

Information extraction from text messages using data mining

Ankit parmar¹, Darshil patel², Soham raval³, Aparna Kumari⁴

Institute of Computer Science and Technology, Institute of Technology, Ganpat University, Ganesh Meridian,
Ground Floor, Chanakyapuri Rd, Vishwas City 1, Chanakyapuri, Ahmedabad, Gujarat 380061

Emails: 1ankitparmar19@gnu.ac.in, 2darshilpatel19@gnu.ac.in, 3sohamraval19@gnu.ac.in



ABSTRACT

We are living in an era of increased pressure and mental disorders. The increased level of stress and pressure results in inclination of the number of people showing suicidal tendencies and thus a larger number of people are committing suicide. Stress can be caused due to family dispute, job dissatisfaction, health issues, etc. In the world of modern computing, people feel free to share their views and feelings over social media with peers and family members via services such as messaging. Due to the reserved nature and busy schedules of people it is extremely difficult to interact with peers and family members in person, therefore social media platforms are considered as the most used platform for personal conversations. The aim of this paper is to estimate the suicidal tendencies of a person by applying data mining techniques to the text messages a person sends to the associated people.

INTRODUCTION

The need to apply data analysis techniques to text messages comes from the ever-increasing levels of suicide in different parts of the world. Saving people's lives is a very important task for the nation. In order to save people's lives their feelings must be known and considered so that the necessary steps can be taken in a timely manner. The best way to find out about a person's feelings is to use data mining techniques in the messages the person is sending. When a person shows signs of severe depression informing the people around that person can help save that person's life. In Text processing is applied to text received from a user. Preliminary processing of text involves making tokens, deleting words and deleting titles and other techniques. Tokenization involves separating text in the form of words called tokens. Tokenization is used to identify keywords in text streams. Stop-word removal is the process of removing words that do not express a specific meaning in the text such as, again, this ... etc. Blocking is done to detect the origin of the data name and to exclude suffixes such as -ing, -ion, etc.

This paper focuses on analyzing emotions to predict a person's level of depression. The guessing model contains SVM and K-NN algorithms. This is done by supplying the system with a set of training program data. This framework can be used in a variety of contexts for other domains.



METHODOLOGY

Sentiment analysis:

In this component the data is assigned a sentiment such as positive or negative and the extent of it by performing data pre-processing using SVM algorithm.

Text Pre-processing: The processes involved in text pre-processing are.

Tokenization: Every new message is split into meaningful words called tokens. Example - “Morning walk is a bliss” is converted to “Morning” “walk” “is” “a” “bliss”.

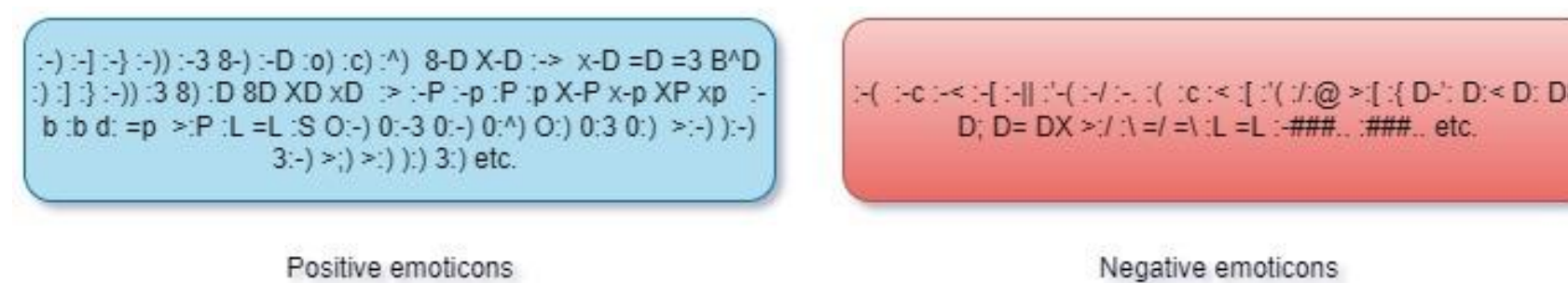
Data standardization: It involves converting all words in the message in standard form, converting all words in lower case.

Example: “The market is near Puneet’s house” is converted to “the market is near puneet’s house“

Emoji conversion: The emoticons present in the text messages are assigned a keyword based on the expression they convey.

The emoticons are classified into following two categories:

Positive emoticons: these are the emoticons which convey positive sentiment and are replaced by positive words based on the symbol.



Negative emoticons: these emoticons reflect the sad or disturbed sentiments of the subject and are thus replaced by negative words.

Stop-word-removal: All the words in the message which do not convey a special meaning are removed like a, the, then, etc[15].

Stemming: It involves obtaining the root word corresponding to every word by dropping suffixes like -ing, -ion, etc[7, 14]. Abbreviation analysis: Replacing the abbreviations present in the message by their full forms. Example FB by facebook, GM by good morning, etc.

RESULTS

The result obtained from the proposed model provides a limited sense of the subject based on the text messages sent by the user. Outcome can be used in many cases, mental disorders and depression levels are limited so if there are “critical” feelings your peers and family members can take steps to encourage, motivate and elevate the emotional state of the subject. thus resulting in the harmony and peace of mind of the subject. Such emotional analysis models are therefore essential to shaping society into a reality.

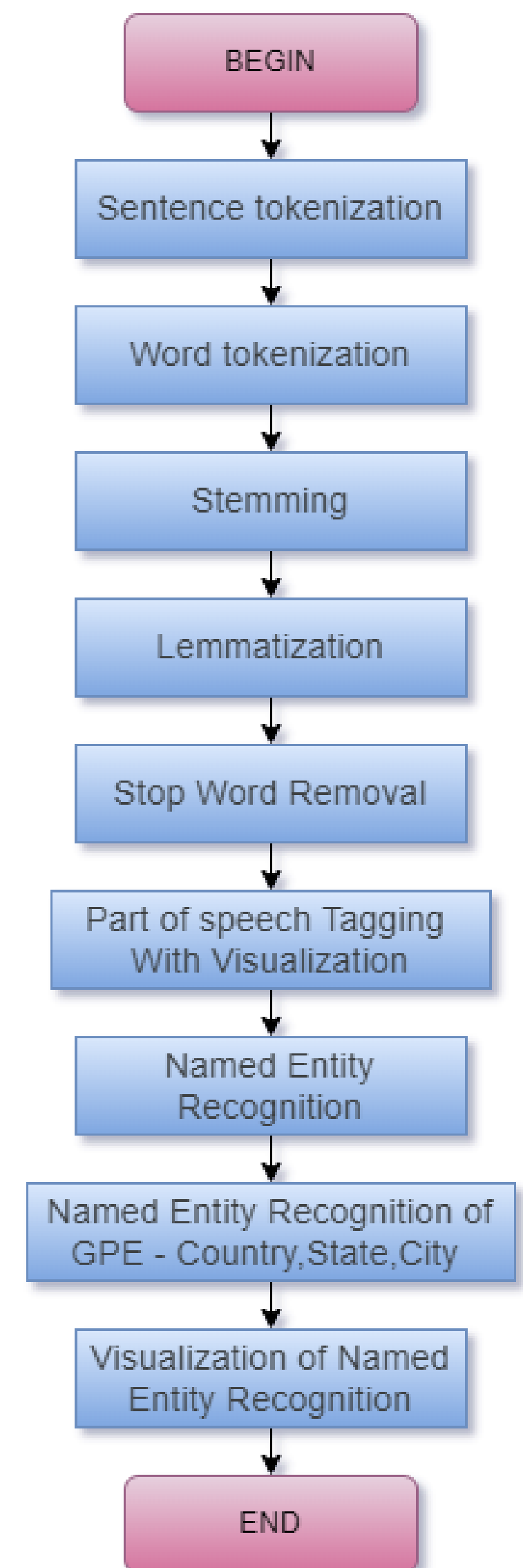
FUTURE SCOPE

The proposed model can be used in situations where emotional analysis is needed to achieve the desired result and for various purposes such as reviews of hotel critics, movies, videos, etc. Emotional analytics techniques have so far been used to identify differences in the thoughts and opinions of all users who enter social media. Businesses are keenly interested in understanding people's thoughts and how they respond to all the products and services around them. Companies use emotion analysis to evaluate their marketing campaigns and improve their products. Companies intend to use such emotional analysis tools in the areas of customer feedback, marketing, CRM, and e-commerce.

REFERENCES

1. K. Tan, Steinbach, Introduction to Data Mining, 2006.
2. D. Lyon and B. Cedex, N-grams based feature selection and text representation for Chinese Text Classification Zhihua WEI, Int. J. Comput. Intell. Syst., 2(4), 365–374, 2009.
3. J.C.B. Christopher, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2(2)(1998), 121–167.

METHOD FLOWCHART



Flow Chart

CONCLUSION

The proposed model takes input from text messages send by the subject. The messages are then preprocessed to obtain the key words from the text. After preprocessing, Thus we propose to give a highly efficient method of finding the sentiment of the person by analysing the text messages and also processing emoticons. Emoticons are very common tokens in any text message in the new world, therefore we must also focus on efficient ways to analyse them. We have converted emoticons to textual form for our computation processes. Thus this model is a requirement and a life saviour in the modern world.

CONTACT INFORMATION: Ankit Parmar, Darshil Patel, Soham Raval