

Project Proposal

Topic : Students' Behavior and Feedback Analysis on Online Learning System during COVID-19

I. Presentation of the project

Feedback review data plays a vital role in business development, particularly in increasing the number of sales in the company. Opinion mining, which is, given a bunch of text, we can computationally study people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes. For example, businesses always want to find public or consumer opinions and emotions about their products and services.

In order to develop an effective learning curriculum, it is important to know the opinions and feedback from students. Proposed by (Rain, 2013), they mainly used Multinomial Naive Bayesian (MNB) and support vector machine as the main classifiers by extending the current work in the field of natural language processing. Naive Bayesian and decision list classifiers were used to classify a given review as positive or negative. Sentiment analysis, however, helps us make sense of all this unstructured text by automatically tagging it. Sentiment analysis helps us to process huge amounts of data in an efficient and cost-effective way.

Instead of receiving the feedback from students and reading them manually one by one, it is very a good contribution to perform Students' Behavior and Feedback Analysis on Online Learning System during COVID-19.

II. Objectives

In this project, we would like to apply machine learning algorithms to perform analysis Students' Behavior and Feedback Analysis on Online Learning System during COVID-19 using the evaluation survey to be collected at the selected universities in Phnom Penh. By the end of this project, we expect to accomplish a scientific paper based on this research work.

III. Planning

This internship duration is 4 months which start from the beginning of August to December. The table below illustrates the internship program:

Tasks	Weeks															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Training																
Define Problem																
Setup environment																
Data Collection																
Exploratory Data Analysis																
Feedback and Customers' Behaviour Analysis																

IV. Dataset

We initially conduct the research on the available methods of performing the sentimental and feedback analysis on the existing dataset consists of 568,454 records from the API of Amazon database, specifically on customer review data. There are a total of 10 attributes describing the product data and customers' feedback and the users' identity are anonymized. Then, we are going to collect our original dataset from the evaluation survey to be conducted in the selected universities in Phnom Penh before using the techniques that obtained from the state-of-the-art to apply in our study.

V. Methods

- **Exploratory Data Analysis :** In this part, we are going to combine the data from the individual datasets of each year and we are going to explore the data in order to gain more insights from the data including removing redundant attributes, missing values, excessive attributes and conduct data visualizations.
- **Feedback Analysis:** For performing sentiment analysis, we will use NLTK package of the Python. WordNet is a large lexical database of English developed by the Princeton University. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a

distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations (Fellbaum, 1998). In other words, Wordnet can be described as online thesaurus. It tells you about word meanings and relationships between word meanings. Polarity is float which lies in the range of $[-1,1]$ where 1 means positive statement and -1 means a negative statement.

- **Students' Behaviour Analysis** : Subjectivity is used for individual sentences to determine whether a sentence expresses an opinion or not. In terms of subjectivity, textual information in the world can be broadly categorized into two main types: facts and opinions. Subjective sentences generally refer to personal opinion, emotion, or judgment whereas objective refers to factual information. Facts are objective expressions about entities, events, and properties. Opinions are usually subjective expressions that describe people's sentiments, appraisals, or feelings toward entities, events, and their properties (Liu, 2010).

VI. References

- Bradford Books. Guibon, G., Ochs, M., & Bellot, P. (2016, June). From emojis to sentiment analysis.
- Rain, C. (2013). Sentiment analysis in amazon reviews using probabilistic machine learning. Swarthmore College.
- Fellbaum, C. (1998). WordNet: An Electronic Lexical Database.
- Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of natural language processing, 2(2010), 627–666.