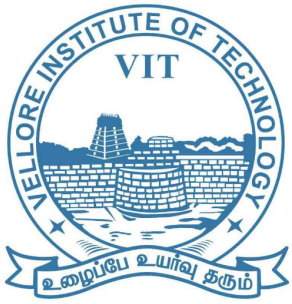

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Automatic sentiment analysis of Twitter messages

Social and Information Networks

Under Guidance of
Professor Manjula R

Team Members,

19BCE2222 - Hemaksh Chaturvedi
19BCE0638 - Mohamamd Ayaazuddin
19BCE0777- Rishabh Raj

Introduction

The medical field in today's day and age has come a long way with providing cures to innumerable numbers of diseases. But still, all these cures have been for physical types of illnesses and not the ones present in our mind. Mental illnesses are still not widely accepted as an actual illness and are thought to be just another part of life that a human goes through. Depression is one of the biggest ones that has affected the world. Although there are known and effective treatments available nowadays for depression, a significant portion of the population never receives such treatment. A few reasons for this are - lack of resources, lack of trained health-care providers and social stigma associated with mental health.

Problem statement

We've all faced different levels of depression or some form of sadness in our life. In this project, we aim to find out those in need by doing a sentimental analysis of twitter messages. Our target is to classify and distinguish people into different risk zones of depression via sentimental analysis.

Objective

With the growing technology, there is a huge volume of data present around us. The Internet these days has become a platform for exchanging learnings, ideas and thoughts . Social media websites like twitter , facebook , instagram are becoming popular as they allow people to share and express views about the things which they want to. This survey focuses mainly on sentiment analysis of twitter data whether the messages (tweets) sound positive ,negative or neutral . We will analyse this by using some of the existing Machine Learning algorithms like Naive Bayes , Support Vector Machines by performing data analysis on the twitter dataset .The main objective is to provide a model with accuracy so that these challenges can solved to the highest possible level and help as many as people as we can.

Scope

The scope of this project is to highlight and flag the tweets of users facing depression or sadness of some form using sentimental analysis and to categorise them into different levels of risk. We hope to find out sooner than later of those at high risk in their depression so that they can get the help and support they require.

Many people are not very vocal about their depression and often stay in denial, so through the help of certain keywords we may be able to flag them and make them confront their emotions and get them the help they need.

Literature Survey

Naïve Bayes (NB) Approach:

Naïve Bayes classifier is a simple probabilistic classifier that uses the concept of mixture models to perform classification. The mixture model relies on the assumption that each of the predefined classes is one of the components of the mixture itself. The components of the mixture model denote the probability of belongingness of any term to the particular component. Thus, they are also known as generative classifiers. Naïve Bayes classifier is a probabilistic classifier that uses the concept of Bayes Theorem and finds the maximum prospect of probability of any word fitting to a particular given or predefined class.

Support Vector Machine (SVM):

Support vector machine (SVM) solves the traditional text categorization problem effectively; generally outperforming Naïve Bayes as it supports the concept of maximum margin. The main principle of SVMs is to determine a linear separator that separates different classes in the search space with maximum distance i.e. with maximum margin. If we represent the tweet using t, the hyperplane using h, and classes using a set Cj € {l, -1} into which the tweet has to be classified, the solution is written as follows equivalent to the sentiment of the tweet.

CNN:

CNN is a type of neural network model which allows us to extract higher representations for the image content. Unlike the classical image recognition where you define the image features yourself, CNN takes the image's raw pixel data, trains the model, then extracts the features automatically for better classification. Although CNNs have been commonly used in computer vision tasks, their ability to recognise high-level spatial features and patterns has made them popular and surprisingly efficient in NLP tasks as well

Overview

In this paper, twitter will provide us with data required and we will perform a thorough sentimental analysis using certain keywords(like depressed,suicide,death,etc.), etc. to flag potential people in risk and try to provide them with help.