JOURNAL OF MARITIME RESEARCH

Vol XXII. No. I (2025) pp 127–136

ISSN: 1697-4840, www.jmr.unican.es

Social Media Behavior Analysis of Marine Engineering Learners using Natural Language Processing and Time Series Analysis

Ken Gorro[1,*]

ARTICLE INFO

Article history:
Received 23 Jun 2024;
in revised from 27 Jun 2024;
accepted 24 Jul 2024.

Keywords:
Social Media, Marine Engineering, Natural Language Processing, Correlation, K-means clustering.

ABSTRACT

Social media undoubtedly has a profound impact on how students communicate and perceive the world. The Philippines is a maritime country and maritime seafarers can also affect the economy of the Philippines. In the Philippines education system, Maritime baccalaureate programs were developed such as marine engineering and marine transportation. Maritime programs are constantly evaluated to the global standard set by the international community. In this study, we value the importance of investigating maritime students' academic performance. This research explores the relationship between marine engineering students' academic standing and social media behavior by employing Bi-term topic modeling on social media texts collected from students of varying academic standings. For marine engineering students with high academic standing, the content of discussions often relates to academic and existential themes. In contrast, marine engineering students with lower academic standing tend to engage more with entertainment and politically oriented material. A significant finding from the correlation analysis is the negative association between the frequency of social media usage and academic performance in mathematics, suggesting that a high frequency of social media usage may adversely affect academic success. Additionally, K-means clustering was used to group Facebook posts, identifying patterns that align with key societal and educational events. These results highlight the complex influence of social media within educational settings, underscoring both its positive features and potential risks. Time series analysis was also performed to check for patterns in the number of students' Facebook posts.

1. Introduction.

This paper is a continuation of what was presented at the International Conference on Business Analytics for Technology and Security (ICBATS), last year (2023) (Gorro et al., 2023). Social networking has become rapidly widespread among teenagers as well as students, the social media has led to a complete overhaul of the way we interact and live within society. As per Meltwater, the figure of the total number of social media users in our country had reached the mark of around 76 million at the time of January 2022 (Meltwater, 2022). Social networks' online communication function and a well-knit connection across the student communities are the major benefits that the students can get out of it. Nevertheless, this convenience also has consequences like ill addiction that results in negative consequences including lack of sleep, eye fatigue, and physical inactivity. As one tends to be engaged with social media content for hours without even realizing that time has passed by, it disrupts sleep patterns and in turn creates additional health problems (Netsweeper, 2021) (Wells & Horwitz, 2021).

The Philippines is a maritime country and maritime seafarers can also affect the economy of the Philippines. Maritime baccalaureate programs were developed such as marine engineering and marine transportation. In this study, we value the importance of investigating maritime students' academic performance.

The topic of social media addiction and its side effects is not

[1]Cebu Technological University, Carmen Campus, Cebu, Philippines.
[*]Corresponding author: Ken D. Gorro. E-mail Address: ken.gorro@ctu.edu.ph.

just limited to adults; students, basically those who have lived with technology for a very long time, are the ones who could easily become addicted. Cyberbullying is also one of the significant cyber problems through which individuals are exposed to objectionable messages and demeaning statements that can negatively impact someone (Ranch, 2015) (stopcyberbullying, 2018).

Given the negative effects of social media use on students, this research intends to investigate the extent to which students' social media activities affect their academic achievement. Marine engineering students were categorized into two groups good academic and bad academic standing. The criteria for the categorization are listed below:

1. Good Academic Standing:

Major Subjects and Mathematics:

Passing Grades: Students must have grades between 1.0 and 3.0. This range indicates that the student is passing the subject.

Relative Performance: Students must be within the top 60% of their cohort based on their GPA for the specific subject. This percentile rank demonstrates that the student is performing better than at least 40% of their peers, reflecting competence and understanding of the subject matter.

2. Bad Academic Standing:

Major Subjects and Mathematics:

Failing Grades: Students with grades from 3.1 to 5.0 are automatically considered to have bad academic standing as these grades indicate failure in the subject.

Low Relative Performance: Students with grades between 1.0 and 3.0 who are in the bottom 40% of their cohort for the subject also fall into this category. Despite passing grades, their lower relative performance suggests challenges in keeping up with the academic demands of the subject.

This division allows for a focused investigation into how different levels of academic achievement correlate with social media usage patterns. The research seeks to achieve the following significant contributions:

Utilizing a bi-term topic modeling algorithm, the study aims to identify themes related to marine engineering students with bad academic standing and their social media activities.

Similarly, the research seeks to uncover themes related to students with good academic standing and their social media activities using the topic modeling algorithm.

This approach ensures a clear understanding of the impact of social media on diverse student groups, providing insights that could inform educational strategies and student support mechanisms. A consent form was given to every student to give us their Facebook profile and allow our Facebook account to be friended with them. Appendix A is the consent form used in this study. This paper is a continuation of what was presented at the International Conference on Business Analytics for Technology and Security (ICBATS), last year (2023). A new method of text analysis was used due to the sparsity of the text and the original text analysis methods being used do not capture potential topics and narratives. The current method of categorization of students was also different in the previous study and more refined experiments were used to capture more context on the social media content.

2. Review of Related Literature.

2.1. Natural Language Processing.

Topic modeling algorithms are unsupervised machine learning algorithms that can uncover narratives and topics within a given corpus and one of those is Latent Dirichlet Allocation (Jelodar et al., 2019) (Barhati, 2018). Gorro et al., use a topic modeling algorithm to understand the corpus of disaster response suggestions (Gorro et al., 2017). In this study, bi-term topic modeling was used to assist in doing qualitative analysis. The result of this study shows that a bi-term topic modeling algorithm can find significant narratives and topics that can help social scientists understand the corpus. Capao et al., use word2vec to understand customer reviews of the establishments in Cebu, Philippines. In this study, word2vec was able to find a semantic relationship of words in the corpus that could lead to a better understanding of the corpus (Capao et al., 2018). Wei and Croft, applied BTM in information retrieval and they used Gibb sampling to improve the accuracy of the BTM model (Wei and Croft., 2006). The result shows a positive and higher accuracy of the information retrieval model. Ancheta et al, use topic modeling algorithms on sparse data such as tweets to get valuable insights during class suspension and the result shows good topic models (Ancheta et al., 2020). Text analysis in understanding education competency is also explored in the study of Shibani et al (Shibahi et al., 2017). In this study, chat dialogue was examined to identify teamwork competencies among students using named entity recognition and text classification algorithms. Nastase et al, uses text mining and techniques to analyze buyer and sell content corpus, and it shows a promising result that could help automate content analysis (Nastase et al., 2007). Another unsupervised machine learning algorithm that can be used to analyze topics within the corpus is the k-means clustering algorithm (Bui et al., 2017). Go Bui et al, uses k-means clustering combined TF-IDF vectorization to cluster disaster response suggestions (Bui et al., 2017).

Khan et al conducted a study where they introduced an initial compute cluster algorithm for K-means clustering. The paper demonstrates that the proposed algorithm leads to improved and consistent solutions (Khan et al., 2004). In a separate study, Wagstaff et al show that by modifying the K-means clustering algorithm to consider the problem domain, there are significant enhancements in cluster accuracy (Wagstaff et al., 2001). Kanungo et al present another study where they introduce the filtering algorithm, a simple and efficient implementation of Lloyd's K-means clustering algorithm. This algorithm demonstrates practical efficiency with improved runtimes and finds application in tasks like color quantization, data compression, and image segmentation (Kanungo et al., 2002). On the other hand, Ding and He establish that principal components serve as the continuous solutions to the discrete cluster membership indicators in K-means clustering, presenting effective techniques for data clustering using K-means (Ding & He., 2004).

Recent academic research on the impact of social media on learners highlights various dimensions. Studies by Mathewson (Mathewson, 2020), Neira & Barber (Neira & Barber, 2020), and O'Dea & Campbell (O'Dea & Campbell, 2011) (Radovic et al., 2017) have delved into the effects on students' mental health, revealing links between social media use and mood disorders, self-esteem, and psychological state. Sampasa-Kanyinga & Lewis (Sampasa-Kanyinga & Lewis , 2015) and Sriwilai & Charoensukmongkol focus on the psychological functioning and emotional exhaustion related to frequent social networking among younger users (Sriwilai & Charoensukmongkol, 2016). The research by Stapel (Stapel., 2007), Tang et al. (Tang et al., 2013), Tsitsika et al. (Tsitsika et al., 2014), and Vernon, Modecki, & Barber extends the understanding to aspects like self-evaluations, social presence, patterns of online social networking in adolescence, and the impacts of problematic social networking on adolescent psychopathology, including sleep disruptions (Vernon, Modecki, & Barber, 2017).

2.2. Social Media Analysis.

The study of Deaton is based on the proximation of the Social Learning Theory by Albert Bandura to social media, which is a recent innovation (Deaton, 2015). This article seems to believe social media platforms like Facebook and Twitter have replaced the old way we used to relate and study. On the other hand, the emergence of these channels offers a place where the transmission of replicative behaviors is not only doable but is encouraged and appears widely. A thesis of the paper is that social media can notably facilitate teaching and learning by shaping a virtual setting in which children as observers, imitators, and models can reproduce behaviors online. This digital present-day setting grants faculty members, students, and researchers to explore the world of the internet that doesn't recognize the physical world's borderlines or temporal boundaries. One of the main concerns of this study is an inadequate examination of negative impacts that might be connected to social media implementation at school for instance distraction, a reduction of interpersonal communication, or sketchy material as a main cause of poor academic performance.

The research by Kolhar et al investigates the influence of university students' social media utilization on studying and learning, college life, and sleep (Kolhar et al., 2021). It is very easy to see that most of the students, who are just 3% of participants, only use social media for chatting and chatting, the other 97% of them are not connected with the books. The study emphasized a high prevalence of social media addiction among students (57%), and students showed a significant devotion to social media instead of to their academic engagement (66%). Our survey found that most studies with a perceived harmful effect of social media on their learning and sleep patterns when students were going to bed late because of social media use. Instagram, Twitter, and Snapchat were the most all used social apps.

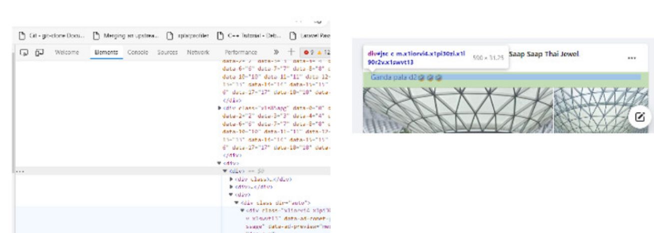
Although the findings are meaningful, in their review, they used first of all self-reported data which may result in biased information since students might not report their usage patterns or the influence entirely correctly. Besides this, the research focuses only on female students from one university, which also makes the results applicable to a different group or the student population across the globe. The analysis was predominantly descriptive, there were no statistical methods to investigate how social network platform affects our learners and the contextual analysis of social media content that people are attached to that interested them. The review suggests a significant gap in the existing literature: social media is an eminent area where there is a need for a comprehensive yet very methodologically grounded approach to this impact on students (Kolhar et al., 2021). Complementing current research by developing a mixed-methods approach, joining quantitative data with qualitative experiences, is valuable for the reduction of the dependence on self-reported data. Bearing this in mind, it would be imperative to apply a multilevel research design that involves a group of diverse students and multiple schools so that the study can provide generalizable results. Besides that, it is necessary to look at the role of newly emerging social media platforms and other possible factors that may in the end tie in the correlation between social media usage and academic performance.

3. Methodology.

3.1. Data Collection.

In this research, Cebu Technological University - Carmen Campus provided a sample of 500 marine engineering students who were enrolled in various math-related courses. We utilized the Facebook API to search for each student's profile and employed selenium and Beautifulsoup to gather their Facebook posts (Beautifulsoup, 2017). From the sample, we were able to capture 400 student Facebook profiles, resulting in a total of 7000 collected Facebook posts. These students were categorized into two groups based on their academic performance, specifically high grades and low grades in mathematics. Among the collected posts, 3000 were from students with good grades, while 4000 were from students with lower grades. Additionally, out of the 2200 posts belonging to students with good grades, 2200 posts comprised purely images and video/video links.

Figure 1: HTML Layout that shows the xpath to capture Facebook posts.



Source: Authors.

To enable the HTML parser's functionality, we utilized selenium automation to log in to a Facebook account. The chosen Facebook account was connected with almost all of the students' Facebook profiles as friends.

3.2. Data pre-processing.

To improve the accuracy of the topic modeling algorithm, the following data pre-processing techniques were applied:

1. *Remove special characters.*
2. *Remove emoticons.*
3. *Remove hyperlinks.*
4. *Remove stop words.*

3.3. Vectorization.

TF-IDF was used to transform all the corpus data for grade students and low grades to a Bi-term topic modeling algorithm.

3.4. Bi-term topic modeling.

Bi-term topic modeling is a method used to discover and analyze topics in a collection of texts or documents (Barhate, 2018). Unlike traditional topic modeling techniques that focus on single words as topics, bi-term topic modeling considers pairs of words (bi-terms) as the fundamental units for topic extraction. Bi-term is synonymous with bi-gram. This approach allows for a more comprehensive understanding of the underlying themes and concepts present in the text data, enhancing the accuracy and granularity of topic identification. By analyzing bi-term associations, the bi-term topic modeling technique can reveal deeper insights and relationships within the textual content, making it a valuable tool for various natural language processing and text mining applications. Bi-term topic modeling was used over LDA in this study for the following reasons.

1. ***Enhanced Contextual Understanding:*** Unlike LDA, whose modeling is focused on the distribution of individual words across documents, a BTM explores bi-terms, i.e. pairs of words. This method, however, becomes more evident when attempting to catch the wide range of word links which make the message clearer and therefore give an in-depth understanding of the main issue. Take for instance ”Marcos golden” which connotes more so than ”Marcos” on the person itself and ”Golden” might be associated with different subjects like ”Golden State” or ”golden retriever”; However, when paired, they distinctly identify content specifically relevant to discussions about the Marcos regime. This specificity enhances the thematic accuracy and depth, providing clearer insights into the subject matter being analyzed.
2. ***Superior Performance in Handling Short Texts:*** The data that we use consists mainly of short sentences which, is a common montage where LDA is relatively weaker due to sparse wordiness appearances that lead to poor inference on topics. However, BTM manages this discussion by exploring the evaluations of the word couples across the whole corpus eliminating the problem of relatively less information being included and providing more such topics that are stable and reasonable.
3. ***Independence from Document Structure:*** Our BTM approach is content-based regardless of sentence structure or length, which makes our corpus that includes different text lengths and structures one reason to choose it. With BTM, we bring in this characteristic of our topic modeling so that as long as our textual documents are consistent in terms of their overall architecture, our topic modeling remains valid and enduring.
4. ***Granularity and Specificity:*** BTM usually reveals thematic representation having more detail and emersion than LDA. This level of detail is very important to us, and it is because while in the process we do closely look into some of the thematic subtlety notices that could otherwise be easily omitted in the case when only single words are put into consideration. The bi-term model's power of demonstrating explicitly the semantic interconnections between the data provides for keen discrimination and understanding of intricate subjects.

1. ***Bi-term topic modeling result for good academic standing students:***
The parameters used to develop the BTM model are the following:
   - Features = 10000
   - # topics models = 5

   The number of topic models was decided based on the coherence score. The coherence score for the 5 topic models in this case is 0.42. Table 1 shows the topic models being discovered by the BTM.

Table 1: BTM Topic Models.

Topic #	Topic Models
0	university, flexible, learning, oxford, leni, president
1	election, may, vote, power, ched, tuition, online, class
2	discovery, youtube, philippines, covid
3	Agoy, birthday, ctu, events, please, share, robredo
4	exam, schedule, midterm, test, hagbong, pasar

Source: Authors.

2. ***Bi-term topic modeling result for bad academic standing students:***
The parameters used to develop the BTM model are the following:
   - Features = 10000
   - # topics models = 5

   The determination of the number of topic models was based on the coherence score, with a value of 0.45 obtained for the 5 topic models in this study. Table 2 illustrates the specific topic models uncovered through the Bi-term Topic Modeling (BTM) process.

Table 2: BTM Topic Models.

Topic #	Topic Models
0	agoy, priso, hahaha, lechon, laag, dilawan, golden, era
1	covid, dilawan, era, doh, uniteam, leni, gordron, redcross
2	joshua, bea, julia, gerald, pep, abs, cbn
3	dota, tiktok, luv, manny, karen, news, balita
4	leni, kakampink, uniteam, marcos, isko, manny, lenivsmarcos, duwag

Source: Authors.

3.5. Open Coding.

Open coding is a manual qualitative technique that examines qualitative codes for analysis.

3.6. K-means Clustering.

The K-means clustering algorithm was utilized to group diverse Facebook posts of students. The generative process in K-means clustering can be described as follows [17]:

1. Input the value of 'k', which determines the number of clusters.
2. Randomly select 'k' data points as initial centroids within the data space and allocate data points to the nearest centroid based on the Euclidean Distance, forming 'k' clusters.

Repeat the following steps:

a. Compute the new centroid as the mean of all points within each cluster, using the Euclidean Distance.
b. Assign data points to the new centroids.
c. Continue until the cluster assignments remain unchanged. To assess the effectiveness of the K-means clustering, the intra-cluster distance was measured using the silhouette score.

The Silhouette Coefficient is defined as follows:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}$$

Several experiments were carried out to determine the optimal number of clusters based on the highest silhouette coefficient.

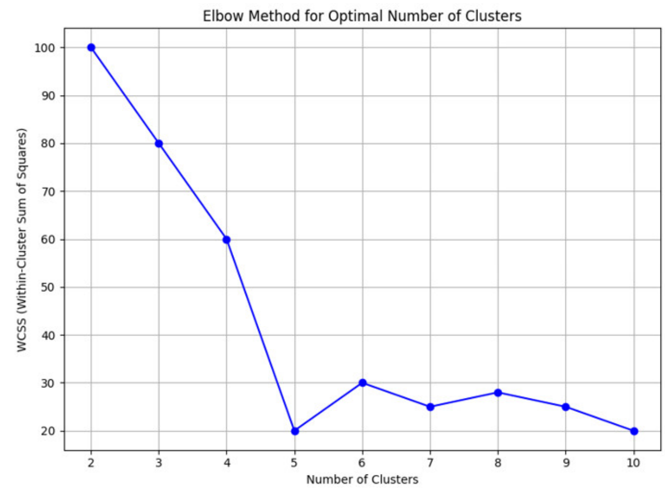
Table 3: K-Means Clustering Experiment.

Experiment #	Number of clusters (k)	Silhouette Score
0	2	0.0031258577715264674
1	3	0.0045089364735821118
2	4	0.0051881844891270432
3	5	0.0120765226548570571

Source: Authors.

The experiment started with k clusters = 2 with a silhouette score of 0.0031258577715264674. It is expected that as the number of k clusters increases, the silhouette score increases as well. Table I shows that some silhouette scores are lower compared to the previous number of k-clusters. As an alternative, an elbow method was conducted to determine the ideal number of clusters. In the context of the elbow method used for determining the optimal number of clusters in a k-means clustering algorithm, WCSS refers to the sum of the squared distances between each data point and its corresponding centroid within a cluster. Figure 2 shows the graph of the elbow method being performed.

Figure 2: Elbow Method Graph.



Source: Authors.

In this study, the non-fluctuating silhouette score that is closest to 1 was chosen. Experiment # 3 shows the optimal number of clusters to be generated. Table 4 shows the cluster models generated from the k-means clustering algorithm.

Table 4: K-Means Clustering.

CLUSTER #	CLUSTER MODELS
0	cebu, bbm, Duterte, uni, team, presidential, president, leni
1	bday, mapriso, trend, hahahaha, debate, lechon
2	Comelec, dq, disq, bbm, leni, isko, debate
3	Agoy, birthday, ctu, events, please, share, robredo, leni, debate
4	enrollment, covid, learning,online,f2f,face,ched
5	exam,schedule, midterm, test, hagbong, pasar

Source: Authors.

3.7. Correlation Analysis.

In this research, to explore the connection between social media activities and student performance, a correlation analysis was conducted using the number of Facebook posts and student performance. Equation 1 represents the formula utilized for this analysis.

$$R_{ij} = \frac{Covirance\ number\ of\ post\ and\ Math\ grades}{\sqrt{Covirance\ number\ of\ post\ .\ Math\ grades}} \quad (1)$$

3.8. Time Series analysis of the number of Facebook posts over time.

All Facebook posts that were gathered also include the date it was posted. A stationary test was conducted using the Augmented Dickey-Fuller test to determine if there was a specific surge of Facebook posts posted on social media for two sets of students.

4. Discussion.

4.1. Narrative Analysis.

Bi-term topic modeling followed by open coding was used to identify themes in the Facebook posts of students with varying academic standings. Additionally, K-means clustering was implemented to further validate and explore these themes.

4.2. Common Themes Across Both Groups.

Trending Memes and Presidential Election:

Both groups exhibited significant engagement with trending memes and the presidential election. This shared interest highlights the pervasive influence of national events and viral culture among the student body.

4.3. Distinct Themes in Each Group.

Good Academic Standing Students:

- **Educational Content**: These students frequently engage with content related to educational updates, school-related announcements, and CHED scholarships, which was supported by the K-means clusters indicating strong engagement with academic and institutional updates.

- **Social Events**: Narratives around personal and social events such as birthdays show a balanced engagement with both academic and social life.

Bad Academic Standing Students:

- **Entertainment and Politic**s: More engagement was noted with entertainment news and political propaganda. Clustering analysis similarly showed a distinct cluster focused on showbiz and political controversies, suggesting a diversion from academic content.

- **Controversial Political Issues**: Specific attention to controversial topics like the disqualification of a presidential candidate indicates a focus on sensational and high-stakes political discourse.

4.4. Analysis of K-means Clustering in Relation to Topic Models.

K-means clustering helped further delineate the themes identified by the topic models, revealing deeper insights into how different clusters of students engage with content:

- Overlap in Political Engagement: Both methods showed that students from both academic standings are heavily engaged in discussions about political events, suggesting that such topics are of universal interest.

- Academic vs. Non - Academic Focus: Clusters among good-standing students often centered around academic - related themes, aligning with the topic models. In contrast, clusters for students with lower academic performance highlighted a greater engagement with non - academic themes.

The dual application of topic modeling and K-means clustering provides a robust methodological framework for understanding student engagement on social media. This approach not only confirms the universality of certain themes but also highlights the distinct ways in which students with different academic performances interact with content. These insights can be leveraged to:

- **Tailor Educational Content**: By integrating academic content with formats and themes that are already popular among students, educational institutions can potentially increase engagement among students, particularly those at risk academically.

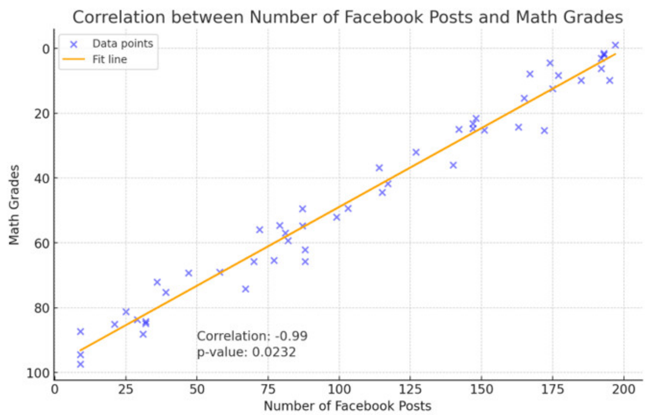
- **Develop Targeted Interventions**: Understanding the specific interests and engagement patterns of students can help in crafting more personalized and effective academic support and interventions.

The cohesion scores for the bi-term topic models, that of the students who score well in their course work is 0.42 and 0.45 for those with lower scores hence good but there is still room for improvement. Generally, score variance reflects the well-developed and situated topic. The performance of the two groups neither differed from each other very much, so there could be a common factor governing social media behavior that works for everyone, including low-performers. Also, the existing social media usage instances are overlapped, indicating that the model may be improved by contrasting the usage applications of students coming from different demographic circles.

4.5. Correlation Analysis.

The findings indicate a negative correlation of 82.2% between the number of Facebook posts per student and their math grades. The p-value is 0.0232 which shows that the null hypothesis is rejected and shows that the relationship is not by chance alone. Figure 3 presents the results of the conducted correlation experiment. To convert the transmuted grades into numerical grades, incomplete grades were translated to 66.

Figure 3: Correlation graph.



Source: Authors.

The interpretation of the result suggests that as students' social media activities increase, their academic performance tends to decline.

4.6. Augmented Dickey-Fuller (ADF) test was performed using Python statistical libraries. Independent ADF tests were conducted for above-average students and below-average students.

The result is as follows:

A. **Group of students with above average grades:**

ADF Statistic (fb posts): -17.02813296727508.

p-value (fb posts): 8.362883266997923e-30.

Critical Values (fb posts):

1%: -3.4524113009049935.

5%: -2.8712554127251764.

10%: -2.571946570731871.

The ADF Statistic for the Facebook posts represents the value of the test statistic, indicating the strength of evidence against the null hypothesis. In this case, the ADF statistic is -17.02813296727508. The more negative the ADF statistic, the stronger the indication against the null hypothesis, which assumes that the data is non-stationary. The p-value associated with the ADF statistic provides insight into the likelihood of obtaining the observed result if the null hypothesis is true. For the Facebook posts data, the p-value is 8.362883266997923e-30, an extremely small value. Typically, when the p-value is smaller than a chosen significance level, such as 0.05, the null hypothesis is rejected. In this instance, the null hypothesis asserting that the data is non-stationary would be strongly rejected.

The Critical Values for the Facebook posts represent thresholds for different levels of significance (1%, 5%, and 10%). They assist in determining the significance of the ADF statistic at these levels. In this case, all the critical values are more negative than the ADF statistic, providing further support for rejecting the null hypothesis.

In summary, with an ADF statistic significantly lower than the critical values and a very small p-value, the null hypothesis is rejected and concludes that the time series data for the Facebook posts is stationary.

B. **Group of students with below-average grades:**

ADF Statistic (fb post): -12.978423039997702.

p-value (fb post): 2.97871987575141e-24.

Critical Values (fb post):

1%: -3.4524859843440754.

5%: -2.871288184343229.

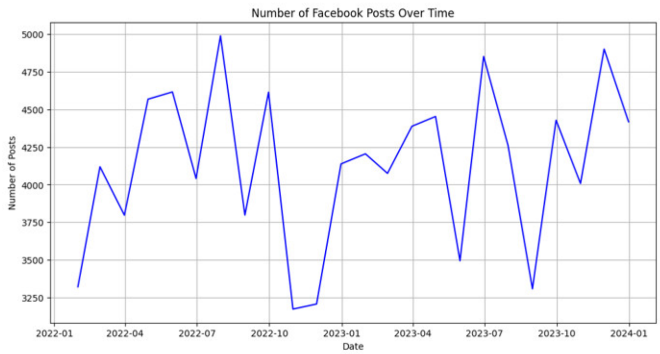
10%: -2.571964047565425.

Upon conducting the Augmented Dickey-Fuller (ADF) test for the Facebook posts data, the analysis yielded notable results. The ADF Statistic, calculated to be -12.978 423039997702, reflects a substantial negative value. This statistic serves as a pivotal indicator in determining the stationarity of the data. As the ADF statistic becomes more negative, it presents stronger evidence against the null hypothesis, which posits that the data is non-stationary.

Moreover, the associated p-value, computed as 2.978719 87575141e-24, emerges as exceedingly small. The p-value signifies the probability of observing the obtained result if the null hypothesis were true. In this case, the diminutive p-value suggests that the null hypothesis of non-stationarity is firmly rejected. Typically, when the p-value falls below a pre-defined significance level, such as 0.05, the rejection of the null hypothesis is warranted.

Furthermore, critical values play a crucial role in validating the significance of the ADF statistic. For the Facebook posts data, the critical values at 1%, 5%, and 10% levels are -3.4524859843440754, -2.871288184343229, and -2.571964047565425, respectively. Notably, all critical values surpass the computed ADF statistic, bolstering the case for rejecting the null hypothesis. Figure 4 shows the time series graph for the total Facebook posts.

Figure 4: Time Series Analysis.



Source: Authors.

In summary, the analysis of the ADF test for the Facebook posts data indicates strong evidence in favor of stationarity. The negative ADF statistic, coupled with the minute p-value and critical values, collectively support the rejection of the null hypothesis, implying that the Facebook posts data exhibits stationarity characteristics.

5. In-depth Analysis.

Interestingly, this study contributes to the literature by analyzing the nature of social media content and its connection to academic success, a significant research issue that we initially marked out. Though the themes identified handled thought awareness through the open coding method is created with a good basis, there is still a lack of analysis on what concurrent impact(s) get to the academic habits and time management of the students. For example, the examination of the factors that are fueling 'Birthday Parties and Events' and 'Presidential Election' predisposition amongst students is mentioned, but the study does not go further to investigate if this or how this affects their learning strategy. Does the act of participating in a similar forum equate to ability in social studies or a loss of time devoted to study?

Being cardinal, our analysis did not use the creation of content by students as a subject of the research, it was an information area we missed. Students' professionalism on social media reflection of content is not a passive activity, but it shows the grade of your interest, seniority, and learning style. A student who joins academic discussions and scholarly pages or uses these materials as additional tools to support classroom learning may also be using the self-directed learning approach. On the contrary, a person's decision to choose entertainment stuff may be the one who could learn best through the conversation or reading of a story.

If we are discussing the backward relationship between Facebook post numbers and math scores then we should be very careful interpreting these outcomes. The findings, as they stand now could bring up misinterpretations in that way where it may seem like a simple cause-effect while in reality, it may be a result of factors that are not captured by our model. For example, a child with poor academic performance could be employing social media as a way to escape pressure or a favorite means of handling other personal problems. Overall, it will affect his performance negatively. This distinction between correlation and causation is critical, and further research is necessary to disentangle these complex relationships.

Moreover, the relatively high coherence scores of the BTM models, while indicative of consistent themes, do not capture the multifaceted impact of these themes. A more nuanced analysis would consider how the identified themes interact with each student's life. It could be that marine engineering students with lower grades engage with political content in a manner that, though intellectually stimulating, competes with academic time, while high-achieving students may effectively integrate such content into their broader learning process.

Conclusion.

This analysis of marine engineering students' social media activity reveals that while students with good academic standing and those with bad academic standing share some commonalities in their online interactions, distinct patterns also emerge that differentiate their engagement on these platforms. For both subgroups, guys mainly spread popular memes, whereas, girls tend to express their views on politics as seen from their Facebook posts. Nevertheless, the main discernible difference is that, in the case of low-achieving students, the conversation leans towards entertainment-based and social issues while the high achievers tend to discuss academic-oriented issues. These results, then, highlight the many-dimensional quality of social media networks which operate as a mirror of the diverse concerns and perspectives of the social media users.

Bi-term topic modeling has done well in finding the main ideas of the corpus, however, this method also revealed the difficulty of distinguishing the first-degree overlapping among topics. Such duplicate may be confusing as to how the themes interrelate and the influence they yield varies. Furthermore, k-means clustering was applied in a bid to deepen the understanding of these topics which was reflected in the five-cluster structure that embodied many of the themes previously identified by the Bi-term topic modeling, however, the silhouette score being somewhat modest could limit the accuracy of the results.

Critical reflection of the found inverse relation between time spent on Facebook and academic success provides some grounds for complex analysis. It points out an apparent tendency for social media involvement to go hand in hand with a falling rate of scholastic accomplishment. Nevertheless, in light of what we have shown from our analyses, the correlation should not be mistaken as a causation. The interplay of social media activity and education is complex and impacts a huge number of aspects beyond the content quality, students' time management, and personal context of the Young learners.

While we take into consideration the following complexities, future research could be useful with a mechanism that monitors the students' social media activities. This, in turn, could assist in the establishment of more discriminating techniques for the evaluation of the effect of post volume on academic performance to reduce subject bias. Through this method, we can have confidence that our research will be more precise and will provide academicians with evidence-based recommendations for students' social media usage.

Additionally, the ADF test did not show any significant seasonal patterns in the number of posts issued by the two groups of students for each group, indicating the same attraction to Facebook throughout the year. This pervasiveness can potentially be harnessed by educational strategies to weld students to their social media life for academic advantage.

Social media represents a double-edged weapon - it may either disrupt or prevent studies from getting done. Looking ahead, the cooperation of educators and researchers who will work together to find the true implications that social networks have in the educational sphere is absolutely necessary. Through this process, we can look upon this powerful instrument favorably, thereby serving student progress rather than weakening it.

Acknowledgements.

The work is supported by Cebu Technological University (CTU). We would like to express gratitude to the CTU-Carmen Research office for their support in this study.

References.

Gorro, K., Ilano, A., Ranolo, E., Pineda, H., Sintos, C., & Gorro, A. J. (2023, March). An Analysis of Social Media Activities and Student Performance using Topic Modeling and Correlation. In 2023 International Conference on Business Analytics for Technology and Security (ICBATS) (pp. 1-5). IEEE. Retrieved from: https://www.meltwater.com/en/blog/social-media-statistics-philippines. Retrieved from: https://www.netsweeper-.com/filter/education-web-filtering/the-positive-and-negative-impact-of-social-media-on-students/35845. Retrieved from: https:-//www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739?mod=hp_-lead_pos7.

Rawhide Boys Ranch. (2015). Teen Cyberbullying and Social Media Use on the Rise [INFOGRAPHIC] Retrieve from http://www.rawhide.org/blog/wellness/teen-cyberbullying-and -socialmedia-use-on-the-rise/.

What Is Cyberbullying," February 7. 2018. Retrieve from https://www.stopbullying.gov/cyberbullying/what-is-it/.

Latent Dirichlet Allocation for Beginners: A high level intuition. Retrieved from https://medium.com/@pratikbarhate/latent-dirichletallocation-for-beginners-a-high-level-intuition-23-f8a5cbad71.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (BTM) and topic modeling: models, applications, a survey. Multimedia Tools and Applications, 78(11), 15169-15211.

Gorro, K., Ancheta, J. R., Capao, K., Oco, N., Roxas, R. E., Sabellano, M. J., ... & Goldberg, K. (2017, December). Qualitative data analysis of disaster risk reduction suggestions assisted by topic modeling and word2vec. In 2017 International Conference on Asian Language Processing (IALP) (pp. 293-297). IEEE.

Capao, K., Gorro, K. D., Gorro, K. D., Sabellano, M. J., Militante, C. L. A. G., & Manalili, J. P. C. (2018, April). Aspect Analysis of Cebu Establishments' Online Reviews using k-means Clustering and word2vec. In 2018 3rd International Conference on Computer and Communication Systems (ICCCS) (pp. 61-66). IEEE.

Wei, X., & Croft, W. B. (2006, August). BTM-based document models for ad-hoc retrieval. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 178-185).

Ancheta, J. R., Gorro, K. D., & Uy, M. A. D. (2020). # Walangpasok on Twitter: Natural language processing as a method for analyzing tweets on class suspensions in the Philippines. In 2020 12th International Conference on Knowledge and Smart Technology (KST) (pp. 103-108). IEEE.

Shibani, A., Koh, E., Lai, V., & Shim, K. J. (2017). Assessing the Language of Chat for Teamwork Dialogue. Educational Technology & Society, 20 (2_A20_), 224–237.

Nastase, V., Koeszegi, S., & Szpakowicz, S. (2007). Content analysis through the machine learning mill. Group Decision and Negotiation, 16(4), 335-346.

K-means (Cluster Analysis). 2016. Retrieved from https://-www.denovosoftware.com/site/manual/cluster_analysis2.htm

Bui, S. M. G., Gorro, K., Aquino, G. A., & Sabellano, M. J. (2017, December). An analysis of DRR suggestions using K-means clustering. In Proceedings of the 2017 International Conference on Information Technology (pp. 76-80). Retrieved from: https://beautiful-soup-4.readthedocs.io/en/latest/

Tan, P.-N., Steinbach, M., Kumar, V., & Karpatne, A. (2005). Introduction to Data Mining. Pearson Publishing

Khan, S. S., & Ahmad, A. (2004). Cluster center initialization algorithm for K-means clustering. Pattern Recognition Letters Volume 25 Issue 11, 1293-1302.

Wagstaff, K., Cardie, C., Rogers, S., & Schroedl, S. (2001). Constrained K-means Clustering with Background Knowledge. Proceedings of the Eighteenth International Conference on Machine Learning, 577-584.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An Efficient k-Means Clustering Algorithm: Analysis and Implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence Volume 24 Issue 7, 881-892.

Ding, C., & He, X. (2004). K-means Clustering via Principal Component Analysis. ICML '04 Proceedings of the twenty-first international conference on Machine learning, 29.

Mathewson, M. (2020). The impact of social media usage on students' mental health. J. Stud. Affairs, 29, 146–160.

Neira, B. C. J., & Barber, B. L. (201networking site use: Linked to adolescents' social self-concept, self-esteem, and depressed mood. Aus. J. Psychol., 66, 56–64.

O'Dea, B., & Campbell, A. (2011). Online social networking amongst teens: Friend or foe? Ann. Rev. CyberTher. Telemed., 9, 108–112.

Radovic, A., Gmelin, T., Stein, B. D., & Miller, E. (2017). Depressed adolescents positive and negative use of social media. J. Adolesc., 55, 5–15.

Sampasa-Kanyinga, H., & Lewis, R. F. (2015). Frequent use of social networking sites is associated with poor psychological functioning among children and adolescents. Cyberpsychol. Behav. Soc. Network., 18, 380–385.

Sriwilai, K., & Charoensukmongkol, P. (2016). Face it, don't Facebook it: Impacts of social media addiction on mindfulness, coping strategies and the consequence on emotional exhaustion. Stress Health, 32, 427–434.

Stapel, D. A. (2007). "In the mind of the beholder: The interpretation comparison model of accessibility effects," in Assimilation and Contrast in Social Psychology, eds D. A. Stapel and J. Suls (London: Psychology Press), 143–164.

Tang, F., Wang, X., & Norman, C. S. (2013). An investigation of the impact of media capabilities and extraversion on social presence and user satisfaction. Behav. Inform. Technol., 32, 1060–1073.

Tsitsika, A. K., Tzavela, E. C., Janikian, M., Ólafsson, K., Iordache, A., Schoenmakers, T. M., et al. (2014). Online social networking in adolescence: Patterns of use in six European countries and links with psychosocial functioning. J. Adolesc. Health, 55, 141–147.

Vernon, L., Modecki, K. L., & Barber, B. L. (2017). Tracking effects of problematic social networking on adolescent psychopathology: The mediating role of sleep disruptions. J. Clin. Child Adolesc. Psychol., 46, 269.

Deaton, S. (2015). Social learning theory in the age of social media: Implications for educational practitioners. Journal of Educational Technology, 12(**??**), 1-6.

Kolhar, M., Kazi, R. N. A., & Alameen, A. (2021). Effect of social media use on learning, social interactions, and sleep duration among university students. Saudi journal of biological sciences, 28(**??**), 2216-2222.