CAPSTONE PROJECT WORK REPORT

Phase II

Stress Detection with Machine Learning

ABINESH G

A report submitted in part fulfilment of the degree of

B.Sc. in Computer Science with Data Analytics

Supervisor: **Mrs. JAYAPRIYA .P ,** M.C.A.,M.E.,(Ph.D)
Associate professor Dept. of CS with DA

Department of Computer Science with Data Analytics

KPR College of Arts Science and Research
(Affiliated to Bharathiar University, Coimbatore)
Avinashi Road, Arasur, Coimbatore – 641 407

NOVEMBER 2022

CAPSTONE PROJECT WORK REPORT

Phase II

Stress Detection with Machine Learning

Bonafide Work Done by

ABINESH G

REG. NO:2028B0001

KPR College of Arts Science and Research
Learn Beyond
Avinashi Road, Arasur, Coimbatore.

Dissertation submitted in partial fulfillment of the requirements for the award of Bharathiar University, Coimbatore-46.

Signature of the Guide **Signature of the HOD**

[Mrs.JAYAPRIYA.P]

Submitted for the Viva-Voce Examination held on _____

Internal Examiner **External Examiner**

CONTENTS

ACKNOWLEDGEMENT

In the accomplishment of completion of my Capstone Project Work Phase -II on **Stress Detection with Machine Learning** I would like to convey my special gratitude to **Dr. S. Balusamy, Principal of KPR College of Arts Science and Research** and **Mrs.JAYAPRIYA.P, Assistant Professor and Department of Computer Science with Data Analytics**. Your valuable guidance and suggestions helped me in phase - II of the completion of this project. I will always be thankful to you in this regard. I am ensuring that this project was finished by me and not copied.

Student Signature

Place:

Date:

KPR COLLEGE OF ARTS SCIENCE AND RESEARCH

(Affiliated to Bharathiar University, Coimbatore)

Avinashi Road, Arasur, Coimbatore – 641 407

ABOUT THE COLLEGE

KPR College of Arts Science and Research is the latest addition to the KPR fleet. The College is located in a picturesque campus of about 11. Acres. The College is run by KPR charities under the leadership of our Chairman Dr. K.P. Ramasamy. The KPR Group is one of the largest industrial conglomerate in the country with interest in Textiles, Sugar, Wind Turbines, Automobiles and Education. The College was established in the year 2019 with a vision of providing top class education and life skills to students and thereby serve the nation and beyond. KPRCAS today offers 12 UG programmes in Management, Commerce and Computer Science streams. The Students of KPRCAS undergo intense training not only in the syllabus and curriculum of the affiliating University but are also trained in various areas. So that they emerge as industry ready graduates to meet the varying demands of the competing industries. Character building and Leadership qualities are inculcated into the students to make them responsible citizens focusing on the development of society and nation. A plethora of Clubs and Events encouraged the students to take part in sports and other cultural activities. KPRCAS offers three years undergraduate courses, which are exclusively for Business, Commerce and Computer Science Stream. The students are equipped with skills and knowledge needed to take up various leadership positions and to develop the society. Beyond Book Teaching help them to be professionals. KPRCAS emphasis on making the students academically brilliant, and also prepare them for the real corporate world. The learning curve begins here for the students of KPRCAS.

ABOUT THE DEPARTMENT

Bachelor of Computer Science with Data Analytics (B.Sc. (CS with DA)) was established in the year 2020. Data Analytics helps to raise the quality of data in the entire business system. The goal of data analytics is to construct the means for extracting business-focused insights from data This requires an understanding of how value and information flows in a business, and the ability to use that understanding to identify business opportunities. The primary aim of a data analyst is to increase efficiency and improve performance by discovering patterns in data. Data analysts exist at the intersection of information technology, statistics and business. They combine these fields in order to help businesses and organizations succeed. The students get exposed to Big Data, Business Intelligence, Data Mining, Data Visualization, Advanced Excel, Predictive Analytics and R Programming.

SYNOPSIS

Stress, anxiety, and depression are threatening the mental health of people. Every person has a reason for having a stressful life. People often share their feelings on social media platforms like on Instagram in the form of posts and stories, and on Reddit in the form of asking for suggestions about their life on subreddits. In the past few years, many content creators have come forward to create content to help people with their mental health. Many organizations can use stress detection to find which social media users are stressed to help them quickly. So if you want to learn how to use machine learning to detect stress on social media posts, this article is for you. In this article, I will take you through the task of stress detection with machine learning using Python.

CHAPTER -I

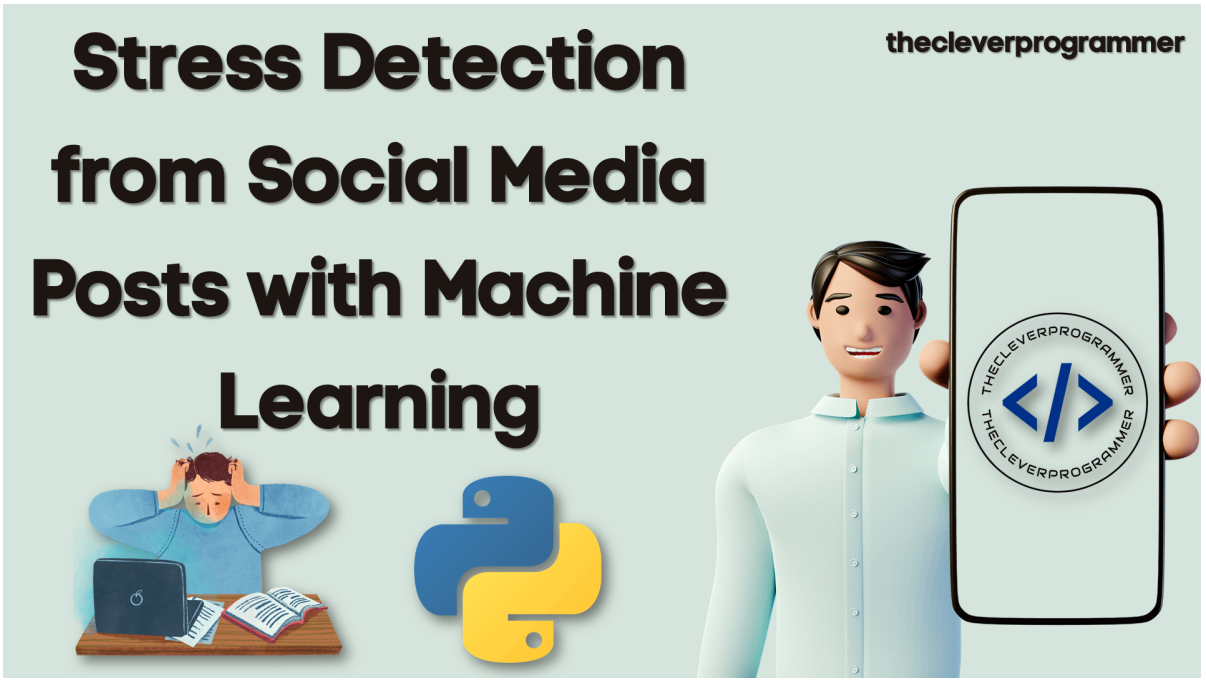
1. INTRODUCTION

1.1 Stress Detection with Machine Learning:

Stress detection is a challenging task, as there are so many words that can be used by people on their posts that can show whether a person is having psychological stress or not. While looking for datasets that I can use to train a machine learning model for stress detection, I found a dataset on Kaggle with 116 columns. We only need to use the text and label column for this task.

The data set I am using for this task contains data posted on sub reddits related to mental health. This data set contains various mental health problems shared by people about their life. Fortunately, this data set is labelled as 0 and 1, where 0 indicates no stress and 1 indicates stress. So in the section below, I will take you through the task of stress detection in social media posts using Python.

.



1.2 Stress impact on society and economy:

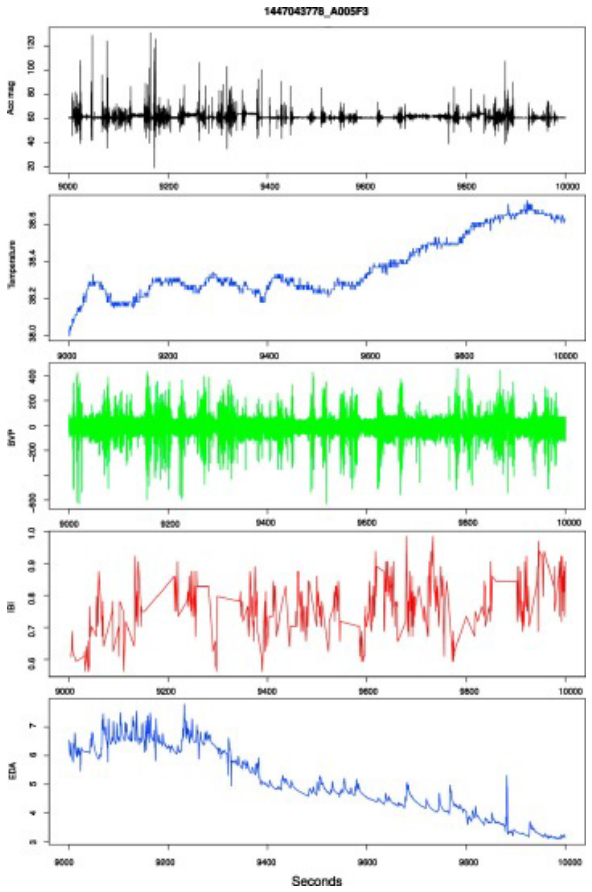
Daily life stress has become a significant issue for the modern society. Offices among other places contribute to the high stress most . The mismatch between job demands and abilities, time pressure and high workloads are general reasons for the office stress. Family-related issues, illnesses and chronic injuries and emotional problems can be listed as off-the-workplace stress causes. Stress is the second most severe work-related health issue in Europe , after musculoskeletal illnesses which can be caused by stress in some cases . The American Institute of Stress reveals that the US spends 300 billion USD per year to diseases caused by stress . In 2013, work-related stress cost 25 billion euros to the EU businesses . Recent public survey revealed that 51% of European workers are exposed to stress at workplaces. It is anticipated that 50–60% of all lost working days in the European business sector caused by work-related stress and psychosocial risks.

There are two types of stress: acute and chronic . Acute stress is more common and the majority of people have experienced this kind of stress. American Psychological Association stated that demands and pressure from recent past and near future cause acute stress . The potential triggers for acute stress can be listed as athletic challenges, test taking, or anxiety when meeting new people. The causes of chronic stress can be counted as long-standing pressures and demands as a result of socioeconomic conditions, difficulties in interpersonal relationships, or an unsatisfying career . The consequences of chronic stress can be destructive if left unmanaged . Since the short-term symptoms of acute stress are more observable, researchers have investigated acute stress more than chronic stress. Furthermore, since the duration of acute stress is shorter, subjects are less affected when they are induced in the experiments when compared to chronic stress.

1.3 Stress Signals:

The Sympathetic Nervous System (SNS) ignites the stress reaction , resulting in psychological, physiological and behavioral symptoms . The psychological way of measuring stress can be self-report questionnaires or being interviewed by a psychologist . Therefore, automatic stress detection topics do not include this class.

The second way to detect stress is by evaluating physiological signals . They include information related to intensity and quality of the affect and experience of the subject . Signals of interest include hormone levels, ElectroCardiogram (ECG), Electroencephalogram (EEG), Electro-Dermal Activity (EDA), Blood Pressure (BP), Skin Temperature (ST), Electromyogram (EMG), Respiration, Blood Volume Pulse (BVP), Pupil Diameter (PD), Eye Gaze and Blinking, Thermal Imaging (TI) and functional Magnetic Resonance Imaging (fMRI). The other two methods are investigating behavioral data and context information which are not investigated thoroughly in the literature



CHAPTER-II

2. SYSTEM SPECIFICATION

2.1. Hardware Configuration

Operating System	Self-Hosted Technical Requirement	Cloud Technical Requirement
Windows	Windows 8.1+	Windows 8.1+
Mac	Mac OS 10.14+	Mac OS 10.14+
Linux	Ubuntu LTS releases 18.04 or later	Ubuntu LTS releases 18.04 or later
RAM	8 GB	
HDD	1 TB	
Processor	64-bit, four-core, 2.5 GHz minimum per core (If your dataset size is significantly larger than the medium dataset, we recommend 8 cores.)	
Mouse	Dell MS116 1000DPI USB Wired Optical Mouse	
Keyboard	Dell KB522 Business Keyboard-Black	
Monitor	Dell 24 Monitor-S2421HN in-Plane Switching (IPS)	

2.3. Software Configuration

IDE	Anaconda
Language Support	Python 3.9
Platform	Jupyter Notebook
Browser	Google Chrome Version 101.0.4951.67
Database	MySQL 8.0.29

CHAPTER-III

3 .SYSTEM STUDY

3.1 Existing System:

Existing systems were designed to detect stress by taking tweets as input from the Twitter or Facebook data set and machine learning algorithms are applied to detect stress from tweets**.**

3.2 Drawbacks:

- Most of the existing system works were on social networking stress data not on body-based sensor data.
- Stress level is calculated based on tweets posted by users.

3.3 Proposed system:

- The proposed system is designed by collecting data from sensors and preparing data set on three features (temperature, heartbeat, age, male or female).
- Using this data set machine learning Decision tree algorithm is applied using and the model is saved.

3.4 Features:

- Data is collected from real-time sensors and a data set is created for different ages and male and female users.
- Data is trained using machine learning which helps automate the process of stress detection.
- The web applications can help users to easily check their stress state based on their features.

CHAPTER-IV

4.SYSTEM DESIGN:

4.1 Stress Detection using Python

Now let's start the task of stress detection with machine learning. I will start this task by importing the necessary Python libraries and the dataset .

4.1.1.Importing Libraries

```
In [1]: import pandas as pd
        import numpy as np
        data = pd.read_csv("file:///E:/ML_projects/Stress_Detect/stress.csv")
        print(data.head())
```

```
          subreddit post_id sentence_range  \
0              ptsd  8601tu       (15, 20)
1        assistance  8lbrx9         (0, 5)
2              ptsd  9ch1zh       (15, 20)
3     relationships  7rorpp        [5, 10]
4  survivorsofabuse  9p2gbc         [0, 5]

                                                text     id  label  \
0  He said he had not felt that way before, sugge...  33181      1
1  Hey there r/assistance, Not sure if this is th...   2606      0
2  My mom then hit me with the newspaper and it s...  38816      1
3  until i met my new boyfriend, he is amazing, h...    239      1
4  October is Domestic Violence Awareness Month a...   1421      1

   confidence  social_timestamp  social_karma  syntax_ari  ...  \
0         0.8        1521614353             5    1.806818  ...
1         1.0        1527009817             4    9.429737  ...
2         0.8        1535935605             2    7.769821  ...
3         0.6        1516429555             0    2.667798  ...
4         0.8        1539809005            24    7.554238  ...
```

4.1.2.Dataset contains any null values or not:

```
In [2]: print(data.isnull().sum())
```

```
subreddit                  0
post_id                    0
sentence_range             0
text                       0
id                         0
                          ..
lex_dal_avg_pleasantness   0
social_upvote_ratio        0
social_num_comments        0
syntax_fk_grade            0
sentiment                  0
Length: 116, dtype: int64
```

4.1.3.Text Coulmn of Dataset:

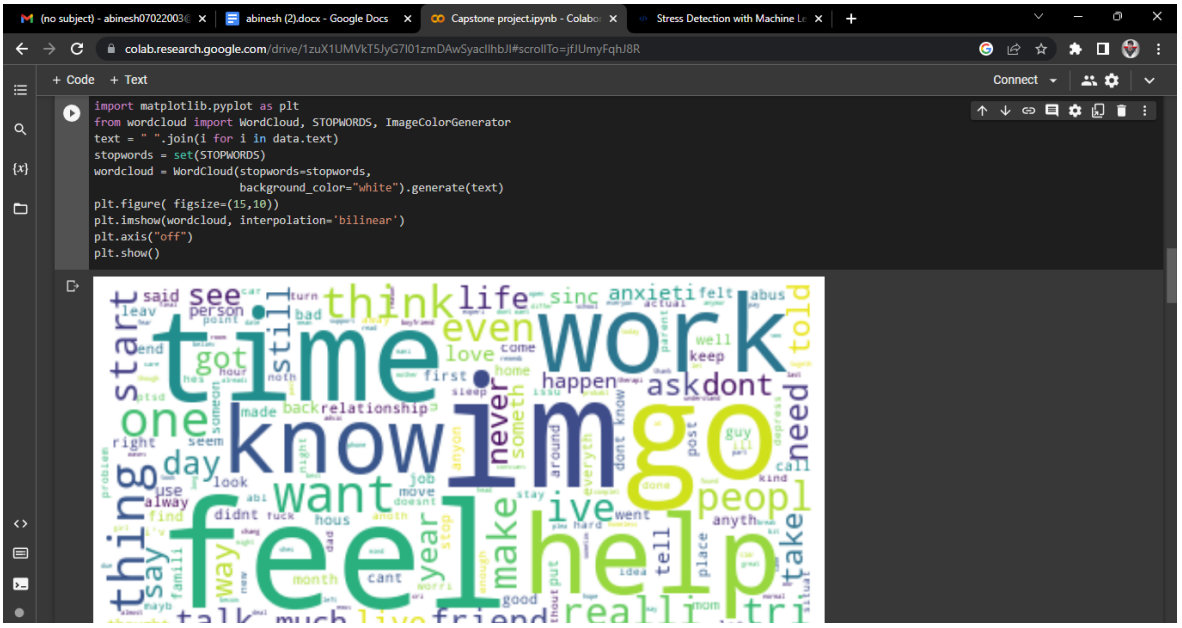
So this dataset does not have any null values. Now let's prepare the text column of this dataset to clean the text column with stopwords, links, special symbols and language errors:

```
In [3]: import nltk
        import re
        nltk.download('stopwords')
        stemmer = nltk.SnowballStemmer("english")
        from nltk.corpus import stopwords
        import string
        stopword=set(stopwords.words('english'))

        def clean(text):
            text = str(text).lower()
            text = re.sub('\[.*?\]', '', text)
            text = re.sub('https?://\S+|www\.\S+', '', text)
            text = re.sub('<.*?>+', '', text)
            text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
            text = re.sub('\n', '', text)
            text = re.sub('\w*\d\w*', '', text)
            text = [word for word in text.split(' ') if word not in stopword]
            text=" ".join(text)
            text = [stemmer.stem(word) for word in text.split(' ')]
            text=" ".join(text)
            return text
        data["text"] = data["text"].apply(clean)
```

4.1.4.Word Cloud of Text Column:

Now let's have a look at the most used words by the people sharing about their life problems on social media by visualizing a **word cloud** of the text column:



4.2 Stress Detection Model:

The label column in this dataset contains labels as 0 and 1. 0 means no stress, and 1 means stress. I will use Stress and No stress labels instead of 1 and 0. So let's prepare this column accordingly and select the text and label columns for the process of training a machine learning model:

4.2.1.Testing:

```
In [5]: data["label"] = data["label"].map({0: "No Stress", 1: "Stress"})
        data = data[["text", "label"]]
        print(data.head())
```

```
                                                text      label
0  said felt way sugget go rest trigger ahead you...     Stress
1  hey rassist sure right place post goe  im curr...  No Stress
2  mom hit newspap shock would know dont like pla...     Stress
3  met new boyfriend amaz kind sweet good student...     Stress
4  octob domest violenc awar month domest violenc...     Stress
```

Now I will split this dataset into training and test sets:

```
In [6]: from sklearn.feature_extraction.text import CountVectorizer
        from sklearn.model_selection import train_test_split

        x = np.array(data["text"])
        y = np.array(data["label"])

        cv = CountVectorizer()
        X = cv.fit_transform(x)
        xtrain, xtest, ytrain, ytest = train_test_split(X, y,
                                                        test_size=0.33,
                                                        random_state=42)
```

4.2.2.Naive Bayes Algorithm:

In the term of machine learning, naive Bayes classifiers consists a group of straightforward "probabilistic classifiers". They work upon the probability, highly scalable. Naive Bayes classifiers are quite adaptable. They require various parameters that are straight in the number of factors (highlights/indicators) in learning issue [10].

As this task is based on the problem of binary classification, I will be using the Bernoulli Naive Bayes algorithm, which is one of the best algorithms for binary classification problems. So let's train the stress detection model:

```
In [7]: from sklearn.naive_bayes import BernoulliNB
        model = BernoulliNB()
        model.fit(xtrain, ytrain)

Out[7]: BernoulliNB()
```

4.2.3.Performance of model on Mental health:

Now let’s test the performance of our model on some random sentences based on mental health:

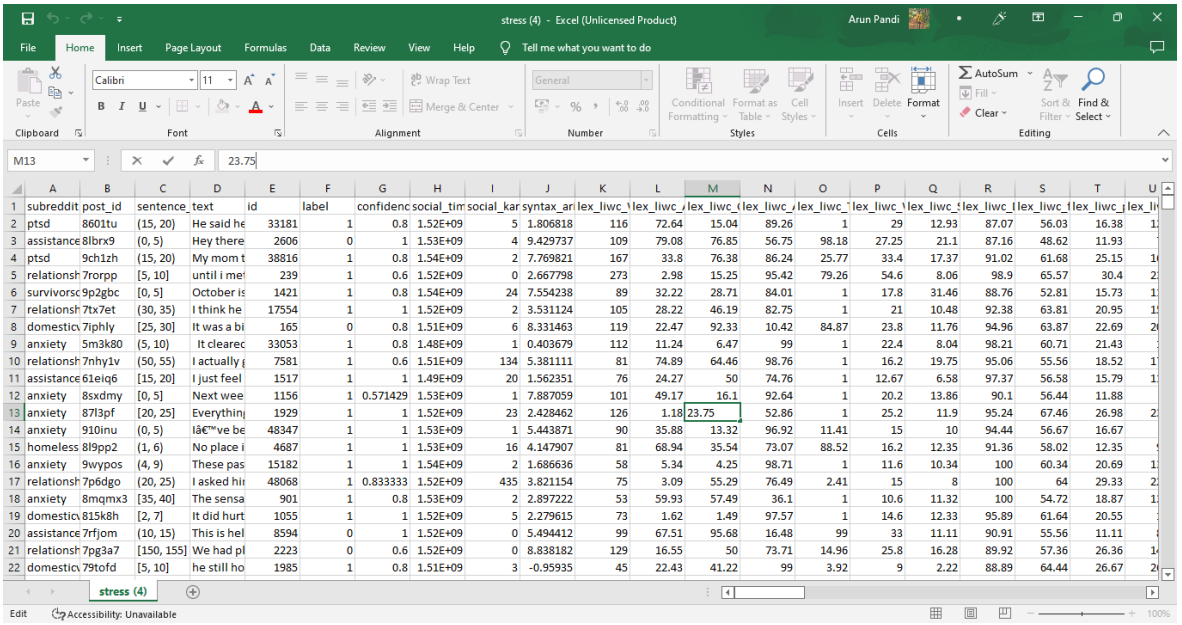
In [10]:
```
user = input("Enter a Text: ")
data = cv.transform([user]).toarray()
output = model.predict(data)
print(output)
```

```
Enter a Text: Sometime I feel like I need some help
['Stress']
```

In [11]:
```
user = input("Enter a Text: ")
data = cv.transform([user]).toarray()
output = model.predict(data)
print(output)
```

```
Enter a Text: Sometime I feel like I need some help
['Stress']
```

4.3 Database Design:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	subreddit	post_id	sentence_	text	id	label	confidenc	social_tim	social_kar	syntax_ar	lex_liwc_	lex_liwc_	lex_liwc_	lex_liwc_	lex_liwc_	lex_liwc_	lex_liwc_	lex_liwc_	lex_liwc_	lex_liwc_	lex_li
2	ptsd	8601tu	(15, 20)	He said he	33181	1	0.8	1.52E+09	5	1.806818	116	72.64	15.04	89.26	1	29	12.93	87.07	56.03	16.38	1
3	assistance	8lbrx9	(0, 5)	Hey there	2606	0	1	1.53E+09	4	9.429737	109	79.08	76.85	56.75	98.18	27.25	21.1	87.16	48.62	11.93	
4	ptsd	9ch1zh	(15, 20)	My mom t	38816	1	0.8	1.54E+09	2	7.769821	167	33.8	76.38	86.24	25.77	33.4	17.37	91.02	61.68	25.15	1
5	relationsh	7rorpp	[5, 10]	until i met	239	1	0.6	1.52E+09	0	2.667798	273	2.98	15.25	95.42	79.26	54.6	8.06	98.9	65.57	30.4	2
6	survivorsc	9p2gbc	[0, 5]	October is	1421	1	0.8	1.54E+09	24	7.554238	89	32.22	28.71	84.01	1	17.8	31.46	88.76	52.81	15.73	1
7	relationsh	7tx7et	(30, 35)	I think he	17554	1	1	1.52E+09	2	3.531124	105	28.22	46.19	82.75	1	21	10.48	92.38	63.81	20.95	1
8	domestic\	7iphly	[25, 30]	It was a bi	165	0	0.8	1.51E+09	6	8.331463	119	22.47	92.33	10.42	84.87	23.8	11.76	94.96	63.87	22.69	2
9	anxiety	5m3k80	(5, 10)	It clearec	33053	1	0.8	1.48E+09	1	0.403679	112	11.24	6.47	99	1	22.4	8.04	98.21	60.71	21.43	
10	relationsh	7nhy1v	(50, 55)	I actually g	7581	1	0.6	1.51E+09	134	5.381111	81	74.89	64.46	98.76	1	16.2	19.75	95.06	55.56	18.52	1
11	assistance	61eiq6	[15, 20]	I just feel	1517	1	1	1.49E+09	20	1.562351	76	24.27	50	74.76	1	12.67	6.58	97.37	56.58	15.79	1
12	anxiety	8sxdmy	[0, 5]	Next wee	1156	1	0.571429	1.53E+09	1	7.887059	101	49.17	16.1	92.64	1	20.2	13.86	90.1	56.44	11.88	
13	anxiety	87l3pf	[20, 25]	Everythin	1929	1	1	1.52E+09	23	2.428462	126	1.18	23.75	52.86	1	25.2	11.9	95.24	67.46	26.98	2
14	anxiety	910inu	(0, 5)	Iâ€™ve be	48347	1	1	1.53E+09	1	5.443871	90	35.88	13.32	96.92	11.41	15	10	94.44	56.67	16.67	
15	homeless	8l9pp2	(1, 6)	No place i	4687	1	1	1.53E+09	16	4.147907	81	68.94	35.54	73.07	88.52	16.2	12.35	91.36	58.02	12.35	
16	anxiety	9wypos	(4, 9)	These pas	15182	1	1	1.54E+09	2	1.686636	58	5.34	4.25	98.71	1	11.6	10.34	100	60.34	20.69	1
17	relationsh	7p6dgo	(20, 25)	I asked hir	48068	1	0.833333	1.52E+09	435	3.821154	75	3.09	55.29	76.49	2.41	15	8	100	64	29.33	2
18	anxiety	8mqmx3	[35, 40]	The sensa	901	1	0.8	1.53E+09	2	2.897222	53	59.93	57.49	36.1	1	10.6	11.32	100	54.72	18.87	1
19	domestic\	815k8h	[2, 7]	It did hurt	1055	1	1	1.52E+09	5	2.279615	73	1.62	1.49	97.57	1	14.6	12.33	95.89	61.64	20.55	
20	assistance	7rfjom	(10, 15)	This is hel	8594	0	1	1.52E+09	0	5.494412	99	67.51	95.68	16.48	99	33	11.11	90.91	55.56	11.11	
21	relationsh	7pg3a7	[150, 155]	We had pl	2223	0	0.6	1.52E+09	0	8.838182	129	16.55	50	73.71	14.96	25.8	16.28	89.92	57.36	26.36	1
22	domestic\	79tofd	[5, 10]	he still ho	1985	1	0.8	1.51E+09	3	-0.95935	45	22.43	41.22	99	3.92	9	2.22	88.89	64.44	26.67	2

5. CONCLUSION:

This is how you can train a machine learning model to detect stress from social media posts. People often share their feelings on social media platforms. Many organizations can use stress detection to find which social media users are stressed to help them quickly. I hope you liked this article on stress detection with machine learning using Python.This research is useful for the well-being of one's mental health. The results are evaluated using various metrics at the macro and micro levels and indicate that the trained model detects the status of emotions based on social interactions.

BIBLIOGRAPHY

- Philip Schmidt, A. Reiss, R. Duerichen, Kristof Van Laerhoven, “Introducing WESAD, a multimodal dataset for wearable Stress and Affect Detection”, International Conference on Multimodal Interaction 2018

- Enrique Garcia-Ceja, Venet Osmani and Oscar Mayora, “Automatic Stress Detection in working environments from smartphones’ accelerometer data: A First Step” , arXiv:1510.04221v1 [cs.HC] 14 Oct 2015.

- Saskia Koldijk , Mark A. Neerincx, and Wessel Kraaij, “Detecting Work Stress in offices by combining unobtrusive sensors”, IEEE Transactions on affective computing 2018.

- Virginia Sandulescu, Sally Andrews, David Ellis, “Stress Detection using wearable physiological sensors”, Springer International Publishing Switzerland 2015.

- Md Fahim Rizwan, Rayed Farhad, Farhan Mashuk, “Design of a biosignal based Stress Detection System using machine learning techniques”,2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST).

- Saskia Koldijk, Maya Sappelli, Suzan Verberne, “The SWELL Knowledge Work dataset for stress and user modeling research”, ICMI 2014.

- Jacqueline Wijsman, Bernard Grundlehner, Hao Liu, ”Towards Mental Stress Detection using wearable physiological sensors”, IEEE 2011.

- G. Giannakakis, M.Pediaditisa, D.Manousos, “Stress and anxiety detection using facial cues from videos”, Elsevier 2016.

- M. Gjoreski, H. Gjoreski, and M. Gams., ”Continuous stress detection using a wrist device: In laboratory and real life.” in UbiComp ’16.

- J. Kim and E. Andr´e, ”Emotion recognition based on physiological changes in music listening”, IEEE Transactions on Pattern Analysis and Machine Intelligence 30, 12 (2008), 2067–2083.

- K. Plarre, A. Raij, and M. Scott. ”Continuous inference of psychological stress from sensory measurements collected in the natural environment”, in 10th International Conference on Information Processing in Sensor Networks (IPSN). 97–108.

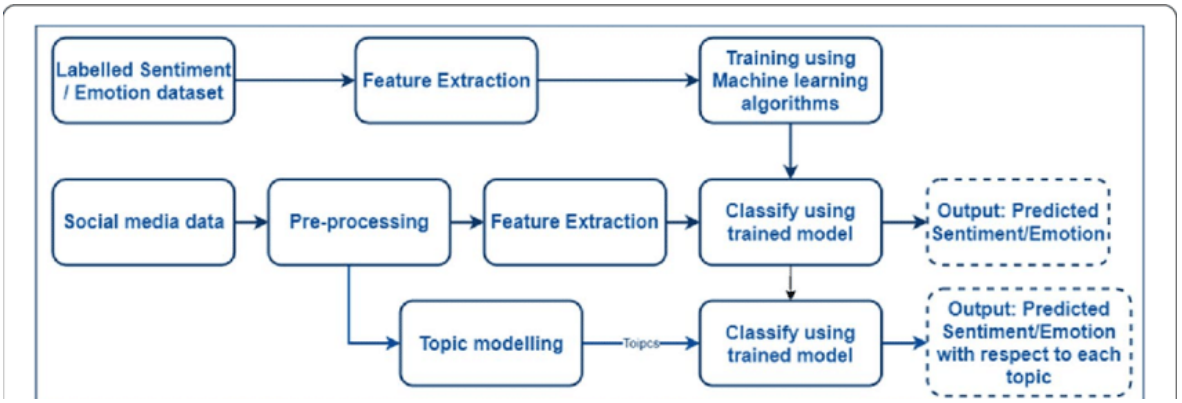
- Alexandros Zenonos, Aftab Khan, and Mahesh Sooriyabandara, ”HealthyOffice: Mood recognition at work using smartphones and wearable sensors”, in PerCom Workshops 2016.

- B. Padmaja, V. V. Rama Prasad and K. V. N. Sunitha, ”A Machine Learning approach for stress detection using a wireless physical activity tracker”, International Journal of Machine Learning and Computing, Vol.8, No. 1, February 2018.
- Global Organization for Stress on stress facts. http://www.gostress.com/stress-facts. Accessed: 2020-27-02.

- HSE on Work-related stress, anxiety or depression statistics in Great Britain, 2019. https://www.hse.gov.uk/statistics/causdis/stress.pdf. Accessed: 2020-27-02.

APPENDICES

A. Data Flow Diagram:



SOURCE CODE

```
 import pandas as pd
import numpy as np
data = pd.read_csv("stress.csv")
print(data.head())
print(data.isnull().sum())
import nltk
import re
nltk.download('stopwords')
stemmer = nltk.SnowballStemmer("english")
from nltk.corpus import stopwords
import string
stopword=set(stopwords.words('english'))

def clean(text):
    text = str(text).lower()
    text = re.sub('\[.*?\]', '', text)
    text = re.sub('https?://\S+|www\.\S+', '', text)
    text = re.sub('<.*?>+', '', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), '', text)
    text = re.sub('\n', '', text)
    text = re.sub('\w*\d\w*', '', text)
    text = [word for word in text.split(' ') if word not in stopword]
    text=" ".join(text)
    text = [stemmer.stem(word) for word in text.split(' ')]
```

```
    text=" ".join(text)
    return text
data["text"] = data["text"].apply(clean)
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
text = " ".join(i for i in data.text)
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords,
                      background_color="white").generate(text)
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
data["label"] = data["label"].map({0: "No Stress", 1: "Stress"})
data = data[["text", "label"]]
print(data.head())
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split

x = np.array(data["text"])
y = np.array(data["label"])

cv = CountVectorizer()
X = cv.fit_transform(x)
xtrain, xtest, ytrain, ytest = train_test_split(X, y,
```

```
                                    test_size=0.33,
                                    random_state=42)
from sklearn.naive_bayes import BernoulliNB
model = BernoulliNB()
model.fit(xtrain, ytrain)
user = input("Enter a Text: ")
data = cv.transform([user]).toarray()
output = model.predict(data)
print(output)
user = input("Enter a Text: ")
data = cv.transform([user]).toarray()
output = model.predict(data)
print(output)
```

*****