Sentiment Analysis of Student Feedback in Online Course

Siti Aisyah Binti Abd Razak
Facuty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, Malaysia
wih190022@siswa.um.edu.my

Keywords: Sentiment analysis, Sentiment classification, Online course, Student feedback, MOOCs

Abstract: Student feedback and reviews are vital information that can provide constructive opinion and useful insights for the improvement of learning and teaching strategy in class. Instructors manually analysing and going through countless feedbacks to get useful insights is tedious and time consuming. Sentiment analysis, a field of natural language processing supplies the mean to analyse, extract and quantify subjective information from textual data. Therefore, predictive models in the application of sentiment analysis would be able to replace the tedious task and solve the problem entirely. In this paper, we are implementing methods for sentiment analysis of student feedback using supervised learning methods, ensemble learning methods and transfer learning. The model used in the respective methods aim to predict state positive, negative, or neutral from feedback in an online course. Comparison between all the machine learning methods were conducted to find the most accurate model with good performance. The evaluation results show that DistilBert in transfer learning give the best result in predicting the sentiment of the student feedback with 91.84 % F-measure score.

1 INTRODUCTION

Over the last decade, online learning has grown significantly as the internet and education have merged to provide people with the ability to learn new skills. This technology has enabled learners with the flexibility, and convenience of learning at any time and from any location in the world (Castle & McGuire, 2010). Popularity of online courses have gained its attention among people recently particularly during the COVID-19 pandemic (Impey & Formanek, 2021). MOOCs platform like Coursera and Udemy supplies thousands of courses in all kinds of topics available for any interested learners. These platforms provided a section in which students are given an option to give reviews and feedbacks about their learning experience on the online course. This piece of information is important as it reflects the effectiveness of the teaching method of the instructor and aid in the improvement of the quality and content of the course (Leckey & Neill, 2001).

Feedbacks left by learners contains sentiment information that can be a useful tool for acquiring valuable insight into the teaching-learning process (Crews et al., 2017). However, due to the enormous number of students enrolled in a MOOCs course, obtaining a thorough overview and general sentiment from the feedback data may be a time-consuming effort. That being the case, the field of sentiment analysis in natural language processing could be a useful tool in extracting user's sentiment. Sentiment analysis is a process of extracting information and identify subjective information in a textual content (Medhat et al, 2014). Applying the technique of sentiment analysis would solve the tedious task of reviewing feedbacks manually and help provide a summary and overview of the student's experience in the course.

Sentiment Analysis has been widely used in various application areas such as consumer products and services, financial markets and brand reputation monitor (Feldman, 2013). This method is widely used in many areas as it has proven useful to provide general opinion on the topic of interests. In the education domain, natural language processing has been applied to various applications such predicting student mental state, give comments on writing, question and answering for education as well as chatbots for answering class related questions.

Therefore, the paper focuses on proposing on approach in predicting the sentiment polarity of textual review to positive, negative and neutral using machine learning techniques. By doing so, instructors can get an overview of the overall sentiment and opinion people have on the online course. Supervised learning methods such as naïve bayes and SVM using

TF-IDF vectorizer as well as random forest and Adaboost for ensemble learning methods and transfer learning using DistilBert are used to implement this approach. The best model will be used as the final model to predict sentiment in the student feedbacks.

The rest of the paper starts by discussing the related work, data used, methodology used in order to complete the project, the analysis and modelling process and proceeds to the result and discussion as well as the conclusion.

2 RELATED WORKS

In the field of natural language processing, many studies have been conducted on sentiment analysis for student feedback in the past few years. The reported work in previous research can be classified into three main approaches: (a) machine learning based (b) lexicon based and (c) hybrid

2.1 Machine Learning Based

Approach in machine learning for sentiment analysis learn a predictive model to predict sentiment using a training set and test it on the test dataset. There are two types of machine learning methods: supervised and unsupervised learning. In supervised learning, sentiment labelled dataset are required while unsupervised learning do not. (ONAN, 2020) used ensemble learning, supervised learning and deep learning approach to achieve high predictive performance on sentiment polarity prediction. Deep learning approach outperformed other methods with the use of word embeddings.

(Kastrati, Imran, et al., 2020) used a weakly supervised framework for aspect-based sentiment analysis on the student feedback. The paper can effectively identify aspect categories and sentiment in unlabeled students' reviews using weak supervision signal. Their approach helped reduce the need in having huge human annotated data for aspect sentiment analysis while providing good results.

Authors in (Wang et al., 2021) used ALBERT-BiLSTM model in deep learning which could generate word vectors in dynamic semantics which help in the problem of polysemous word and heighten significance of familiar words in the predicting sentiment of the reviews. (Dsouza et al., 2019) done a comparative analysis between Multinomial Naïve Bayes, Random Forest and Support Vector Machine performance on sentiment analysis of student feedback using data collected in a survey they conducted. Study conducted by Kandhro et al. (2019) using Long Short Term Memory Model and word embedding as an input to map the words in the reviews shown promising result as it capable to collect semantic and syntactic information that is significant to make prediction and overcome the flaws of traditional methods. (Moreno-Marcos et al., 2018) used lexicon approach and machine learning approach to compare the performance of both methods. Findings from the study suggest that supervised algorithms performed best with Random Forest as the best algorithms beating SentiWordNet and the lexicon created by the author. A study conducted by (Kastrati, et al., 2020) analysed the aspect-based sentiment from manually annotated student's feedback from Coursera. The study uses machine learning to extract aspect categories in a feedback and find out the sentiment of the aspect. (Sangeetha & Prabha, 2020) proposed a method that used word embedding and multi-head attention layer in LSTM using different dropout rates to predict the sentiment. The method helps in focusing on more attention to the influence word on the emotion and shown improve result in the classification task.

2.2 Lexicon Based

Sentiment Analysis using lexicon-based approach used sentiment lexicon to evaluate the sentiment of a textual content. Lexicon or can be known as dictionary is a collection of words that have sentiment polarity linked to them. The lexicon can be generated in manual as specified or automatically. (Aung & Myo, 2017) created a database of English words as the lexical reference. Words containing intensifier extracted from the text are also analyzed to make prediction on the opinion of the feedback. In a study conducted by Rajput et al. (2016) MPQA corpus is used as the sentiment dictionary to tag each words in the feedback with its polarity. Sentiment score are computed in each feedback and thus determine its polarity.

2.2 Hybrid

Hybrid approach in sentiment analysis make used both machine learning and lexicon-based methods to make prediction on the sentiment. (Nasim et al., 2017) used both lexicon-based features and TF-IDF feature vector to analyze the sentiment in the feedbacks. Comparative analyses are done by the author and the proposed methods are found to be the best.

3 DATA USED

The dataset used in this paper has 21,940 reviews gathered from 15 courses on Coursera, the leading Massive Open Online Course platform (Kastrati et al., 2020). All of the reviews are in English language with an average of 25 words per review and a range of 1 to 554 words per reviews. The dataset was manually annotated with the sentiment label of positive, negative, and neutral. Table 1 shows few examples of the reviews and its label.

Table 1: Sample Dataset

No	Student Feedback	Sentiment Label
1	not making enough effort to help students make sense of the convoluted implementations that they had used.	Negative
2	Universally applicable, presented in an entertaining manner, and quite thorough. even experienced writers will pick up on one or two aspects they usually overlook. highly recommended!	Positive
3	some irritating errors, but overall, it was okay.	Neutral

4 METHODOLOGY

This section explains the steps taken to complete the task for sentiment analysis of student feedback in online course.

4.1 Business Understanding

The phase of business understanding concerns with understanding the project's objectives and requirements. This involves determine whether resource is available and determine the tools that will be used to complete the project. Gaining a thorough understanding of the issue that needs to be address is critical (Saltz, 2021).

4.2 Data Collection

This phase focused on identifying and finding the data required to execute the project. The paper uses data from a research conducted by (Kastrati, et al., 2020) for aspect based and sentiment analysis of student feedback. The data contains 10,914 reviews with its sentiment label collected from Coursera, an online open learning platform. The dataset is highly skewed towards the positive reviews, with low example in negative and neutral reviews. Table 2 shows the distribution of the sentiment polarity label for the dataset.

Table 2: Distribution of Label

Sentiment	Frequency	Percentage (%)
Positive	18478	84.22
Negative	2317	10.56
Neutral	1145	5.21

4.3 Preprocessing

Preprocessing step is a crucial technique used to clean noises and unwanted elements that could disrupt our machine learning model from performing well. Preprocessing techniques done to textual data have huge influence in setting up the analysis to success (Anandarajan et al., 2018). Student feedback data are raw text that is considered to be unstructured data. It may have punctuations, spelling errors, grammatical mistakes, and URLs thus removing these unimportant element is a crucial step before proceeding to modelling. A few of the preprocessing steps that was done to the data using python NLTK library and regex.

1. removing html tags, links, punctuations, symbols, numbers and lower casing words.

2. Removing stop word: They are unnecessary and unimportant words which does not supply any significant value in the feature vector

3. Lemmatization: Shorten the words to its root form.

4.4 Feature Extraction

Student feedback are raw textual data which cannot work directly with most machine learning algorithms. To provide output for the test data, Machine Learning algorithms learn from a pre-defined collection of features from the training data. Therefore, to transform the text into format than can work with machine learning algorithms, they need to be transformed into a matrix of feature vectors, which require the use feature extraction techniques.

TF-IDF is used to convert the textual feedback by computing a measure that evaluate how related a word to a document in a collection of documents. The value in the vector works by increase in proportion to the number of times the word appears in a document but is counter by the number of the occurred words in the collection of the documents. TF-IDF vectorizer will be used in the traditional machine learning method while DistilBERT tokenizer is used for feature extraction for DistilBERT model.

DistilBERT tokenizer used word piece tokenization in which it runs an end-to-end tokenization, punctuation splitting and word piece on the input text data, generate input_ids and attention masks to be fed into the fine-tuning model. The maximum length for the input token for the tokenizer is set at 512 as the maximum word in a review is at 554 and it is the maximum length the tokenizer can be set. All the sentences are also padded and truncate to ensure that they have equal length.

4.5 Data Analysis/Modelling

4.5.1 Supervised Learning Methods

Support Vector Machine is a supervised learning algorithm that can both be used in regression and classification task in machine learning. It works by finding the optimum decision boundary between vectors that belong to one group and vectors that don’t. (Vapnik, 2000)

Naïve Bayes is a classification method in supervised machine learning that is based on the Bayes’ Theorem and the premise of predictor independence. The algorithm believes that the existence of one feature in a class is unrelated to the presence of any other feature

4.5.2 Ensemble Learning Methods

Adaboost is a boosting ensemble learning algorithm that is prevalent in classification task. It works well by combining a few weak learners into strong learners in helping it make better prediction.

Random Forest is an algorithm that utilizes ensembling technique by combining many decision trees. The algorithm is trained using bagging and bootstrapping aggregating. Predictions outcomes are made using the algorithm by taking the average from all the outputs from the combination of the decision tree.

Catboost is a fast scalable highly performance gradient boosting algorithm that can work well in most regression and classification problem.

4.5.3 Transfer Learning

Transfer Learning is a machine learning technique in which a model that is created for one job is utilized as the base for a model on a different task. In this project, DistilBERT is used as the pretrained model.

DistilBERT is a pretrained language model that used the method of knowledge distillation to approximate Google’s BERT language model particularly the BERT base uncased (Sanh et al., 2019). The model was a scaled down version of BERT with 40% fewer parameters and runs 60% faster while maintaining 95% performance of BERT’s model. The model is pretrained with large corpus of unlabelled text making it good in performance for a wide range of NLP tasks.

4.5 Model Evaluation

Making evaluation to the trained model is a necessary step in order to find out if the model make good predictions on unseen data. Due to the nature of the dataset which is highly imbalanced in each class of the output label. Accuracy is not used as the evaluation metric of the model. The evaluation is performed on the test dataset using python sklearn library for evaluation metric. Precision, recall and f-measure are used as the metric. The interpretation for the metric is as follows:

- True Positive (TP) is when the actual and predicted value are positive
- False Positive (FP) is when the actual value is negative, but the predicted value is positive
- True Negative (TN) is when the actual and predicted value are negative
- False Negative (FN) is when the actual value is positive, but the predicted value is negative

The following evaluation metric were used:

1) Precision

Precision can be defined as the ratio between correctly predicted positive input to the total predicted positive input. Higher number to 1 indicates that the model is great in making prediction for the positive class.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

2) Recall

Recall is an evaluation metric that count the number of positive class prediction that is made out of all the positive example in the dataset. This metric would show representation on how the model missed the positive predictions

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

3) F-measure

Making prediction on the sentiment of student feedback is a multi-class classification task. F- measure is a metric that is commonly used in this type of task. F- measure provides the geometric mean of precision and recall. Our dataset which is highly imbalance in nature would be a suitable metric to measure the performance.

$$F - measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

In this paper, weighted F-measure, precision and recall are used to compare the performance of this method with the result obtained from (Kastrati et al., 2020)

4.6 Deployment

The final phase of the project includes the deployment of the machine learning model into a web application. The best performing model will be deployed into a web application to predict customer input student reviews sentiment from the end user and contains a dashboard from the visualization formed from the data used in this project. The deployed web application uses the framework from Dash and programming language Python.

5 ANALYSIS/MODELLING

This section contains the detailed process taken for the analysis and modelling approach. Two approach are taken in which one is used for the traditional method involving traditional machine learning algorithms while the other is used for the analysis and modelling process for DistilBERT model. Both model used the same raw text data containing student's feedback from Coursera with its sentiment labels that is going to be fed to the model for training and evaluation purposes.

5.1 DistilBERT Model

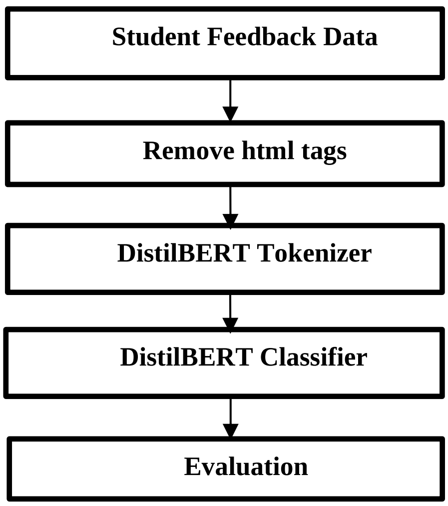

Figure.1: Diagram of Deep Learning Approach Distil Bert Model

Figure 1 shows the process taken for using DistilBERT approach. The dataset is split into 80% training set, 10% validation set and 10% test set. Before fed the data into the model, html tags in the reviews will be removed and preprocessing is done using Distil Bert tokenizer which will perform tokenization for punctuation and word piece and convert it into input ids and attention mask. Other cleaning steps are not done to the data for DistilBERT model as the DistilBERT tokenizer will perform the preprocessing step required for training the model. The produced outputs from the tokenizer are then translated to a TensorFlow dataset object to convert it into the right format before model is fine tuned.

Then, the model is fine-tuned using training set with 5 epoch and validation set will confirm the performance during training. Distil BERT model that is fine-tuned will then make prediction on the specified test dataset and results were received. Student feedbacks were then classified into positive, negative and neutral by the model. The performance was evaluated using precision, recall and f-measure.

5.1 Traditional Machine Learning Method

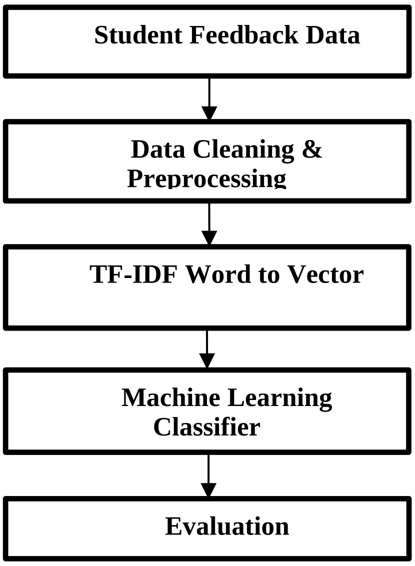

Figure.2: Diagram of Traditional Machine Learning Approach

Figure 2 above shows the process done for the modelling approach in traditional model. Student feedback data which has unimportant elements such as punctuation, html links, stop words and noises is cleaned and preprocessed. Cleaned data is then undergo feature extraction that will convert it into a matrix of feature vector that can be fed directly to machine learning algorithms. TF-IDF vectorizer in sklearn library is used to convert the raw text format into vectors. After TF-IDF the training data will train naïve bayes, support vector machine, CatBoost and random forest to make prediction on the sentiment label of the student’s feedback.

6 RESULTS & DISCUSSION

The following section will discuss the results obtain from the evaluation of the machine learning algorithms used. Precision, recall and F-measure is used as the evaluation metrics to show indication on how well our model make predictions on unseen data for the sentiment of the feedbacks.

Table 3: Performance of Techniques

Technique	Precision (%)	Recall (%)	F-measure (%)
DistilBERT	92.03	91.70	91.84
SVM	84.64	89.70	86.91
Naïve Bayes	81.36	84.37	77.82
Random Forest	70.03	83.68	76.25
Adaboost	82.02	87.15	83.52
CatBoost	70.03	83.68	76.25

Table 3 shows the overall result obtained from all the methods used in this project. Having number closer to 1 would show an increase in performance of the model using the metric of precision, recall and f-measure. From the table it can be seen that the best performing model is DistilBert followed by SVM, Adaboost, Naïve Bayes, Random Forest and Catboost. Highest F-measure by Distilbert achieved 4.93% difference with the second highest performer which is Support Vector Machine. Based on the result obtained from the best model, the model shows better result than the best result obtained by (Kastrati et al., 2020). The comparison of the results is shown in Table 4. Our best model achieved 3.17% improvement from the previous study.

Table 4: Comparison of Result

Technique	Precision (%)	Recall (%)	F-measure (%)
DistilBERT	92.03	91.70	91.84
Decision Tree using TF-IDF	88.72	88.61	88.67

4 CONCLUSIONS

In this project, we have implemented a few machine learning methods in supervised learning, ensemble learning and transfer learning to predict the sentiment in the student feedback. The result shown that the best model, DistilBERT has performed quite well obtaining the best score of 91.84%. However, given the nature of the dataset in which it is highly skewed to positive label and the obtained result which is not much of a different from the other techniques, further efforts need to be done to obtain balanced dataset for the sentiment label and further investigate the topic and improve more on the result.

In future work, aspect-based sentiment analysis can be done for student reviews so that key aspects related to the reviews can be identified while identifying the overall sentiment of the topic. Dataset containing textual feedback of student which is balanced in all classes of the polarity label should be used in future works so that better model that can predict each sentiment label can be trained.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Kasturi Dewi A/P Varathan for guiding and helping me complete this project.

REFERENCES

Anandarajan, M., Hill, C., & Nolan, T. (2018). Text Preprocessing. Practical Text Analytics, 45–59. https://doi.org/10.1007/978-3-319-95663-3_4

Aung, K. Z., & Myo, N. N. (2017). Sentiment analysis of students' comment using lexicon-based approach. 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). https://doi.org/10.1109/icis.2017.7959985

Castle, S. R., & McGuire, C. (2010). An Analysis of Student Self-Assessment of Online, Blended, and Face-to-Face Learning Environments: Implications for Sustainable Education Delivery. International Education Studies, 3(3). https://doi.org/10.5539/ies.v3n3p36

Crews, T., Bordonada, T., & Wilkinson, K. (2017, June 5). *Student Feedback on Quality Matters Standards for Online Course Design*. EDUCAUSE. https://er.educause.edu/articles/2017/6/student-feedback-on-quality-matters-standards-for-online-course-design

Dsouza, D. D., Deepika, Nayak, D. P., Machado, E. J., & Adesh, N. D. (2019). Sentimental analysis of student feedback using machine learning techniques. International Journal of Recent Technology and Engineering, 8(1 Special Issue 4), 986-991.

Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82–89. https://doi.org/10.1145/2436256.2436274

Kastrati, Z., Imran, A. S., & Kurti, A. (2020). Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs. IEEE Access, 8, 106799–106810. https://doi.org/10.1109/access.2020.3000739

Kastrati, Z., Arifaj, B., Lubishtani, A., Gashi, F., & Nishliu, E. (2020). Aspect-Based Opinion Mining of Students' Reviews on Online Courses. Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence. https://doi.org/10.1145/3404555.3404633

Kandhro, I. A., Wasi, S., Kumar, K., Rind, M., & Ameen, M. (2019). Sentiment Analysis of Student′s Comment by using Long-Short Term Model. Indian Journal of Science and Technology, 12(8), 1–16. https://doi.org/10.17485/ijst/2019/v12i8/141741

Leckey, J., & Neill, N. (2001). Quantifying Quality: The importance of student feedback. Quality in Higher Education, 7(1), 19–32. https://doi.org/10.1080/13538320120045058

Lundqvist, K., Liyanagunawardena, T., & Starkey, L. (2020). Evaluation of Student Feedback Within a MOOC Using Sentiment Analysis and Target Groups. The International Review of Research in Open and Distributed Learning, 21(3). https://doi.org/10.19173/irrodl.v21i3.4783

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113. https://doi.org/10.1016/j.asej.2014.04.01

Moreno-Marcos, P. M., Alario-Hoyos, C., Munoz-Merino, P. J., Estevez-Ayres, I., & Kloos, C. D. (2018b). Sentiment analysis in MOOCs: A case study. 2018 IEEE Global Engineering Education Conference (EDUCON).

https://doi.org/10.1109/educon.2018.8363409
Nasim, Z., Rajput, Q., & Haider, S. (2017). Sentiment analysis of student feedback using machine learning and lexicon-based approaches. 2017 International Conference on Research and Innovation in Information Systems (ICRIIS). https://doi.org/10.1109/icriis.2017.8002475
ONAN, A. (2020). Sentiment analysis on massive open online course evaluations: A text mining and deep learning approach. Computer Applications in Engineering Education, 29(3), 572–589. https://doi.org/10.1002/cae.22253
Rajput, Q., Haider, S., & Ghani, S. (2016). Lexicon-Based Sentiment Analysis of Teachers' Evaluation. Applied Computational Intelligence and Soft Computing, 2016, 1–12. https://doi.org/10.1155/2016/2385429
Saltz, J. S. (2021). CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps. 2021 IEEE International Conference on Big Data (Big Data). https://doi.org/10.1109/bigdata52589.2021.9671634
Sangeetha, K., & Prabha, D. (2020). Sentiment analysis of student feedback using multi-head attention fusion model of word and context embedding for LSTM. *Journal of Ambient Intelligence and Humanized Computing*, *12*(3), 4117–4126. https://doi.org/10.1007/s12652-020-01791-9
Vapnik, V. N. (2000). The Nature of Statistical Learning Theory. Statistics for Engineering and Information Science. https://doi.org/10.1007/978-1-4757-3264-1
Wang, C., Huang, S., & Zhou, Y. (2021b). Sentiment analysis of MOOC reviews via ALBERT-BiLSTM model. MATEC Web of Conferences, 336, 05008. https://doi.org/10.1051/matecconf/202133605008
.