ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 7, Issue 2, pp: (43-64), Month: April - June 2019, Available at: www.researchpublish.com

Mining Social Media Data of Philippine Higher Education Institutions Using Naïve Bayes Classifier Algorithm

Joey Aviles[1], Rosanna Esquivel[2]

[1]Panpacific University North Philippines – Tayug Campus Inc., Philippines

[2]Angeles University Foundation, Philippines

Abstract*: Philippine Higher Education Institutions (HEIs) integrate social media (SM) platform like Facebook to digitally market their position and brand, post announcements, news and other pertinent information about their institution. Though SM is used in education, it is noted that its primary purpose is for users' online social engagement. SM users react to posts, comment and give their sentiments on a particular topic to discuss and share their daily life in an informal and casual manner. Students even create unofficial Facebook pages associated to their institution and post their experiences and sentiments. These digital footprints of SM users could be a good source to use Data Mining (DM) techniques to understand HEI stakeholders' experiences and form decision-making interventions to improve quality education. The main objective of this study is mining social media data to support Philippine HEIs identify their stakeholders' SM behaviors and make proactive and knowledge-driven decisions. Specifically, the objectives of this study are: to develop a workflow of analyzing social media data; to use Naïve Bayes Classifier in identifying the SM behaviors of Philippine HEI stakeholders; and to interpret the results of DM and propose intervention measures to improve quality education. Using Agile Methodology CRISP-DM model, this study centers on mining users' post and comments in Philippine HEI Facebook pages.

***Keywords:* Data Mining, Social Media, Higher Education Institutions, Philippines, Naïve Bayes.**

1. INTRODUCTION

The higher education landscape has evolved in many years. This change has been conveyed by many factors, one of which is embracing technology. Education today is digitally revolutionized with the rapid use of Information and Communications Technology (ICT).

Now, students have access to tons of information online. Collaboration among them and their teachers has now been limitless on boundaries. This learning community brings them to communicate and exchange ideas and thoughts anywhere and anytime around the globe. The 21st century learning made students to be independent and creative allowing them to take risks and learn deeply. Brought about by technology, online classes and lectures are now conducted virtually. These days, teachers are more innovative in shaping and inspiring young minds. They adopt new media of learning and incorporate technology in classrooms [1]. Indeed, ICT has changed the landscape of education especially on the increasing utilization of the interconnection of networks worldwide, the internet.

Internet technology has been successful in education because it answers the basic needs for information exchange, communication and collaboration.

The most popular internet technology is the use of Social Media (SM) with high user engagement rates due to increased worldwide usage of smartphones and mobile devices. In 2019, statistics shows that out of 4.2 billion internet users, there are 3.397 billion active social media users [2].

Moreover, SM is an important source of learning of opinions, sentiments, subjectivity, assessments, approaches, evaluation, influences, observations, feelings, borne out in text, reviews, blogs, discussions, news, remarks, reactions, or some other documents [3].

In the 2018 report of We Are Social and Hootsuite, Philippines is the country that spends most time on Social Media. At average, almost 4 hours daily are spent by Filipino SM users on different platforms followed by Brazilians and Indonesians. Among the SM platforms, Facebook remains on the lead with 2.32 billion users.

SM is becoming more and more widespread in Higher Education Institutions (HEIs). The trend can be observed in many higher education institutions around the world [4]. It does not only offer the academic community to communicate beyond local or social boundaries, it also offers possibilities to share user-generated content like pictures, videos, opinions, interests and other forms of expressions.

With this, many academic institutions have started to integrate SM platform to digitally market their position and brand. Aside from having their own website, many HEIs create their Facebook Pages to reach their stakeholders in SM by posting announcements, news and other pertinent information about their institution. Traditional media such as television, newspaper, radio and magazines are one-way, static broadcasting technologies [5] unlike SM where everyone can give feedback, create and distribute their own content.

This moving landscape makes teachers explore its significance to drastically transform the pedagogical basis of their teaching experience, giving them the tools in order to create truly adapted and flexible learning experiences for students [6].

Studies suggest that the high take up of SM applications as an addition to formal educational settings offers new opportunities for innovating and modernizing education institutions and for preparing learners for the 21st century [4].

Though SM is used in education, it is noted that its purpose is for users' social engagement online. In fact, 71% of teachers use it for personal use [7]. Furthermore, a consumer insight service mentioned that more than 98% of college-aged students use SM [8].

SM users react to posts, comment and share their sentiments on a specific topic. Students for example discuss and share their daily life in an informal and casual manner in different SM sites [9]. Students' casual discussion on SM focused on their educational experience, mind-set, and worry about the learning procedure [10]. They even create unofficial Facebook Page associated to their institution and post their experiences and sentiments both positive and negative.

In 2013, confessions of students from the University of the Philippines were raised on 'The Diliman Files Facebook Page'. Since then, similar pages have emerged in other colleges and universities in the country garnering hundreds and thousands of likes. These unofficial Facebook pages served as a means for students to talk about their lives as students. Rappler, an online news website in the Philippines, spoke to Secret Files page administrators form the University of Sto Tomas, De la Salle University Manila and Ateneo de Manila University on how university communities are being reshaped through mass exchange of secret messages. Many of the posts talked about were love confessions and issues about income bracket-based discrimination on campus. Each page has encountered fare share problems. The page administrator of UST Secret files for example had trouble aligning the nature of the page with the identity of UST. They were not able to consider that UST is a catholic school and putting up stories with mature content would not be nice for the reputation of the school. De la Salle University Manila secret files also had problems with the University administration and its alumni. A senior student of the university said that the page is not being in line with the 'Lasallian core values', it ruined the reputation of the school, and risked their own names associated with malicious posts. Though the page was requested for shut down, they still refused to back down because of the encouragement of its readers to continue the page.

These Facebook page secret files let SM users exercise their freedom of speech. However, the anonymity encouraged gossips among students and faculty. Female students became target of lascivious post. Universities felt helpless to go after these pages because they continue to hide on fictitious names and accounts [11].

These digital footprints of SM users could be a good source for researchers and HEIs to use Data Mining (DM) techniques to understand their stakeholders' experiences and form decision-making interventions to improve quality education.

DM is used in the education sector for several reasons. One way is to effectively address the challenges for improving the quality of education to provide new knowledge related to the educational processes. This knowledge can be extracted from historical and operational data that reside in the educational organization's databases using DM [12].

DM in education is used to develop models for improving both learning experiences and institutional effectiveness [13]. When it comes to answering the issues of predictions of student's performance and their profiling, DM tools and techniques can be effectively used [14]. Also, it can be used for classifying and predicting the students' behaviour, performance, dropouts as well as teachers' performance [15].

Furthermore, DM techniques have shown to be capable of mining big data generated on SM sites. This is made possible by way of extracting information from large data set generated on SM and transforming them into understandable structure for further use (Olowe, Gaber, & Stahl, n.d.).

Though there are many applications of DM in education, there is little empirical research conducted to analyze amorphous SM data in this sector. [16] recommended in their study that the area of SM calls for more profound research that takes into account accurate implementation of DM techniques in the academic sector. It is in this context that this study was synthesized.

This study centers on mining SM data specifically on users' post from Facebook (FB) pages among selected HEIs in the Philippines. The challenge for this research is on how to analyze such unstructured data and transform them into valuable results to help HEIs understand its stakeholders' behaviors in SM and eventually make some interventions and improve quality education. Through this study, HEIs could augment their level of instruction like developing a curriculum that addresses the holistic development of a learner. HEIs could also improve their policies. As said by [11], standards of discipline as set by the school policy should be higher for conduct in social media and cyberspace.

Statement of the Problem

The main problem of this study is how to perform data mining on social media data associated with Philippine HEIs.

Specifically, it deems to answer the following queries:

1. What could be the workflow to analyze social media data?

2. What are the behaviors of HEI stakeholders in SM based on the result of using Naïve Bayes Classifier?

3. What could be the proactive and knowledge-driven decision intervention measures to improve quality education for HEIs in the Philippines?

Objectives of the Study

The main objective of this study is to apply data mining on social media data associated with Philippine HEIs.

Specifically, the objectives of this study are:

1. to develop a workflow of analyzing social media data

2. to use Naïve Bayes Classifier in identifying the behaviors of Philippine HEI stakeholders in SM, and

3. to interpret the results of the data mining and propose intervention measures to improve quality education for Philippine HEIs.

Significance of the Study

This study attempts to analyze the social media data associated to higher education institutions in the Philippines. The contribution of this research undertaking shall benefit the following:

Higher Education Institutions (HEIs). The result of the data mining shall be an instrument for HEIs in the Philippines to understand their stakeholders' behavior in social media and later develop strategies and interventions to improve quality education.

Stakeholders. With the proposed measures in improving quality education, services shall be value-added for stakeholders.

Other Researchers. This study shall serve as a reference for other researchers exploring data mining social media in education.

Scope and Limitation of the Study

This study is bound to identify the social media behaviors of higher education institution stakeholders in the Philippines. Posts from Facebook pages associated to selected universities from at most the past three years or at most 1000 posts and comments observations shall be the main source for data mining regardless of the number of characters per post or comment. Facebook is the target SM site of this study because unlike Twitter, Facebook has 63,206 characters for every status update and 8,000 characters for comments. Only text posts and comments shall undergo data mining. Pictures, videos, and sound files in the Facebook pages are not part of the data mining.

2. REVIEW OF RELATED LITERATURE

Data Mining Social Media

It is very significant to research on mining social media because of the increasing amount of publicly available data such as customers' sentiments. Individuals, companies, organizations and even the government of different countries track the activities of their audience on SM. They do so in order for them to get knowledge on how their audience reacts to postings that concerns them [17]. They are conscious on the significance of the sentiments of their stakeholders as posted in SM to protect their image and for further development of their products and services.

SM can surpass the boundaries of the physical world to learn human relationships [18] and help measure common social and political sentiment associated to regional populations without using unambiguous surveys [19]. The enormous data constantly collected on these sites make it difficult for traditional methods like the use of field agents, clipping services, and ad hoc research to handle SM Data [20]. Therefore, it is necessary to employ tools in data mining that are capable of analyzing the expression of sentiments in SM. These sentiments expressed on SM could be analyzed using various data mining techniques [21].

The core functionalities of DM include applying various methods and algorithms to discover useful patterns among the data. It is capable of handling three dominant disputes with SM data which are size, noise and dynamism [22]. SM data sets are very voluminous. As data mining requires huge data sets to mine, SM appears to be perfect sites to work on especially where opinion or sentiment expression is involved [23]. Users of SM like Facebook and Twitter post and tweet respectively irrelevant data making the characteristics of the data sets to be noisy. DM is also very much applicable to SM since the data sets are versatile and evolving rapidly overtime causing them to be dynamic in nature. DM is examining the data stored in very huge sources which are analyzed from multiple perspectives. The discovered knowledge is the result that is summarized to become useful information [24]. Its core functionalities in order to discover useful patterns of stored data include the application of various methods and algorithms in order to preprocess, classify, cluster, and associate the data [25]. To extract meaningful information and knowledge from amorphous data, DM techniques are used. From huge data sets, DM extracts patterns, changes, relations, and variances [26]. Its goal is to find valid, novel, potentially useful and understandable associations and patterns in the existing data [27]. DM is a confluence of multiple fields including statistics, machine learning, databases, pattern discovery and visualization [28]. In practice, prediction and description are the primary goals of DM. Predictive DM creates the model of the system from the given data [29] which involves predicting unknown or future values of interest by using some variables or fields in the database. Descriptive DM on the other hand spawns noteworthy data sets from the current data which focuses on finding patterns defining the data.

Ever-increasing interest is being put into attention on multi-lingual data mining: the ability to extract information across multiple languages and group similar items from different linguistic sources based from their meaning. Text data mining is the process of deriving high-quality information from text. It involves the application of techniques from areas such as information retrieval, natural language processing, information extraction and data mining [30].

How to clean the data is one of the major challenges in data mining. Data cleaning also called data cleansing or scrubbing [31] deals with detecting and removing errors and inconsistencies from data in order to improve the quality of data [27]. This is very much needed in the study in order for the unstructured data from SM to be prepared for analysis, especially that these concerns the V's of big data namely: velocity, volume, value, variety and veracity. Velocity of data in the SM giant Facebook is very fast. Every 60 seconds on Facebook, there are 510,000 comments posted, 293,000 statuses updated, and 136,000 photos uploaded [32]. When it comes to volume, 1.52 billion people on average, log onto Facebook daily and are considered as daily active users. Personal user data is worth to Facebook because it is a digital ad broker of user data for targeting ads [33]. Data are noticeably unstructured. 80% of all world's data are photos, videos and social media updates. Veracity is also a concern on social media data. In Facebook for instance, there are 83 million fake profiles. Fake or not, these are still consumers of social media which should be taken into account when implementing data mining.

Data can be mined with the following methods as mentioned by [34]: characterization, classification, regression, associattion, clustering, change detection, deviation detection, link analysis and sequential pattern mining. Among the data mining methods, classification shall be used in this study which is explicitly discussed in the succeeding discussions.

Using Naïve Bayes Classifier in Identifying Sentiments in SM

Sentiment Analysis (SA) or sometimes referred as Opinion Mining [35] is a type of DM that measures the opinions of people through Natural Language Processing (NLP), computational linguistics and text analysis. Such are used to extract and analyze subjective information from the web – mostly SM (techopedia.com). SA is a prevalent study these days because of SM where users are free to express their thoughts, feelings and impressions online [36]

The fields of opinion mining and sentiment analysis are distinct but deeply related [37]. Opinion mining focuses on the positivity, negativity or neutrality of what is being mined or so called polarity detection. Sentiment analysis on the other hand involves emotion recognition. Since identifying the polarity of text is often a phase in sentiment analysis, the two fields are usually pooled under the same umbrella. Sentiment Analysis (SA) is a type of natural language processing for tracking the moods and sentiments of the public about a particular service, product or topic. It may involve building a system or automated method of collecting and examining opinions about the product made through blogs posts, comments, reviews or tweets [38]. SA is the detection of attitudes [39]. In the Scherer Typology of Affective States, attitudes refer to enduring, affectively colored beliefs and dispositions towards objects or persons. Other states are emotion, mood, interpersonal stances and personality traits. In order to get the sentiments, you need to identify the holder or source of attitude, the target or the aspect of the attitude, the type of the attitude and the text of the attitude. It is necessary to implement sentiment analysis in SM because it basically recognizes the potential drift in the society as it concerns the attitudes, observations, sentiments and other expectations of stakeholder or populace and to make prompt necessary decisions [22].

There are different studies conducted to analyze data using data mining. [40]used tweets ending in positive and negative emoticons, then build models using Naïve Bayes, MaxEnt and SVM. [41] also reviewed several techniques in mining SM data. The Study illustrated some applications of DM, specifically in telecommunication industry, to support customer satisfaction and maintain customer relationship. The techniques used were text mining, clustering and visualization.

A similar study was also conducted by [42]. It focuses on SA to estimate the Filipino internet customers' satisfaction related to the quality of the service provided by the Internet Service Providers (ISPs). Data were collected from Blog comments shared with online social media. Using the word pair set {"Good" and "Slow"} as initial seed for the word dictionary, automatic word seed selection was applied. The Naïve Bayes method was used to identify the frequent words used to express the sentiments of customers and to determine the polarity of their opinions. The downside of the study were various factors that are causal to low performance. These are language translation tools, uncertainty of synonyms and antonyms of words in online dictionary, use of urban words, colloquial language, Short Message (SMS), and long words.

[16] identified 19 DM Techniques as shown in the next Figure. Among the DM Techniques, Support Vector Machine (SVM), Bayesian Network (BN), and Decision Tree (DT) were identified as the most applied techniques in the area of social media with a percentage of 51% of the selected articles reviewed.

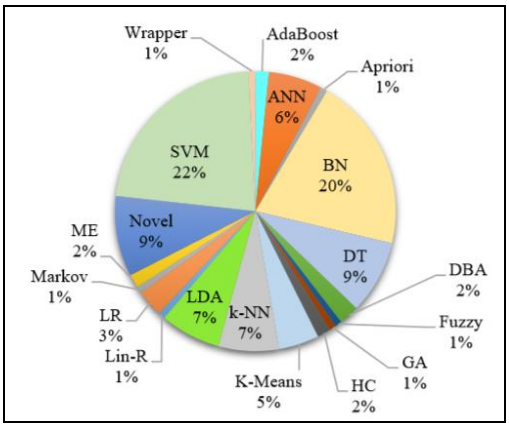

Fig 1: Data Mining Techniques among Selected Papers

Moreover, the study identified that social network data analysis and business and management were the most active domains that require mining of social media data while the education sector was the least as shown Table 1.

Table 1: Domains in Mining SM data

Sector	Percentage
Business and Management	17%
Education	1%
Finance	3%
Government and Public	9%
Medical and Health	8%
Social Networks	62%

Additionally, Table 3 shows the sentiment analysis tools in SM as reviewed in the study of [17] entitled: A Survey of Data Mining Techniques for Social Media Analysis, ranging from unsupervised, semi-supervised to supervised learning.

The study revealed that different levels of success have being made with the use of these techniques either singularly or combined. Furthermore, it revealed that there are different ways in which the opinion or sentiment of SM users can be considered when making valuable decisions whether individual, organizational or governmental level. The review concluded that although some studies considered data mining techniques like Association Rule Data Mining, Decision Tree, KNN and Neural Network techniques, they have not gained popularity as much as SVM, Naïve Bayes and Maximum Entropy. Moreover, the researchers of the study expect that future researches will make use of both currently used and yet-to-be-explored data mining techniques to investigate deeper into mining the ever growing online data created on SM.

Likewise, the results of the investigations are expected to support various entities in retrieving vigorous information on SM and consequently using this information as decision support tools.

[43] conducted a performance study of algorithm based on lowest computing time and accuracy. It concluded that Naïve Bayes is a superior algorithm because it takes lowest time and at the same time providing highest accuracy as shown in the next table.

Table 2: Performance of Algorithms based on computing time and accuracy

Classifier (699 instances)	Algorithms implemented	Time Taken (Sec)	Correctly classified instances (%)	Incorrectly classified instance (%)	Kappa Statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)	Root relative squared error (%)
Decision Tree	Random Forest	0.11	95.9943	4.0057	0.9112	0.0574	0.1744	12.6928	36.6836
Bayes Classification	Naïve Bayes	0.02	95.9943	4.0057	0.9125	0.0394	0.1953	8.7175	41.0975
K-Nearest Neighbor	ibk	0.02	94.9928	5.0072	0.8886	0.0487	0.2161	10.7796	45.4728

Table 3: Data Mining Techniques Used in SM

Unsupervised	Semi-Supervised	Supervised
Binary Distinction of Positive and Negative	Transductive Support Vector Machines (TSVMs)	Naïve Bayes
Pattern Extraction	Randomized Mincuts	Decision Tree
Bootstrapping	Graph-based Semi supervised Learning	KNN
Blank Slate Method		Neural Network
Semantic Orientation		Review and Rating (RnR)
Unsupervised Lexical Classifier		Latent Aspect Rating Analysis (LARA)
Opinion Definition		Text Mining

		Aspect-rating of Short Text
Opinion Summarization		CHAID
K-Means		Bayesian Networks
Hierarchical agglomerative clustering		Association Mining Rule
Hierarchy-based mood classification		

Another study using Naïve Bayes Classifier from random sample of streaming Facebook statuses was used in the study conducted by [44]in using SA for language learning. The result of the experiment shows that the accuracy in analyzing the sentimental state of Facebook users, using the Naive Bayes Classifier, is really high.

[45] also take a Naïve Bayes multinomial approach to collect and classify tweets. Similarly, [46], worked on sentiment analysis of English Tweets using Rapidminer. They collected the dataset that are in natural language and applied text mining techniques to build sentiment classifier. Likewise, Twitter microalgae and various classifiers like Naive Bayes, SVM, K-Nearest Neighbour were used by [47] to do sentiment analysis on Arabic Tweets.

Furthermore, Sentiment classification of social media with the help of data mining techniques is done on the research conducted by [48]. They used three classifiers - K-NN, Naive Bayes and Decision Tree. The result shows that the accuracy of K-NN, Naive Bayes and Decision Tree wich are 77.50%, 80% and 78% respectively.

The study of [10] pays attention on engineering students’ tweets on Twitter to know the latter’s problems and troubles in their educational practices. It is found out that engineering students encounter problems such as heavy learning load, lack of social meeting, and sleep deficiency. Based on this outcome, Naïve Bayes Multi-label Classifier algorithm is applied to categorize tweets presenting student’s problems. Decision tree algorithm is applied to make more accurate result.

Similar study was conducted by [9]. The study explored on understanding engineering students’ experiences by integrating both qualitative methods and large-scale data mining techniques. Through Qualitative content analysis and Naïve Bayes Multi-Label Classifier, the study found out that heavy study load is mostly the biggest problem of engineering students of Purdue University. This problem further leads to many consequences including lack of social engagement, sleep problems and other psychological and physical health problems. Table 2 shows the top most probable words in each category ranked using the conditional probability.

Whith these literatures cited, it could be noted that the best classifier to be used with social media dataset is Naïve Bayes. However, the data of most researches in data mining social media are from Twitter. Thus, this study delved on gathering Facebook data specific to HEI pages then eventually do sentiment analysis using Naïve Bayes Classifier.

In Naive Bayes classifier, the result of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. Naïve Bayes algorithm uses conditional probabilities by counting the occurrence of values and combinations of values in the historical data and is therefore called the probabilistic method. There are four applications of the Naïve Bayes Algorithm according to Analytics Vidhaya, the provider of a community-based knowledge portal for Analytics and Data Science professionals. These are real time prediction, multi class prediction, text classification / spam filtering / Sentiment Analysis and Recommendation System.

Table 4: Top 25 Most Probable Words in Engineering Problems

Category	Top 25 words
Heavy Study Load	hour, homework, exam, day, class, work, problem, study, week, too,much, all, lab, still, out, time, page, library, spend, today, long, school, due, engineer, already, negtoken
Lack of Social Engagement	Friday, homework, out, study, work, weekend, life, class, engineer, exam, drink, break, Saturday, people, social, lab, spend, tonight, watch, game, miss, party, sunny, beautiful, negtoken
Negative Emotion	hate, f***, shit, exam, week, class, hell, engineer, suck, study, hour, homework, time, equate, FML, lab, sad, bad, day, feel, tired, damn, death, hard,negtoken

Sleep Problems	sleep, hour, night, bed, allnight, exam, homework, nap, coffee, time, study, more, work, class, dream, ladyengineer, late, week, day, long, morning, wake, awake, nosleep,negtoken
Diversity Issues	girl, class, only, guy, engineer, Asia, professor, speak, English, female, hot, kid, more, toomuch, walk, people, teach, understand, chick, China, foreign, out, white, black,negtoken

Additionally, the Naïve Bayes algorithm is one of the most used supervised learning in Sentiment Analysis as reviewed by [17]. It is a classification technique based on Baye's theorem with an assumption of independence among predictors [49]. In multinomial naive Bayes' classifier, we find P(c|d) for each class c in C: the probability of returning class c given that our observation is d. Next is, we find ĉ, the maximum of {P(c1|d), P(c2|d), ..., P(cn|d)}. ĉ is "our prediction of the correct class".

$$\hat{c} = \max_{c \in C} \Pr(c \mid d)$$

As cited in [50]in simplifying assumptions, first we treat the text as a bag of words. Second, we treat the relative frequency with which a word appears in a document as a feature (multinomial).

$$\hat{c} = \max_{c \in C} \Pr(c) \Pr(f_1, f_2, ..., f_n \mid c)$$

Third, is Naïve Bayes assumption that each feature is independent from the others.

$$\Pr(f_1, f_2, ..., f_n \mid c) = \Pr(f_1 \mid c) \cdot \Pr(f_2 \mid c) \cdot ... \cdot \Pr(f_n \mid c)$$

For text classification specifically, we assume a feature is just the existence of a word in a document, so we can find $P(w_i|c)$ by iterating through every word in d.

$$\hat{c} = \max_{c \in C} \Pr(c) \prod_{i=1}^{n} \Pr(w_i \mid c)$$

3. METHODOLOGY

Using Applied research, Agile Methodology was used in this study following the CRISP-DM model which stands for "Cross Industry Standard Process for Data Mining", a proven method for the construction of a data mining model. This methodology conceptualized data science as cyclical process.

The methodology provides a framework that includes stages as shown in figure 2.

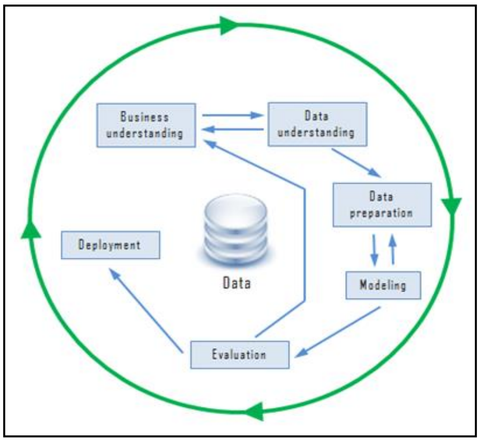

Figure 2: Agile - CRISP - DM

Business Understanding. This is the preliminary phase for understanding the objectives and requirements of the project from a business perspective. The converted knowledge is then transformed into a data mining problem definition.

In this phase, the researcher identified that there is a problem most HEIs face in the behaviors of their stakeholders in SM. The view that mining data generated on Facebook pages associated to Philippine HEIs was ideated.

Data Understanding. This phase is the data collection. It involves processes on data familiarization, identify the problems, and discovering first insights or detecting interesting subsets to form hypotheses for hidden information.

In this phase, the researcher collected the data from Philippine HEI facebook pages. A letter was written and submitted to the Commission on Higher Education (CHED) in Region 1 to ask permission for the data collection. After which, CHED forwarded an indorsement letter to all HEIs in the Region to allow the researcher for data collection subject to the Data Privacy Act and Institutional Policies.

Data Preparation. This phase covers all activities to construct the final data set.

In this phase, the data collected are prepared for data cleansing.

Modeling. In this phase, a modeling technique is selected and applied.

Naïve Bayes Classifier was used in this phase to classify the behaviors of the HEI stakeholders.

Evaluation. A high-quality model shall be built in this phase from a data analysis perspective.

A training data set shall be built for testing the other data sets.

Deployment. The model shall be used to increase knowledge of the data.

The selected Facebook pages associated to HEIs underwent data mining specifically using Naïve Bayes classifier. Qualitative analysis shall also be undertaken to gather in-depth understanding of the sentiments of the users.

Population and Locale of the Study

Of the 17 regions in the Philippines, the collection of data focused in the Ilocos Region (Region 1) where there are 75 private and public colleges and universities. With 5% of margin error and 95% confidence level, the sample size for the data collection is 63 HEIs.

4. DISCUSSION OF FINDINGS

This chapter presents the findings of the study. The following discussions cover the social media data associated to HEIs as provided by its respective stakeholders. Specific discussions cover the workflow on how the social media data were analyzed, the application of Naïve Bayes Classifier on these data and the proposed intervention measures to improve quality education.

Workflow Used In Analyzing Social Media Data

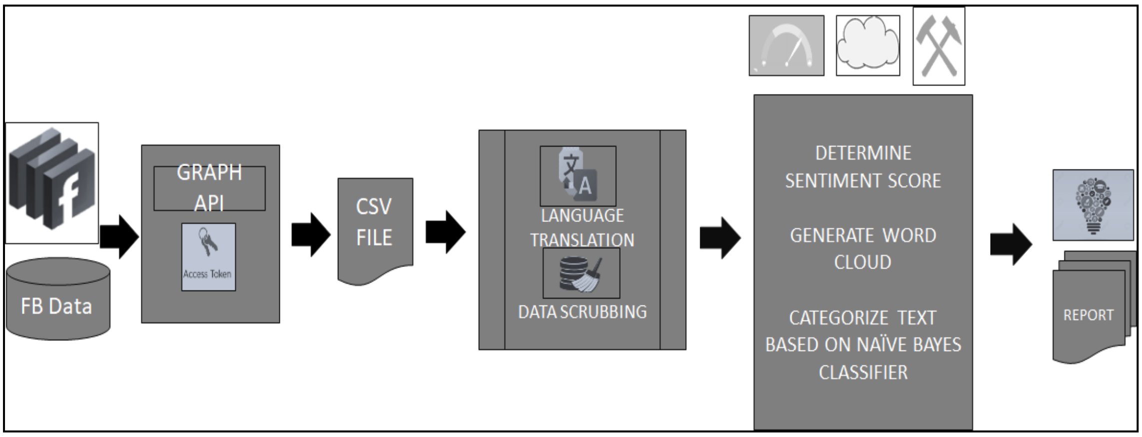

Figure 3: Workflow Used in Analyzing SM Data

A web-based application was developed in this study to dynamically classify the posts and comments in Facebook pages of HEIs. Figure 3 illustrates how the data were collected, cleansed and transformed into a valuable information.

Using access tokens that conform to the OAuth 2.0 protocol, the Facebook data are fetched through Facebook Graph API v. 3.2. The Graph API is HTTP based and is used to write to the Facebook social graph. Since this project is in its inception, Page APIs have restrictions where the application can only be used by users who have a role on the application. To address this limitation, the researcher created a Facebook developer account and made used of Facepager, a system made for collecting publicly available data from Facebook and other JSON-based APIs.

25725	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Ikaw na lang sana tumakbo last election huhu sayang	2017-10-07T14:24:14+0000
25726	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Ang dami mong problema ang dami mong reklamo gi	2017-10-07T13:55:00+0000
25727	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Ano iyan sa kadahilanang hindi matanggap??? Ikaw r	2017-10-08T03:49:43+0000
25728	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Mark Labaton Damasco ang jejemon mo talaga kung	2017-10-08T00:13:38+0000
25729	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	pagkain lang ba habol mo dun ? haha. or walang nag	2017-10-08T08:04:55+0000
25730	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	hahaha. andami mong Alam Baka academics nga ba	2017-10-08T08:04:12+0000
25731	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	pasakit andami mong Alam! DI SANA IKAW NA ORG	2017-10-08T08:02:37+0000
25732	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Hello po makikiopinyon lang po.. Laht naman tayoo ,	2017-10-08T15:23:11+0000
25733	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	SHOUT OUT DIN SAYO SENDER! SANA IKAW NA L	2017-10-07T12:19:13+0000
25734	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Just Wow! Mas gigil kami sayo! Jusme. Kakulo ng du	2017-10-07T12:15:20+0000
25735	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Gusto mo lang walang klase eh?? Hahahaha.	2017-10-07T01:08:29+0000
25736	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	HAHAHA. GOD BLESS YOU MAN! WE LOVE YOU. I	2017-10-07T12:36:30+0000
25737	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Sa sender , SANA PO IKAW NAG ORGANIZE NG LA	2017-10-07T11:56:25+0000
25738	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	lakas pa ng loob mong imention . As in wow!????	2017-10-07T12:19:13+0000
25739	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Baka nagtipid ganun lng un.	2017-10-07T04:46:47+0000
25740	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Wala kayong sports club girl ?	2017-10-07T03:54:36+0000
25741	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	PAWER SI BESYEE????	2017-10-07T12:55:59+0000
25742	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Damn savage??	2017-10-08T10:27:34+0000
25743	17759	2	52813535420030	data	fetched (200)	48:54.8	Facebook:<post>/comments	Parang tanga ung ngpost neto ??	2017-10-07T15:48:21+0000

Figure 4: Sample Queried Data with 13076 Observations

The collected .csv data contains the following fields: id, parent_id, level, object ID, object_type, query_status, query_time, uery_type, message, created_time, updated_time and error.message. Under message field are posts and comments in English and Filipino language, and colloquial terms. For the machine learning algorithm to better understand the data, the researcher used translate my sheet add-on [51] to translate the queried messages and posts to English. The translated data formed the corpus for classification. However, a limitation of this process is that, there are colloquial words that cannot be translated to English and therefore remained as it is in the system.

	A
635	"Throwback is just a big winner, and it's a good day to come back and I'll be back." Stop, Look, and Listen -Libnos Stop, when it's time to come up with a good start, it's a rigat ken saem. , it's a good idea to keep it in t
636	
637	"Wrong Timing" First time I saw you, I feel sorry for someone else. Second year I, Third Year you. I do not know much like you, but I love you more. After many months seems like I lost my look to you, then suddenly
638	I'm here in person. Kunti nalang kami pre 😂😂😂😂👍👍👍
639	You just made it! 👏👏 Youll just be her forever friend. ☺☺ atleast there you can prove there forever hahahahaha
640	the pain Amor Luis Cacal III, I feel like I'm here
641	Charles Estabillo yan kase haist
642	Ouch. May forever: friendship.
643	Awww bilib nk talaga knka .. 😉 Buruk
644	
645	"Fixed Marriage is REAL :(- - - I just want to ask for advice Please take this seriously Close the family of our two Oh btw, he's a half pinoy, half british .. So, it all started nung year 2011. First year I'm here for a couple
646	For goodness sakes. You still do not know the Briton e. But that does not mean it's a reason for you to stop that marriage. Anyway, you're tied up e. If so, you can not do it anymore. But I just advice you, try to give hi
647	Just like in wattpad lng .. Should have happy ending ,,, jejeje
648	I know that feeling. You're just one of those who do not really love you. Just talk to brit boy that you do not love her. If you really love him. He will let you go. He will also talk to your parents. Say what you want to be
649	more difficult to be single tea..ahaha de joke lng .. think that you are good .. give yourself a chance..try you to love in a way you hnd think that nka arraged-marriage you .. if hnd tlaga, then talk to him .. he's kinda ov
650	You may just want me to be Mr. British? Hahahahaha TULUNGAN MO AKO !! n Marc Jeramy Rosete Abuluyan
651	Am'ammok pay 😂
652	Fight what is right. ☺ Tell your parents about your feelings. for sure maiintindihan ka po nila.
653	You can at least try, give the both of you a chance. If not, tell her. He'll understand maybe one way or another, and then you can ask his help to call it off. You 're parents will not hold it against you because you've trie
654	Bigte 😂😂 haha dejoke kaya mo yan: *
655	ate grab n yan ... kung ako yan go lng ng go da more chances of winning .. may lahi pa .... foreigner is 🤍🤍

Figure 5: Sample csv data of translated posts and comments

After translation, the corpus undergoes data cleansing. The texts were transformed to lowercase. Numbers, punctuation marks, stop words and white spaces were also removed. Stop words are words that have the same likelihood occurring in the documents that are not relevant to the query [52].

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 7, Issue 2, pp: (43-64), Month: April - June 2019, Available at: www.researchpublish.com

```
dvantage man can read true a life can either accepted changed accepted must changed cha
nged must accepted â best way friend one welcome freshmen transferees colleagues school
 opening hope enjoyed vacations hope enjoy stay dmmmsu epecially new comers â learn sch
ool lifeâ seneca roman philosopher mid st century ad difference school life school re t
aught lesson given test life re given test teaches lesson tom bodett civil engineering
po ba kayo mechanical electrical po ah salamat po one door happiness closes another ope
ns often look long closed door see one opened us helen keller duration enrollment dmsu
say hats graduates class congratulations graduates welcome real world cramming graduati
ng students patience key finish requirements good lucks guys sooner later discover impo
rtant moments life advertised ones birthdays graduations weddings great goals achieved
real milestones less prepossessing come door memory unannounced stray dogs amble sniff
around bit simply never leave lives measured say feel mind don matter matter don mind d
r seuss haha really like times just put period something end just settle comma whay cau
se time will come ll realize nice see complete sentence rather seeing phrase completely
 hanging doesn even make sense christino patacsil bsme â remember oâ man dust dust shal
l returnâ genesis ash wednesday also known dies cinerum means day ashes marks start len
t continues days easter ash wednesday said day repentance people dusted ashes expressed
 sorrow regarding sins committed sometimes need experience everything learn easy way li
ving life live cry laugh crazy don miss chances life giving important things things tim
e people making life worth living charmielyn april velasco bsm graduate mluc ahahahaha
dd http shine yahoo com event momentsofmotherhood laughing baby loves ripping paper peo
ple don know satisfied without knowing sometimes wanting means losing everything messag
e christino patacsil bsee mluc exactly ouchhh true sometimes cramming last minute prepa
rations exam don just miss days bet everyone looking forward summer vacation excited su
mmer classes ve decided attend thinking expenses life lots people watching next move de
sperately waiting us fail even praying us give don mind just doubles fun living survivi
ng let keep mind failures problems things scared lead breakthroughs great discoveries e
njoy life â congratulations nursing graduates passed board exam tentative dates graduat
ion march prefer centralized decentralized graduation half trouble life can traced sayi
ng yes quickly saying soon enough yes easily said don want disappoint people yes answer
s actually just need courage saying big sounding let pride drive behavior will always i
ncrease chance losing trying keep make sense pride kills many people drives insecurity
later without knowing creating false noise malign people reputation just accept one goo
d smart close people around power mean control people mind life intelligence power mean
s helping people grow inspiring song want dedicate graduates class next line http www y
outube com watch zkcyzzq years old mom always told happiness key life went school asked
 wanted grow wrote happy told didn understand assignment told didn understand life stil
```

Figure 6: Sample of Data that underwent Data cleaning

From the cleaned data, sentiment analysis is performed to determine the positivity and negativity of the posts and comments as shown in the next figure. The opinion lexicon used in the study for positive and negative sentiments are based on the study of [53].

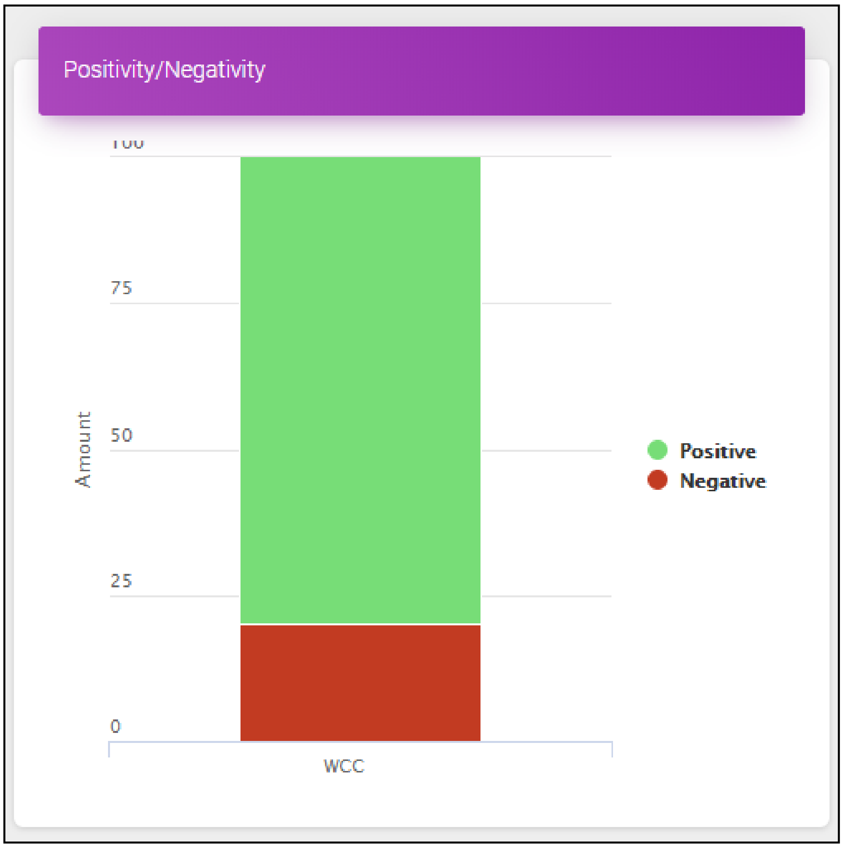

Figure 7: Visualization on the Polarity of Posts and Comments

The next step in the workflow is generating a word cloud which shows the most frequent words that are posted and commented in the respective HEI Facebook pages. The next figure shows a word cloud of one HEI facebook page.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 7, Issue 2, pp: (43-64), Month: April - June 2019, Available at: www.researchpublish.com

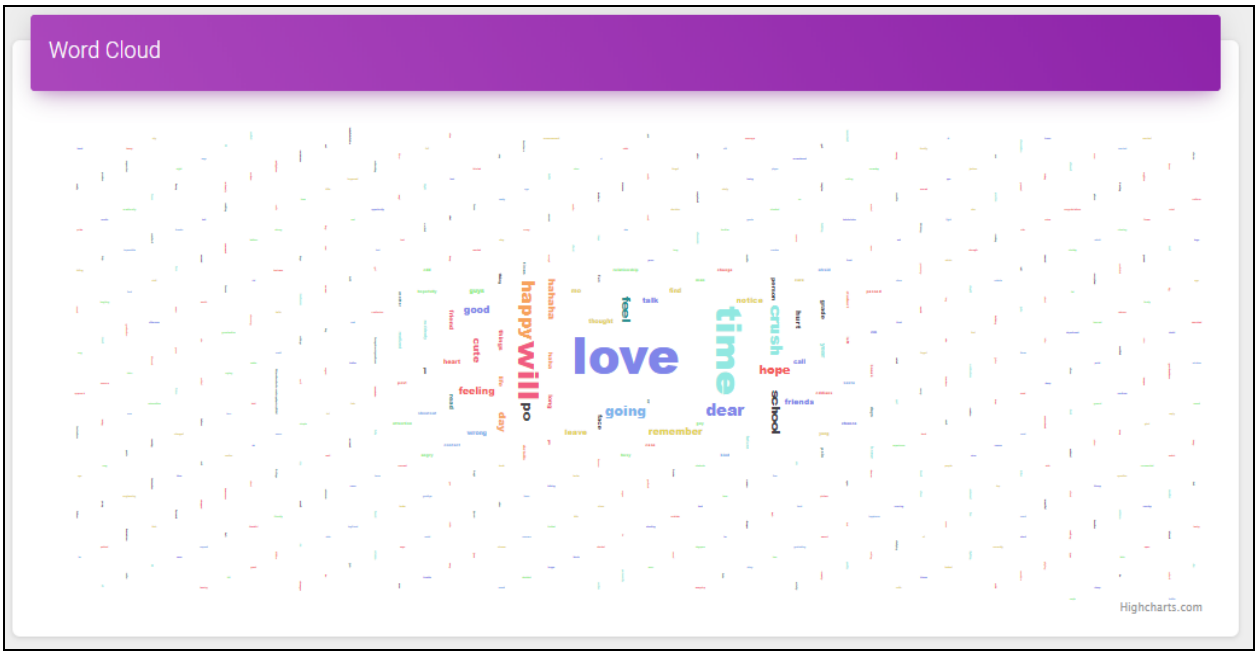

Figure 8: Generated word cloud for an HEI

The final step is classifying the words using the Naïve Bayes Algorithm. The classification was based on the study of [9]. For this study, the classifications are on academics, social engagement, emotions, finances, policies, and health. The words associated to these classifications were set as the training data sets. The Test data set is the cleaned corpus. A visualization report reflects to which classification the data falls into as shown in Figure 9. Since the corpus could fall into more than one category, multi-label classification is used.

Based on the development of the model, it could be inferred that the Naïve Bayes Classifier is highly scalable. It requires a number of parameters linear in the number of features in a learning problem.

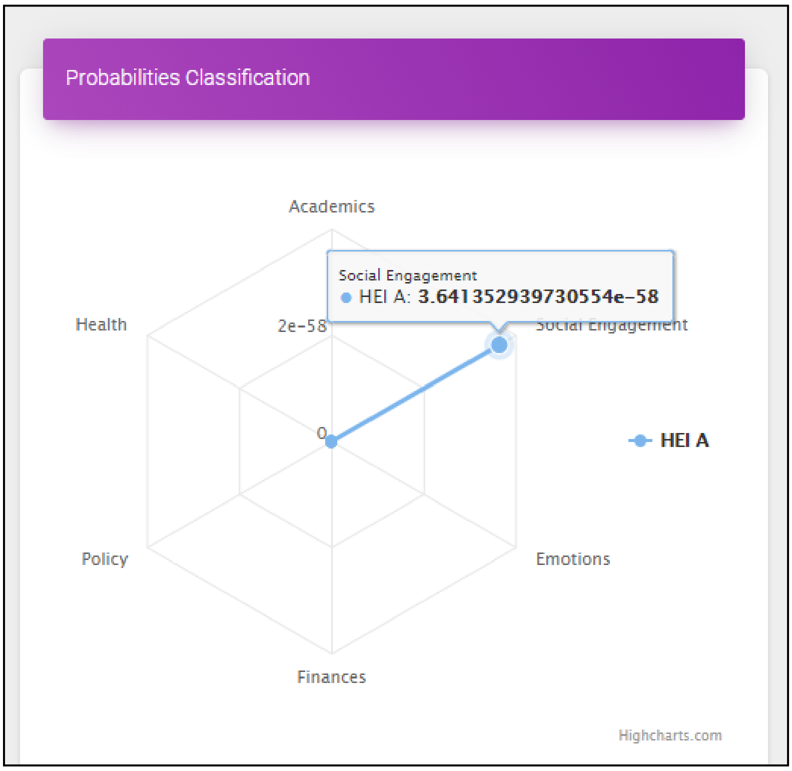

Figure 9: Visualization of Classification Probability

The first equation is determining the priors or the probability of a class which is the number of documents with that class over the total number of documents.

$$P(c) = \frac{Nc}{N}$$

Next is determining the likelihood of a word given a class by dividing the count of words in the class plus 1 smoothing divided by the count of all the words in the class plus the number of unique words.

$$P(w|\ c) = \frac{count\ (w, c) + 1}{count\ (c) + |V|}$$

In predicting the class where the test data falls into, we get the priors multiplied by all of the features or conditional probabilities of a specific class.

$$P(c \mid d) \propto P(c) * P(w|c) * (P\ (w|c)\ )n$$

Although we could already determine the probability of the words when the naïve bayes algorithm is applied, we still should understand deeper as to why the data falls in a specific classification by applying qualitative analysis like in the study of [9]. Further discussions on these are in the next section of this paper.

Behaviours of SM Users

Using the Naïve Bayes Algorithm, the data are classified according to which the posts or comments are related to academics, social engagement, emotions, finances, policies and health. Table 5 shows the corresponding value of each classification with respect to an HEI. Most of the posts and comments are positive in nature with a probability of 0.02827 while negative posts and comments averaged 0.01123 probability.

Social Engagement.

Out of the sample population, there were 65.07% or 41 HEIs with unofficial Facebook pages where the data were collected. Upon applying the multinomial Naïve Bayes algorithm, we could infer that the result may fall into one or more classifications. However, we could identify that most posts and comments are related to social engagement with an average probability of 4.30E-01. Most of the SM users are students whose intention in using social networking sites like Facebook is for social relationships [54]. This finding supports the study of [55] that states that social engagement can provide support for releasing stress and is helpful in learning.

The following are statements related to social engagement and are extracted from different HEIs:

a. "When we can not get what we love, we must love what is within our reach"

b. "Gifts of time and love are surely the basic ingredients of a truly merry Christmas"

c. "… I just realized that no matter how imperfect this world to live on. I am still lucky … Because we struggle through this battle …"

d. "College might not be what we imagine. Sometimes you have to let people go. There is a time for everything. And being challenged to grow is a good thing"

e. "Do you know which is the best part of life? It's when your family understands you as a friend and your friends supports you as family"

f. "Surround yourself with good people, and good things will happen… sure good things will happen to you bestie! Because of cool teachers and super fun Student Life…"

Students use SM as an outlet to release their stress from school. They use SM to provide them better opportunities to interact with their peers and build positive relationship with them.

Having a positive relationship with their peers can be important to create learning communities where they can work collaboratively, share opinions and respond constructively to the ideas of their fellow students. These are a necessary requirement to develop critical thinking that supports learning to occur [56]. In addition, these types of interaction in social media is essential on how they learn as humans [57].

Moreover, these students are mostly in their adolescent stage where group interactions, [58]and social relationships are particularly important [59].

Furthermore, reports revealed that using online technology like social media increases self-esteem, social support, social capital, safe identity experimentation and opportunity for self-disclosure [60] .

The next figure shows a word cloud related to social engagement in one HEI.

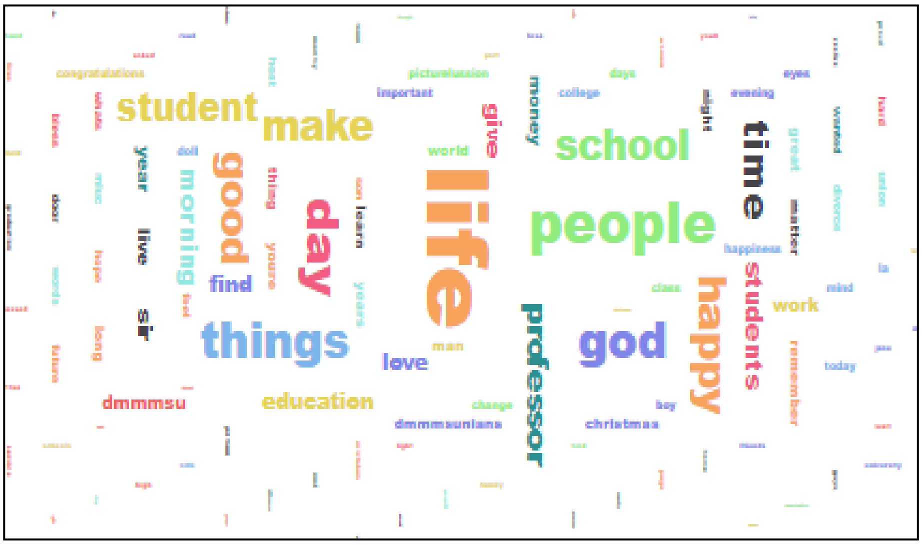

Figure 10: Word cloud related on Social Engagement

These results also prove the developmental theory asserting that adolescents display cross-situational continuity in their social behaviors and recommend that the continuity of this conceptualization may be stretched into the online domain [61].

This however is in contrast to public anxiety as mentioned by [62] that social media distracts from education and reduces the social skills of young people-despite an exemplary body of prior research that reject such simple conclusions. As asserted by [63]possible threats associated with SM are too boundless. Student engaging in a private relationship online solicits for the beginning of inappropriate SM behavior.

Academics

Posts and comments which fall under Academics classification have an average prediction score of 3.76E-01. Social Media like Facebook allow people identify other SM users with whom they have connection; they read, react and comment to postings; and send and receive private or public messages. In this study, these SM users are mostly students, teachers, and other stakeholders of HEIs. They connect with other SM users via social media as a platform of discussions for their homework, schedules, announcements and other information related to their course work. These have supports from [67] stating that SM provides students a direct medium to publicly evaluate and comment their campus environment, policies of their school, classes, professors, school administrators and fellow students.

Students these days use social media to communicate with other students more so than to do their coursework [68]. However, they also post their academic concerns in social media because this type of platform poses liberal environment for them to discuss their views and opinions on issues that otherwise would not have been possible in a face-to-face classroom [69]. Below are statements related to academics as posted by SM users:

a. “Ok..kmi in byad..pro grade wlapa ngayon.january up until now ... more pgbayad alert kyo ....”

b. “Time comes, hope hnd ka nlng nag effort or never give your best because at the end of the sem di mo deserve ung grades they will give… Kasumpa sumpa??????”

c. “Remove teachers who are earning yet they are not teaching…”

d. “Proud to be one of the million teachers that will provide the students for the SY 2012-2013 with adequate knowledge and skills”

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 7, Issue 2, pp: (43-64), Month: April - June 2019, Available at: www.researchpublish.com

HEI	OBS	ACA	SEN	EMO	FIN	POL	HEA
A	951	1.09E-65	9.32E-56	7.11E-61	2.12E-68	3.23E-73	6.62E-70
B	36	1.06E-09	1.66E-11	8.12E-15	6.50E-14	6.50E-14	1.27E-16
C	317	1.68E-39	4.10E-43	3.82E-52	1.96E-49	3.91E-49	2.39E-53
D	21	0.000017	0.000035	2.73E-07	8.52E-09	3.41E-08	8.52E-09
E	84	0.0000087	0.0000087	5.45E-07	6.81E-08	6.81E-08	1.70E-08
F	17	0.1429	0.0011	0.00028	0.00028	0.0022	0.00028
G	572	2.12E-68	7.28E-58	4.45E-62	2.07E-71	1.62E-73	1.01E-74
H	19	0.00028	0.00007	0.0000088	0.00028	0.00007	0.000017
I	9	0.0179	0.00056	0.00003	0.00003	0.0006	0.00007
J	37	4.26E-09	1.33E-10	4.06E-15	3.25E-14	6.50E-14	2.03E-15
K	808	7.71E-80	3.23E-73	1.97E-77	4.71E-84	9.19E-87	1.47E-85
L	53	1.33E-10	1.70E-08	1.06E-09	2.08E-12	5.20E-13	2.60E-13
M	16	3.10E-20	3.17E-17	3.10E-20	1.94E-21	3.03E-23	6.05E-23
N	46	4.06E-15	3.33E-11	5.08E-16	3.97E-18	4.96E-19	1.98e1-18
O	8	0.0003	0.000035	0.000017	0.000017	0.0000087	0.0000011
P	112	7.39E-27	1.51E-23	5.77E-29	7.04E-33	1.13E-31	4.51E-31
Q	132	1.15E-28	3.78E-24	3.69E-27	5.63E-32	2.82E-32	7.04E-33
R	948	3.86E-80	2.71E-66	1.65E-70	2.94E-85	4.70E-84	1.21E-81
S	219	1.44E-29	3.87E-21	2.95E-26	1.80E-30	2.82E-32	4.51E-31
T	190	1.59E-17	1.62E-14	2.54E-16	1.24E-19	6.20E-20	9.91E-19
U	2	0.1429	0.2857	0.1429	0.1429	0.1429	0.1429
V	219	8.60E-37	5.63E-32	4.30E-37	4.20E-40	3.28E-42	1.31E-41
W	160	6.05E-23	6.34E-17	9.68E-22	2.95E-26	3.69E-27	2.36E-25
X	259	1.95E-49	1.31E-41	4.00E-46	4.77E-53	5.97E-54	7.46E-55
Y	812	2.84E-60	3.06E-51	1.86E-55	5.56E-63	1.73E-64	1.08E-65
Z	110	1.44E-29	9.68E-22	3.69E-27	9.02E-31	3.52E-33	2.25E-31
AA	229	1.60E-45	1.68E-39	2.56E-44	1.95E-49	7.63E-52	6.11E-51
BB	186	2.75E-35	4.51E-31	5.64E-32	1.31E-41	4.20E-40	6.72E-39
CC	327	1.86E-55	4.20E-40	3.13E-48	1.17E-56	4.55E-59	2.33E-56
DD	29	8.52E-09	0.0000087	6.81E-08	2.66E-10	3.33E-11	2.13E-09
EE	336	2.62E-41	5.63E-32	6.72E-39	1.60E-45	1.28E-44	3.20E-45
FF	29	7.74E-21	3.17E-17	1.93E-21	7.56E-24	4.73E-25	1.52E-23
GG	321	1.80E-30	9.23E-28	5.50E-35	4.30E-37	8.60E-37	4.20E-40
HH	18	3.78E-24	3.10E-20	9.68E-22	2.36E-25	1.48E-26	7.39E-27
II	125	9.77E-50	4.10E-43	3.13E-48	4.66E-56	2.91E-57	7.46E-55
JJ	437	5.13E-44	4.40E-34	4.20E-40	1.25E-47	3.13E-48	1.28E-44
KK	52	6.34E-17	1.30E-13	1.59E-17	6.20E-20	9.68E-22	2.48E-19
LL	918	8.60E-37	1.41E-32	1.05E-40	4.10E-43	2.56E-44	1.64E-42
MM	279	1.66E-70	7.28E-58	9.10E-59	1.62E-73	4.04E-74	8.28E-71
NN	13	7.56E-24	2.48E-19	6.20E-20	5.91E-26	2.36E-25	9.45E-25
OO	3	0.0714	0.1429	0.1429	0.0357	0.0357	0.0357
Total/Ave	9459	3.76E-01	4.30E-01	6.98E-03	4.37E-03	4.43E-03	4.47E-03

Table 5: BEHAVIORAL PROBABILITY OF SM USERS OF HEIs

Legend: HEI – Higher Education Institution

OBS – observations (Number of post and comments regardless of the words)

ACA – Academics

SEN – Social Engagement

EMO – Emotions

FIN – Finance

POL – Policies

HEA – Health

POS – Positive

NEG – Negative

e. “the life of the dorm, the life of the forest he! He’s the teachers / professors that they are really supportive to that’s students, and the best thing especially in bacnotan campus without child pobre or children rich in the same”

f. “#Prof Why oh why? Ganern why? Yes somewhat no spoon-feeding the college but to the point that you do not understand in class? Do not understand why. Kakawa naman po kami. Huhuhuhu. Okay we would have nagegets those examples but too vague eh then concentrated on a book reference. Nakakaloka -asdfghjkl 20 **”

Below is a sample wordcloud related to academics.

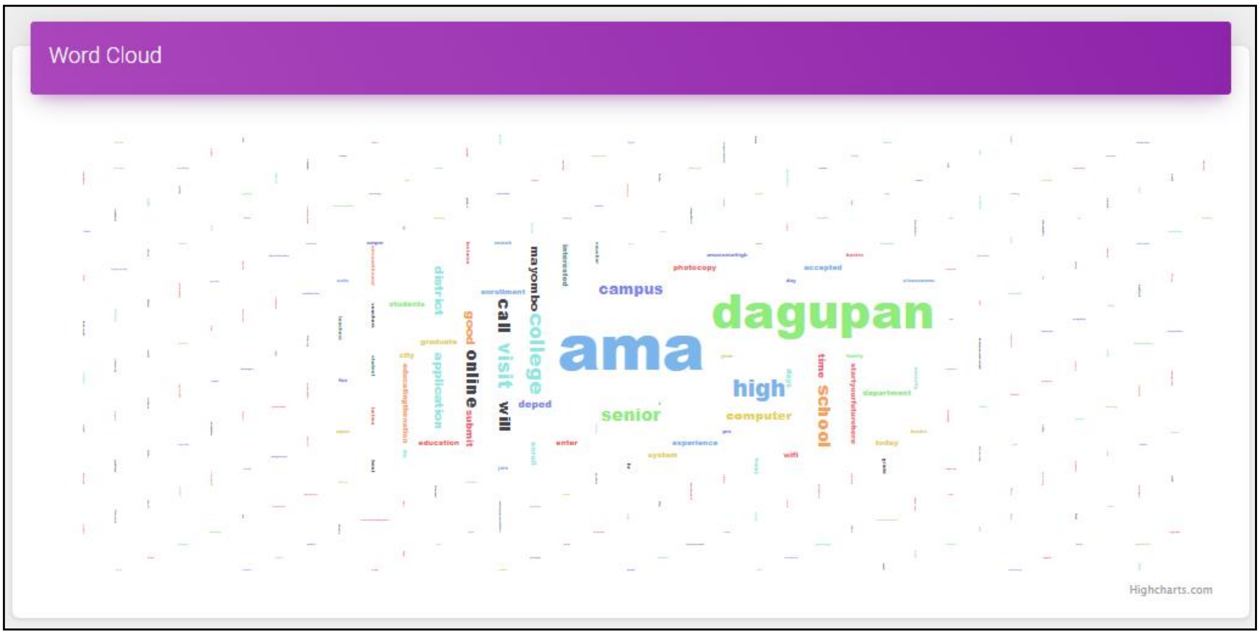

Figure 11: Word cloud related on academics

Emotions.

The result of the average prediction under ‘Emotions’ classification is 6.98E-03. Social media is a platform for users to post their feelings and emotions. Student post positive feelings like showing their affection to someone. With this, feedback on their profiles enhance their self-esteem and well being [64]. They feel overjoyed when they get some reactions like ‘likes’ and ‘comments’. As mentioned by [62]SM emphasizes more temporary pleasures creating stress over public appearance.

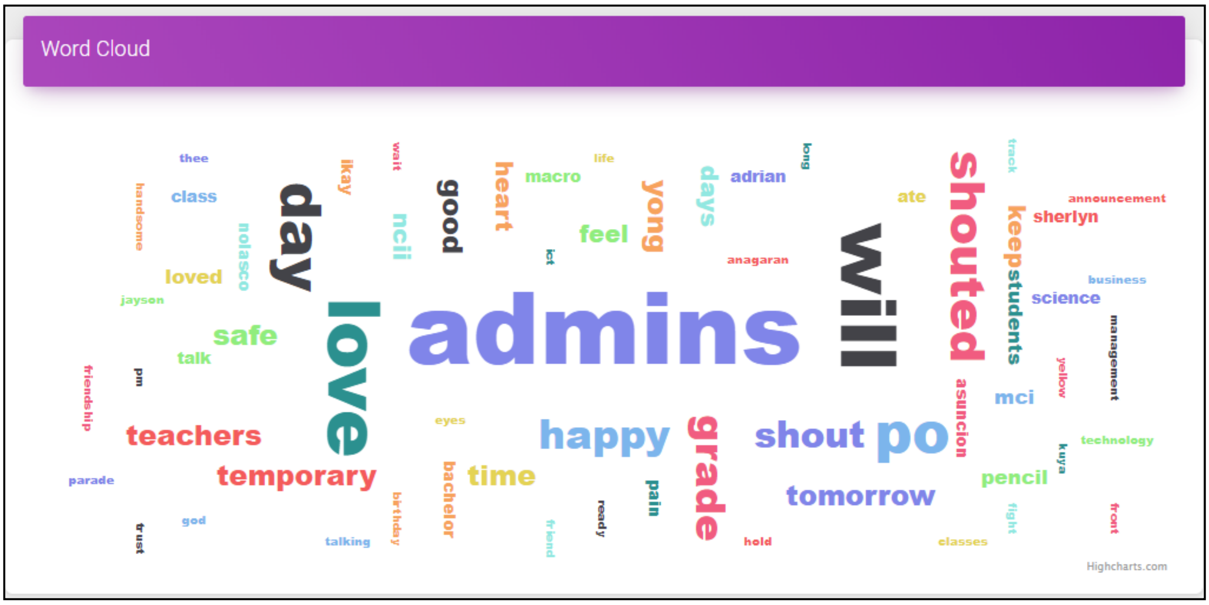

Figure 12: Word Cloud related on emotions

However, sometimes, emotional strain is caused by social relationships. Emotional exhaustion is seen as an affective reaction and emotion. It describes feelings of being emotionally overextended due to the usage of social media like Facebook [65]

Some of the posts and comments related to 'Emotions' classification are the following:

a. "You know that I went f***boy eh bat pregnant ka…"

b. "My unforgettable experience in ***** was I got incomplete grade… I got mad to madam… she made me cry bulls****"

c. "You broke my heart, but I still love you with all the pieces…"

d. "I feel you po. Nangayri din saakin yan. When I was grade 10 first grading palang yun aa. 73 ba naman yung grades ko don sobrang iyak non. Hate na hate kasi ang math…"

e. "oo masaya ako ng malaman kung gina**ago ako. Oo masaya ako sa katalksh*tan mo! Oo masaya ako sa paggawa ng sariling katangahan ko!' Oo masaya ako kahit alam kong wala na akong Halaga sayo…"

f. "Admin officer no consideration to her students!!!!"

While others find it difficult to express their feelings in words, many are vocal about it. These days, most people especially the young ones are very impulsive and have taken to social media sites like Facebook to vent out their frustrations. As mentioned by [66] emotional outburst are more in number as compared to posts containing happy moments. In fact, 65% of people immediately post online after being hurt or angered. They post on social media because it is simple and quick, it has a wider reach, it ensures public support, and the response is flattering. These impulsive posts are obviously not backed with reason and thought.

Health

The predicted average score under Health classification is 4.47E-03. As cited by [70], students reported that stress and sleep difficulties negatively impacted their academic performance. An example of which are the following posts:

a. "Sleep late tomorrow tonight sleep last night sleep every night"

b. I really missed those times! Sleep and lunch at Bucasa's eatery!

Moreover, many college students engage themselves in unhealthy behaviors, making them at risk for developing serious problems in health in the future.

Students post their concerns on their health and how to live life. As revealed by [71] ,college students show symptoms constant with depression on social media sites like Facebook. Given the frequency of depression symptoms, SM sites could be a platform to combat stigma surrounding metal health conditions or for identifying risks of students towards depression. Some statements concerning health of SM users are as follows:

a. "Shoutout for myself. Please, myself kayanin mo lahat ng ibinabato sayong problema. Okay? Keep your mindset healthy. Wag mong tatangkaing kunin ang patalim o ang cutter"

b. "If you think that you pinakamalas person. Give up. But if you open my eyes and think how the homeless. No food at the table. No more tuition money. There is no provision that spending on luxury and needs him / them … Sleep and rest and tomorrow get up and smile because you woke up yet to feel a mixture of sadness and joy, grace and problems…"

c. "Maging healthy lahat kaming pamilya. At biyayaan pa ng madaming blessing. Yung iba landi ang inuuna eh. Tss ???? -Mushroom ??"

Policies

'Policies' classification has an average prediction of 4.4eE-03. Attendance, discipline, and grades are among the policies that are implemented in schools. Such policies and practices affect student's experiences. Instructional decisions for example affect students' feelings about the curriculum [74].

Below are some statements related to policies:

a. “NO CLEARANCE, NO PERMIT, NO EXAM!! 1st Trimester is about to end.. whooa!!”

b. “Attention ! Scc students, tomorrow will be the FIRST FORMAL EXAMINATION. September 04, 2018. Please settle your accounts. No haircut, no exam. (For boys) Wear your COMPLETE UNIFORM WITH SCHOOL ID”

c. “Whatever you do do, they know how right or wrong. Parekla can not complain especially if you’re just Intern”

Finances

4.37E-03 is the average prediction score under “Finances” classification. SM users post their financial related concerns. As mention by the study of [72], likelihoods of self-worth distinctively contribute to academic and financial problems experienced by freshmen college students beyond level of self-esteem and other personality variables which uniquely contribute to later social difficulties. Moreover, Bigger financial problems may lead students to reduce coursework or drop out of school [73]. Below are some statements related to finances as posted by SM users.

a. “why are the tuition fees so expensive?”

b. “No tuition and fee-grad…AprilFools”

c. “Rice or Tuition”

d. “Kano is the tuition to freshman education student ..... full load?”

e. “What about fees for books and uniform. I have a niece looking for a good school.”

PROPOSED PROACTIVE AND KNOWLEDGE-DRIVEN DECISION INTERVENTION MEASURES

Although the use of the internet specifically social media is to exercise freedom of speech, HEIs should make ways on how its stakeholders share information online. Whatever a student, teacher, staff or any personnel related to an HEI does inside and outside the university, offline or online, he or she still carries the name of the institution s/he is related to. Thus, the privacy of information is uncontrolled. It is in this context, that Institutions should be proactive. With the results of this study, the following intervention measures of HEIs are suggested:

a. A framework on digital strategy for higher education can be researched on and further used where all the departments of an institution are related and integrated and move towards online presence

b. Rules and Regulations of the University should be amended to Integrate expected student behavior online most especially in social media

c. Unofficial pages should be reported to Facebook. HEIs should strengthen their social media presence to highlight their branding and secure their information.

d. HEIs should integrate socio-relational activities in the curriculum across programs. They should educate students to professionally perform their role in society.

e. Teachers should use teaching-learning strategies to improve social relationships

f. Enhance personality development through the new General Education courses: Understanding the Self and Ethics under CMO no. 20 s.2013.

g. Student services should be strengthened by creating programs that enriches socialization

h. HEIs should widen their scope in monitoring student’s performance: academically, physically, socially, emotionally and spiritually.

5. CONCLUSIONS AND RECOMMENDATIONS

Conclusions

From the findings of the study are the following conclusions:

1. The developed workflow could be used by HEIs in identifying their stakeholders’ behavior in social media. Furthermore, qualitative research methodology should be employed for deeper understanding of the SM data.

2. Social engagement is the main SM data concerning HEIs in Region 1 followed by Emotions, Academics, Health, Finances and Policies. Moreover, it could be inferred that SM behaviors differ from one HEI to another.

3. Proactive and Intervention Measures could be used by HEIs to holistically develop their stakeholders especially their students.

Recommendations

With the findings and conclusions stated, the following are further recommended:

1. The system should be submitted to Facebook for review.

2. Factors affecting student SM behaviors could further be researched. In addition, deeper qualitative studies should be conducted like effects of social media on social skills among young people.

3. Other data mining algorithm specifically text classification may be applied and compared.

4. The training models may be enhanced.

ACKNOWLEDGEMENTS

The researcher is extremely grateful to the following people who in one way or another helped in the realization of this study.

The administrators of PUNP Tayug campus namely: Dr. Donna P. Taguiba, the University President; Dr. Amapola C. Padilla, the Academics Vice President; and Mr. Edward C. Taguiba, the Operations Vice President; for their trust and financial support.

The faculty of the College of Computer Studies and Professional Schools of Angeles University Foundation especially Dr. Gilbert Tumibay, Dr. Rosanna Esquivel, Dr. James Esquivel, Dr. Ruel Reyes, Dr. Menchita Dumlao, and Dr. Shaneth Ambat for their expertise in helping the improvement of this study.

The family, friends and colleagues of the researcher namely: Analyn, Yheojhann Xavier, Yhanna Xia, Vilma, Benjamin, Jaypee, Jerny, Jester, Efren, Dr. Marlyn Quiming, Dr. Jundy Raga, Mylene, Noel, Gab, Ariel, and Lei who have given courage, motivation, moral support and help to the researcher.

And above all, to the Almighty for giving, the researcher, strength and faith to finish this study.

REFERENCES

[1] S. Paladhi, "Ed Tech Review," 8 July 2017. [Online]. Available: http://edtechreview.in/news/2849-factors-21st-century-education-technology.

[2] K. Smith, "123 Amazing Social Media Statistics and Facts," 1 March 2019. [Online]. Available: https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/.

[3] B. Liu, "Sentiment Analysis and Opinion Mining," AAAI, 2011.

[4] L. Bexheti, B. Ismaili and B. Cico, "An Analysis of Social Media Usage in Teaching and Learning," Recent Advances in Electrical and Computer Engineering. Research Gate, May 2017.

[5] D. Zarella, "The Social Media Marketing Book," Canada, O'Reilly Media Inc., 2010.

[6] C. Redecker, "Review of Learning 2.0 Practices: Study on the impact of Web 2.0 Innovations on Education and Training in Europe," JRC Scientific and Technical Report, 2009.

[7] K. Seat, "Surprising Statistics about Educators' Social Media Use," Direct Network, 2 August 2016.

[8] R. Griffin, "Social Media is Changing How College Students Deal with Mental Health, For Better or Worse," www.huffingtonpost.com, 21 July 2015.

[9] X. Chen, M. Vorvoreanu and K. Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences," IEEE Transactions on Learning Technologies, pp. 246-259, 2014.

[10] P. Pagare, "Analyzing Social Media for Understanding Student's Problem," International Journal of Computer Applications. Innovations and Trends in Computer and Communication Engineering, 2014.

[11] J. Estrada, "Balancing Students' freedom of speech and expression in social media and the school's right to discipline," The Manila Tmes, January 2018.

[12] K. Alaskar and P. Tandale, "Data Mining Applications in Higher Education," National Conference on Emerging Trends: Innovations and Challenges in IT, 2013.

[13] G. Kashyap and E. Chauhan, "Review on Educational Data Mining Techniques," 2nd International Conference on Recent Innovations in Science, Engineering and Management, 2015.

[14] J. Srivastava and A. Srivastava, "Data Mining in Education Sector: A Review," National Conference on Cloud Computing and Big Data, n.d..

[15] M. Ali, "Role of Data Mining in Education Sector," International Journal of Computer Science and Mobile Computing, 2013.

[16] M. Injadat, F. Salo and A. Nasif, "Data Mining Techniques in Social Media: A Survey," Neurocomputing http://dx.doi.org/10.1016/j.neucom.2016.06.045, 14 June 2016.

[17] M. Olowe, M. Gaber and F. Stahl, "A Survey of Data Mining Techniques for Social Media Analysis," School of Computing Science and Digital Media, n.d..

[18] H. Lauw, J. Shafer, R. Agrawal and N. A., "Homophily in the digital world: A live journal case study," Internet Computing, IEEE,14(2):15 –23, 2010.

[19] G. Barbier, "Data Mining in Social Media," Social Network Data Analytics, 2011.

[20] B. Pang and L. Lee, "Using Very Simple Statistics for Review Search: An Exlporation," in International Conference on Computational Linguistics, 2008.

[21] C. Aggarwal, "An Introduction to Social Network Data Analytics," Springer U.S., 2011.

[22] M. Olowe, M. Gaber and F. Stahl, "A Survey of Data Mining Techniques for Social Media Analysis," School of Computing Science and Digital Media. Robert Gordon University. School of Systems Engineering. University of Reading, n.d..

[23] J. Cortizo, F. Carrero, J. Gomez, B. Monsalve and E. Puertas, "Introduction to Mining SM," 1st International Workshop on Mining Social Media, 2009.

[24] D. Kumar and D. Bhardwaj, "Rise of Data Mining: Current and Future Application Areas," in IJCSI International Journal of Computer Science Issues, Vol. 8, Issue No.5, 2011.

[25] U. Fayadd, G. Piatesky-Shapiro and P. Smyth, From Data Mining to Knowledge Discovery Databases, AAAI Press. The MIT Press, Massachusetts Institute Of Technology. ISBN 0–26256097–6, 1996.

[26] M. Chen, J. Han and P. Yu, "Data Mining: An Overview from Database Perspective," IEEE Transactions on Knowledge and Data Eng, 1999.

[27] A. Paidi, "Major Research Challenges in Data Mining," International Journal of Trend in Research and Development, 2015.

[28] J. Han and M. Kamber, "Data Mining: Concepts and Techniques. Morgan-Kaufman Series of Data Management Systems," San Diego: Academic Press, 2001.

[29] M. Vedanayaki, "A Study of Data Mining and Social Network Analysis," Indian Journal of Science and Technology Vol. 7 , November 2014.

[30] S. Ghosh, S. Roy and S. Bandyopadhyay, "A Tutorial Review on Text Miing Algorithms," International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 4, 2012.

[31] J. Han and J. Gao, "Research Challenges for Data Mining in Science and Engineering".

[32] D. Noyes, "The top 20 Valuable Facebook Statistics," March 2019. [Online]. Available: https://zephoria.com/top-15-valuable-facebook-statistics/amp/.

[33] Navin, "This is exactly how much your personal user data is worth to Facebook," 2019. [Online]. Available: https://techgrabyte.com/personal-user-data-worth-facebook/.

[34] J. Han and M. Kamber, "Data Mining: Concept and Techniques," 2nd Edition. Morgan Kauffmann, 2006.

[35] B. Liu, "Sentiment Analysis and Opinion Mining," Morgan & Claypool Publishers, 2012.

[36] C. Virvou, J. Espinosa, K. Llaguno and J. Caro, "Sentiment Analysis of Facebook Statuses using Naive Bayes Classifier for Language Learning," n.d..

[37] D. Yadav, "Sentiment Analysis Based on Data Mining and Natural Language Processing," International Journal of Advanced Research in Computer Engineering and Technology (IJARCET), November 2015.

[38] P. Chesley, B. Vincent, L. Xu and R. Srihari, "Using verbs and adjectives to automatically classify blog sentiment," Proc. AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs, Stanford University, CA. AAAI Press, TR SS-06-03, pp.27-29, March 2006.

[39] D. Jurafsky, "Sentiment Analysis," Stnaford University. Natural Language Processing, n.d..

[40] A. Go, R. Bhayani and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," Technical Report. Stanford, 2009.

[41] V. Tundjungsari, "Business Intelligence with Social Media and Data Mining to Support Customer Satisfaction in Telecommunication Industry," International Journal of Computer Science and Electronics Engineering. Vol. 1. Issue 1. ISSN 2320-4028, 2013.

[42] F. Patacsil, A. Malicdem and P. Fernandez, "Estimating Filipino ISPs Customer Satisfaction Using Sentiment Analysis," Computer Science and Information Technology 3(1): 8-13,2015. DOI: 10.13189/csit.2015.030102 , 2015.

[43] C. Shah and A. Jivani, "Comparison of data mining classification algorithms for breast cancer prediction," in International Conference on Computing, Communications and Networking Technologies, 2013.

[44] C. Troussas, M. Virvou, K. Espinosa, L. K. and J. Caro, "Sentiment Analysis of Facebook Statuses using Naive Bayes Classifier for Language Learning," Department of Informatics. University of Piraeus, Greece. Department of Computer Science, University of the Philippines. Philippines, n.d..

[45] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," Seventh International Conference on Language Resources and Evaluation, 2010.

[46] P. Tripathi, S. Vishwakarma and A. Lala, "Sentiment Analysis of English Tweet Using Rapidminer," International Conference on Computational Intelligence and Communication Networks, 2015.

[47] M. Ibrahim and N. Salim, "Sentiment Analysis of Arabic Tweets:With Special Reference Restaurant Tweets," IJCST, vol. 4, no. 3, June 2016.

[48] F. Laeeq and M. Nafiz, "Sentimental Classification of Social Media using Data Mining," International Journal of Advanced Research in Computer Science, June 2017.

[49] R. Sunil, "6 Easy Steps to Learn Naive Bayes Algorithm," September 2017. [Online]. Available: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/.

[50] "Naive Bayes' Classifier: How to Build a Sentiment Analysis Program," [Online]. Available: https://triton.ml/blog/sentiment-analysis.

[51] J. Bronner, "Translate my sheet".

[52] W. J. Wilbur, "The automatic identifiation of stop words," Journal of Information Science, vol. 18, no. 1, pp. 45-55, 1992.

[53] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2004.

[54] C. Cheung, P. Chiu and M. Lee, "Online Social Networks: Why do students use facebook," Computers in Human Behavior, 2011.

[55] S. Cohen and T. Wills, Stress, Social Support,and the Buffering Hypothesis, 1985.

[56] A. Chickering and Z. Gamson, " Seven principles for good practice in undergraduate," American Association for Higher Education & Accreditation Bulletin, 1987.

[57] A. Pardo, "Social learning graphs: combining social network graphs and analytics to represent learning experiences," in Int. J. Social Media and Interactive Learning Environments, 2013.

[58] R. Lerner and L. Steinberg, Handbook of Adolescent Psychology, New York: Wiley, 2004.

[59] S. Choudhury, S. Blakemore and T. Charman, "Social Cognitive during the adolescence," in Social Cognitive and Affective Neuroscience, 2006.

[60] P. Best, R. Manktelow and B. Taylor, "Online communication, social media and adolescent well-being : A systematic narrative review," Children and Youth Services Review, 2014.

[61] A. Mikami, D. Szwedo, J. Allen, M. Evans and A. Hare, "Adolescent peer realationships and behavior problems predict young adults' communication on social networking websites," Developmental Psychology, 2010.

[62] D. Miller, E. Costa, N. Haynes, T. McDonald, R. Nicolescu, J. Sinanan, J. Spyer, S. Venkatraman and X. Wang, How the World Changed Social Media, UCL Press, 2016.

[63] H. Gurcan, "Contribution of Social Media to the Student's Academic Development," International Journal of Information and Education Technology, 2015.

[64] R. Boateng and A. Amankwaa, "The Impact of Social Media on Student Academic Life in Higher Education," Global Journal of Human-Social Science: G Linguistics and Education, 2016.

[65] S. Hu and G. Kuh, "Computing experience and good practices in undergraduate education: Does the degree of campus wiredness matter?," Annual Meeting of the American Educational Research Association, Seattle, 2001.

[66] R. Boateng and A. Amankwaa, "The impact of social media on student academic life in higher education," Global journal of human-social science : G linguistics and education, 2016.

[67] P. Valkenburg, J. Peter and A. Schouten, "Friend Networking Sites and their Relationship to Adolescent's Well-being and Social Self-Esteem," CyberPsychology and Behavior, 2006.

[68] C. Maeir, S. Laumer, A. Eckhardt and T. Weitzel, "When Social Networking Turns to Social Overload: Explaining the stress, emotional exhaustion, and Quitting behavior from Social Network Sites' users," n.d..

[69] S. Shinde, "Why do people vent on social media," 2014. [Online]. Available: https://timesofindia.indiatimes.com/life-style/relationships/love-sex/Why-do-people-vent-on-social-media/articleshow/40579845.cms. [Accessed April 2019].

[70] D. Baldwin, K. Towler, M. Oliver II and S. Datta, "An examination of college student wellness: A research and liberal arts perspective," Health Psychology Open, 2017.

[71] M. Moreno, L. Jelenchick, K. Egan, E. Cox, H. Young, K. Gannon and T. Becker, "Feeling Bad on Facebook: depression disclousures by college students on a social networking site," Depression and Anxiety, 2011.

[72] C. Danielson, Enhancing Student Achievement, 2002.

[73] J. Crocker and R. Luhtanen, "Level of Self-Esteem and Contingencies of Self-Worth: Unique Effects on Academic, Social, and Financial Problems in College Students," Personal and Social Pyschology Bulletin, 2003.

[74] S. Joo, D. Durband and J. Grable, "The Academic Impact of Financial Stress in College Students," Journal of College Retention: Research, Theory & Practice, 2009.