INTERNSHIP COMPLETION**

Date: **2nd September, 2022**

Name: **Vaishnavi R**

USN: 1MV19IS062

**Dear Student,**

We would like to congratulate you on being selected for the **Machine Learning with-Python(Research Based)**Internship position with **Varcons Technologies Pvt Ltd**, effective Start Date **2nd September, 2022**, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With-Python (Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!.

Sincerely,

Spoorthi H C

**Director**

Varcons Technologies Pvt Ltd

213, 2st Floor,

18 M G Road,Ulsoor, Bangalore-560001

**ACKNOWLEDGEMENT**

This Internship is a result of accumulated guidance, direction and support of several important persons. We take this opportunity to express our gratitude to all who have helped us to complete the Internship

We express our sincere thanks to our Principal, for providing us adequate facilities to undertake this Internship

We would like to thank our Head of Dept – branch code, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank our (Lab assistant name) Software Services for guiding us during the period of internship

We express our deep and profound gratitude to our guide, Guide name, Assistant/Associate Prof, for her keen interest and encouragement at every step in completing the Internship.

We would like to thank all the faculty members of our department for the support extended during the course of Internship.

We would like to thank the non-teaching members of our dept, for helping us during the Internship.

Last but not the least, we would like to thank our parents and friends without whose constant help, the completion of Internship would have not been possible.

**VAISHNAVI R**

**1MV19IS062**

**ABSTRACT**

Predicting stock market movements is a well-known problem of interest. Now-a-days social media is perfectly representing the public sentiment and opinion about current events. Especially, twitter has attracted a lot of attention from researchers for studying the public sentiments. Stock market prediction on the basis of public sentiments expressed on twitter has been an intriguing field of research. Previous studies have concluded that the aggregate public mood collected from twitter may well be correlated with Dow Jones Industrial Average Index (DJIA). The thesis of this work is to observe how well the changes in stock prices of a company, the rises and falls, are correlated with the public opinions being expressed in tweets about that company. Understanding author’s opinion from a piece of text is the objective of sentiment analysis. The present paper have employed two different textual representations, Word2vec and Ngram, for analyzing the public sentiments in tweets. In this paper, we have applied sentiment analysis and supervised machine learning principles to the tweets extracted from twitter and analyze the correlation between stock market movements of a company and sentiments in tweets. In an elaborate way, positive news and tweets in social media about a company would definitely encourage people to invest in the stocks of that company and as a result the stock price of that company would increase. At the end of the paper, it is shown that a strong correlation exists between the rise and falls in stock prices with the public sentiments in tweets.

**Table of Contents**

| **Contents** | **Page no,** |
| --- | --- |
| Abstract | 6 |
| Table of Content | 7 |
| Company profile | 8 |
| About the company | 9 |
| Introduction | 10 |
| System Analysis | 11 |
| Requirement Analysis | 12 |
| Implementation | 13-15 |
| Snapshots | 16-21 |
| Conclusion | 22 |
| Reference | 23 |

**CHAPTER 1**

**COMPANY PROFILE**

**A Brief History of Varcons Technologies Pvt.Ltd**

Varcons Technologies is a leading provider of cutting-edge technologies and services, offering scalable solutions for businesses of all sizes. Founded by a group of friends who started by scribbling their ideas on a piece of paper, today we offer smart, innovative services to dozens of clients. We develop SaaS products, provide Corporate Seminars, Industrial training and much more.

Varcons Technologies, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the client's demand better. Varcons Technologies work with their clients and help them to define their exact solution requirement. Sometimes even they wonder if they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group composed of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence ” Technology helps you to Delight your Customers” and that is what we want to achieve.

**CHAPTER 2**

**ABOUT THE COMPANY**

Varcons Technologies is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ.The organization where they have a right mix of professionals as a stakeholders to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solutions. Motto of our organization is to “Collaborate with our clients to provide them with the best Technological solution hence creating a Good Present and Better Future for our client which will bring a cascading positive effect in their business shape as well”. Providing a Complete suite of technical solutions is not just our tagline, it is Our Vision for Our Clients and for Us, We strive hard to achieve it.

**Services provided by Varcons Technologies.**

* Core Java and Advanced Java
* Web services and development
* Python
* Conference / Event Management Service
* Academic Project Guidance
* On The Job Training

**CHAPTER 3**

**Introduction to ML**

Machine learning (ML) is a branch of artificial intelligence (AI) that enables computers to “self-learn” from training data and improve over time, without being explicitly programmed. Machine learning algorithms are able to detect patterns in data and learn from them, in order to make their own predictions. In short, machine learning algorithms and models learn through experience.

In traditional programming, a computer engineer writes a series of directions that instruct a computer how to transform input data into a desired output. Instructions are mostly based on an IF-THEN structure: when certain conditions are met, the program executes a specific action.

Machine learning, on the other hand, is an automated process that enables machines to solve problems with little or no human input, and take actions based on past observations.

While artificial intelligence and machine learning are often used interchangeably, they are two different concepts. AI is the broader concept – machines making decisions, learning new skills, and solving problems in a similar way to humans – whereas machine learning is a subset of AI that enables intelligent systems to autonomously learn new things from data.

Instead of programming machine learning algorithms to perform tasks, you can feed them examples of labeled data (known as training data), which helps them make calculations, process data, and identify patterns automatically.

Put simply, Google’s Chief Decision Scientist describes machine learning as a fancy labeling machine. After teaching machines to label things like apples and pears, by showing them examples of fruit, eventually they will start labeling apples and pears without any help – provided they have learned from appropriate and accurate training examples.

Machine learning can be put to work on massive amounts of data and can perform much more accurately than humans. It can help you save time and money on tasks and analyses, like solving customer pain points to improve customer satisfaction, support ticket automation, and data mining from internal sources and all over the internet.

**Project Scope :**

Real-Time Twitter Sentiment Analysis for stocks based on which the future movement of the market is predicted Goal: Understand the working of Sentiment analysis and Improve the accuracy.

**CHAPTER 4**

**SYSTEM ANALYSIS**

* **Existing System**

Earlier studies on stock market prediction are based on the historical stock prices. Later studies have debunked the approach of predicting stock market movements using historical prices. Stock market prices are largely fluctuating. The efficient market hypothesis (EMH) states that financial market movements depend on news, current events and product releases and all these factors will have a significant impact on a company’s stock value . Because of the lying unpredictability in news and current events, stock market prices follow a random walk pattern and cannot be predicted with more than 50% accuracy

With the advent of social media, the information about public feelings has become abundant. Social media is transforming like a perfect platform to share public emotions about any topic and has a significant impact on overall public opinion. Twitter, a social media platform, has received a lot of attention from researchers in recent times. Twitter is a micro-blogging application that allows users to follow and comment on other users' thoughts or share their opinions in real time . More than a million users post over 140 million tweets every day. This situation makes Twitter like a corpus with valuable data for researchers .Each tweet is 140 characters long and speaks public opinion on a topic concisely. The information exploited from tweets are very useful for making predictions,

* **Proposed System**

The most well-known publication in this area is by Bollen . They investigated whether the collective mood states of public (Happy, calm, Anxiety) derived from twitter feeds are correlated to the value of the Dow Jones Industrial Index. They used a Fuzzy neural network for their prediction. Their results show that public mood states in twitter are strongly correlated with Dow Jones Industrial Index. Chen and Lazer derived investment strategies by observing and classifying the twitter feeds. Bing et al. Studied the tweets and concluded the predictability of stock prices based on the type of industry like Finance, IT etc. Zhang found out a high negative correlation between mood states like hope, fear and worry in tweets with the Dow Jones Average Index. Recently, Brian et al. Investigated the correlation of sentiments of public with stock increase and decreases using Pearson correlation coefficient for stocks. In this paper, we took a novel approach of predicting rise and fall in stock prices based on the sentiments extracted from twitter to find the correlation. The core contribution of our work is the development of a sentiment analyzer which works better than the one in Brian’s work and a novel approach to find the correlation. Sentiment analyzer is used to classify the sentiments in tweets extracted.The human annotated dataset in our work is also exhaustive.

**CHAPTER 5**

**Requirement Analysis**

**Hardware Requirement Specification**

* 128MB Ram
* Memory: 15.6 GB
* Hard Disk: 40 GB

**Software Requirement Specification**

* pandas
* numpy
* matplotlib.pyplot
* seaborn
* nltk
* logistic regression

**CHAPTER 6**

**Implementation**

**Import modules**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import re

import seaborn as sns

import string

import nltk

import warnings

%matplotlib inline

warnings.filterwarnings('ignore')

**loading the database**

from google.colab import files

files.upload()

df=pd.read\_csv('stock\_market\_crash\_2022.csv',encoding="ISO-8859-1")

df.head()

df1=df.drop(['media','inReplyToUser','replycount','retweetcount','hashtags','retweetedTweet','quotedtweet','username','quotecount','language','likecount','inReplyToTweetId','inReplyToUser','mentionedUsers','user followers count'],axis='columns')

df1.head()

df1.info()

**Preprocessing the dataset**

def remove\_pattren(input\_txt,pattren):

  r = re.findall(pattren,input\_txt)

  for word in r:

    input\_txt=re.sub(word,"",input\_txt)

  return input\_txt

df1['clean\_txt']=np.vectorize(remove\_pattren)(df1['text'],"@")

df1.head()

df1['clean\_txt']=df1['clean\_txt'].str.replace("[^a-zA-Z#]"," ")

df1.head()

 df1['clean\_txt']=df1['clean\_txt'].apply(lambda x: " ".join([w for w in x.split() if len(w)>3]))

df1.head()

**Exploratory data analysis**

!pip install wordcloud

all\_words=" ".join([x for x in df1['clean\_txt']])

from wordcloud import WordCloud

wordcount = WordCloud(width=800,height=500,random\_state=42,max\_font\_size=100).generate(all\_words)

plt.figure(figsize=(15,8))

plt.imshow(wordcount, interpolation='bilinear')

plt.axis('off')

plt.show()

all\_words=" ".join([x for x in df1['clean\_txt'][df1['text\_sentiment']=='Positive']])

from wordcloud import WordCloud

wordcount = WordCloud(width=800,height=500,random\_state=42,max\_font\_size=100).generate(all\_words)

plt.figure(figsize=(15,8))

plt.imshow(wordcount, interpolation='bilinear')

plt.axis('off')

plt.show()

all\_words=" ".join([x for x in df1['clean\_txt'][df1['text\_sentiment']=='Negative']])

from wordcloud import WordCloud

wordcount = WordCloud(width=800,height=500,random\_state=42,max\_font\_size=100).generate(all\_words)

plt.figure(figsize=(15,8))

plt.imshow(wordcount, interpolation='bilinear')

plt.axis('off')

plt.show()

def hashtags\_extract(tweet):

  hashtags=[]

  for word in tweet:

    ht=re.findall(r"#(\w+)",tweet)

    hashtags.append(ht)

  return hashtags

ht\_postive=hashtags\_extract(df1['clean\_txt'][df1['text\_sentiment']==0])

ht\_negative=hashtags\_extract(df1['clean\_txt'][df1['text\_sentiment']==1])

**input split**

from sklearn.feature\_extraction.text import CountVectorizer

bow\_venctorized=CountVectorizer(max\_df=0.90,min\_df=2,max\_features=1000,stop\_words='english')

bow =bow\_venctorized.fit\_transform(df1['clean\_txt'])

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

x\_train, x\_test,y\_train,y\_test = train\_test\_split(bow,df1['text\_sentiment'],random\_state=42, test\_size=0.25)

**model training**

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import f1\_score, accuracy\_score

model = LogisticRegression()

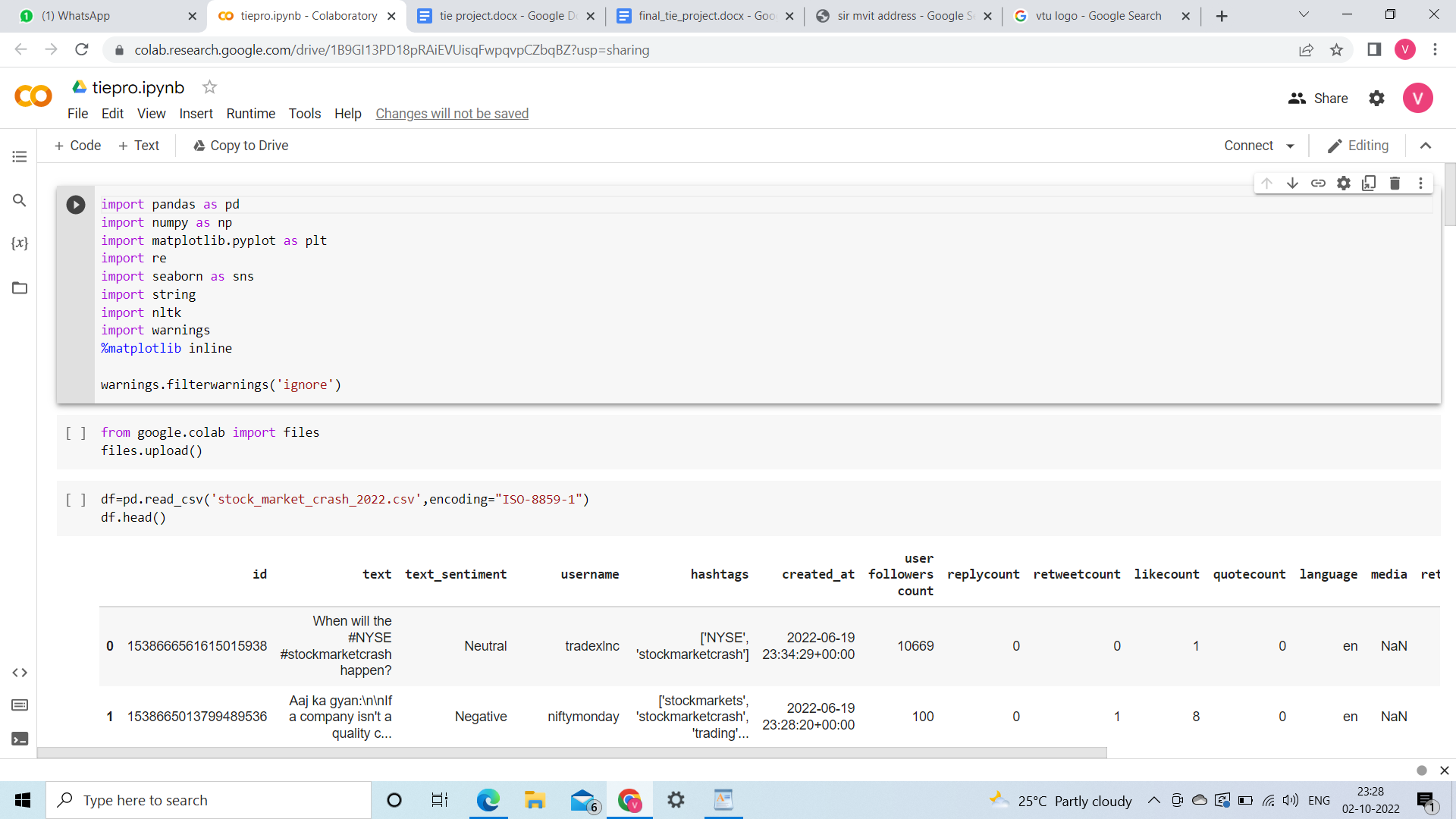
model.fit(x\_train,y\_train)

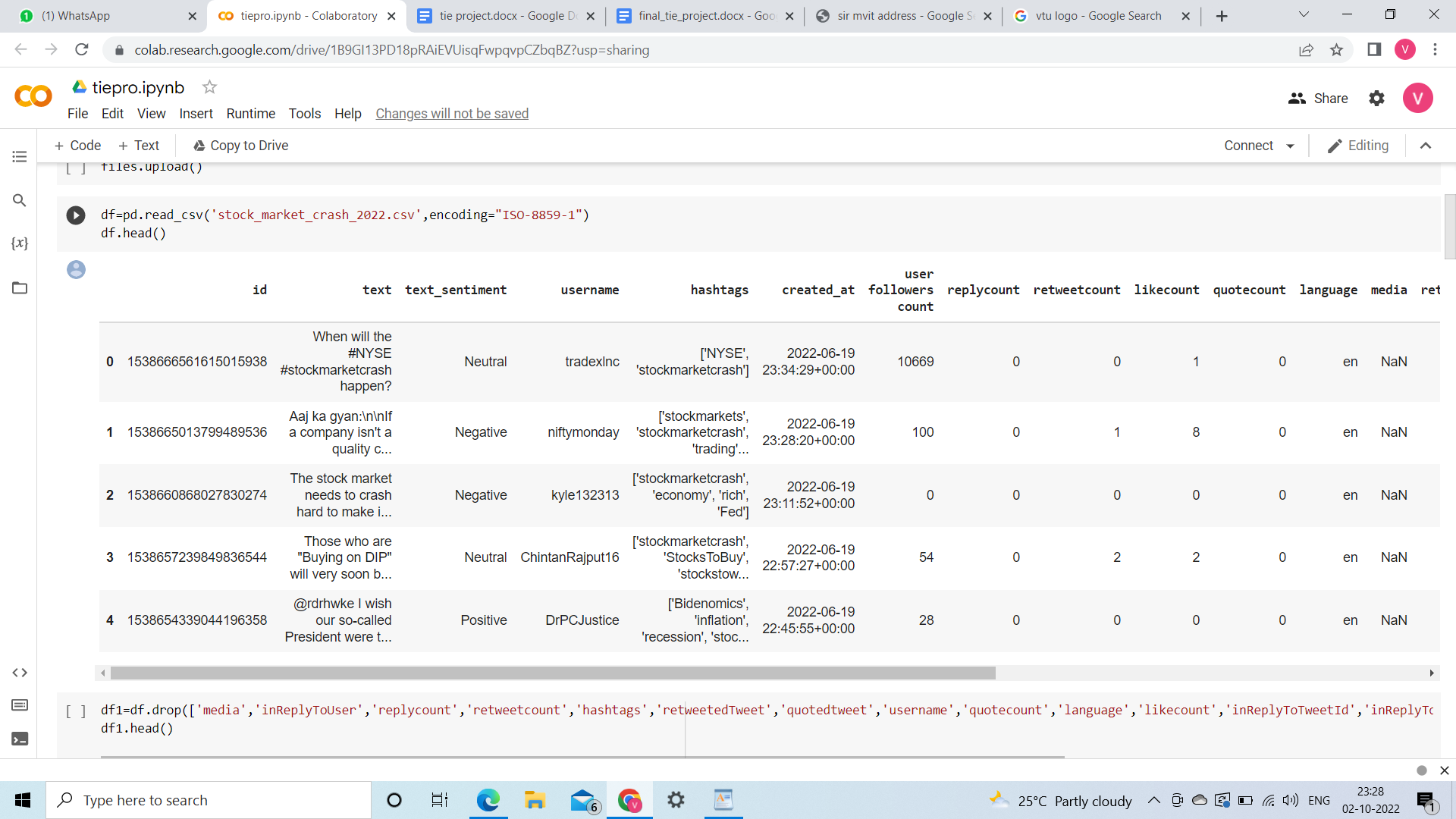
pred = model.predict(x\_test)

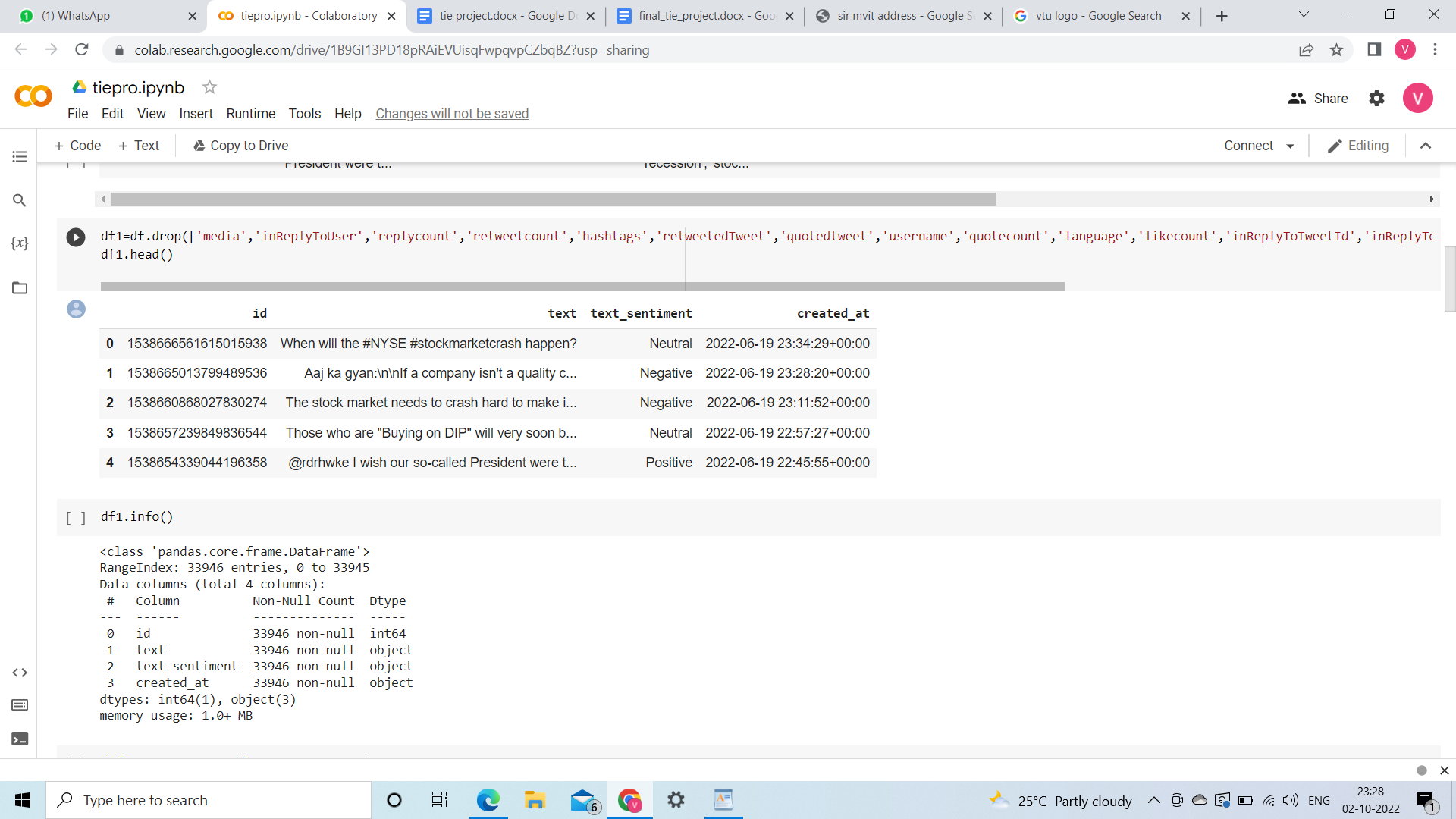
accuracy\_score(y\_test,pred)\*100

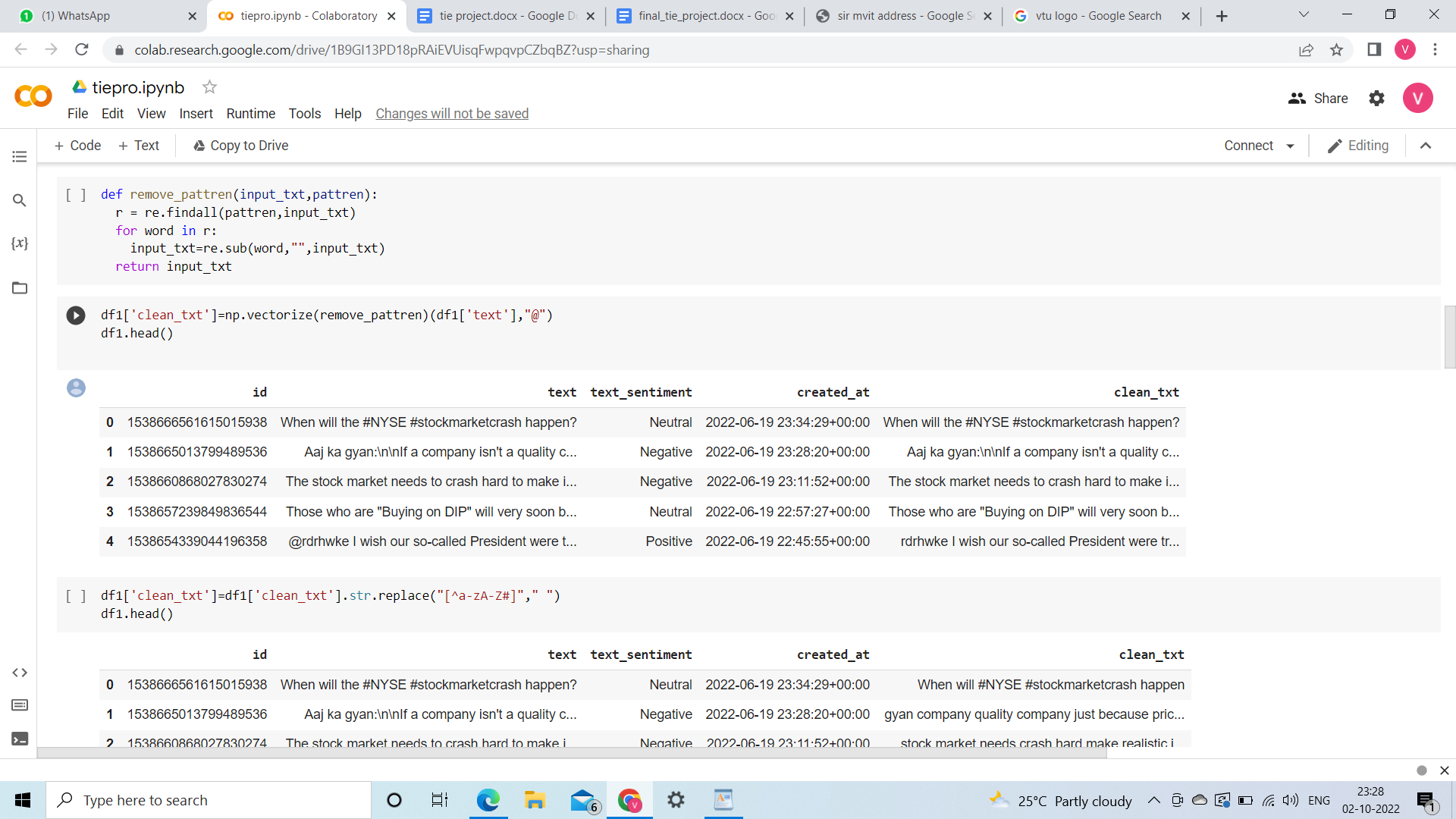
**CHAPTER 6**

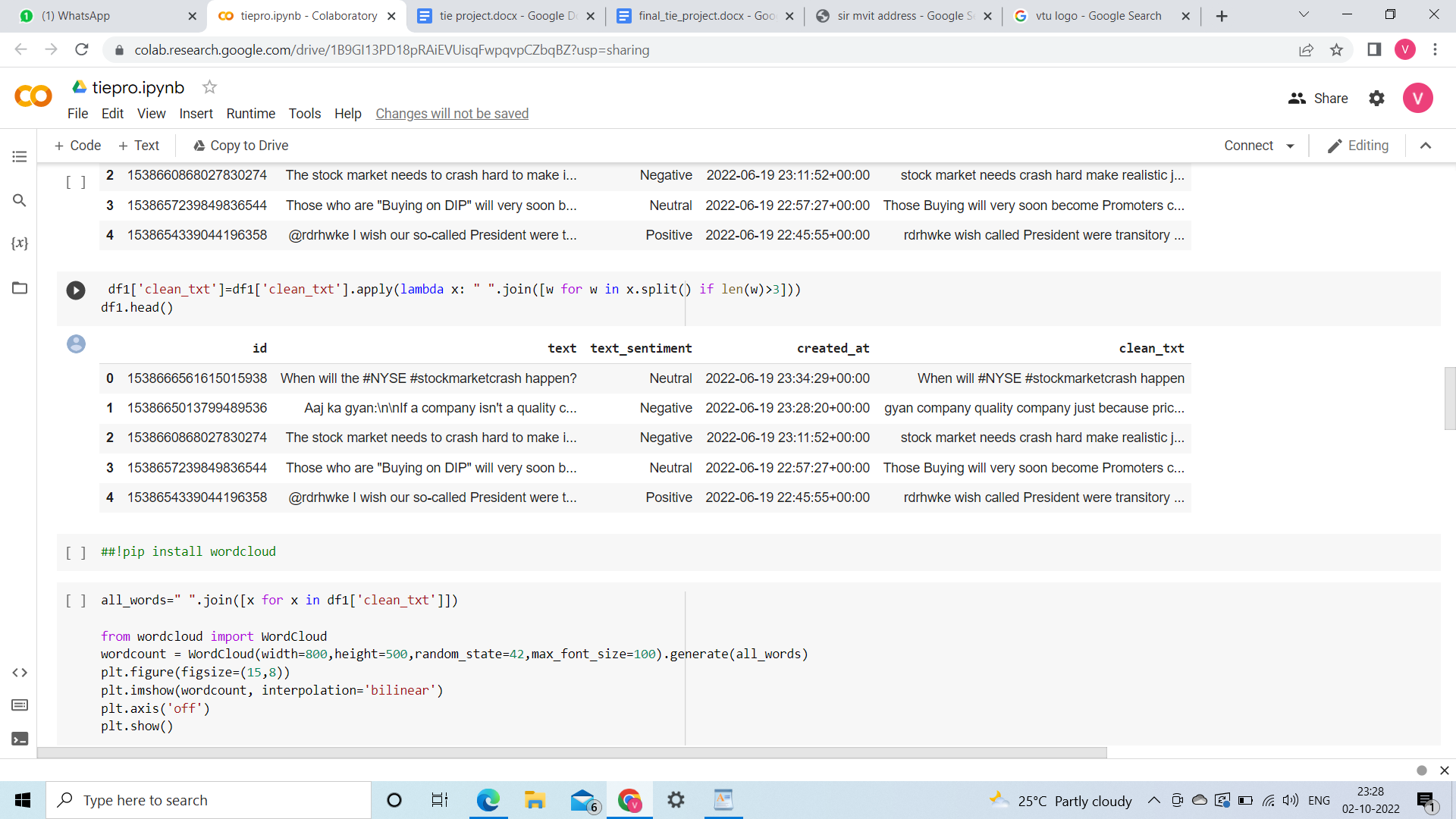
**SNAPSHOTS**

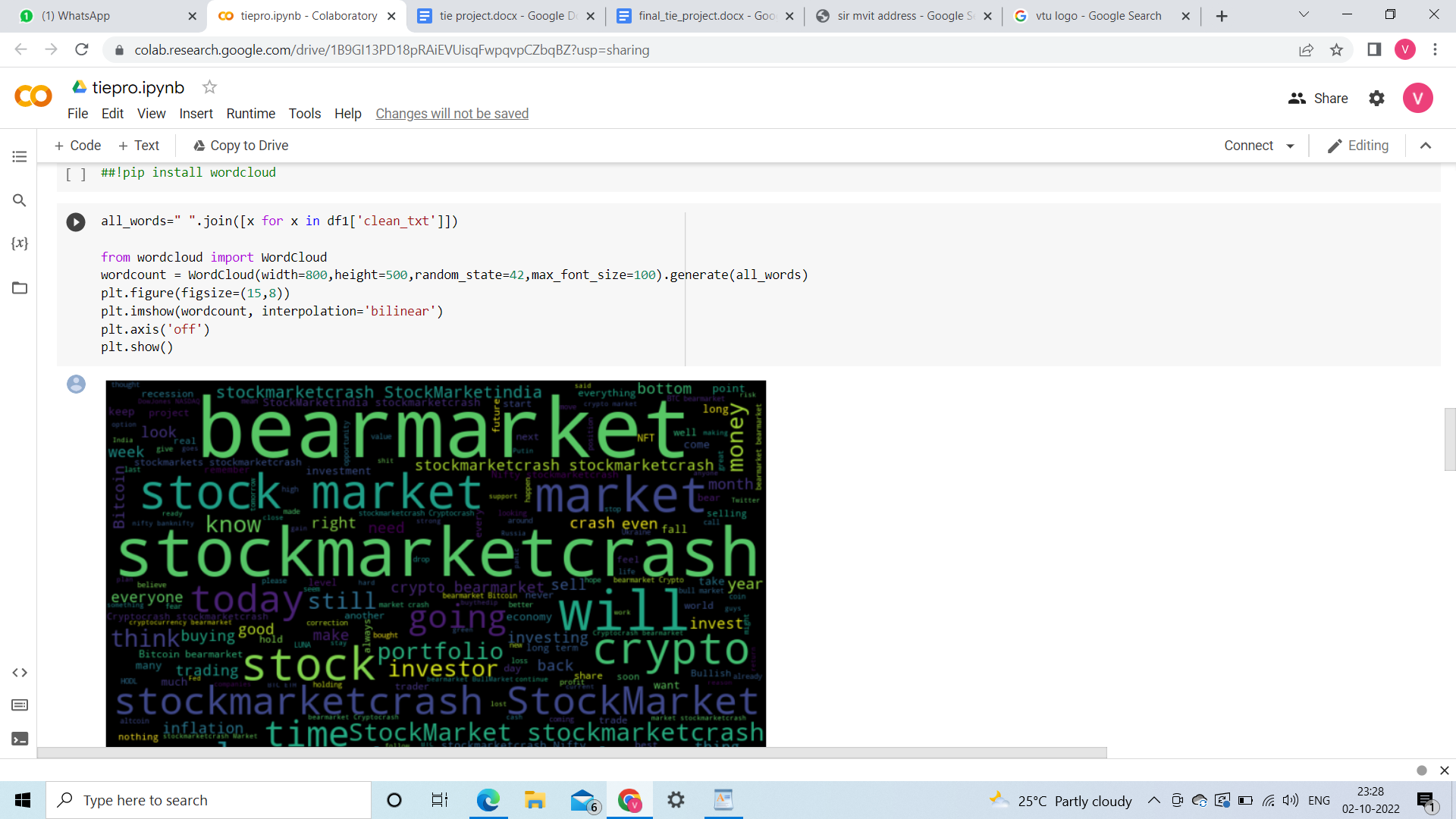
****

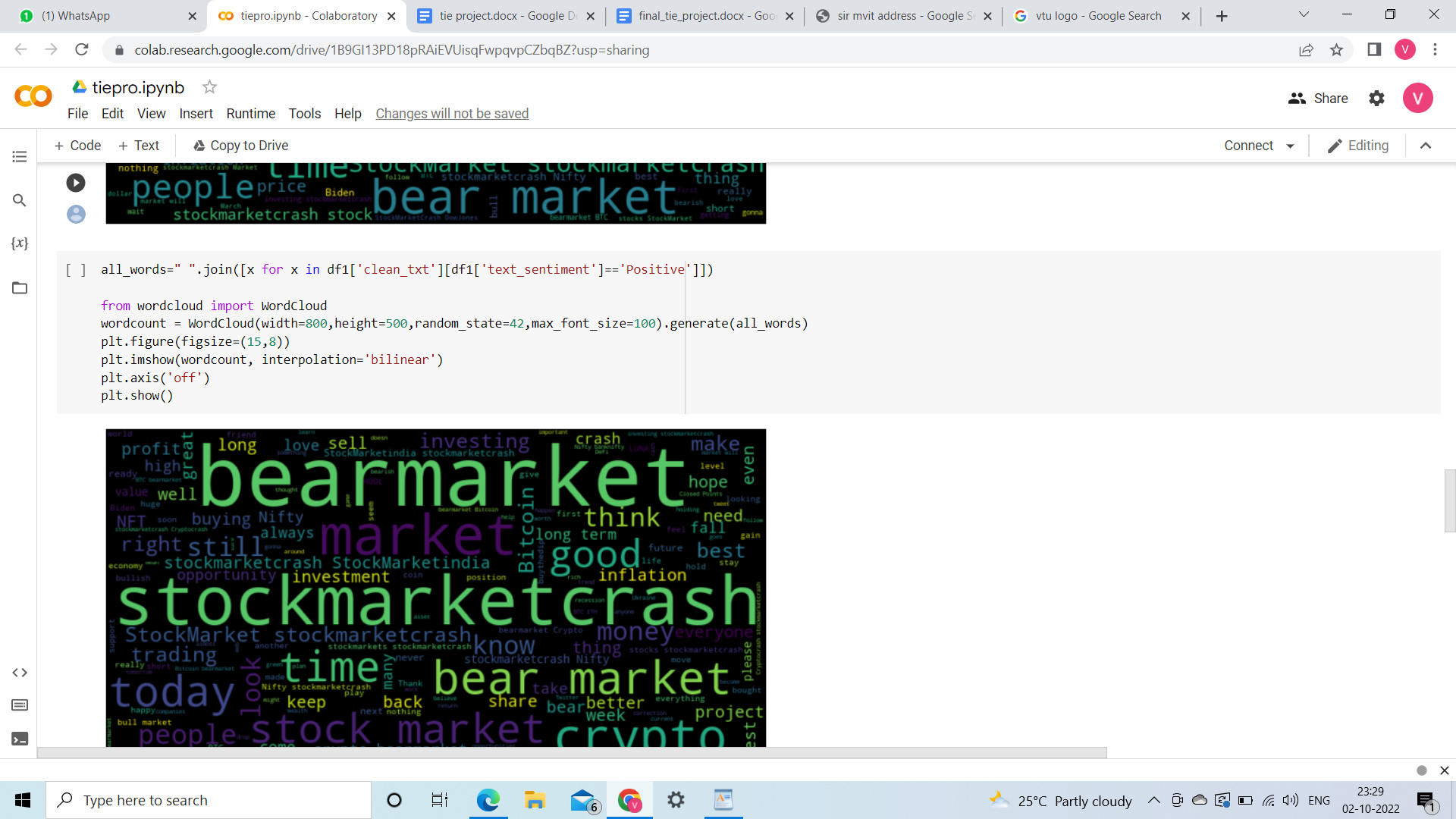


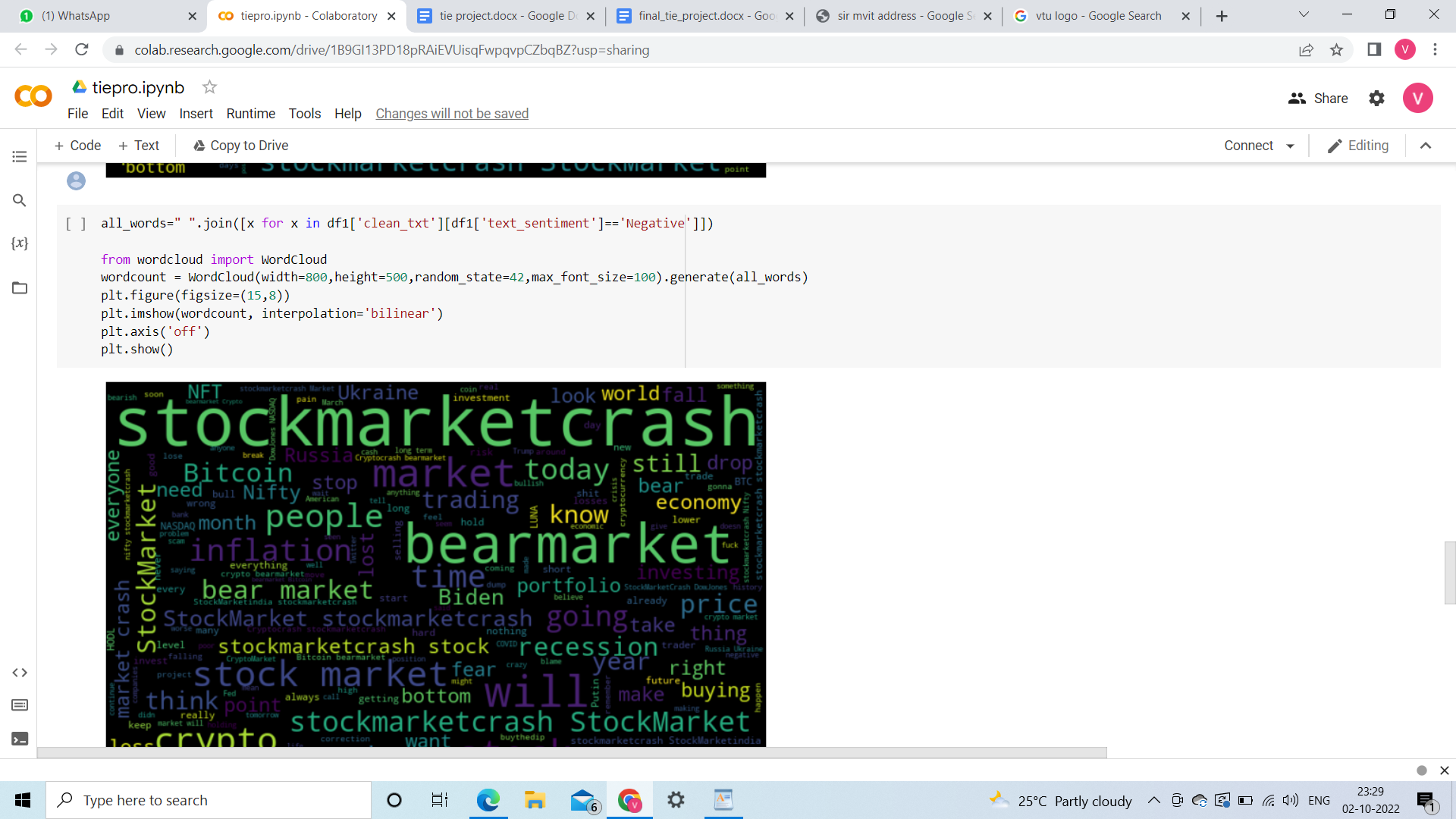


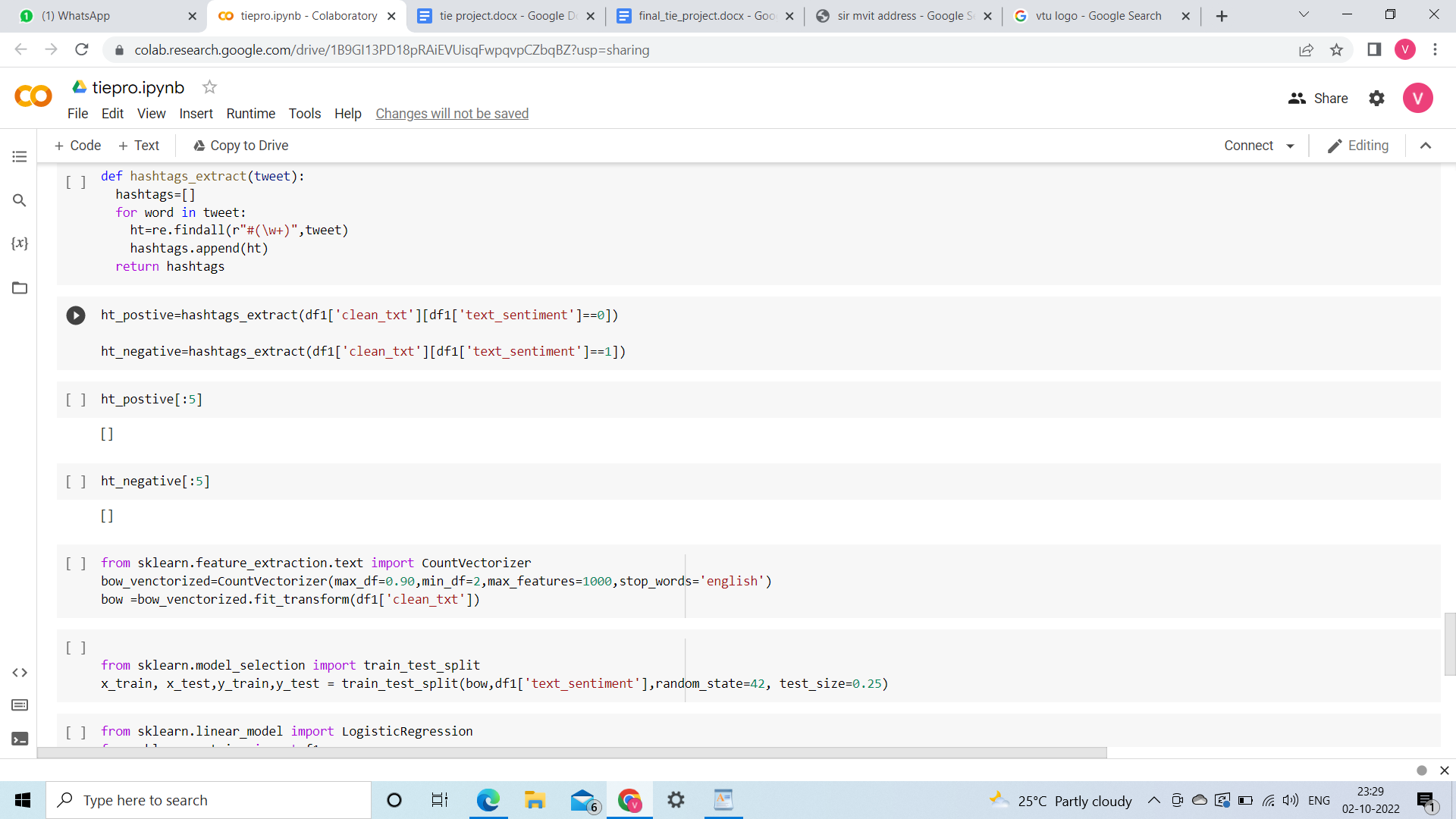


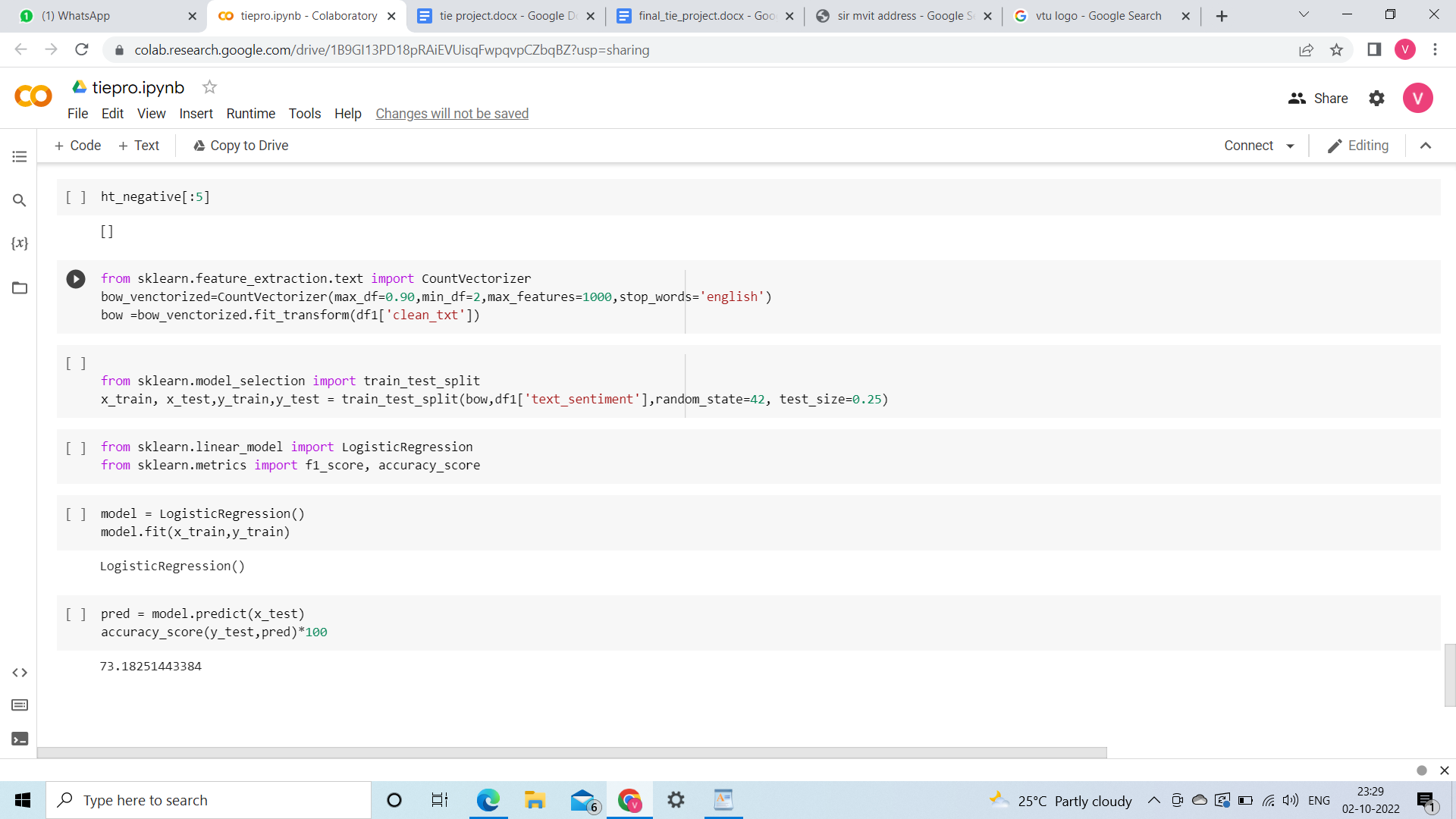












**CHAPTER 7**

**Conclusion**

In this report, we have shown that a strong correlation exists between the rise/fall in stock prices of a company to the public’s opinions or emotions about that company expressed on twitter through tweets. The main contribution of our work is the development of a sentiment analyzer that can judge the type of sentiment present in the tweet. The tweets are classified into three categories: positive, negative and neutral. At the beginning, we claimed that positive emotions or sentiment of the public in twitter about a company would reflect in its stock price. Our speculation is well supported by the results achieved and seems to have a promising future in research.

**CHAPTER 8**

**Reference**

* <https://www.kaggle.com/datasets/tejasurya/huge-stock-market-crash-2022>
* <https://youtu.be/4OlvGGAsj8I>
* <https://youtu.be/RLfUyn3HoaE>