Depression Prediction Using KNN and Social media usage Frequency

Labiba Ibnat[1*] **and Tasnim Ferdous Dima**[1+]

[1]AUST, Computer Science and Engineering, Dhaka, postcode, Bangladesh
[*]15.01.04.143
[+]15.01.04.145
[1]these authors contributed equally to this work

ABSTRACT

Psychologists have been studying personality for a very long time and it is also a very important topic in data analysis.In our research we tried to analysis whether there is a link between depression and soical media usage time. We also tried to analyse the fact that if a person spends more time on social media and if the person also sleeps overtime make any difference on his or her mental health.

Introduction

Every person need to be healthy physically as well as mentally. Humans are competitive by nature. It's natural to feel low if one person sees another person's life much more accomplished then them. In the era of social media rising it's easy to see other people's success . We try to see if use of social media usage make any difference in people's mental health or not.

To get rid of depression we wanted to see the border line of social media usage time above which if a person spends time on social media he will be classified as a depressed person .

For finding the link between sleep frequency and social media usage of a person, we have applied KNN algorithm.We make some features to predict those attitude. Than, we extract the features from the dataset. We have asked three question to determine the mental health of a person.The questions are:

1. How much time you spends on social media
2. Sickness frequency in a month
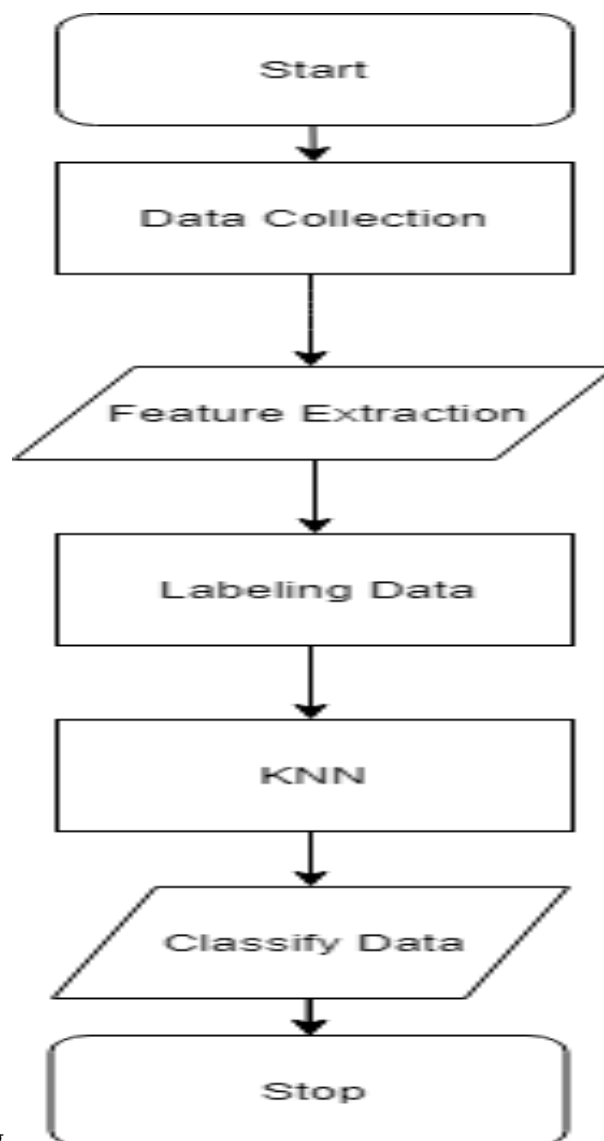3. Do you feel low seeing people's achievement.

Methods

Diagram.png

1 Step to solution

a) Creating data frame with necessary features.
b) Fixing dependent and independent variables.
c) Taking the dependent variables as the output.
d) Calculating accuracy.

2 Theoretical Background

2.1 KNN

KNN is an non parametric lazy learning algorithm. That is a pretty concise statement. When you say a technique is non parametric , it means that it does not make any assumptions on the underlying data distribution. This is pretty useful , as in the real world , most of the practical data does not obey the typical

theoretical assumptions made (eg gaussian mixtures, linearly separable etc) . Non parametric algorithms like KNN come to the rescue here.

It is also a lazy algorithm. What this means is that it does not use the training data points to do any generalization. In other words, there is no explicit training phase or it is very minimal. This means the training phase is pretty fast . Lack of generalization means that KNN keeps all the training data. More exactly, all the training data is needed during the testing phase. (Well this is an exaggeration, but not far from truth). This is in contrast to other techniques like SVM where you can discard all non support vectors without any problem. Most of the lazy algorithms – especially KNN – makes decision based on the entire training data set (in the best case a subset of them).

The dichotomy is pretty obvious here – There is a non existent or minimal training phase but a costly testing phase. The cost is in terms of both time and memory. More time might be needed as in the worst case, all data points might take point in decision. More memory is needed as we need to store all training data.

Results

Boxplot grouped by Do you feel low seeing other people's achievement on social media
Daily spent time in social media (hour)

[Do you feel low seeing other people's achievement on social media]

K=3

Predicted

True