DATA 698 Research Project Proposal

Topic: Student Counselling Software

Irene Jacob

Introduction

Today students face various problems like stress, homesickness, exam fear and so on. This affects their studies as well as makes student life which is to be spent cheerfully filled with stress and sadness. But in most institutions, there is only one counsellor for the entire student body. With the help of a student counselling software, that is freely and easily accessible with a click of a button, students will be able to identify their issues and problems.

Relevant research

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document.

Students encounter problems such as heavy study load, lack of social engagement, and sleep deprivation. Chronic student problems can have lasting effects. Student counselling is used to encourage conversation among students. It can also help to process feelings of being excluded or rejected. Students may feel supported and encouraged to manage changes with the help of the student counselling software.

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes, and emotions toward an entity. The entity can represent individuals, events, or topics. These topics are most likely to be covered by reviews. The two expressions SA or OM are interchangeable. They express a mutual meaning.

Opinion Mining extracts and analyses people's opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyses it. Therefore, the target of SA is to find opinions, identify the sentiments they express, and then classify their polarity
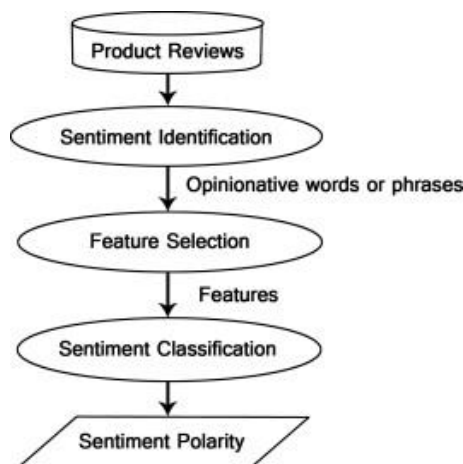
The social network sites and micro-blogging sites are considered a very good source of information because people share and discuss their opinions about a certain topic freely. They are also used as data sources in the SA process.

There are many applications and enhancements on SA algorithms that were proposed in the last few years.

Paper Details

Paper 1: X. Chen, M. Vorvoreanu and K. Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences," in IEEE Transactions on Learning Technologies, vol. 7, no. 3, pp. 246-259, July-Sept. 2014, doi: 10.1109/TLT.2013.2296520.

> This paper tries to understand engineering student's experiences with the help of different large-scale data mining techniques and qualitative methods. Massive study load causes a lot of issues with these students, and they do not

have a proper system in place to help them sort through these consequences. Culture stereotypes and conflicts are also dominant among these students.

Paper 2: Walaa Medhat, Ahmed Hassan, Hoda Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams Engineering Journal, Volume 5, Issue 4, 2014, Pages 1093-1113, ISSN 2090-4479, https://doi.org/10.1016/j.asej.2014.04.011.

This paper presents the different sentiment analysis techniques and applications in the related fields. It categorizes many articles and illustrates the current trends of sentiment analysis. It is a very helpful survey for entry-level analysts in this field to get familiarized with the concepts. The latest fields where sentiment analysis is applied includes Transfer Learning, Emotion Detection and Building Resources. These are also discussed in this survey.

Paper 3: C. Kaur and A. Sharma, "Social Issues Sentiment Analysis using Python," 2020 5th International Conference on Computing, Communication and Security (ICCCS), 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9277251.

This paper analyses how people express their feelings using social media platforms. It uses a number of Machine learning algorithms to classify the posts as negative, positive or neutral. It combines both supervised and unsupervised learning algorithms to classify the emotions.

Paper 4: Kharde, Vishal & Sonawane, Sheetal. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications. 139. 5-15. 10.5120/ijca2016908625.

This paper provides a survey on the different methods of opinion mining such as machine learning and lexicon-based approaches. The conclusion made from this survey suggests that Naïve Bayes and SVM are most accurate. It was observed that clean data provided more accurate results.

Research Question

The primary objective of this project is to develop counselling software using sentiment analysis. This can help reduce tensions among students and make school a happy abode.

The secondary objective of this project is to learn programming in python and use of NLTK (Natural Language Tool Kit) as well as to gain real world experience in developing a project.

1. To know and understand the comparison between accuracy/results of different data mining techniques
2. To find out why twitter is the most preferred platform compared to other social media platforms like facebook or Instagram for data mining
3. Why is python useful for analysis such as sentiment analysis?

Methodology

The software takes the text the user enters describing his issues as the input. It processes this text to identify the problems faced by the student to give counsel.

Here sentiment analysis is used to build student learning. It is to be done in python using NLTK. A person in need of counselling enters his/her grievances in a paragraph or two. This text acts as the input to the program. The program is expected to generate the necessary advice for the client by identifying his problem (say home sickness) from the input. The model is trained with a training set of data obtained from twitter and is expected to produce output when it encounters a new input.

The first stage is about processing natural language, and the second about training the model. The first stage processes text in a way that, when we are ready to train our model, we already know what variables the model needs to consider as inputs. The model itself is responsible for learning how to determine the sentiment of a piece of text based on these variables. Output of this software will be the prediction of the cause or root of the described issue (text input to software from student).