



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 Issue: IV Month of publication: April 2023

DOI: <https://doi.org/10.22214/ijraset.2023.51000>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Identification of Suicidal Intent Using Machine Learning Techniques over Twitter Data

Shivam Kadam¹, Sohum Kulkarni², Saakshi Padamwar³, Prof. Shubhangi Bhagat⁴

Abstract: Machine learning based on categorical classification has integrated usages in a variety of fields like prediction, finance, supply chain management, sales and operations as well as product analytics. This study shows how Support Vector Machine Learning Model from the Supervised Learning sub-branch of Machine Learning predicts the suicidal intent of a person's "tweet" on the social media platform 'Twitter'. This model basically indicates whether the text tweeted or posted by a person may be suicidal or not. Regularised data set for Modeling is divided into test data set and training data set at the rate of 3:7. Machine Learning for this study needs to use Pandas, Numpy, Pillow, ScikitLearn, Textblob and Nltk frameworks. Massive amount of actual Twitter data is used as a dataset for the training and testing purpose so the model can analyse the text with maximum accuracy.

I. INTRODUCTION

- 1) In today's world the cargo work has become a pointer in people's daily routine and lifestyle which is leading to some of the extreme steps like committing suicide.
- 2) Technology has some positive fundamentals on human aspects while some are turning out to be negative towards the human psychology like depression, suicidal steps and many more life changing decisions
- 3) Nowadays developers have come across various concepts which are in progress towards the remedy for the listed problems like suicide and depression.
- 4) The reference of the mentioned problems have a serious impact towards the life of the humans living in this modern era. However with the help of some of the pioneer methods residing in Computer Technology like Machine Learning, Deep Learning with the help of the methodology assigned as Sentiment Analysis it has become easy to prioritize this issue with some Machine Learning terminology and Algorithms
- 5) The problems were studied and it was associated with the help of some of the powerful definitions of Deep Learning and Data Science frameworks with the empowerment of strong functioning of libraries and tools.
- 6) The interface was designed in such a way with interactive KPIs which made it easy for the end users to accumulate their data fetched from the users which had been projected on the Twitter Platform with some of the keywords associated with some negative thoughts like die, unhappy, sad and many more.
- 7) The study was implemented with vigorous discussions and thought exchange programs with the teammates and came up with the working solution of implementing a Machine Learning Model which was used for identification of suicidal intent with a help of huge Dataset which was further Trained and Tested with great accuracy.

A. Background

Machine Learning algorithms are used for the classification of large data into smaller chunks. We aim to order the extremity of the tweet where it is either suicidal or not. On the off chance that the tweet is sad but not suicidal then the model would identify it as not suicidal. Basically the more predominant estimation sentiment to be picked. Various machine learning algorithms can be used to extract the features from the data.

B. Statement

- 1) The solution was implemented with the help of an interactive Machine Learning Model and also with some of the highly advanced Research Publications in the relevant field as a reference and the findings were executed on the working of the model.
- 2) The Implementation was performed with some block building algorithms which were developed from scratch towards some final end result. The advanced concepts from Deep Learning subject were used as a prerequisite towards the working model for the advancement of the results for better working interface and accurate results towards the end users.

C. Motivation

A Machine Learning model which can accurately identify the suicidal intent of a person just by taking input the 'Tweets' he/she made on the Twitter platform is the main objective. This application can be used by all the therapists to track the virtual behaviour of the patient and can treat him/her accordingly. There are many Sentiment Analysis or Text Analysis models readily available, but a model specially tailored to address the issue of suicidal patients is far from practically or commercially available in a ready-to-use state for the therapists/psychiatrists.

D. Challenge

The following are the challenges of this project:

- 1) Developing a Scientifically sound, time and cost effective Machine Learning model for predicting the suicidal intent of a person using his/her Twitter data.
- 2) Differentiating tweets into just sad or suicidal accurately.
- 3) Raising awareness on how social media acts as a medium to share suicidal thoughts for people in depression or facing anxiety.
- 4) Collecting and obtaining a massive Twitter dataset to accurately train the model.
- 5) Constantly working on increasing the dataset size to make it more and more accurate along with prevention of overfitting of the data.
- 6) Getting feedback from therapists on whether the model practically works just as expected.

II. SYSTEM PLANNING

A. Literature Review

In the modern era where advancement has been carried out in various sectors including Industries, Technology and Hospitality but the workload towards the human brain has also been building day by day due to the massive work functioning and hectic daily routine.

The authors have taken some initial steps towards the solution of the human problems which are causing some serious damage towards their health and their balanced lifestyle, However the remedy includes not only the analysis of the thoughts but also their guidance throughout their important life phases with a constant support from external psychological community builders to make the audience feel positive during their low feelings.

Lastly the authors have gone through the internal aspects of the working model and made use of a recursive trial and error technique to make it more accurate towards the result with constant aim linked with maximum correct result with minimum time consumption.

III. FEASIBILITY STUDY

In this phase, the project's practicality is evaluated, and a business proposal is presented that outlines a broad plan for the project and provides some cost approximations. In the system analysis phase, the feasibility of the suggested system is assessed to ensure that it does not impose an excessive burden on the company. It is necessary to have a basic understanding of the primary system requirements to conduct a feasibility analysis.

Three key components of the feasibility analysis are

- 1) *Technical Feasibility*: The purpose of this study is to assess the technical feasibility, which refers to the technical specifications that the system must meet. It is essential to develop a system that does not place a significant strain on the existing technical resources because it may result in overburdening the client. The system should have modest technical requirements, meaning that only minimal or no changes need to be made to implement it.
- 2) *Economical Feasibility*: This study aims to assess the economic impact of the system on the organization. The company's ability to invest in research and development for the system is limited, and all expenses must be justified. Therefore, the system was developed within budget constraints, and this was made possible by utilizing freely available technologies. Only customized products needed to be purchased to complete the system.
- 3) *Social Feasibility*: The purpose of this study is to evaluate the system's user acceptance level, which involves training the user to utilize the system effectively. It is essential that the user does not perceive the system as a threat, but rather accepts it as a necessary tool. The user's acceptance level is determined by the approach used to educate and familiarize them with the system. The user's confidence must be boosted so that they can offer constructive criticism, which is encouraged as they are the ultimate end-users of the system.

IV. REQUIREMENTS

A. User Requirements

- 1) Tweets chosen by the user to test can be pasted on the main User Interface of the app after signing up and logging in into the software.
- 2) In this project a set of available libraries has been used. These libraries include the stopwords as well which are eliminated from the tweet content and only core target words are focused for prediction.
- 3) User just has to click a button and the software will pop-up a message saying whether the tweet is suicidal or not.

B. Non-functional Requirements

In Computer engineering and Requirement engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific aspects. They are into two opposite directions with functional requirements that define specific behavior or functions. Non-functional requirements add massive value to business development. It is commonly misled by a lot of people. It should be mandatory for business stakeholders, and Clients to clearly develop the requirements and their high expectations in measurable terminology. If the non-functional requirements are not scalable then they should be revised or interpreted again to gain better clarity. For example, User stories help in dissolving the gap between developers and the user stakeholders in Agile Methodology.

C. Usability

Prioritize the End users tweet as per the sentiment analysis terminology and then segmenting the tweets into diversified importance while simultaneously working with the dataset.

D. Reliability

Reliability refers to the level of confidence in a system that is established over time through its use. It indicates the extent to which software can perform its intended functions without encountering errors or issues within a specific timeframe. The number of issues and bugs encountered while execution of the working model were fixed with the help of reliability behavioral testing and also through exclusive discussion panels throughout the functioning of the model. Your goal should be to create point to point dense algorithms for the machine learning model and which makes the model easy to implement and familiar to the user of the working directory.

E. Performance

Under what circumstances and at what specific peak times, such as during stress periods like the end of the month or during payroll disbursement, should system response times be measured from any point? Additionally, are there times when the load on the system will be abnormally high?

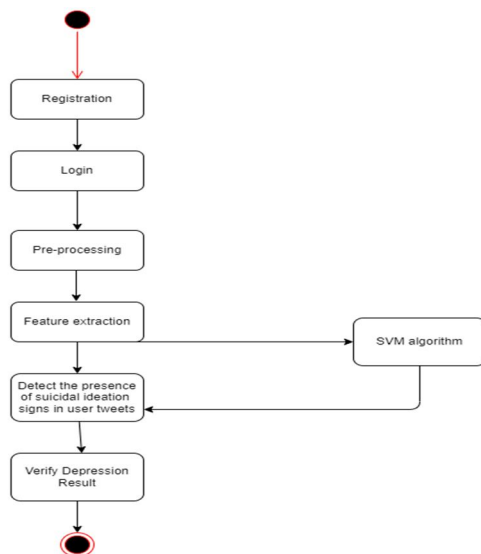
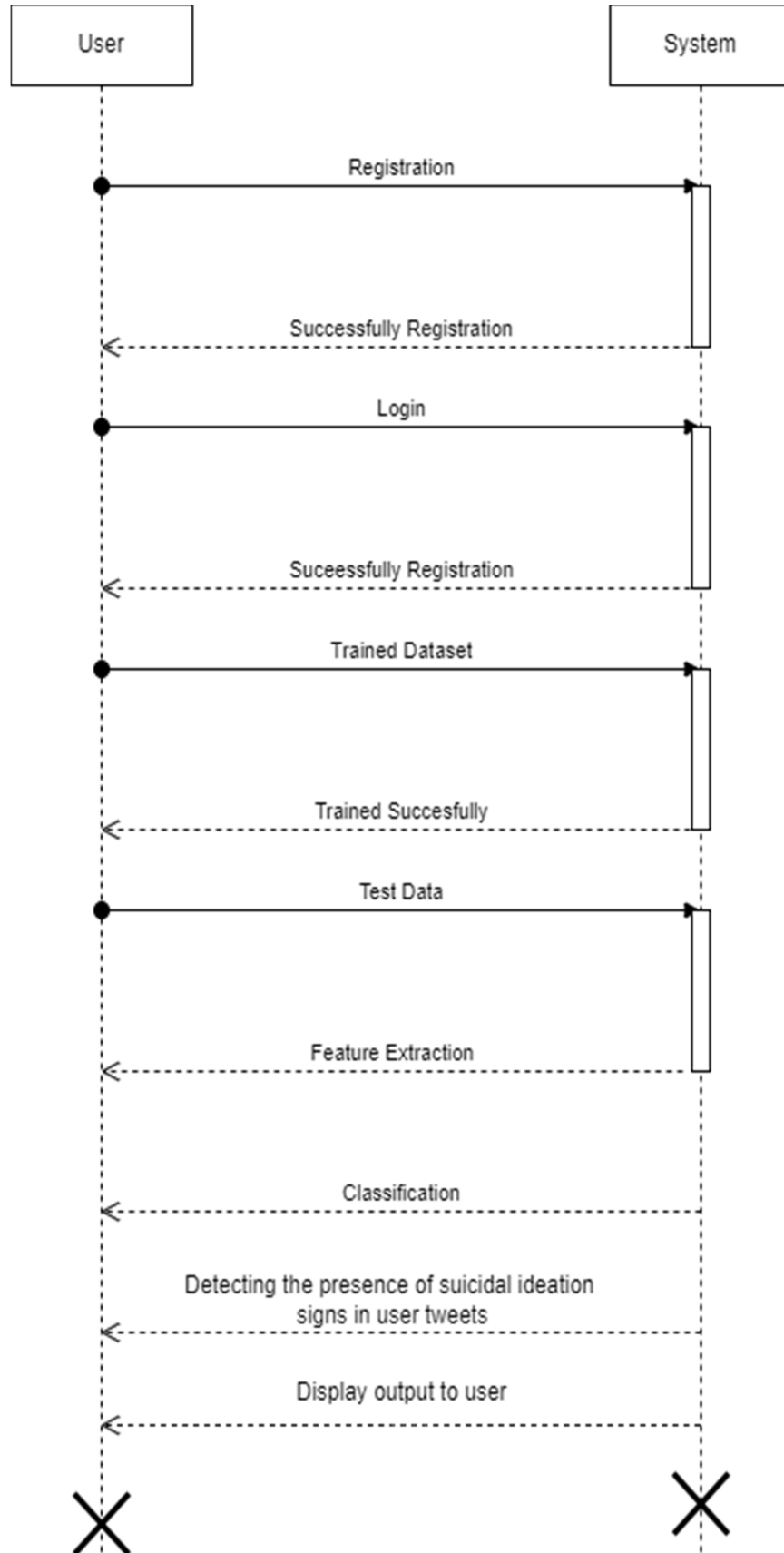


Fig: Activity Diagram





V. IMPLEMENTATION OF SYSTEM

A. Existing System

The increasing number suicides is not a new issue and many projects have been conducted and implemented in the past to address this issue and tried to find a practical solution which will contribute in eventually reducing the number of suicides. Most of these projects involve models which make use of Sentiment Analysis or Text Analysis to segregate the texts into negative and positive behaviours.

B. Disadvantages

The major disadvantages of these models include

These models fail to find a difference between the tweets which are just sad and those which are actually suicidal and may prove the person to have a suicidal intent.

C. Proposed System

This research proposes a Machine Learning model which will identify the intent of a person who may or may not be having thoughts about committing suicide using the tweets he/she posts on Twitter. This model can be implemented practically by therapists to analyse the behavioural patterns of their patients and treat them with proper medications accordingly. A Support Vector Machine (SVM) technique is used for the classification purpose. It involves mapping the input data points into a higher dimensional feature space, where a hyperplane is then used to separate the classes or predict the target values. It aims to find the optimal hyperplane that maximizes the margin, which is the distance between the hyperplane and the closest data points from each class.

D. Advantages of Proposed System

- 1) The main advantage of the proposed system over the already existing similar models is that this model is tailored specifically to identify the suicidal intent and not just to identify general emotions like happy or sad. The accuracy can be achieved by using actual twitter data of people in depression in large amounts so the model can be trained accurately.
- 2) SVM is the most accurate technique amongst all the other techniques which predicts the intent with maximum accuracy.

VI. MODULE DESCRIPTION

A. Machine Learning and its Categories

Machine learning is a subfield of artificial intelligence that involves developing algorithms and statistical behavioural models that enable computer systems to inculcate from data and make decisions without being explicitly programmed. In understandable terms, machine learning involves using statistical techniques to enable Virtual Machines to improve their performance on a specific task module by learning from data. Machine learning algorithms are used in diversified applications, such as image and speech recognition, natural language processing, predictive analytics, and autonomous vehicles. To function, machine learning algorithms require three key components: data, models, and optimization. The data provides the algorithm with the information necessary to learn and make decisions. The model is the algorithm that is trained on the data, enabling the system to recognize patterns and make predictions. Optimization is the process of adjusting the model to improve its accuracy and effectiveness. The increasing amount of available data has made machine learning more critical in recent years. It has the potential to transform numerous industries, from healthcare and finance to manufacturing and transportation, by providing more accurate predictions and faster decision-making. However, it also raises significant ethical and social concerns, including privacy, bias, and job displacement, that require careful consideration and resolution.

B. Supervised Machine Learning

Supervised learning involves training a model on a labelled dataset, where the output is known for each input, while unsupervised learning involves training a model on an unlabeled dataset to discover patterns and structures in the data. Supervised learning is a commonly used machine learning algorithm that involves training a model on a labelled dataset. In supervised learning, the model is provided with input data and the corresponding correct output data, allowing it to learn the relationship between the two. Once the model is trained, it can be used to make predictions on new input data by applying the learned relationship to the new data. This makes supervised learning a useful tool for tasks such as classification and regression. In this example, the input data are the number of bedrooms, bathrooms, square footage, and location, while the output data is the actual selling price of the house. The model is trained using the labelled dataset, and once it has learned the relationship between the features and the selling price, it can be used to predict the selling price of new houses that it has not seen before.

C. Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a sort of machine learning algorithm that is conventionally employed for grouping and regression analysis. It entails projecting the input data points into an augmented dimensional feature space, where a hyperplane is then utilised to segregate the categories or prognosticate the target values. The SVM aspires to seek out the most favourable hyperplane that maximises the gap, which is the interval between the hyperplane and the nearest data points from each category. The SVM is taught exploiting a set of labelled data and can subsequently be employed to foretell the labels or target values of recent, unlabeled data points. The formula has been broadly employed in numerous applications, comprising image identification, text classification, and bioinformatics. In the suggested framework, the Twitter data employed to train the SVM is amassed from bona fide sources (institutions) in which factual tweets from individuals who are grappling with anxiety or depression issues are employed, and a dichotomous target variable is attributed to each tweet, with 0 or 1 representing non-self-destructive or self-destructive, respectively.

VII. CONCLUSION

Machine Learning algorithms like Supervised Learning approaches are highly valuable in resolving pragmatic issues. In this study, the primary aim is to establish how the suicidal intention of an individual can be detected by scrutinizing the tweet he/she disseminated through the utilization of the SVM model. The corpus of authentic tweets is utilised to achieve this objective.

VIII. FUTURE WORK

This study has enormous possibilities in enhancing the techniques of premature recognition of an individual devising or contemplating suicide. There are numerous means to enhance the current model. Utilisation of Real-time Tweets can be performed by leveraging the Twitter Developer account to extract real-time data and scrutinise it.

One of the uses of this inquiry can culminate in producing a program for physicians/psychotherapists, wherein they can merely integrate their patients' Twitter accounts into the software to track everyday musings and trigger warnings if any tweet is vital. It can be employed for a vast-scale tracking of patients and also the precision of the model will rise as the corpus becomes more voluminous. This model can also be fused with different novel technologies like Artificial Neural Networks or Deep Learning methods.

REFERENCES

- [1] J. T. Fiquer, P. S. Boggio, and C. Gorenstein, "Talking bodies: Nonverbal behavior in the assessment of depression severity," *Journal of affective disorders*, vol. 150, no. 3, pp. 1114–1119, 2013.
- [2] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, "A review of depression and suicide risk assessment using speech analysis," *Speech Communication*, vol. 71, pp. 10–49, 2015.
- [3] J. F. Cohn, T. S. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. T. Padilla, F. Zhou, and F. De la Torre, "Detecting depression from facial actions and vocal prosody," in *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*. IEEE, 2009, pp. 1–7.
- [4] L.-S. A. Low, N. C. Maddage, M. Lech, L. Sheeber, and N. Allen, "Influence of acoustic low-level descriptors in the detection of clinical depression in adolescents," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE, 2010, pp. 5154–5157.
- [5] J. C. Mundt, P. J. Snyder, M. S. Cannizzaro, K. Chappie, and D. S. Geralts, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (ivr) technology," *Journal of neurolinguistics*, vol. 20, no. 1, pp. 50–64, 2007.
- [6] S. Alghowinem, "From joyous to clinically depressed: Mood detection using multimodal analysis of a person's appearance and speech," in *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*. IEEE, 2013, pp. 648–654.
- [7] Y. Yang, C. Fairbairn, and J. F. Cohn, "Detecting depression severity from vocal prosody," *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 142–150, 2013.
- [8] T. R. Almaev and M. F. Valstar, "Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*. IEEE, 2013, pp. 356–361.
- [9] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. M. Mavadati, Z. Hammal, and D. P. Rosenwald, "Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses," *Image and vision computing*, vol. 32, no. 10, pp. 641–647, 2014.

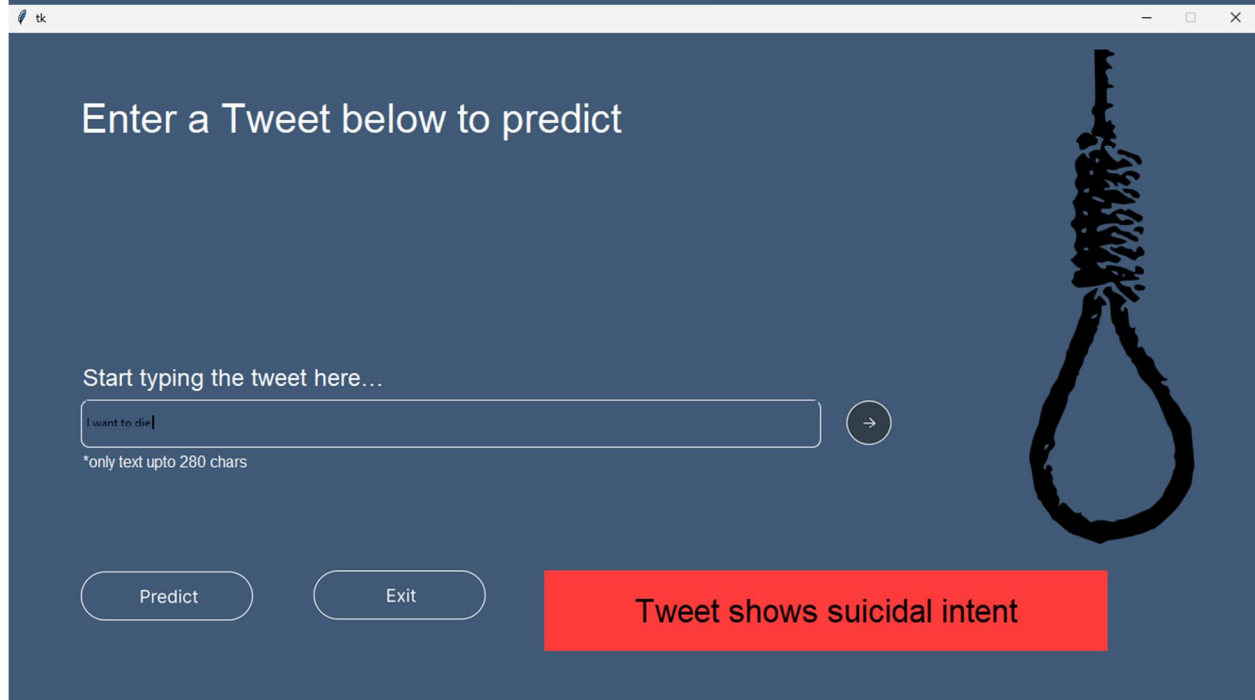
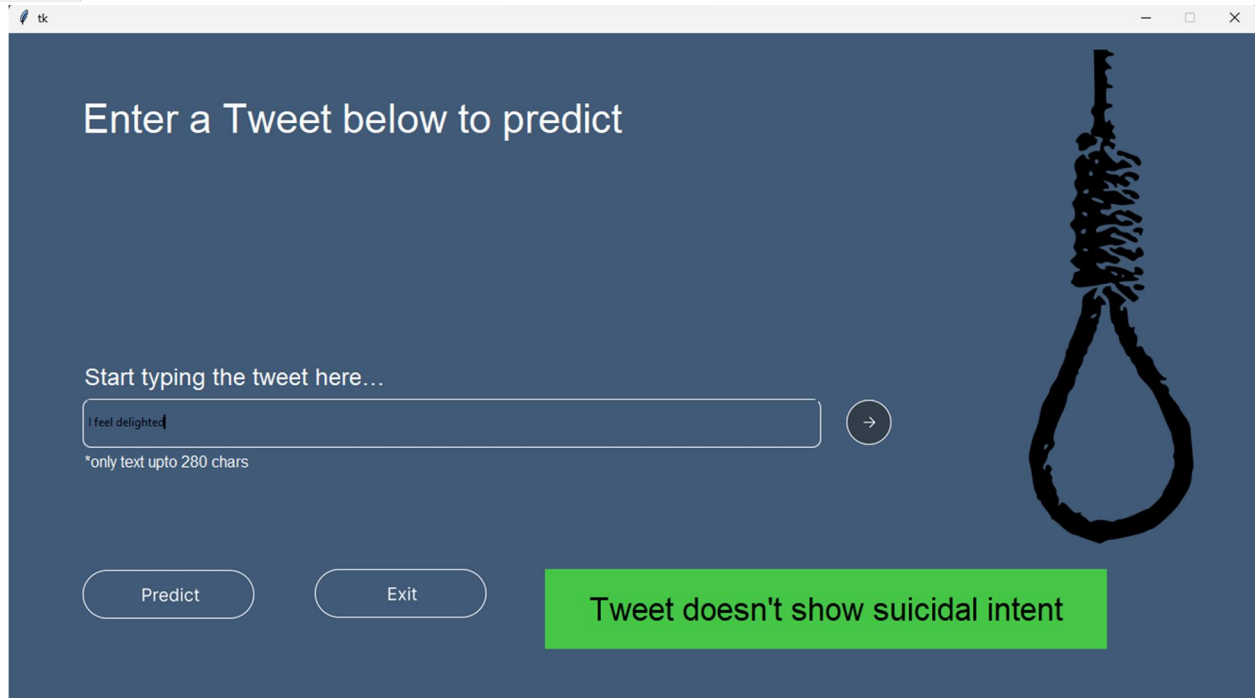
RESULTS

```
1 import pandas as pd, numpy as np, re
2 from sklearn.metrics import classification_report, accuracy_score, confusion_matrix
3 from sklearn.model_selection import train_test_split
4 import tkinter as tk
5 from sklearn import svm
6 from PIL import Image, ImageTk
7 from tkinter import ttk
8 from joblib import dump, load
9 from sklearn.feature_extraction.text import TfidfVectorizer
10 from textblob import TextBlob
11 from nltk.corpus import stopwords
12 from sklearn import metrics
13 from sklearn.model_selection import GridSearchCV
14 import pickle
15 import nltk
16 #####
17 nltk.download('stopwords')
18 stop = set(stopwords.words('english'))
19 nltk.download('punkt')
20 nltk.download('averaged_perceptron_tagger')
21 #####
22
23 from pathlib import Path
24 from tkinter import Tk, Canvas, Entry, Text, Button, PhotoImage
25
26 OUTPUT_PATH = Path(__file__).parent
27 ASSETS_PATH = OUTPUT_PATH / Path(r"/Users/sohum/College/BE Project/New Main/build/assets/frame0")
28
29
30 def relative_to_assets(path: str) -> Path:
31     return ASSETS_PATH / Path(path)
32
33 window = Tk()
34 window.geometry('1280x700')
35 window.configure(bg = "#415A77")
36
37 canvas = Canvas(
38     window,
39     bg = "#415A77",
40     height = 700,
41     width = 1280,
42     bd = 0,
43     highlightthickness = 0,
44     relief = "ridge"
45 )
46
47 canvas.place(x = 0, y = 0)
48 canvas.create_rectangle(
49     859.0,
50     570.0,
51     905.0,
52     616.0,
53     fill="#415A77",
54     outline="")
55
56 canvas.create_text(
57     76.0,
58     64.0,
59     anchor="nw",
60     text="Enter a Tweet below to predict",
61     fill="#FFFFFF".
```

```
189
190     print("=" * 40)
191     print("=====")
192     print("Classification Report : ",(classification_report(label_test, pred)))
193     print("Accuracy : ",accuracy_score(label_test, pred)*100)
194     accuracy = accuracy_score(label_test, pred)
195     print("Accuracy: %.2f%%" % (accuracy * 100.0))
196     ACC = (accuracy_score(label_test, pred) * 100)
197     repo = (classification_report(label_test, pred))
198
199     label4 = tk.Label(window,text =str(repo),width=35,height=10,bg='khaki',fg='black',font=("Tempus Sanc
200     label4.place(x=205,y=100)
201
202     label5 = tk.Label(window,text ="Accuracy : "+str(ACC)+"%\nModel saved as SVM_MODEL.joblib",width=35,he
203     label5.place(x=205,y=320)
204
205     dump (clf, "/Users/sohum/College/BE Project/SVM_MODEL.joblib")
206     print("Model saved as SVM_MODEL.joblib")
207
208     entry_image_1 = PhotoImage(
209         file=relative_to_assets("entry_1.png"))
210     entry_bg_1 = canvas.create_image(
211         454,
212         400.0,
213         image=entry_image_1
214     )
215     entry = Entry(
216         bd=0,
217         bg="#415A77",
218         fg="#000716",
219         highlightthickness=0
220     )
221     entry.place(
222         x=81.0,
223         y=376.0,
224         width=749.0,
225         height=45.0
226     )
227
228
229     #####
230     def Test():
231         predictor = load("SVM_MODEL.joblib")
232         Given_text = entry.get()
233         #Given_text = "the 'roseanne' revival catches up to our thorny po..."
234         vec = open('vectorizer.pickle', 'rb')
235         tf_vect = pickle.load(vec)
236         X_test_tf = tf_vect.transform([Given_text])
237         y_predict = predictor.predict(X_test_tf)
238         print(y_predict[0])
239         if y_predict[0]==0:
240             label4 = tk.Label(window,text ="Person is not in depression",width=20,height=2,bg='#46C646',fg='l
241             label4.place(x=450,y=550)
242         else:
243             label4 = tk.Label(window,text ="Person is in depression",width=20,height=2,bg='#FF3C3C',fg='black
244             label4.place(x=450,y=550)
245
246     #####
247     def exit():
248         window.destroy()
```

```
129 def train():
130
131     result = pd.read_csv(r"tweets.csv",encoding = 'unicode_escape')
132
133     result.head()
134
135     result['headline_without_stopwords'] = result['tweet'].apply(lambda x: ' '.join([word for word in x.s
136
137
138     def pos(headline_without_stopwords):
139         return TextBlob(headline_without_stopwords).tags
140
141
142     os = result.headline_without_stopwords.apply(pos)
143     os1 = pd.DataFrame(os)
144     #
145     os1.head()
146
147     os1['pos'] = os1['headline_without_stopwords'].map(lambda x: " ".join(["/".join(x) for x in x]))
148
149     result = result = pd.merge(result, os1, right_index=True, left_index=True)
150     result.head()
151     result['pos']
152     result_train, result_test, label_train, label_test = train_test_split(result['pos'], result['target']
153                                     test_size=0.2, random_state
154
155     tf_vect = TfidfVectorizer(lowercase=True, use_idf=True, smooth_idf=True, sublinear_tf=False)
156
157     X_train_tf = tf_vect.fit_transform(result_train)
158     X_test_tf = tf_vect.transform(result_test)
159
160
161
162
163
164
165
166     clf = svm.SVC(C=10, gamma=0.001, kernel='linear')
167     clf.fit(X_train_tf, label_train)
168     pred = clf.predict(X_test_tf)
169
170     with open('vectorizer.pickle', 'wb') as fin:
171         pickle.dump(tf_vect, fin)
172     with open('mlmodel.pickle', 'wb') as f:
173         pickle.dump(clf, f)
174
175     pkl = open('mlmodel.pickle', 'rb')
176     clf = pickle.load(pkl)
177     vec = open('vectorizer.pickle', 'rb')
178     tf_vect = pickle.load(vec)
179
180     X_test_tf = tf_vect.transform(result_test)
181     pred = clf.predict(X_test_tf)
182
183     print(metrics.accuracy_score(label_test, pred))
184
185     print(confusion_matrix(label_test, pred))
186
187     print(classification_report(label_test, pred))
188
```

```
255
256 button_image_2 = PhotoImage(
257     file=relative_to_assets("button_2.png"))
258 button_2 = Button(
259     image=button_image_2,
260     borderwidth=0,
261     highlightthickness=0,
262     command=Test,
263     relief="flat"
264 )
265 button_2.place(
266     x=76.0,
267     y=550.0,
268     width=176.0,
269     height=51.0
270 )
271
272 button_image_1 = PhotoImage(
273     file=relative_to_assets("button_1.png"))
274 button_1 = Button(
275     image=button_image_1,
276     borderwidth=0,
277     highlightthickness=0,
278     command=exit,
279     relief="flat"
280 )
281 button_1.place(
282     x=314.0,
283     y=550.0,
284     width=176.0,
285     height=51.0
286 )
287
288 image_image_1 = PhotoImage(
289     file=relative_to_assets("image_1.png"))
290 image_1 = canvas.create_image(
291     1130.0,
292     270,
293     image=image_image_1
294 )
295
296
297 button_image_5 = PhotoImage(
298     file=relative_to_assets("button_5.png"))
299 button_5 = Button(
300     image=button_image_5,
301     borderwidth=0,
302     highlightthickness=0,
303     command=lambda: print("button_5 clicked"),
304     relief="flat"
305 )
306 button_5.place(
307     x=859.0,
308     y=376.0,
309     width=46.0,
310     height=46.0
311 )
312
313 canvas.create_text(
314     78.0,
```





10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)