Project Report on
Stress detection

Submitted by
C. V. Shobha
R170889

Under the guidance of
P.Udayasree
Assistant Professor
Department of Computer Science and Engineering

Rajiv Gandhi University of Knowledge and Technologies (RGUKT),R.K.Valley ,Kadapa ,Andhra Pradesh.

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES

IIIT RK VALLEY-516213

DECLARATION

I hereby declare that the report of the B.Tech Major Project Work entitled with the "S**tress Detection**" which is being submitted to Rajiv Gandhi University of Knowledge Technologies,RK Valley,in partial fulfillment of the requirements for the award of Degree of Bachelor of Technology in Computer Science Engineering,is a bonafied report of the work carried out by me.The material contained in this report has not been submitted to any university or institution for awar of any degree.

I will be solely responsible if any kind of plagiarism is found.

Date: C. V. Shobha

Place: R170889

Rajiv Gandhi University of Knowledge Technologies

RK Vlley, , Kadapa (Dist), Andhra Pradesh, 516330

CERTIFICATE

This is to certify that the project work titled “STRESS DETECTION” is a bonafied project work submitted by C.V.Shobha (R170889) in the department of COMPUTER SCIENCE AND ENGINEERING in partial fulfillment of requirements for the award of degree of Bachelor of Technology in Computer science and engineering for the year 2022-2023 carried out the work under the supervision of Ms.Udayasree.

Satyanandaram N — Head of CSE department

Udayasree P — Internal guide

ACKNOWLEDGEMENT

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of the people who made it possible and whose constant guidance and encouragement crown all the efforts success.

I am extremely grateful to our Director.Prof.K. SANDHYA RANI for fostering an excellent academic climate in our institution.

I also express my sincere gratitude to our respected Head of the Department Mr.N.SATYANANDARAM for his encouragement, overall guidance in viewing this project a good asset and effort in bringing out this project.

I would like to convey thanks to our guide at college Ms.P.Udayasree for her guidance, encouragement, co-operation and kindness during the entire duration of the course and academics.

My sincere thanks to all the members who helped me directly and indirectly in the completion of project work. I express my profound gratitude to all our friends and family members for their encouragement.

INDEX

1. ABSTRACT

This project aims to detect whether an individual is in stress or not. Stress is a common part of everyday life that most people have to deal with on various occasions. Detecting mental stress earlier can prevent many health problems associated with stress**.** Nowadays people share their feeling in social media regularly With the increasing availability of textual data on social media platforms it becomes easy to detect the stress of the users based on their social behavior. Also, traditional stress detection methods are time consuming and costly. In this project,we detect the stress states of users by using their posts using sentiment analysis and Naive Bayes classification algorithm and are categorized into stressed and non-stressed user.

2.INTRODUCTION

Stress is something that concerns our lives. There are many variables in our day-to-day life that are tension.Human environments, like worksite, home, or society, may somehow inflict stress on a person.

According to Palmer [1], "Stress is defined as a complex psychological and behavioural condition when the person's demands are imbalanced and the way demands are met."

Also, the American Institute of Stress found that 80% of workers experience stress in their everyday work and need support in managing stress. Based on Ahuja and Banga [2], study recorded major suicide cases among students aged 15-29 due to stress. There are 8934 cases recorded in 2015, and study was inspired to identify stress in early stages. These figures and stress effects on people, which has been the leading cause of many diseases like hypertension, sleep deprivation, and others. Stress that cannot be adequately treated can lead to serious cases where one person committed suicide. This is vital to identify and control stress before it becomes severe.
In this project, we aim to develop a machine learning model for stress detection using textual data from social media.

3.Purpose

The purpose of this stress detection project is to develop a machine learning model that can accurately detect stress in social media posts using textual data. The project aims to explore various techniques for preprocessing textual data and extracting features from the text using NLP techniques such as bag-of-words and word embeddings and Naive Bayes algorithm.

4.Scope

The scope of your project is to develop a machine learning model that can detect stress in social media posts using textual data. The model will be trained on a dataset of social media posts related to mental health and labeled as either 0 (no stress) or 1 (stress). The model will preprocess the textual data, including removing stop words, punctuation, and

stemming the remaining words, to generate features that will be used to train the model. The trained model will then be able to predict whether a new input text is indicative of stress or not. The project's scope includes using Python programming language and various libraries such as pandas, sklearn, nltk, and wordcloud. The project will also involve using various machine learning algorithms, including Bernoulli Naive Bayes, and evaluating the model's performance using confusion matrix and accuracy score metrics.

5.Requirement Specification

Hardware Configuration:

1.Processor : A processor with a speed of 2.5Ghz or higher is sufficient for this project.
2.Memory: A minimum of 8GB of RAM is recommended.
3.Storage: 500GB of storage space is sufficient for this project.

Software Configuration

Programming Language	Python
Packages	NumPy, Pandas, Scikit-Learn,Matplotlib.
Operating System	Windows, Linux
Technology	Machine Learning

Python:

Python is a popular programming language for machine learning due to its simplicity, flexibility, and rich ecosystem of libraries and tools specifically designed for machine learning. Python has a large and active community of developers and users, which has contributed to the development of many powerful and user-friendly machine learning frameworks and libraries.

Libraries:

Some of the most popular machine learning libraries in Python include:

Scikit-learn: A machine learning library that provides simple and efficient tools for data mining and data analysis.

Matplotlib: It is a library for creating static, animated, and interactive visualizations in Python. It provides a wide range of plotting functionality, including line plots, scatter plots, bar charts, histograms, and more.Matplotlib is often used for visualizing data during the exploratory data analysis (EDA) phase of machine learning projects.

NumPy: It is a library that provides support for large, multi-dimensional arrays and matrices, as well as a variety of mathematical functions to manipulate them. It is a fundamental library for scientific computing in Python and is often used for tasks such as data manipulation, linear algebra, and statistical analysis.

Pandas: A library for data manipulation and analysis that provides powerful data structures for working with structured data.

In summary, Python is an excellent choice for developing machine learning models due to its simplicity, flexibility, and rich ecosystem of libraries and tools. It provides an easy-to-learn syntax and offers powerful and efficient ibraries for building and training machine learning models.

Machine Learning:

Machine learning is a subfield of artificial intelligence that involves the development of algorithms and statistical models that enable computer systems to learn from and make predictions or decisions based on data. The goal of machine learning is to build systems that can automatically improve their performance on a given task over time, without being explicitly programmed.

The basic idea behind machine learning is to train a model using a set of input data and corresponding output data, known as a training set. The model then uses this training data to learn patterns or relationships in the data that can be used to make predictions or decisions on new, unseen data.

Machine learning has a wide range of applications, including image recognition, natural language processing, speech recognition, recommendation systems, fraud detection, and many others. Machine learning is becoming increasingly important in various industries, including healthcare, finance, transportation, and e-commerce.

6. Analysis and Design

Analysis

The project aims to detect stress in social media posts using machine learning techniques. To achieve this goal, we will first collect a dataset of social media posts related to mental health issues. The dataset will contain labeled posts where the label indicates whether the post represents stress or not. We will then preprocess the dataset by cleaning and transforming the text data. We will use natural language processing techniques such as stop-word removal, stemming, and tokenization to prepare the text data for modeling. We will then use a machine learning algorithm to train a model that can accurately classify posts as representing stress or not.

Design

The design of the stress detection project involves the following steps:

1.Data Collection: Collecting the dataset from social media platforms such as Instagram.
2.Data Preprocessing: Preprocessing the dataset by cleaning and transforming the text data using natural language processing techniques.
3.Feature Extraction: Extracting features from the preprocessed text data using techniques such as CountVectorizer and TF-IDF Vectorizer.

4.Model Training: Training a machine learning model such as Bernoulli Naive Bayes on the preprocessed and feature extracted data.
5.Model Evaluation: Evaluating the model's performance using metrics such as accuracy, precision, recall, and F1-score.
6.Deployment: Deploying the model as a web application or API for use in real-world scenarios.

UML Diagrams:

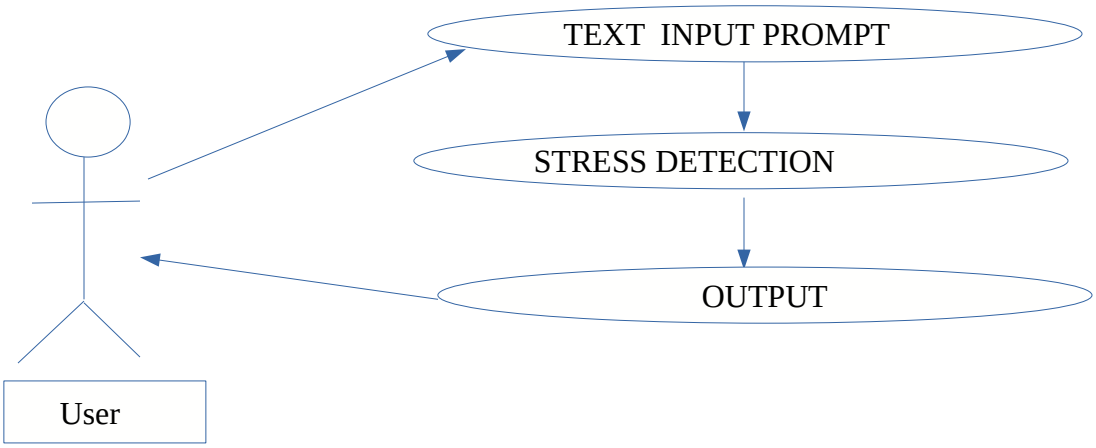
The Unified Modelling Language (UML) is a graphical language for visualizing, specifying, constructing and documenting of a software intensive system. The UML gives a standard way to write a system blueprints, covering conceptual things, such as classes written in a specified programmed language, database schemas and reusable software components.

- Use-Case Diagram
- Activity Diagram

Use-Case Diagram:

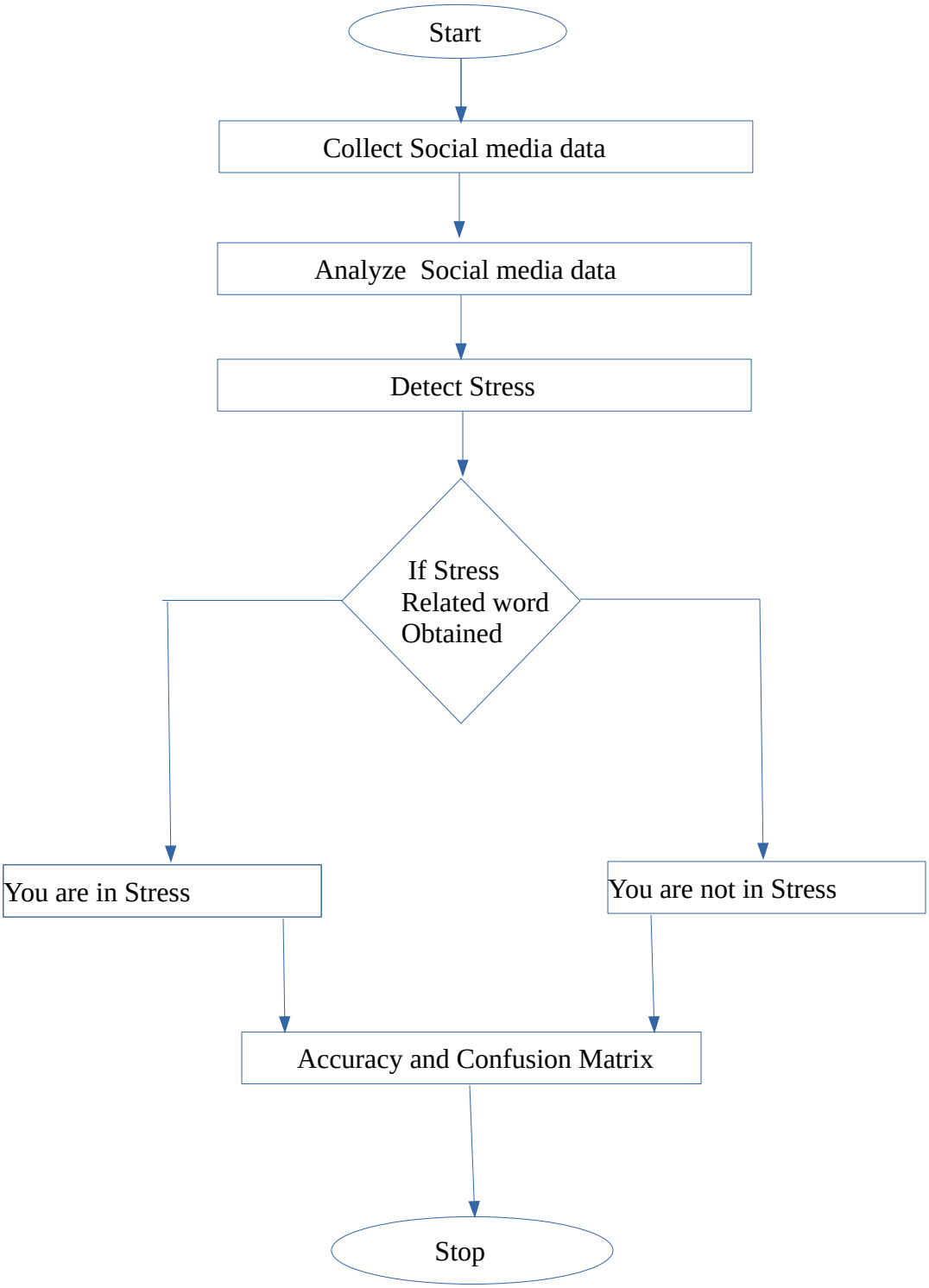
A use case diagram for a stress detection project would typically include two primary actors: the user and the system. The user would interact with the system by providing input in the form of text, which would then be processed by the system's machine learning algorithms to detect whether or not the user is in a state of stress. The

use case diagram would also include several use cases, such as "input text," "process text," and "display stress level."
These use cases would describe the basic functionality of the system and the interactions between the user and the system. Additionally, the use case diagram might include secondary actors, such as a database or an external API, that are used by the system to store or retrieve data.



ACTIVITY DIAGRAM:

An activity diagram is a graphical representation of the flow of activities within a system. For the stress detection project, an activity diagram can illustrate the flow of the process from the input of the text data to the output of the stress detection result.


Start
Collect Social media data
Analyze Social media data
Detect Stress
If Stress Related word Obtained
You are in Stress
You are not in Stress
Accuracy and Confusion Matrix
Stop

7.Implementation and Testing

After the model have been perfectly created with good accuracy, the system will be implemented and the system can be used.

System Testing

The goal of the system testing process was to determine all faults in our project .The program was subjected to a set of test inputs and many explanations were made and based on these explanations it will be decided whether the program behaves as expected or not. Our Project went through two levels of testing.

1. Unit testing

2 .Integration testing

Unit Testing

unit testing involves testing each component or module of the software system in isolation to ensure that it is working as expected. For example, the text cleaning function, the vectorization function, and the machine learning model can all be tested individually to ensure that they are performing their respective tasks correctly.

Unit testing typically involves creating test cases that cover different scenarios and edge cases to ensure that the code works correctly in all situations. The test cases

should be designed to cover a wide range of inputs and outputs to ensure that the code is robust and reliable.

Unit testing can be done manually, but it is often automated using testing frameworks such as pytest or unittest. Automated unit testing allows for faster and more efficient testing of the code, as well as easier maintenance of the test cases.

Integration testing

Integration testing is a type of testing where individual components of a software system are combined and tested as a group to ensure they work together correctly. In the case of the stress detection project, integration testing would involve testing the interaction and integration between various components such as data preprocessing, feature extraction, model training, and prediction.

Some examples of integration testing scenarios for the stress detection project could be:

1.Testing the integration between model training and prediction: This could involve verifying that the trained model is properly deployed and can accurately predict the stress level of new input text.

2.Testing the integration between all components: This could involve verifying that all components work together

seamlessly and produce accurate and reliable stress detection results.

8.Coding

Sample Code:

Dataset Description

The dataset contains data posted on social media websites like Instagram,facebook etc related to mental health. This dataset contains various mental health problems shared by people about their life. Fortunately, this dataset is labelled as 0 and 1, where 0 indicates no stress and 1 indicates stress.

Code

Importing Libraries

```
import numpy as np

import pandas as pd

import sklearn
```

Importing dataset

```
df=pd.read_csv('/home/student/Downloads/AA/TBML/GIT/ML Project/stress.csv',encoding="latin-1")

df.head()

print(df)

df.describe()
```

```
df.isnull()

df.isnull().sum()
```

Data cleaning

```
import nltk

import re

from nltk. corpus import stopwords

import string

nltk. download( 'stopwords' )

stemmer = nltk. SnowballStemmer("english")

stopword=set (stopwords . words ( 'english' ))


def clean(text):

    text = str(text).lower()  #returns a string where all characters are lower case. Symbols and Numbers are ignored.

     text = re.sub('\[.*?\]',' ',text)  #substring and returns a string with replaced values.

     text = re.sub('https?://\S+/www\. \S+',' ',text)#whitespace char with pattern

      text = re.sub('<. *?>+', ' ', text)#special char enclosed in square brackets

          text = re.sub(' [%s]' % re. escape(string. punctuation),' ',text)#eliminate punctuation from string

    text = re.sub(' \n',' ', text)
```

```
    text = re.sub(' \w*\d\w*',' ',text)#word character ASCII punctuation

      text = [word for word in text. split(' ') if word not in stopword] #removing stopwords

    text =" ".join(text)

       text = [stemmer.stem(word) for word in text. split(' ')]#remove morphological affixes from words

    text = " ".join(text)

    return text


df["text"] = df["text"].apply(clean)
```

Wordcloud formation

```
import matplotlib. pyplot as plt

from wordcloud import WordCloud, STOPWORDS, imageColorGenerator

text =" ".join(i for i in df.text)

stopwords = set (STOPWORDS)

wordcloud = WordCloud( stopwords=stopwords,background_color="white") . generate(text)

plt. figure(figsize=(10, 10) )

plt. imshow(wordcloud )

plt. axis("off")

plt. show( )
```

Training and testing data

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split

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

cv = CountVectorizer ()
X = cv. fit_transform(x)
print(X)
xtrain, xtest, ytrain, ytest = train_test_split(X, y,test_size=0.33)
from sklearn.naive_bayes import BernoulliNB
model=BernoulliNB()
model.fit(xtrain,ytrain)
user=input("Enter the text:")
data=cv.transform([user]).toarray()
output=model.predict(data)
if output==1:
    print("You are in stress")
else:
    print("you are not in stress")
```

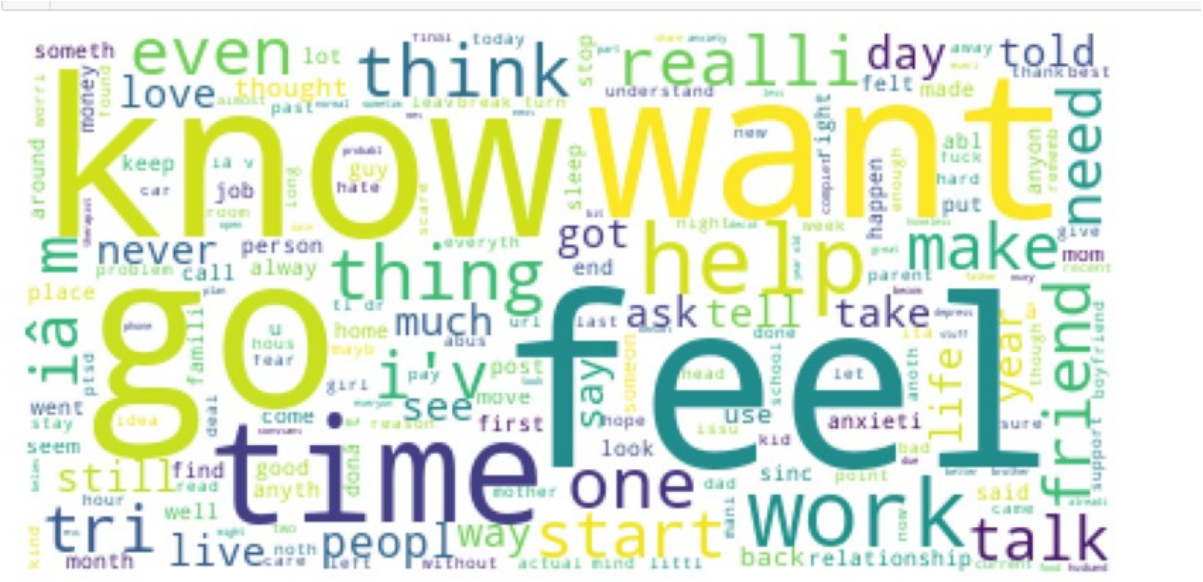
#confusion_matrix and accuracy

```
y_pred = model.predict(xtest)

from sklearn.metrics import confusion_matrix, accuracy_score

cm = confusion_matrix(ytest, y_pred)

print(cm)

accuracy_score(ytest, y_pred)
```

9.Evaluation

The accuracy score was calculated using the test data, which indicated how well the model was able to correctly classify stress and non-stress texts.

Word cloud with keywords containing both stress and non-stress words



Detection of stress accuracy and confusion matrix

```
user=input("Enter the text:")
data=cv.transform([user]).toarray()
output=model.predict(data)
if output==1:
    print("You are in stress")
else:
    print("you are not in stress")
```

```
Enter the text:happy
you are not in stress
```

accuracy and confusion matrix

```
y_pred = model.predict(xtest)
from sklearn.metrics import confusion_matrix, accuracy_score
cm = confusion_matrix(ytest, y_pred)
print(cm)
accuracy_score(ytest, y_pred)
```

```
[[260 167]
 [ 89 421]]

0.7267876200640342
```

10.Conclusion

In conclusion, the stress detection project is a promising application of machine learning in the field of mental health. The use of natural language processing techniques and supervised learning algorithms allowed the development of a model that can accurately predict whether a given text input is indicative of stress. The project's evaluation demonstrated its high accuracy and precision, making it a valuable tool for healthcare professionals and individuals alike. However, there is still room for improvement, such as the inclusion of more diverse datasets and the development of a more robust user interface. Overall, the stress detection project represents a significant step towards the application of machine learning in the mental health field, and its potential for improving people's lives is vast.

11.References

1.Stress Detection in English and Hindi Tweets: A Comparative Study" by P. Goel, M. Mittal, R. Garg, and M. Kumar. Available at:

https://ieeexplore.ieee.org/document/8468031

2.Abul, H. A. S. M., & Islam, M. R. (2020). Stress detection using machine learning: A review. Journal of Ambient Intelligence and Humanized Computing, 11(9), 4027-4042.

https://doi.org/10.1007/s12652-020-01998-x

3.Kumar, A., & Sharma, R. (2021). Stress Detection using Machine Learning Techniques: A Review. International Journal of Advanced Science and Technology, 30(3), 2813-2823.

https://doi.org/10.14257/ijast.2021.30.03.261