**Data Mining Tools for the Detection of Suicide IN Twitter**

**Abstract**

In the last days the rate of suicides has been increased a lot and we want to get red of this problem . we want to introduce about how we can use artificial intelligence , machine learning and other techniques to get solution to reduce suicides.

This work main to create a program, which will be capable of detecting suicide in social networks "Twitter ". we will work to compare the results of this work with previous work, and will be setting a machine classification which means that the classification will be automatic, and perhaps we could integrate another data mining tool in our work. We could develop a machine classification method, which could automatically distinguish between text containing suicidal ideation.

**Introduction**

there is a lot of data generated everywhere around us include texts tweets and changing Facebook connections ,a lot of data arrive sequentially . the entire data set cannot be stored but short batches can be . data are often distributed and stored over multiple locations and we need to collect and analysis data to a central location .

All electrical engineering and computer science are benefited hugely from collecting data using data science and other machine learning .

The purpose of our discussion is to provide an overview about how we can use artificial intelligence

To reduce suicides . phenomenon and its potential in both positive and negative .

Machine learning is a field of artificial intelligence that has been applied in many social media platforms to detect patterns in data and dealing with large amount of data collected from social media with various formats . when the data contain substantial volume, variety or velocity may be not accurate .

we discuss specifically how we can collect the data by data analytics Defining big data can be tricky as there isn’t one simple, agreed-on definition big data is refers to multiple data sources being integrated with very different types of data. A data set that was large and would easily overwhelm available computing resources 20 years ago can be easily handled by the

typical desktop computer or smartphone. Similarly, a computer science big data application may

typically deal with hundreds of terabytes but a much smaller volume of data would easily overwhelm..

Since an ever-increasing part of the population makes use of social media in our dialy life.

Big data analytics has recently emerged as an important research area due to the popularity of the Internet social media data is being analysis in many different disciplines. The social media analytics process involves four steps, data discovery, collection, preparation, and analysis.

The most common applications of big data for social media are, social media analytics,

sentiment analysis, and opinion mining. And we can use these properties in our topic to reduce suicides.

Social media can be used to improve suicide risk surveillance., with emerging data indicating that online information may provide detection and surveillance.

patient-generated data, combined with advances in predictive analytics (i.e., techniques using data mining and machine learning), may provide the information needed to drive improvement in suicide risk prediction by applying machine learning algorithms.

 Data Science started with statistics, and has evolved to include concepts/practices such as Artificial Intelligence Machine Learning, and the Internet of Things, to name a few. As more and more data has become available, first by way of recorded shopping behaviors and trends, businesses have been collecting and storing it in ever greater amounts. With growth of the Internet, the Internet of Things, there has been a flood of new information or Big Data. Once the doors were opened by businesses seeking to increase profits and drive better decision making, the use of Big Data started being applied to other fields, such as medicine, engineering, and social sciences.

the model was created in 1949 by Donald Hebb depending on some theory about human brain cells .in 1957 Frank Rosenalatt developed these theories and reach to the first successful neuro computer which has been constructed for image recognition .from 1970 and 1980 Artificial Intelligence had focused on using logical , knoweldge-based approaches rathe than algorithms , neural network research was abandoned by computer science and AI researches .this caused Artificial Intelligence and machine learning take a separate paths .Machine learning had been used as a training for Artificial program Intelligence . another technique Data Science started with statistics, and has evolved to include concepts such as Artificial Intelligence Machine Learning, and the Internet of things . seeking to increase profits and making better decision, the use of Big Data started applied to other fields , such as medicine, engineering, and social sciences. Big data analytics has recently appear as an important research area due to the popularity of the Internet Defining big data can be tricky as there isn’t one simple, agreed-on definition big data is refers to multiple data sources being integrated with very different types of data.

The positive side in Our program has the role, which is to detect suicidal people only from their publication into social networks like Facebook , Tweeter. In this search paper we are going to focus in the social network twitter because almost every user’s tweets are completely public, also we want to note that when we detect the suicidal person then we could prevent it’s suicide, we can talk to him about his conditions, and perhaps he does not know that he is doing things which they are signs of suicide, or even we could recommend some authority or divisions or organization all over the world that take good cares of suicidal people . but the negative is Privacy lives saved this problem is the trade-off between a person’s right to privacy and the widely agreed-upon moral imperative to act on information that may save lives. For a simple and direct yet difficult example decision, consider that a hallmark feature of risk is being withdrawn in many cases, the very act of reaching out can be perceived as an invasion of privacy.

**Why is Twitter Sentiment Analysis important?**

Twitter sentiment analysis systems allow you to sort large sets of tweets and detect the polarity of each statement automatically. And the best part, it’s fast and simple, saving teams valuable hours and allowing them to focus on tasks where they can make a bigger impact.

Twitter allows users to collect tweets with the help of Twitter API (the term stand s for Application Programming Interface ). this premium API enables you to receive real-time streams of Twitter data, listing to up to 250,000 keywords, hashtags, mentions, locations, and user IDs.

**Methodology**

* **Data Collection**

We will use [Python script](https://github.com/feconroses/gather-tweets-from-stream) because Python allows you to gather

* Tweet content: text of the tweet
* Matched keywords: the keywords that were a match to the search in the Twitter stream
* Date: date and hour of the tweet
* User: name of the author of the tweet
* Source used to post the tweet (for example, Twitter Web Client or Buffer)
* Tweet ID
* Tweet URL

Then , To use Twitter API we must first have a twitter account. It can be easily created by filling the sign up details in twitter.com website. After this you will be provided with a username and password which is use for login purpose. Once your account is created, you can now read and send tweets on any topic you want to explore. Twitter provider a platform from which we can access data from twitter account and can use it for our own purpose. For this we have to login with twitter credentials in this website on the [Twitter apps site](https://apps.twitter.com/).

but the activation of the app might take more than weeks due to twitter's new strict rules about COVID-19

so We Will use an unofficial API called tweet\_scraper

**Now we create a Python script which will be used to fetch tweets from twitter .**

**Before creating this script we first install a library in Python called tweet\_scraper.**

tweet\_scraper:

it is a library is one of the open source python library which enable python to communicate with twitter and very easy to use and enables you to gather Tweets automatically.

To install tweet\_scraper, just provide a command ‘pip install tweet\_scraper’ in command prompt .

USING PIP

Simply type pip install tweepy into your terminal.

First, let’s import tweet\_scraper

**import** tweet\_scraper

we will fetch thousands of tweets from some celebrities and then use our second librariy which is monkeylearn

and we will install it using pip too,

pip install monkeylearn

we will use this API to classify the data whether it is Positive or Negative

but some data can be neutral so we took that in consideration too

we will explain the code in the main.py file comments

the program is divided into 2 files which are main.py and tokens.txt

main.py contains the main code to run the program while tokens.txt is the file that has the access tokens used in monkeylearn website