Predicting Student Performance Based on Moodle Forum Interaction Logs

Maya Louise E. Asuero
Department of Computer Science
Xavier University-Ateneo de Cagayan
Cagayan de Oro City, Philippines
masuero@xu.edu.ph

Angelu Ferdinand A. Garcia
Department of Computer Science
Xavier University-Ateneo de Cagayan
Cagayan de Oro City, Philippines
angelu.ferdinand.10@gmail.com

Julius Czar Makiling
Department of Computer Science
Xavier University-Ateneo de Cagayan
Cagayan de Oro City, Philippines
juliusczarmakiling@gmail.com

Jun Rangie C. Obispo
Department of Computer Science
Xavier University-Ateneo de Cagayan
Cagayan de Oro City, Philippines
jobispo@xu.edu.ph

ABSTRACT

This paper predicted students' performance on a Moodle course based on their forum interaction logs and the community of inquiry model. The research aimed to identify key attributes of students' forum interactions, based on the community of inquiry model's indicators of social presence, that can be used to create models that predict a student's course performance. To conduct this study, the researchers used a methodology grounded in data mining which involved data collection, data preprocessing, training, and model evaluation. The researchers used tools and algorithms provided by the machine learning software WEKA in creating the model. Results of the study showed that all indicators of social presence are present among the forum discussion logs, with *continuing a thread* significantly affecting the final grade. However, this may have been influenced by the large volume of messages classified as continuing a thread in comparison to the other indicators. Generally, the indicators of social presence can predict whether the student will merit a final grade that is high grade and low grade to a limited extent.

CCS CONCEPTS

• Theory of computation → Theory and algorithms for application domains → Machine learning theory → Machine learning theory

KEYWORDS

data science; data mining; Moodle logs; forum logs; student performance; computer science

PCSC2024, May 2024, Laguna, Philippines

https://doi.org/10.1145/1234567890

ACM Reference format:

Maya Louise Asuero, Angelu Ferdinand Garcia, Julius Czar Makiling and Jun Rangie Obispo. 2024. Predicting Student Performance Based on Moodle Forum Interaction Logs. In *Proceedings of Philippine Computing Science Congress (PCSC2024).* Laguna, Philippines, 8 pages.

1 INTRODUCTION

Online learning has seen a dramatic increase in the midst of the global COVID-19 pandemic [19]. Due to stay-at-home orders, educational institutions have shifted from face-to-face course delivery to distance learning. However, online learning is not new. In fact, the market for online learning and educational technologies has been steadily growing for the past 10 years [18, 19]. Among the many technologies and modes used in delivering online courses, learning management systems (LMS) are one of the most used. Massive open online courses and in-person universities have integrated the use of LMS in its courses which allows them to create and organize lessons and other materials used in class such as quizzes [20]. This study is focused on Moodle, an open-source LMS, that aims to provide educational institutions a single robust, secure, and integrated system [1].

Moodle hosts notable local universities such as Ateneo de Davao University, Mindanao State University - Iligan Institute of Technology, University of the Philippines – Baguio, University of the Philippines Open University, and Xavier University – Ateneo de Cagayan. These institutions combine or blend the use of Moodle and the traditional face-to-face classes for their approach to education. However, in response to the COVID 19 pandemic, the universities adapted to full online education, wherein face-to-face classes are substituted with synchronous and recorded video calls, and usual classroom activities are done on Moodle. Forums have also been used by instructors to stimulate discussion and collaboration in asynchronous learning where students and instructors can initiate discussion threads and interact with one another.

In the context of online education, the Community of Inquiry Model is important to understand the process required for effective learning by considering its three key elements: Teaching Presence, Social Presence, and Cognitive Presence. This model serves as a framework for text-based interactions and has been used in exploring forum logs generated by Learning Management Systems [4].

While numerous studies guided by the Community of Inquiry model have been conducted to investigate the relationship between LMS interaction logs and student performance, studies focusing on forum interaction logs and its relationship to student performance remain limited. This leaves the relationship between forum activities and course performance unclear. A better understanding of this relationship is needed to enhance the learning experience, especially in an online-learning environment.

This study aims to investigate forum interaction logs and its relationship with student performance by using data mining techniques to find patterns from the collected data, create models to predict the students' final grades, and analyze its results. Grounded by the studies of [4] and [8] which underlines the theory of Community of Inquiry and its key element, Social Presence, it is assumed that the interaction between the students manifests Social Presence. With these considerations, this study aims to answer the following research questions:

1. What indicators of social presence are present among the forum discussion logs?
2. Which indicators of social presence features from the forum discussion logs significantly affect the students' final grade?
3. To what extent is social presence in forum discussion logs a good predictor for students' final grade?

2 RELATED LITERATURE

In this section, the researchers discussed topics relevant to ground the study. Aside from Educational Data Mining, and Mining LMS Data for Student Performance, this section focuses on studies that use interaction logs from LMSs in relation to student performance, and the use of Social Presence as a predictor.

2.1 Mining Forum Interaction Logs for Student Performance

There are several notable studies that focus on mining forum interaction logs. A study investigated the course forums of Shanghai Fudan University's sociology program [7]. The forums studied included interactions and the contents of the posts between students and teacher's assistants in relation to the performance of the students in the course. Results of the study showed that students who are more active in forum discussions tend to have higher grades. The researchers assumed that those who got higher grades have better learning attitudes whereas those who received lower marks did not have enough time in the course.

2.2 Social Presence as a Predictor of Student Performance

Social Presence is one of the three key elements of the Community of Inquiry Model. [4] explains that Community of Inquiry serves as a framework for educators in using text-based computer conferencing in delivering a successful higher educational experience to their students. What makes this model important is the fact that research shows that there is a relationship between the three key elements (Teaching Presence, Social Presence, and Cognitive Presence), and the students' perceived learning, satisfaction with the course, satisfaction with the instructor, actual learning, and sense of belonging [3, 5, 13].

Among the studies that associate social presence with students' grades, and perhaps the one that is most similar to the current study is authored by [9]. Their study investigated the use of indicators of social presence as predictors of final grades of a Master's level Computer Science online course. To determine the association between the indicators of social presence and the grades, descriptive statistics and multiple regression analysis were applied. Correlation analysis showed that *continuing a thread*, *asking questions*, *complimenting*, *expressing appreciation* and *vocatives* indicators were strongly positively correlated with the final grade, while the correlation with *self-disclosure* tends to be marginally significant. On the other hand, the results of multiple regression analysis showed that *continuing a thread* is positively associated, and *complimenting*, *expressing appreciation* is negatively correlated with grades.

[22] emphasized the importance of Social Presence in Online Forums among Distance Learners. Their methodology involved manually reviewing and coding forum posts that contained indicators of social presence. Learners found that lack of engagement and participation among themselves, and this made them anxious about presenting their ideas. [14] found that a high degree of social presence affects learners' perceived learning and increases their satisfaction in an online learning experience. Therefore, through a good sense of social presence it could bring the learners to a greater emotional satisfaction.

A separate study [23] adapted a similar methodology as the one previously mentioned. This focused on exploring the relationship between learners' social presence in MOOC forums and learners' prestige with the use of automated content analysis and social network analysis. The study adopted the use of a revised framework for social presence [16] from the first proposed framework [4]. From 4,650 posts with an equivalent of 23,755 sentences, the researchers randomly chose 3,500 to manually code using the categories of social presence. Three researchers participated in the coding process with each coder being randomly assigned 1,200 sentences to code. To ensure correctness in coding, the coders constantly compared codes until they reach an agreement rate of 100%. This was then used to train and validate the text classification model. Results of the study showed that certain indicators of social presence have positive correlations with learners' prestige such as *asking questions*, *expressing gratitude*, *self-disclosure*, *sharing resources* and *using vocative*. One the other hand, *disagreement/doubts/criticism* and *negative emotions* had a negative correlation with learners' prestige.

Although there have been numerous studies that attempt to relate student performance to general interaction logs, only a few investigations made use of forum interaction logs against student performance. Even so, none of these studies have given a special focus to classifier algorithms to investigate student performance.

3 METHODOLOGY

The research design used by this study leans towards the direction of data mining (Figure 1). Thus, the following steps were adapted: data collection, preprocessing, feature engineering, training, and evaluation [15].

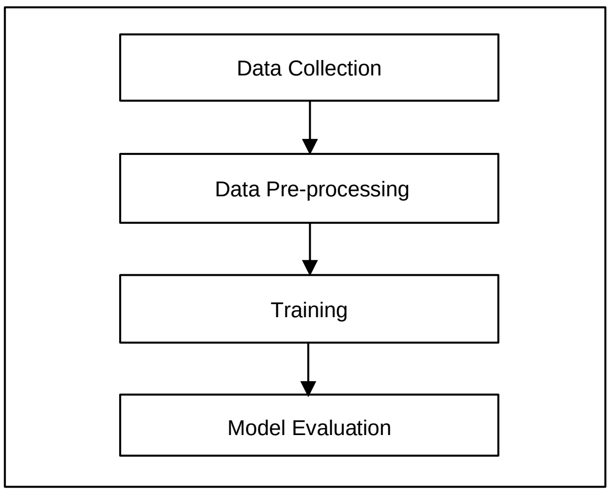

Figure 1. Research Design

To provide a brief overview, firstly, data will be collected from the Moodle LMS. Then, data is prepared through preprocessing. The preprocessing phase starts with the data cleaning, manual classification of forum messages, and feature engineering. After that, the machine learning algorithm will use the data to produce models. Finally, the performance of the resulting models is evaluated through statistical methods.

3.1 Data Collection

The datasets were taken from a university learning management system implemented on Moodle. In terms of the selection of courses, a criteria was applied similar to a study conducted on Coursera MOOC [21] and as suggested by the study of [11]. This was done to ensure that there are enough data points from a course to conduct this study. The criteria were decided based on the indicators of social presence since these courses have an increased level of meaningful interactions between students [9]. Specifically, the selected courses must show the presence of the indicators of social through the forum discussion logs and have variations of the said indicators.

The dataset includes data from six courses of the school year 2020 - 2021. To avoid overfitting, the study selected courses with a distinct composition from the other selected course based on year level and program. Each course spans a quarter of the academic year which is approximately two months. Students in the dataset were anonymized. The datasets include the students' grades, which reflect the breakdown of scores in activities and tests as well as the final grade in that course, forum discussion logs which show the contents and data submitted on discussion forums, and general interaction logs of the students in the course.

Moodle has five forum types. This study used forum discussion logs regardless of the type of forum.

3.2 Preprocessing

Before the data can be processed, it must be prepared to match the requirements of the mining software. This includes converting the data into the desired format. The datasets were presented in an MS Excel spreadsheet and were converted into a file format recognized by WEKA. Specifically, the datasets were converted into single-sheet xlsx format and uploaded to the mining software using extension packages found in WEKA.

3.2.1 Manual Classification

To distinguish the quality of the posts, the messages in discussion forums were classified into categories grounded in the open communication indicators of social presence [8]. Specifically, this study makes use of six categories that correspond to the indicators. Messages were coded or classified following the definitions of each category. It is worth noting that the posts or messages were manually classified by the researchers, similar to the studies of [6, 9]. Each message was classified as a single unit, therefore the researchers took into consideration the general idea and intent of the message based on its content.

Posts that fall into multiple categories were classified according to the value of the overall forum message. These will be ranked in the following order of precedence:

1. Quoting from others' messages,
2. Referring explicitly to others' messages,
3. Asking questions,
4. Expressing agreement.
5. Complimenting, expressing appreciation,
6. Continuing a thread

Along with the order of precedence, to further minimize the subjectivity of the classification, each entry was cross validated among the researchers; the messages were carefully reread, and their classification was made sure to be correct at least three times. To do this, the forum discussion logs which contained the messages were divided into three sets of messages. Each set was evaluated and classified by three different researchers.

Once all the forum discussion logs were classified, the researchers then checked the classification. If all evaluators gave the same classification for a message, then the message is classified according to the said classification. However, if there are differences or conflicts in classification, the evaluators will have to discuss the said message and come to a unanimous conclusion as to its final classification.

3.2.2 Feature Engineering

An important step of data preprocessing is the identification of variables from the datasets which will be mined. The variables, in this case, refer to spreadsheet columns. Therefore, the columns that contain data that are not relevant to the study were removed. From the data set, the columns that are relevant to the study are *userid* to uniquely identify the student, *message* which shows the content of the forum post, *parent* which distinguishes replies from discussion starters, *wordcount* of the message, and course total in percentage.

Table 1. Features

Features	Description
total_OCt	Total number of forum posts classified as *Continuing a thread*
total_wordcount_OCt	Total number of words used in total_OCt
total_OQ	Total number of posts classified as *quoting from others' messages*
total_wordcount_OQ	Total number of words used in total_OQ
total_OR	Total number of posts classified as *referring explicitly to others' messages*
total_wordcount_OR	Total number of words used in total_OR
total_OA	Total number of forum post classified as *asking questions*
total_wordcount_OA	Total number of words used in total_OA
total_OCa	Total number of forum posts classified as *complimenting, expressing appreciation*
total_wordcount_OCa	Total number of words used in total_OCa
total_OE	Total number of forum posts classified as *expressing agreement*
total_wordcount_OE	Total number of replies received, including indirect replies
replies_received	Total number of words used in replies_received
replies_wordcount	Total number of characters used in replies_received
replies_charcount	Total number of times a thread was clicked and opened
Disc_viewed	Total number of times a course module (which may contain discussions/threads) was clicked and opened
Module_viewed	Total number of forum posts classified as *Continuing a thread*

3.2.3 Analysis

Analysis involves making use of the built-in tools of WEKA that can be specifically found in its the Experiment Environment. This automates the creation of statistical tests and the running of multiple learning schemes on the data sets. For instance, the test facility allows statistical significance test of different learning schemes.

Further, Spearman's Rank Correlation Coefficient and Point-Biserial Correlation were used to understand the relationship between each of the different features found in Table 1 and the final grade. A correlational matrix was also used to show the relationship between the features. Histograms were used to show the distribution of the data between different features.

3.3 Training

The study used the mining tool WEKA to process the datasets. This tool was developed by the University of Waikato in New Zealand. It is a software that has a collection of machine learning algorithms and visualization tools that can be used for data mining. WEKA's built-in Experiment Environment was used to run and train the prediction models.

The data was processed using Naive Bayes and Logistic Regression classification algorithms. Classification algorithms are used to model data and its relationship to a certain result. The Naive Bayes classifier assumes that features are independent given a class, and was found to be widely used and effective in practice [14]. After evaluation of the results, Logistic Regression was used for the High-Low Grade classification. The algorithm Logistic Regression can be used by predicting binary outcomes such as the high-low grade classification.

3.4 Model Evaluation

To test the accuracy of the models, k-fold cross-validation was used. For this study, 10-folds cross-validation was used where 90% of the data points are trained and 10% are tested with over 10 different runs.

The results of the experiment made on WEKA are taken directly from the data mining software. To identify the attributes which have the greatest effect on final marks, the researchers used feature-selection algorithms. Finally, to rank the attributes or features, the frequency of selection for each attribute as decided by each algorithm is recorded. Completion of this step will give a list of features sorted according to how much influence they have on the final course percentage.

4 RESULTS AND FINDINGS

Initially, data from 14 classes were released to the researchers. These were then filtered to reduce overfitting. An overfit could occur in this context when the same student is found in at least two sections or blocks. From this point onwards, a block (referred to as a class in normal circumstances) should be understood as a group of students studying the same subject. Furthermore, two blocks may or may not study the same subject.

Table 2. Summary of Forum Interaction Logs from 11 Blocks

Summary	Value
Number of students	291
Total Messages	1513
Threads Started by Students	1304
Threads Started by Teachers	37

This filtering process reduced the number of blocks to 11, with an average of 26 students. Table 2 shows a general summary of the forum interaction logs from the final set of blocks. With 291 students, most of the messages are thread-starters created by the students, while only an ample number are created by teachers.

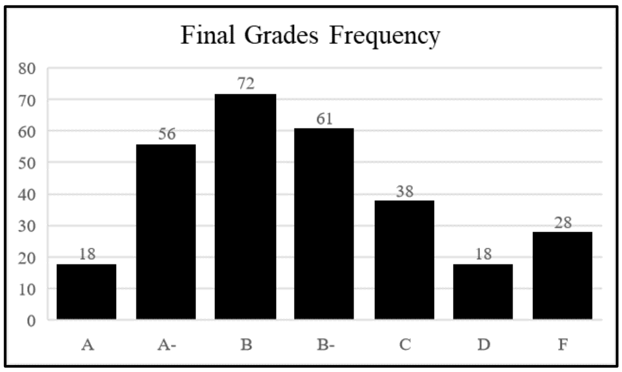

Figure 2. Histogram of Letter Grades

The quantity of the final grades is shown in Figure 2. As observed, most of the students have a letter grade of either A-, B, or B-. This creates an imbalance between the other letter grades which could mean that the algorithm will have fewer data points to learn from.

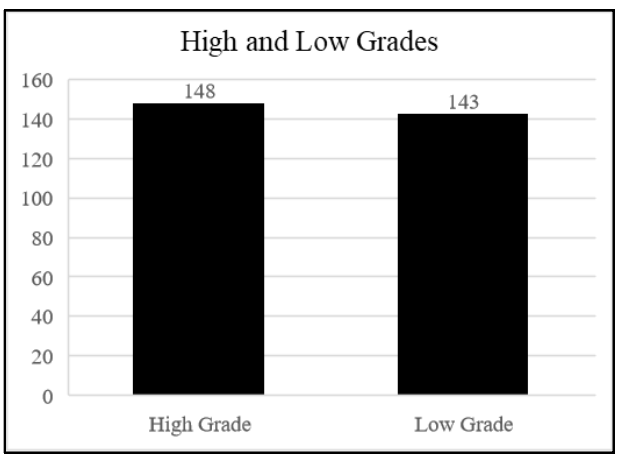

Figure 3. Histogram of High-Low Grades

To mitigate the imbalance, the researchers have decided to classify the students as having low grades or low performance and high grades or high performance. The median was used to determine how the data would be classified as high grade or low grade. With a median of 76.18, final grades of A, A-, and B were classified as high grade while final letter grades of B-, C, D, and F were classified as low grade. Figure 3 shows the histogram of the high and low grades classification.

In the previous sections, the researchers posed three research questions for this study. The succeeding sections will discuss the findings of the study and answers to these questions.

4.1 What indicators of social presence are present among the forum discussion logs?

Manual classification showed all the indicators of social presence can be found among the forum discussion logs. Although complete, the frequency for each indicator is imbalanced. Table 3 lists each indicator according to its frequency as classified in the messages.

Table 3. Message Classification

Indicators of Open Communication	Count
Continuing a thread	1132
Asking a question	241
Expressing agreement	75
Complimenting, expressing appreciation	48
Referring explicitly to others' messages	14
Quoting from others' messages	3

This result is consistent with other studies, wherein, likewise in online forums, all the types of indicators of social presence under open communication can be found [9, 22].

The social presence category of open communication allows for reciprocal and respectful exchanges, especially the mutual awareness and recognition of others' contributions [4]. Since tools that are integrated into these systems are designed to overcome the limitations of distance learning, social presence prevails naturally because it promotes and encourages such interaction [10].

The teacher's module design in the LMS has a major influence on the students' activity in the forums. For example, messages classified as *continuing a thread* are mostly responses to graded activities. Thus, the ranking of each indicator gives a good insight into how the teacher utilizes the LMS.

However, it is best to take the interpretation with caution. Following the order of precedence during the manual classification of messages means that when it fails to be captured by the other indicators, the default classification will always be *continuing a thread*. With this thought, a convincing argument as to why there is an overwhelming number of messages classified as *continuing a thread* is because the latter can be broken down into finer subcategories of indicators, or simply because the correct classification does not exist in the set of indicators of social presence.

4.2 Which indicators of social presence features from the forum discussion logs significantly affect the students' final grades?

Although the study explored two ways of classifying student performance through letter grade classification and high-low grade classification, *continuing a thread* as an interaction and its word count were present in the feature selection results for both types of classification. This may have been influenced by the large volume of messages classified under this indicator. Moreover, the results of the analysis showed that *continuing a thread*, as an interaction and its word count, also has the highest correlation and a strong relationship with the class feature. However, these two features were determined as highly correlated which prompted the researchers to drop the feature for continuing a thread word count from the feature selection.

Aside from continuing a thread, *referring to others' messages* was also selected from the feature selection process as having a significant contribution to the model. However, it was only considered for the high-low grade classification and not the letter grade classification. This non-inclusion can be accredited to the limited range and variation of the data in this feature which can make it a challenge to distinguish for the 7-letter grade classification. Close inspection of the dataset shows that students with *referring to others' messages* were classified as having A, A-, or B letter grade or high grade for high-low classification.

Notably, the word count for the indicator *asking questions* was included in the results of the feature selection. However, the total interaction of this indicator was not selected. This could indicate that, for instance, the actual number of questions asked does not matter, but rather the content of the question being asked. However, without a method for scoring the content of the message, it cannot be generalized that longer messages have more substantial content; only that longer messages are predictive of high performance.

Similar studies have shown similar results for the indicators [5, 9, 11, 22] and the correlation of word counts [12]. Other studies have reported on other indicators, such as *asking questions*, *quoting others' messages*, and *complimenting others*, have better relationships with student performance [17]. The results of this dataset mainly highlight *continuing a thread* and *asking questions* word count as the indicator that can significantly affect the students' performance. Moreover, replies word count, and discussions viewed are also predictive of student performance. *Referring to others' messages* can be considered predictive of the student performance; however, its significance diminishes as the number of classes increases.

4.3 To what extent is social presence in forum discussion logs a good predictor for students' final grades?

Given the model evaluation results, the social presence of a student can be used as a predictor for students' final grades to a limited extent. Between the two methods of grade classification, the model has a higher performance rate when predicting high grades vs low grades in comparison to the university marking system classification. This is proven by observing the increased model precision, recall, and accuracy when classification decisions are limited to only high and low seen in Table 4.

Looking into the histogram of student final grades, a convincing explanation is that there exists an imbalance in the representation. This result is similar to [2]'s results where the model could not accurately predict 5 to 10-grade categories, but could accurately predict failing students from other classifications.

Table 4. Summary of Evaluation Results

Metric	Letter Grades (Naive-Bayes)	High-Low (Naive-Bayes)	High-Low (Logistic Regression)
Accuracy	29.21%	69.76%	69.42%
Kappa Statistic	.1733	.3988	.3899
Precision	.332	.731	.701
Recall	.292	.698	.694

The resulting models of the 2-class datasets using both Naïve Bayes and logistic regression revealed that both models perform very similarly. This can be proven by observing the close metrics for its accuracy, kappa statistic, precision, and recall. Considering that both models perform unsatisfactorily at accurately predicting student performance, it further strengthens the previously mentioned remarks regarding the lack of variety in student forum post types to correctly classify high-low grades.

In summary, the resulting models perform unsatisfactorily in predicting students' performance using forum interactions following the indicators of social presence, particularly for open communication. For the courses included in the dataset, participation in forum discussion counts toward the final grades. However, this comprises only a small portion of the final grade, so other factors might come into play. This includes how well the student performed in other activities, quizzes, and exams. In addition, the data imbalance for the letter grade classification limits the performance of the model as some classes only have limited data in comparison to others. From the results gathered, it may be sufficient to say that the social presence in forum discussion logs contributes as a predictor for final grade only to a limited extent of accuracy, but more data is needed for it to be counted as a reliable result.

5 CONCLUSION

Based on the results, data analysis, and model evaluation, the researchers reached three conclusions.

First, all indicators of social presence for the open communication category are present among the forum discussion logs. Manual classification of the messages showed that all the indicators of social presence can be found among the forum discussion logs.

Subsequently, *continuing a thread* significantly affects the students' final grade. Analysis on both the 7-class and 2-class features consistently showed that *continuing a thread* has a strong positive correlation with the students' final grade. A possible explanation for this is that majority of the messages are classified to this indicator. It could be that messages that fall into this category are mostly responses to questions posted by instructors and are therefore graded.

Finally, indicators of social presence can predict whether the student will merit a final grade that is above average or below average to a limited extent. Predicting the university-standard letter grades resulted in an upsetting prediction performance. However, predicting whether the student will merit a grade that is high (above the median) or low (below the median) resulted in a much better performance.

In furthering the study conducted, the researchers recommend four avenues to explore:

5.1 Wider Scope

To provide a better model with increased performance, a larger dataset is essential to provide a more reliable result. Due to the manual classification nature of preprocessing the forum post type, the researchers were not able to work on a much larger scale to work around the time constraint. The researchers recommend exploring other programs and colleges to see if similar results can be obtained.

The profile of the data used in this study has shown the large number of messages classified as *continuing a thread* which may have led to its correlation with student performance. The addition of more data may be able to balance the representation of different indicators with a more generalized audience. In turn, working with a bigger dataset can include more instances that may balance the distribution letter grades and may improve the model.

5.2 Enhanced Preprocessing

Machine Learning offers a wide variety of tools and techniques. To come up with different results, the researchers recommend exploring other methods particularly in the preprocessing stage such as incorporating natural language processing (NLP). The results of the study focused more on the statistical metrics of the forum messages. Use of NLP may be used to explore the semantics of forum messages and capture patterns not presented in this study.

Future iterations of this study may also consider other factors or attributes (assignments, exams, time between posts, etc.) to a certain extent with a focus on forum logs, and other categorization methods.

5.3 Exploring Methods

Using different algorithms may affect the results of the study. The researchers recommend exploring other algorithms, especially more advanced algorithms, to see if the results are better. This may include not only classification but regression and clustering models as well. A combination of different algorithms is also worth exploring. Specifically, the researchers recommend the exploration of random trees and support vector machines (SVM) classifiers since both are also widely used in related studies.

5.4 Include Other Categories of Social Presence

This study limited the use of the indicators of social presence to the indicators under the open communication category. Use of the indicators under the categories interpersonal communication and cohesive communication may reveal more insight into the context of the messages posted by students. The researchers recommend including both categories in future studies to gain more understanding of the dataset.

ACKNOWLEDGMENTS

This research was supported by the Xavier University – Ateneo de Cagayan College of Computer Studies and the Kinaadman University Research Office. With the guidance of the research adviser, Mr. Jun Rangie Obispo, along with the research panelists, Engr. Gerardo S. Doroja and Ms. Shayryl Mae R. Sabal, and the special help of EDM and learning analytics expert, Dr. May Talandron-Felipe, this would not have been possible.

REFERENCES

[1] About Moodle - MoodleDocs: *https://docs.moodle.org/311/en/About_Moodle*. Accessed: 2021-07-19.

[2] Ademi, N., Loshkovska, S. and Kalajdziski, S. 2019. Prediction of Student Success Through Analysis of Moodle Logs: Case Study. *ICT Innovations 2019: Big Data Processing and Mining*. Springer Nature Switzerland AG. 27–40.

[3] Akyol, Z. and Garrison, D.R. 2008. The Development of a Community of Inquiry over Time in an Online Course: Understanding the Progression and Integration of Social, Cognitive and Teaching Presence. *Journal of Asynchronous Learning Networks*. 12, (Dec. 2008), 3–22.

[4] Anderson, T., Garrison, D.R. and Archer, W. 1999. Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*. 2, 2 (Mar. 1999), 87–105. DOI:https://doi.org/10.1016/S1096-7516(00)00016-6.

[5] Arbaugh, J.B. 2008. Does the Community of Inquiry Framework Predict Outcomes in Online MBA Courses? *International Review of Research in Open and Distributed Learning*. 9, 2 (2008), 1–21. DOI:https://doi.org/10.19173/irrodl.v9i2.490.

[6] Bhatia, S., Biyani, P. and Mitra, P. 2012. Classifying User Messages For Managing Web Forum Data. *WebDB* (2012).

[7] Feng, Y., Chen, D., Zhao, Z., Chen, H. and Xi, P. 2015. The Impact of Students And TAs' Participation on Students' Academic Performance in MOOC. *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (New York, NY, USA, Aug. 2015), 1149–1154.

[8] Garrison, D.R. 2011. *E-Learning in the 21st Century: A Framework for Research and Practice*. Routledge.

[9] Joksimović, S., Gašević, D., Kovanović, V., Riecke, B.E. and Hatala, M. 2015. Social presence in online discussions as a process predictor

of academic performance. *Journal of Computer Assisted Learning*. 31, 6 (2015), 638–654. DOI:https://doi.org/10.1111/jcal.12107.

[10] Kear, K. Social presence in online learning communities. 9.

[11] Liu, S.Y., Gomez, J. and Yen, C.-J. 2009. Community College Online Course Retention and Final Grade: Predictability of Social Presence. (2009), 19.

[12] López, M.I., Luna, J.M., Romero, C. and Ventura, S. 2012. Classification via clustering for predicting final marks based on student participation in forums. (Jan. 2012), 4.

[13] Richardson, J. and Lowenthal, P. 2017. Instructor Social Presence: Learners' Needs and a Neglected Component of the Community of Inquiry Framework. *Social Presence in Online Learning: Multiple Perspectives on Practice and Research*. (Jan. 2017).

[14] Richardson, J. and Swan, K. 2003. An Examination of Social Presence in Online Courses in Relation to Students' Perceived Learning and Satisfaction. *Journal of Asynchronous Learning Network*. 7, 1 (Feb. 2003).

[15] Romero, C. and Ventura, S. 2010. Educational Data Mining: A Review of the State of the Art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 40, 6 (Nov. 2010), 601–618. DOI:https://doi.org/10.1109/TSMCC.2010.2053532.

[16] Shea, P., Hayes, S., Vickers, J., Gozza-Cohen, M., Uzuner, S., Mehta, R., Valchova, A. and Rangan, P. 2010. A re-examination of the community of inquiry framework: Social network and content analysis. *The Internet and Higher Education*. 13, 1 (Jan. 2010), 10–21. DOI:https://doi.org/10.1016/j.iheduc.2009.11.002.

[17] Taghizadeh, M. and Vaezi, S. 2011. Exploring Social Presence in Virtual Learning Environment. *World Applied Sciences Journal*. 15, (Jan. 2011), 120–128.

[18] The Evolution of Distance Education in 2020: 2020. *https://educationonline.ku.edu/community/distance-education-evolution-in-2020*. Accessed: 2021-06-24.

[19] The Rise Of Online Learning: 2020. *https://www.forbes.com/sites/ilkerkoksal/2020/05/02/the-rise-of-online-learning/*. Accessed: 2021-06-24.

[20] What is a Learning Management System: 2017. *https://www.cmswire.com/digital-workplace/what-is-a-learning-management-system/*. Accessed: 2021-06-24.

[21] Xing, W., Tang, H. and Pei, B. 2019. Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. *The Internet and Higher Education*. 43, (Oct. 2019), 100690. DOI:https://doi.org/10.1016/j.iheduc.2019.100690.

[22] Zaini, S.H. and Ayub, A.F.M. Exploring Social Presence in Online Forums among Distance Learners. 16.

[23] Zou, W., Hu, X., Pan, Z., Li, C., Cai, Y. and Liu, M. 2021. Exploring the relationship between social presence and learners' prestige in MOOC discussion forums using automated content analysis and social network analysis. *Computers in Human Behavior*. 115, (Feb. 2021), 106582. DOI:https://doi.org/10.1016/j.chb.2020.106582.