VISVESVARAYA TECHNOLOGICAL UNIVERSITY



BELAGAVI – 590018,Karnataka

INTERNSHIP REPORT

ON

"Sentiment Analysis Of Lockdown In USA During Covid-19 A Case Study On Twitter using ML"

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

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

Submitted by:

Sinchana R

4GW20CS095



Conducted at Varcons Technologies Pvt Ltd

GSSS INSTITUTE OF ENGINEERING AND TECHNOLOGY FOR WOMEN DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GSSS Institute of Engineering & Technology for Women KRS Road, Metagalli, Engineering Campus, Mysore- 570016

GSSS INSTITUTE OF ENGINEERING AND TECHNOLOGY FOR WOMEN

GSSS Institute of Engineering and Technology for Women KRS Road, Metagalli, Engineering Campus, Mysore -570016



CERTIFICATE

This is to certify that the Internship titled "Sentiment Analysis Of Lockdown In USA During Covid-19 A Case Study On Twitter using ML" carried out by Sinchana R, a bonafide student of GSSS Institute of Engineering and Technology for Women in partial fulfillment for the award of Bachelor of Engineering, in Computer Science and Engineering, under Visvesvaraya Technological University, Belagavi, during the year 2023-2024. It is certified that all corrections/suggestions indicated have been incorporated in the report.

Signature of Guide	Signature of HOD	Signature of Principal
	F 4 1 1 7.°	
	External Viva :	
Name of the Examiner		Signature with Date
1)		2)

DECLARATION

I, Sinchana R (4GW20CS095), final year student of Computer Science and Engineering, GSSS Institute of Engineering and Technology for Women-570 016, 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 Computer Science and Engineering, during academic year 2023-2024.

Date: September 21, 2023

Place: Bangalore.

USN: 4GW20CS095

NAME: Sinchana R





Date: 11th August, 2023

Name: Sinchana R USN: 4GW20CS095

Dear Student,

We would like to congratulate you on being selected for the Machine Learning With Python (Research Based) Internship position with Varcons Technologies, effective Start Date 11th August, 2023, 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

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.

I express our sincere thanks to our Principal Dr. Shivakumar M, for providing usadequate facilities to undertake this Internship.

I would like to thank our Head of Dept CSE Dr. Raviraj P, for providing us an opportunity to carry out Internship and for his valuable guidance and support.

I express our deep and profound gratitude to our guide, for her keen interest and encouragement at every step in completing the Internship.

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

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

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

Sinchana R 4GW20CS095

ABSTRACT

The COVID-19 pandemic of 2020 prompted governments worldwide to implement lockdown measures to curb the spread of the virus. This study focuses on analyzing public sentiment during the lockdown period in the United States, utilizing Twitter data and machine learning techniques. The objective is to gain insights into how people expressed their sentiments on the social media platform in response to these unprecedented circumstances.

The methodology involves a multi-step process, starting with the collection and preprocessing of a substantial corpus of tweets related to the COVID-19 lockdown. The text data is cleaned, tokenized, and transformed into numerical features, enabling the application of various machine learning models. These models are trained and evaluated on labeled sentiment data to accurately classify tweets into positive, negative, or neutral sentiment categories.

The results of this sentiment analysis provide a comprehensive view of how the public's emotions evolved throughout the lockdown period. By analyzing sentiment trends over time, geographical variations, and specific lockdown-related topics, this study sheds light on the emotional impact of the pandemic on the American population.

TABLE OF CONTENTS

SI no	Description	Page no
1	Company Profile	1
2	About the company	2
3	Introduction	4
4	System Analysis	5
5	Flow Chart	7
6	Code	8
7	Snapshots	11
8	Advantages and disadvantages	14
8	Conclusion	15
9	Reference	16

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 trainings and much more. At VCT, they make sure every product/service that they offer is built keeping in mind the practical usability of the product/Service, they are a startup focused on Creativity and Customizability, and they also provide subscription models for Software that they have already built, Since the application is already configured, the user has a ready-to-use application. This not only reduces installation and configuration time but also cuts down the time wasted on potential glitches linked to software deployment. Smart solutions are at the core of all that we do at VCT. Our main goal is to find smart ways of using technology that will help build a better tomorrow for everyone, everywhere. SaaS offers a variety of advantages over traditional software licensing models and We here at VCT tend to include the key features of SaaS in everything we build. They understand that the best desired output can be achieved only by understanding the clients demand better. Varcons Technologies work with their clients and help them to define their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solution consulting group comprising 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.

2. ABOUT THE COMPANY



At Varcons Technologies Pvt Ltd, they make sure every product/service that they offer is built keeping in mind the practical usability of the product/Service, they are start-up focused on Creativity and Customizability, and They also provide subscription models for Software that they have already built, Since the application is already configured, the user has a ready-to-use application. This not only reduces installation and configuration time but also cuts down the time wasted on potential glitches linked to software deployment.

they create API's and tools that help you automate any process with a host of features pertaining to the Device. They believe that with the Right Software, Service and Analytics, Great Things Can Happen

Services of Varcons Technologies Pvt.Ltd

Website as Software

They develop websites that behave and interact similar to sophisticated software.

Branding as Design

We offer professional Graphic design, Brochure design & Logo design. We are experts in crafting visual content to convey the right message to the customers. They also design custom wraps for your products(also known as package designing).

Search Engine Optimization

We help you manage your SEO campaign more efficiently and effectively. We help you gain market share by leveraging our expertise. our holistic approach to identify anything that may be hurting your traffic or rankings and show you just how to outrank the competition

Analytics and Research

They analyze the way your users/customers interact with you/your business by gathering, studying and understanding the consumer voice and their perception of the product/service.

Comprehensive Customer Support

With a comprehensive range of services, we guarantee your technology needs are not just met, but exceeded. We shall work with your customers/users closely to understand the way your users/customers use/make use of products/services.

Embedded Systems and IOT

We work with Consumer Electronics, Lighting, Home Automation, Metering, Sensor-Technology, Home Appliance and Medical Device companies to help them create smart and connected products. Through its integrated Embedded and IoT services, VCA helps build intelligent & connected devices that can be remotely monitored and controlled while leveraging edge and cloud computing for a host of intelligent applications and analytics.

3. INTRODUCTION

The outbreak of COVID-19 caused heavy disruption to the everyday lives of people across the globe. In a country like with a large, diverse population like India, there are bound to be instances of mass hysteria and panic which are further fueled by unreliable and sometimes inaccurate data. Gauging the feelings/emotions of the citizens would provide insights into the public mindset and would pave the way for the government and many organizations to address these situations by providing them with the right data and information, eradicating fake news, thereby helping in suppressing unnecessary panic among the people. Social media acts as the bridge between the people, the government, and such organizations. The scope of this project lies in the application of sentiment analysis to the views expressed by people on social media, twitter, in this case, to analyze the trends in the dynamic mood of the population. Usually, the terms "fight" and "positive" are used in a negative and positive context respectively, but we observe a role reversal in this situation. The identification of such terms and their usage according to the context would be an essential part of the project. Also, the scope of the project can be found in stopping the spread of fake news related to the pandemic, creating an interactive dashboard that delivers information about the current situation, real-time sentiment analysis of tweets, trend analysis of various COVID-19 related hashtags, engagement on Twitter, overall sector-wise polarity score of the tweets and the public emotion charts.

PROBLEM STATEMENT

Built a python application that asks for a keyword and you need to identify the sentiment of that keyword using an open source dataset.

4. SYSTEM ANALYSIS

Initially, the IEEE Coronavirus (COVID-19) Tweets Data set was downloaded from their website. Upon inspection, it was found that many tweets did not have geo-locationtags, and also many were in different languages apart from English. Due to this challenge in obtaining proper data, a new data set named Geo-Tagged Coronavirus (COVID-19) Tweets Data set was obtained from the same website. These tweets were then hydrated using the "Hydrater" software and also a few python commands. Then, tweets in "English" and tweets from "India" were randomly chosen and a new dataset was created. Further, other data set containing COVID-19 related tweets from India were obtained from Kaggle. This data set was then cleaned and normalized to make it useful for further analysis.

This cleaned data were then subjected to further analysis by extracting bigrams, trigrams, and plotting frequency bar graphs, Word Clouds, Relationship Nexus, Boxplots, etc. This was done using both Python and R. Some Interactive Plots were also plotted. This complete process is termed as Exploratory Data Analysis.

After Exploratory Data Analysis is completed, the tweets are then Tokenized and are made in a format suitable for the Language Model. In this Step, two models were used:

- 1) Roberta Model: Transfer learning methods were implemented to carry out sentiment analysis. Sentiment Analysis of Tweets was carried out by integrating and using both the Hugging face Transformer Library and Fast AI. Further Slanted Triangular Learning Rates, Discriminate Learning Rate and even Gradual Unfreezing were used, as a result of which, state-of-the-art results were obtained rapidly without even tuning the parameters. The tokenized data was then passed through the RoBERT a model to perform Sentiment Analysis. This yielded a model with an accuracy of 97% over the data set. The Tweepy API was used to scrape tweets in real-time which were then passed through the model to obtain the sentiments.
- 2) RoBERTa-CNN Sentiment Extractor: After the completion of the sentiment analysis the data was further explored for the sentiment triggers in the tweets. Hugging Face transformers don't have a TFRoberta For Question Answering, for this purpose, a TFRobertaModel was created to convert trained data into arrays that the Roberta model can interpret. While training the Sentiment Extractor model, 5 stratified KFolds were used in such a way that, in each fold, the best model weights were saved and these weights were reloaded before carrying out testing and predictions. Roberta with CNN head was used for Twitter Sentiment Extraction. Thus after passing the data through this model we Internship report 2023-2024

obtained a new column of the extracted text for the sentiments which was also used to plot certain graphs.

3) Now the entire process pertaining to the data and Model building is completed. Now, the Flask APP is built for the purpose of Deployment. First, the application is deployed on the localhost and debugged and then we move on to deploying on the WebServers.

A flask app was used for setting up website routing. It is used to integrate the back end machine learning models with the dashboard. Then Socketio (web sockets) were used for dynamic implementations on the website, namely the Real-Time Plot Generators and Twitter live feed. The basic functionality of the Flask Socketio lies in running background threads when the client is not connected to the website thereby enabling dynamic plotting. The above built Dashboard was deployed on the Local Machine and debugged for any possible errors. The scraping rate and other parameters were monitored and corrected accordingly models with the dashboard. Then Socketio (web sockets) were used for dynamic implementations on the website, namely the Real-TimePlot Generators and Twitter live feed. The basic functionality of the Flask Socketio lies in running background threads when the client is not connected to the website thereby enabling dynamic plotting. The above built Dashboard was deployed on the LocalMachine and debugged for any possible errors. The scraping rate and other parameters were monitored and corrected accordingly.

REQUIREMENT ANALYSIS:

HARDWARE

System : 4Core Processors

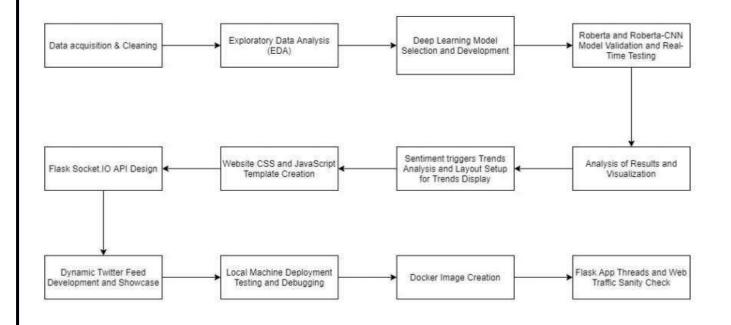
• Hard Disk : 142 GB

• Ram : 4 GB

SOFTWARE

- Jupyter Notebook
- Visual Studio Code
- Python Version 3.11.0

5. FLOW CHART



6. CODE:

#import libraries import numpy as np import pandas as pd import matplotlib.pyplot as plt import seaborn as sns import string from wordcloud import WordCloud import nltk nltk.download('all',quiet=True) from PIL import Image #Model libraries from sklearn.metrics import * from sklearn.model_selection import train_test_split **from** sklearn.linear_model **import** LogisticRegression from sklearn.tree import DecisionTreeClassifier from xgboost import XGBClassifier from sklearn.model_selection import GridSearchCV from sklearn.svm import SVC from sklearn.neighbors import KNeighborsClassifier **from** sklearn.metrics **import** confusion_matrix,classification_report from sklearn.model_selection import cross_val_score import warnings warnings.filterwarnings('ignore') #checking Head df1.head()

import string

def remove_punctuations(text):

for punctuation **in** string.punctuation:

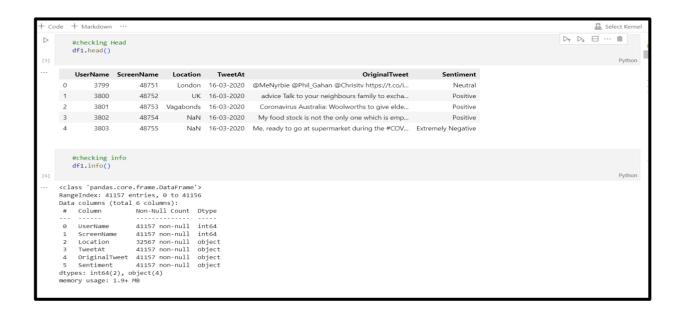
Internship report 2023-2024

```
text = text.replace(punctuation, ")
return text
df["clean_tweets"] = df['OriginalTweet'].apply(remove_punctuations)
df['clean_tweets']
from nltk.stem.porter import *
stemmer = PorterStemmer()
#function for stemming
def stemming(text):
  text = [stemmer.stem(word) for word in text]
  return (" ".join(text))
df['stemmed'] = df['clean_tweets'].apply(lambda x: stemming(x))
#result
df.stemmed.head()
# replacing values
replace_values = { "Sentiment": { 'Extremely Negative': 'Negative', 'Extremely Positive': } }
df = df.replace(replace_values)
sentiment_count1 = df['Sentiment'].value_counts().reset_index()
sentiment_count1.columns = ['Sentiment','count']
sentiment_count1
df['OriginalTweet'][0]
df['OriginalTweet'] = df['OriginalTweet'].str.replace('http\S+|www.\S+', ", case=False)
import string
def remove_punctuations(text):
  for punctuation in string.punctuation:
     text = text.replace(punctuation, ")
  return text
df["clean_tweets"] = df['OriginalTweet'].apply(remove_punctuations)
df['clean_tweets']
```

```
# Plotting the piechart for Sentiments distribution
sentiment_count1 = df['Sentiment'].value_counts().to_list()
labels=['Positive','Negative','Netural']
plt.figure(figsize=(10,8))
plt.pie(x=sentiment count1,explode=[0.04,0.04,0.1],shadow=
True,labels=labels,autopct="%.2f%%",radius=1.1)
plt.title("Proportion Of Sentiments", fontsize=20)
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left', borderaxespad=0)
plt.show()
#seperating the sentiments for word cloud
neutral = pd.DataFrame(df[['stemmed','lemmed']] [df['Sentiment'] == 'Neutral'])
positive = pd.DataFrame(df[['stemmed', 'lemmed']] [df['Sentiment'] == 'Positive'])
negative = pd.DataFrame(df[['stemmed','lemmed']] [df['Sentiment'] == 'Negative'])
In [40]:mask = np.array(Image.open("/content/drive/MyDrive/Colab Notebooks/Capstone Projects
/Twitter Sentiment Analysis/toppng.com-transparent-background-twitter-logo-943x800.png"))
wc = WordCloud(background_color='white',mask =
mask,contour_width=1,contour_color='steelblue')
wc.generate(str(neutral['lemmed']))
plt.figure(figsize=(20,20))
plt.imshow(wc,interpolation='bilinear')
plt.axis("off")
plt.show()
#Plotting Confussion matrix
cf1= (confusion_matrix(y_test,pred_lr_cv))
plt.figure(figsize=(8,5))
ax= plt.subplot()
sns.heatmap(cf1, annot=True, fmt=".0f",ax = ax)
# labels, title and ticks
ax.set_xlabel('Predicted labels', fontsize=15)
ax.set_ylabel('Actual labels', fontsize=15)
ax.set_title('Confusion Matrix (Logistic Regression with CV)', fontsize=20)
ax.xaxis.set_ticklabels(labels)
ax.yaxis.set_ticklabels(labels)
```

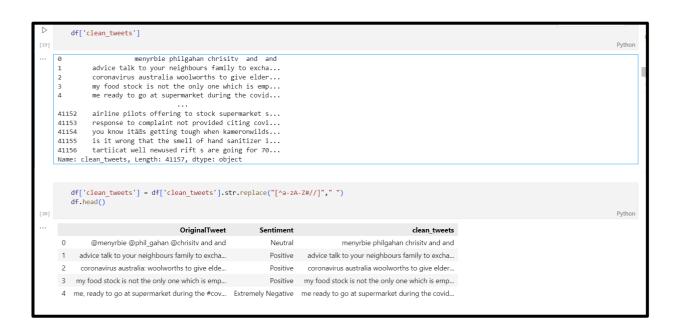
Internship report 2023-2024

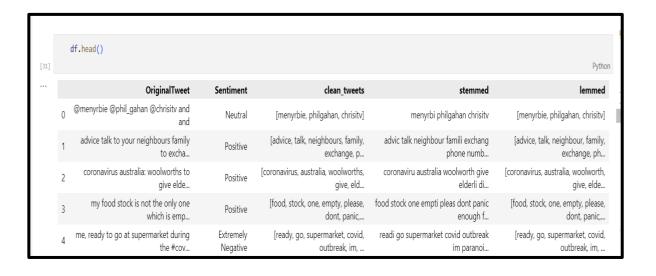
7. SNAPSHOTS:





Internship report 2023-2024





```
sentiment_count = df['Sentiment'].value_counts().reset_index()
sentiment_count.columns = ['Sentiment','count']
sentiment_count

Sentiment count

Python

Negative 11422

Negative 9917

Negative 9917

Negative 6624

Extremely Positive 6624

Extremely Negative 5481
```

```
▷, ▷, □ … ⑪
    # Accuracy
cv_score_dt_cv= cross_val_score(dt_cv,cv_X_train,y_train, cv=5)
print("Accuracy: {}" .format(np.mean(cv_score_dt_cv)))
                                                                                                                                                                   Python
Accuracy: 0.6914852657105394
    # Classification report of Performance metrics
    label=['Neutral','Positive','Negative']
    print(classification\_report(y\_test,pred\_dt\_cv))
               precision recall f1-score support
    Negative
Neutral
                     0.69
                               0.68
                                           0.68
                                                       4620
                     0.61
                                0.66
                                            0.64
                                                       2314
    Positive
                              0.74
                     0.75
                                          0.75
                                                       5414
    accuracy
                                            0.70
                                                      12348
macro avg
weighted avg
                     0.69
                                0.69
                                                      12348
                                            0.69
                                                      12348
                     0.70
                               0.70
                                            0.70
```

8. ADVANTAGES AND DISADVANTAGES:

ADVANTAGES:

- Real-time Insights: Twitter provides a vast amount of real-time data, allowing you to gauge
 public sentiment as events unfold, which can be crucial during a rapidly evolving situation like
 a pandemic.
- Large Dataset: Twitter has a massive user base, providing a large dataset for analysis, which can improve the accuracy of sentiment analysis models.
- Understanding Public Opinion: Sentiment analysis helps in understanding how people perceive and react to lockdown measures, which can inform policymakers and public health authorities.
- Monitoring Trends: You can track sentiment trends over time to see how opinions change in response to different phases of the lockdown or government policies.
- Targeted Response: Identifying negative sentiment can help authorities address concerns and improve communication to increase compliance with lockdown measures.

DISADVANTAGES:

- Noise in Data: Twitter data can be noisy due to spam, bots, or irrelevant content, which can negatively impact the accuracy of sentiment analysis.
- Bias and Sampling: Twitter users are not representative of the entire population, and there can be sampling bias, as not everyone uses the platform.
- Limited Context: Twitter's character limit constrains the depth of opinions shared, making it challenging to capture nuanced sentiment accurately.
- Ambiguity and Sarcasm: Sentiment analysis may struggle to interpret sarcasm, irony, or ambiguity in tweets, leading to misclassification.
- Lack of Demographic Information: Twitter data often lacks demographic information, making it difficult to understand sentiment variations among different groups.
- Ethical Concerns: Analyzing personal data on a public platform without consent raises privacy and ethical concerns, which must be carefully addressed.
- Model Dependence: The accuracy of sentiment analysis heavily depends on the quality of the machine learning model and training data, which can be challenging to optimize.

9. CONCLUSION

Sentiment analysis or opinion mining is a hot topic in deep learning. There is still a long way to go before sentiments can be accurately detected from texts, because of the complexity involved in the English language, and even more when other languages like Hindi are considered. Though the Roberta model developed as a partof this project has predicted and classified the sentiments of the test data set into positive, negative and neutral categories with an accuracy of 97%, by making necessary modifications and additions to the model, sentiment analysis can be donewith greater accuracy by taking the language complexities into consideration.

10. REFERENCE

- [1] https://www.hindawi.com/journals/cin/2022/8898100/
- [2] https://www.frontiersin.org/articles/10.3389/fpubh.2021.812735/full
- [3] https://iopscience.iop.org/article/10.1088/1742-6596/1828/1/012104
- [4] https://ieeexplore.ieee.org/document/9612431