Journal of Theoretical and Applied Information Technology
31st March 2022. Vol.100. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ANALYSIS STUDY TO DETECT STUDENT LEARNING PROBLEMS IN ONLINE LEARNING USING TEXT MINING

[1]YEFTA CHRISTIAN
[1] Department of Information System, Universitas Internasional Batam, Indonesia
E-mail: [1]yefta@uib.ac.id

ABSTRACT

How big the level of comfort and level of satisfaction of students in participating in learning needs to be known by the organizers of learning to improve the quality of learning and increase student interest in learning. One of them is by providing a facility in the form of an Academic Guidance System that students can use to express their complaints and opinions regarding the learning activities carried out. In this study, the author will classify with a text mining algorithm on any complaints or opinions given by students. The method used in this research is Lexixon Based Sentiment Analysis using two tools, namely Flair NLP Framework Application and KNIME Application. These two tools can classify every complaint or opinion into two categories, namely positive and negative. Positive can mean that there are no problems related to the lectures being carried out while negative means the opposite. The results of the application of the text mining algorithm on the KNIME Workflow that were applied in this study succeeded in mapping positive and negative sentiments from student learning, with an accuracy of 70%, still less than the desired standard of 85%.

Keywords: *Students, Learning Problems, Online Learning, Text Mining, Lexicon Based*

1. INTRODUCTION

Buildings, classrooms, air conditioners, computers, sports facilities, and laboratories are facilities for students to support their learning activities, but for now online learning is one of the most widely used learning tools. The COVID-19 pandemic is the main cause of online learning activities, almost all learning institutions conduct online learning. Based on the results of the study, online learning affects students' interest in learning [1]. The better and more interesting the learning media provided in online learning activities, the higher the student's interest in learning. Interest in learning is closely related to the success of students in pursuing education, the higher the interest in learning, the better the learning outcomes obtained [2].

Online learning on the one hand can keep interest in learning because of the ease of access, more flexible time and no need for face-to-face meetings, but on the other hand many obstacles arise compared to offline learning activities. The material is not clear, it is difficult to do practical learning, it is difficult to contact lecturers and there are still quite a lot of obstacles caused by online learning activities. It is undeniable that these obstacles affect and reduce student interest in learning.

At University X, students are given the means to tell the obstacles they face when doing online learning through the Academic Guidance System. Through this system, students can inform various things related to lecture constraints, including those related to learning, to academic supervisors.

In the Academic Guidance System students can convey the obstacles faced in the form of text, then the lecturer will read and provide responses according to the obstacles faced by the student, but lecturers often have difficulty distinguishing students who really face obstacles and require further handling. Lecturers often assume that some students do not experience problems and do not require further treatment, even though these students need better solutions related to the obstacles they face. Lecturers usually have difficulty because there are quite a lot of students who need to be guided and only get a description of the constraints in the form of text.

Based on these constraints, it is necessary to have an automation system that can help lecturers to distinguish which students really experience problems in learning and which students do not. Data in the form of text can be processed and mapped using text mining. In this study, these texts will be classified and given a level to distinguish which texts

are closer in meaning and reflect an obstacle in the education process at one of the campuses in Indonesia. In Indonesia, such research is still rarely done and it is quite difficult to find Indonesian-language references for research that focuses on the use of text mining to improve the learning process.

2. LITERATURE REVIEWS

Research conducted by Dwi Smaradahana and Budi Sentosa focuses on data obtained from Shopee Indonesia's twitter [3]. The data obtained from Twitter is taken using certain functions and then processed according to the text mining stages of Feldman and Sanger and the K-Means clustering algorithm. The result of this research is the grouping of each word that appears most often along with its type into 28 clusters. This research is only able to process 140 characters due to the limited number of characters on Twitter.

Vladimer B. Kobayashi, Stefan T. Mol, Hannah A. Berkers, Gábor Kismihók and Deanne N. Den Hartog applied text mining to job vacancies [4]. The researchers first divided the stages of text mining into three stages, namely text pre-processing, application of text mining and post-processing. Each stage is carried out with the aim of producing an analysis of the job details of each available job vacancy. After the analysis stage is carried out, classification is carried out according to the type of work available. It can be concluded from the research conducted by the researcher that text mining is very well used to analyze the types of jobs available in job vacancies. Even so, there is a need for research on the application of text mining in other types of organizations.

Research conducted by Anu Priya, Shruti Garg and Neha Prerna Tigga focuses on applying machine learning to detect a person's mental health [5]. There are three elements of mental health measured in this study, namely level of consciousness, level of depression and level of stress. The three elements are grouped back into 5 levels. After the results are obtained and fit to be applied, the initial measurement results are applied to several machine learning classification algorithms. The results of this study state that the Naïve Baiyes algorithm is the most suitable algorithm for use in this form of data.

Kang Woo Lee, Da Young Lee and Hyun Ju Hong applied text mining to analyze teachers' reports regarding the suicide rate of their students [6]. Text mining is used to find links between words that often appear verbally and in writing from students who commit suicide. The result is that 30% of students who have attempted suicide have said or written words that also emerged from the results of this study. However, this research is only applied to 1 or 2 words and has not been applied to related words or to more complex sentences.

Research conducted by Meisy Fortunatus, Patricia Anthony and Stuart Charters combines several text mining features to detect cyberbullying on social media [7]. There are 5 things that are measured in this study, namely emoji sentiment score, emoticon sentiment score, text aggregation score, positive word score and sentiment score. These five things are measured and classified with a classification algorithm. This study proves that the classification algorithm used can detect cyberbullying well, but in this study the stages of text mining have not been carried out in the form of interrelated words and in the form of sentences.

From the results of the studies above, the researcher tries to continue text mining research at the next stage. This study attempts to overcome the obstacles in previous studies such as processing constraints on the limited amount of text, not detecting the link between words in a sentence and also comparing tools that are considered to be more appropriate for doing text mining. Researchers will apply text mining to related words or to more complex sentences according to the stages in previous research. In more depth, text mining will be applied to the constraint sentences written in the UIB Academic Guidance System.

2.1 Learning Problem

According to the Big Indonesian Dictionary (KBBI) the meaning of the word problem is an actor or condition that limits, hinders, or prevents the achievement of targets; the force that forces the cancellation of the implementation [8], while the meaning of the word learning is the process, method, act of making learning [9]. Based on what is stated in the KBBI, it can be interpreted that learning problems are conditions that limit someone when learning something. Research conducted by Hasmiana Hasan concluded that learning problems often arise from several factors including the lack of educators mastering the material, the difficulty of educators in applying various teaching methods and the inaccuracy of educators in using teaching aids [10].

2.2 Online Learning

Online learning is an activity that is carried out remotely using internet media. Online learning is very effective, especially during the Covid-19 pandemic. Lecturers can keep face-to-face with students using various available online learning tools. Online learning is effective in preventing the spread of the Corona virus because there is no direct interaction between students and lecturers [11].

2.3 Text Mining

To process and retrieve information from data, certain techniques are needed which are often referred to as data mining. Specifically in data mining, processing, and retrieval of information in the form of text is referred to as text mining [3]. Text mining is used to group certain words and sentences so that they can be classified based on certain classifications.

2.4 Sentiment Analysis

Sentiment analysis is a technique that focuses on opinion mining in the form of identification and classification of unstructured sentence data or words such as comments or product reviews, sentiment analysis is widely used in cases of brand review assessments, political campaigns, marketing analysis, or criticism and suggestions. from customers [12].

Sentiment analysis can also be described as a computational study that is useful for assessing attitudes, judgments and opinions of people, problems, topics, events, and products with their attributes [13].

2.5 Lexicon Based Sentiment Analysis

Lexicon is a dictionary of words that have been classified into positive words and negative words that are used to determine the polarity of sentiment based on the tendency of messages from the content of a collection of words, the sentence will be assessed based on the number of positive words and the number of negative words, so that it can be judged that the sentence is a sentence. negative or positive sentences [14].

2.6 Flair NLP Framework

The Flair Framework is an NLP (Natural Language Processing) framework designed to facilitate the training and distribution of sophisticated sequence labeling, text classification, and language models, the main purpose of the Flair Framework is to abstract from the specific engineering challenges posed by different types of word embedding [15].

The purpose of the Sentiment Analysis Framework is to estimate and generalize mapping functions so that when new input data is presented, the model can accurately predict the output variables for that data. In general, the features or attributes used to build the ML model are predefined by the user. Figure 1 shows the general framework for Sentiment Analysis using Supervised Machine Learning according to [12].

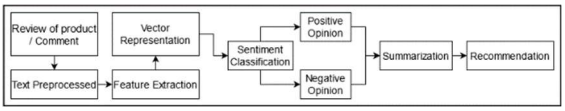


Figure 1: Sentiment Analysis Framework

2.7 KNIME

KNIME is a modular data analysis environment software developed by the University of Konstanz in Germany [16]. KNIME is widely used to perform data mining simulations. By using knime the user can model workflows, which consist of nodes that process data, transported via connections between those nodes. One of the functions that are mostly performed using KNIME is text mining. In this study, the text mining function in KNIME will be used to perform Lexicon Based Sentiment Analysis as a comparison of data from the Flair NLP Framework.

2.8 Related Work

Based on research [16] it was found that sentiment analysis can classify sentiment opinions into three classes, namely positive, neutral and negative. In this reseach, the researcher needs a larger lexicon scale to have more efficient result. Researcher [14] write about lexicon-based machine learning for sentiment analysis. In this research, the reseacher found accurate accuracy but the results for efficient results require a larger lexicon scale. Based on research [12] one of the prominent approaches to Sentiment Analysis is to use Supervised Machine Learning (SML), an algorithm that uses data sets with class labels determined based on learning mathematics from training data sets.

3. METHODOLOGY

This research is a qualitative research that will answer the research problem with the research

results obtained [17]. The research flow is represented in Figure 2.

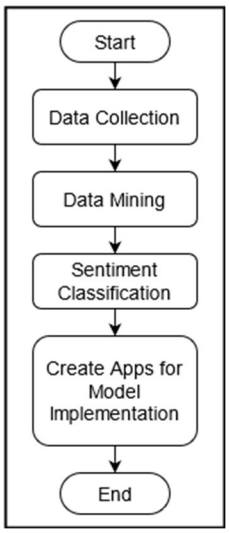


Figure 2: Research Flowchart

3.1 Data Collection

Data Collection is the first step in this research, data collection is represented in Figure 3. Academic guidance data is obtained from the input of student guidance through the Academic Guidance System in accordance with the mentoring periods that have been determined by the Academic Development Center Bureau.

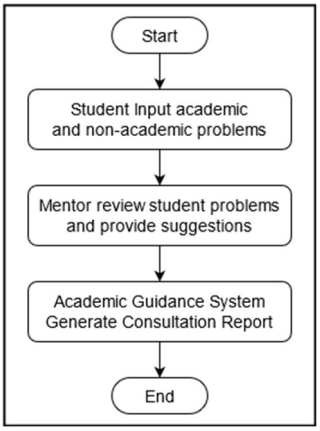


Figure 3: Data Collection Flowchart

3.2 Data Mining

Data mining stages refer to the Information Flow in Data Mining [18] as shown in Figure 4.

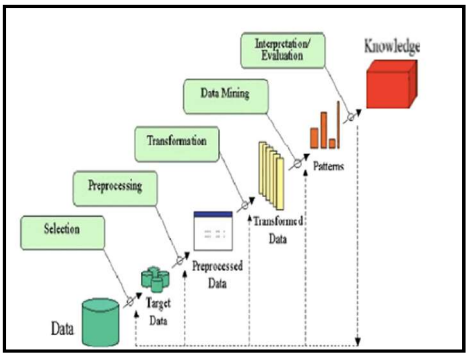


Figure 4: Information Flow of Data Mining

The data that has been collected will enter the data selection stage by sorting and selecting according to the desired initial data criteria. After the data is selected, the data will enter the data pre-processing stage by separating related sentences and words into several predetermined segments. The next stage is the data transformation stage, at this stage the words that have been divided into segments will be converted into mathematical form before they will be processed using a text mining algorithm.

After the data is converted into mathematical form, the existing data will enter the data mining stage (text mining) using an algorithm. The algorithm used will classify words into certain pre-determined classifications. In the last stage, the existing data will be evaluated and reprocessed so that the existing data can be understood by all parties who need the data. It is hoped that the existing data can map student constraints based on how severe the obstacles are.

3.3 Sentiment Classification

The Machine Learning method for Sentiment Analysis usually relies on supervised classification methods, where tagged/labeled data is used for the approach. The function of the feature extractor is to transfer the text input from the previous step to the feature vector, where a matrix of text tags is built and then the feature vector and tags (positive, negative, or neutral examples) are inserted into the Machine Learning code/algorithm that will generate the model.

At the prediction stage, the feature extractor work is to change the input text that is not patterned into a feature vector. This feature vector is then entered into the model which will produce a predicted tag that is positive, negative, or neutral. Figure 5 shows an overview of the Architecture of the Sentiment Classification process.

Journal of Theoretical and Applied Information Technology
31st March 2022. Vol.100. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

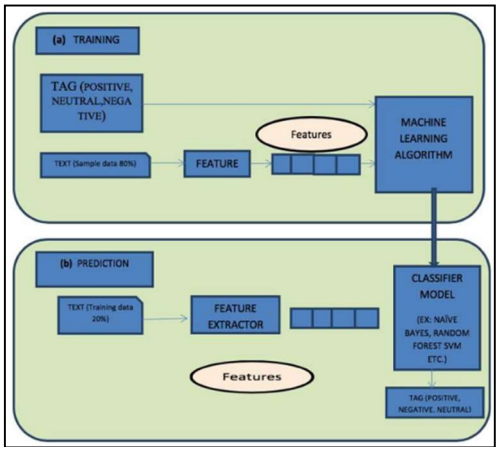


Figure 5: Sentiment Analysis

3.4 Sentiment Classification Application

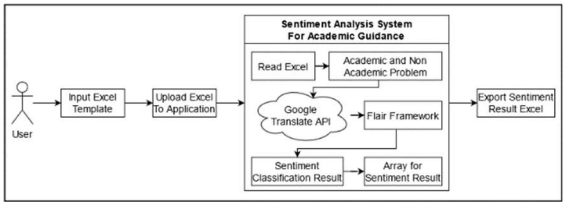


Figure 6: Sentiment Classification Application

The use of the application starts from the user upload stage of data in excel format that has been formatted according to a predetermined template, the uploaded excel will go through the process of retrieval of Academic and Non Academic Problem data, the data that has been obtained will be processed to translate using the Google Translate API, the results from the Google Translate API Sentiment Classification will be processed using the Flair Framework, the results of the Sentiment Classification will be stored in a storage array to be reprocessed so that the Output is in the form of Excel that has been filled with the results of the Sentiment Classification.

4. RESULT

Application development begins with the development of the most effective Sentiment Classification model to use, the following is an example of developing the Sentiment Classification model described in Figure 7.

```
from flair.models import TextClassifier
from flair.data import Sentence

translate_text = "I Am Healthy"

classifier = TextClassifier.load('en-sentiment')
sentence = Sentence(translate_text)

classifier.predict(sentence)

# print sentence with predicted labels
print('Sentence above is: ', sentence.labels)
```

```
2021-09-08 01:14:26,314 loading file C:\Users\IT\
```

Figure 7: Sentiment Classification Model Example

For adjustment to the Sentiment Classification model, a Google Translate API is needed to translate Academic and Non-Academic Problem students into English, following the use of the Google Translate API is depicted in Figure 8.

```
from google_trans_new import google_translator

translator = google_translator()
translate_text = translator.translate('Saya Sehat',lang_tgt='en')

print('Saya Sehat : ', translate_text)
```

```
Saya Sehat :  I am healthy
```

Figure 8: Google Translate API Example

The following is an example of an Excel template that is used as input in the application as shown in table 1.

Table 1: Excel Template Example

ID	Period	Date
2152	Semester Ganjil 2020 - PA 1	14/09/2020
2151	Semester Ganjil 2020 - PA 1	14/09/2020
2149	Semester Ganjil 2020 - PA 1	14/09/2020
2148	Semester Ganjil 2020 - PA 1	14/09/2020
2147	Semester Ganjil 2020 - PA 1	14/09/2020
2146	Semester Ganjil 2020 - PA 1	14/09/2020
2145	Semester Ganjil 2020 - PA 1	14/09/2020

Table 2: Excel Template Example

ID	Academic Problem	Non-Academic Problem

2152	Terkadang penyampaian dosen terputus karena jaringan, seperti layar freeze, tiba-tiba suara hilang dan lain sebagainya	Apakah tidak ada keringanan uang kuliah di masa pandemi ini?
2151	Saya ad masalah sedikit dalam proses pembelajaran kuliah dengan dosen yg mengajar mata kuliah Akuntansi Manajemen, sy merasa sangat tertekan dengan proses pembelajaran kuliah dengan beliau, dikarenakan pada proses kuliahnya beliau meminta mahasiswa membuka kamera selama 4 jam, bagi sy sebagai mahasiswa sangat tertekan dan berat, terutama karena sy tidak memiliki fasilitas wifi, Sy hny ingin memohon dosen dapat memperhatikan mahasiswa yg kurang mampu dalam masalah fasilitas dan ekonomi, lagipun fasilitas dan ekonomi mahasiswa jg pribadi punya, jadi tidak bisa selalu sesuaikan dgn keinginan dosen yg ad.Terimakasih.	Sy ingin tanya kapan nilai semester 2 sy bisa terinput semua, terimakasih.
2149	Tidak ada masalah.	Tidak ada masalah.
2148	Tidak ada masalah.	Kegiatan mahasiswa yang terbatas diakibatkan oleh online.
2147	Belum ada.	Belum ada.
2146	Terkait perkuliahan saya ada sedikit masalah, antara menyeimbangkan tugas tugas yg ada dgn kondisi fisik saya yg memang sedikit lemah, jadi terkadang saya sering down tiba tiba.	Tidak ada masalah.
2145	Untuk semester baru ini, belum ada masalah yang menghambat perkuliahan saya. Kalau ada pun, masalah jaringan internet yang membuat penjelasan dosen terbata-bata dan tidak jelas.	Masalah non-akademik bagi saya belum ada masalah, sejauh ini baik. Pengumpulan SA juga mudah kalau kita mengetahui prosedur pengumpulan. Layanan lms yang menurut saya agak susah ketika byk yg mengakses dan apalagi ketika di hari-h dosen ada kuis mendadak, lms loading sangat lama.

The following is the display of the Sentiment Classification Application when carrying out the Sentiment Classification process.

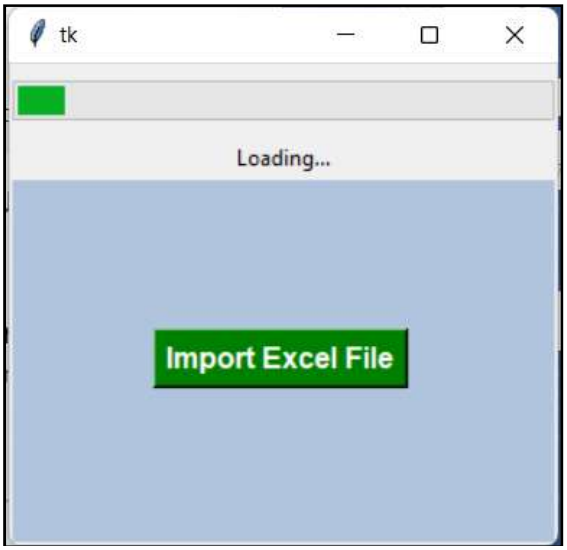


Figure 9: Sentiment Classification on progress

If the application has problems with the Sentiment Classification process, the display in Figure 10 will appear.

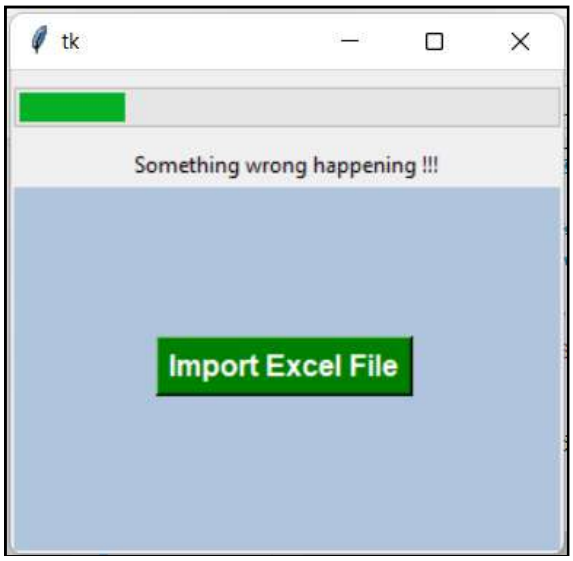


Figure 10: Sentiment Classification Error

Conditions that cause errors can be categorized into two, namely failure in the Sentiment Classification model and failure when carrying out the Translate process using the Google Translate API, where there is a limit for API calls to 1 IP Address, when testing the API limit found with the Delay configuration between rows. excel data for 1 second, is 2000 times the Google Translate API call.

The following is the display of the Sentiment Classification Application when the process has been completed and the results of the Sentiment Classification have been exported to excel with the name hasil.xlsx.

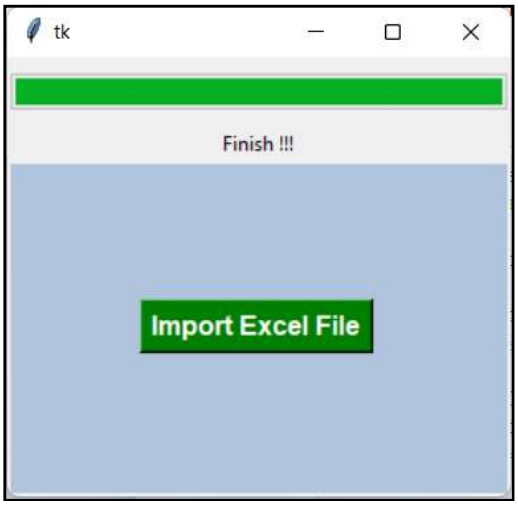


Figure 11: Sentiment Classification Done

Table 3: Result Example

No	Gelombang	Tanggal PA
1	Ganjil 2020 - PA 1	2020-09-14 00.00.00
2	Ganjil 2020 - PA 1	2020-09-14 00.00.00
3	Ganjil 2020 - PA 1	2020-09-14 00.00.00
4	Ganjil 2020 - PA 1	2020-09-14 00.00.00
5	Ganjil 2020 - PA 1	2020-09-14 00.00.00

Table 4: Result Example

Masalah Akademik	Penilaian Masalah Akademik	Masalah Non Akademik	Penilaian Masalah Non Akademik
Untuk semester baru ini, belum ada masalah yang menghambat perkuliahan saya. Kalau ada pun, masalah jaringan internet yang membuat penjelasan dosen terbata-bata dan tidak jelas.	[POSITIVE (1.0)]	Masalah non-akademik bagi saya belum ada masalah, sejauh ini baik. Pengumpulan SA juga mudah kalau kita mengetahui prosedur pengumpulan. Layanan lms yang menurut saya agak susah ketika byk yg mengakses dan apalagi ketika di hari-h dosen ada kuis mendadak, lms loading sangat lama.	[POSITIVE (1.0)]
Terkait perkuliahan saya ada sedikit masalah, antara menyeimbangkan tugas tugas yg ada dgn kondisi fisik saya yg memang sedikit lemah, jadi terkadang saya sering down tiba tiba	[POSITIVE (0.7815)]	Tidak ada	[POSITIVE (1.0)]
Belum ada.	[POSITIVE (0.9777)]	Belum ada.	[POSITIVE (0.9777)]
Belum memiliki masalah	[POSITIVE (0.9956)]	Kegiatan mahasiswa yang terbatas diakibatkan oleh daring online.	[NEGATIVE (0.9955)]
Sejauh ini belum ada masalah.	[POSITIVE (1.0)]	Sejauh ini belum ada masalah.	[POSITIVE (1.0)]

Journal of Theoretical and Applied Information Technology
31st March 2022. Vol.100. No 6

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Terkadang penyampaian dosen terputus karena jaringan, seperti layar freeze, tiba-tiba suara hilang dan lain sebagainya.	[NEGATIVE (0.9995)]	Apakah tidak ada keringanan uang kuliah di masa pandemi ini?	[POSITIVE (1.0)]

After the classification carried out using the Flair NLP framework is complete, the next step is to classify Lexicon Based Sentiment Analysis using KNIME. This step is done as a comparison of the first way to use the application. The use of KNIME begins with adapting the workflow obtained from the KNIME Hub [19] which is tailored to the needs of existing data and text classifications. Figure 12 below shows the workflow.

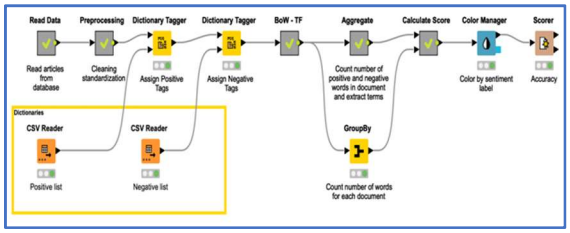


Figure 12 : KNIME Workflow

After the workflow is done, the next step is to enter the positive list and negative list taken from research [20] into the workflow that has been created. The list is entered into the CSV Reader in the existing workflow as shown in Figure 13. This list is needed to match positive and negative words from the problems presented by students through the Academic Guidance System.

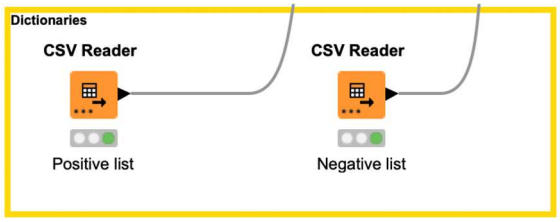


Figure 13: Positive and Negative List Reader

The data reading process is carried out after all stages of the workflow preparation are completed. The process of reading the data is carried out in several stages. The first stage is to pull the problem sentence that appears from the database and then save it into an excel file. After that, the data that is still presented in Indonesian is entered into a google sheet and translated into English, then the existing data is downloaded and converted to CSV.

Data that is already in English and in CSV form is then inputted into the existing workflow as shown in Figure 14 below.

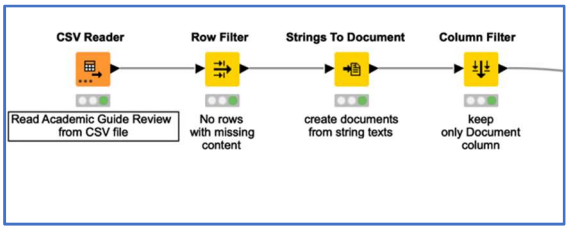
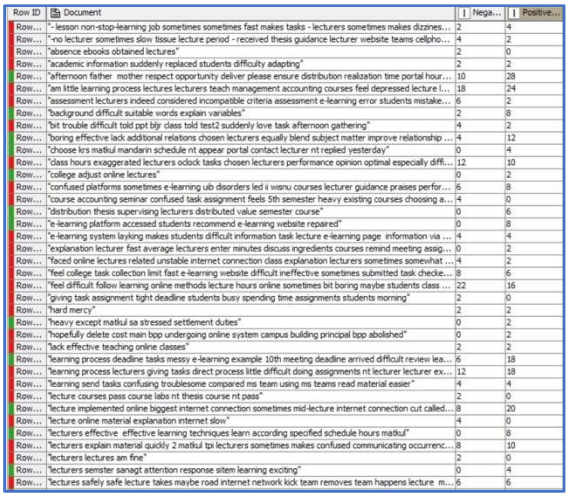


Figure 14:Data Reading in KNIME

The workflow in Figure 14 is in the Read Data workflow collection. After the data is inputted, the next step is to run the existing workflow and see the results. At run time, the existing data will be matched with the list and go through the next stage to be classified and calculated for accuracy.

The results of the classification can be seen in Figure 15 and Figure 16 below.



Row ID	Document	Nega...	Positive...
Row...	"- lesson non-stop-learning job sometimes sometimes fast makes tasks - lecturers sometimes makes dizzines...	2	4
Row...	"-no lecturer sometimes slow tissue lecture period - received thesis guidance lecturer website teams cellpho...	4	2
Row...	"absence ebooks obtained lectures"	2	0
Row...	"academic information suddenly replaced students difficulty adapting"	2	2
Row...	"afternoon father mother respect opportunity deliver please ensure distribution realization time portal hour...	10	28
Row...	"am little learning process lectures lecturers teach management accounting courses feel depressed lecture l...	18	24
Row...	"assessment lecturers indeed considered incompatible criteria assessment e-learning error students mistake...	6	2
Row...	"background difficult suitable words explain variables"	2	8
Row...	"bit trouble difficult told ppt bljr class told test2 suddenly love task afternoon gathering"	4	2
Row...	"boring effective lack additional relations chosen lecturers equally blend subject matter improve relationship ...	4	12
Row...	"choose krs matikul mandarin schedule nt appear portal contact lecturer nt replied yesterday"	0	4
Row...	"class hours exaggerated lecturers oclock tasks chosen lecturers performance opinion optimal especially diffi...	12	10
Row...	"college adjust online lectures"	0	2
Row...	"confused platforms sometimes e-learning ub disorders led ii wisnu courses lecturer guidance praises perfor...	6	8
Row...	"course accounting seminar confused task assignment feels 5th semester heavy existing courses choosing a...	4	0
Row...	"distribution thesis supervising lecturers distributed value semester course"	0	6
Row...	"e-learning platform accessed students recommend e-learning website repaired"	0	8
Row...	"e-learning system layking makes students difficult information task lecture e-learning page information via ...	4	4
Row...	"explanation lecturer fast average lecturers enter minutes discuss ingredients courses remind meeting assig...	0	2
Row...	"faced online lectures related unstable internet connection class explanation lecturers sometimes somewhat ...	4	2
Row...	"feel college task collection limit fast e-learning website difficult ineffective sometimes submitted task checke...	8	6
Row...	"feel difficult follow learning online methods lecture hours online sometimes bit boring maybe students class ...	22	16
Row...	"giving task assignment tight deadline students busy spending time assignments students morning"	2	0
Row...	"hard mercy"	2	2
Row...	"heavy except matikul sa stressed settlement duties"	0	2
Row...	"hopefully delete cost main bpp undergoing online system campus building principal bpp abolished"	0	2
Row...	"lack effective teaching online classes"	2	2
Row...	"learning process deadline tasks messy e-learning example 10th meeting deadline arrived difficult review lea...	6	18
Row...	"learning process lecturers giving tasks direct process little difficult doing assignments nt lecturer lecturer ex...	12	18
Row...	"learning send tasks confusing troublesome compared ms team using ms teams read material easier"	4	4
Row...	"lecture courses pass course labs nt thesis course nt pass"	2	0
Row...	"lecture implemented online biggest internet connection sometimes mid-lecture internet connection cut called...	8	20
Row...	"lecture online material explanation internet slow"	4	0
Row...	"lecturers effective effective learning techniques learn according specified schedule hours matikul"	0	8
Row...	"lecturers explain material quickly 2 matikul tpi lecturers sometimes makes confused communicating occurrenc...	8	10
Row...	"lecturers lectures am fine"	2	0
Row...	"lecturers semster sanagt attention response sitem learning exciting"	0	4
Row...	"lectures safely safe lecture takes maybe road internet network kick team removes team happens lecture m...	6	6

Figure 15: Classification Result in KNIME

I all Words	D Sentim...	S Sentim...	S Docum...
36	0.056	NEG	NEG
32	-0.062	NEG	NEG
8	-0.25	NEG	NEG
14	0	NEG	NEG
140	0.129	POS	NEG
122	0.049	NEG	NEG
24	-0.167	NEG	NEG
12	0.5	POS	POS
28	-0.071	NEG	NEG
28	0.286	POS	NEG
26	0.154	POS	NEG
74	-0.027	NEG	NEG
8	0.25	POS	POS
42	0.048	NEG	NEG
38	-0.105	NEG	NEG
16	0.375	POS	NEG
16	0.5	POS	NEG
44	0	NEG	NEG
56	0.036	NEG	NEG
26	-0.077	NEG	NEG
54	-0.037	NEG	NEG
130	-0.046	NEG	NEG
24	-0.083	NEG	NEG
4	0	NEG	NEG
14	0.143	POS	NEG
26	0.077	NEG	NEG
10	0	NEG	NEG
46	0.261	POS	NEG
90	0.067	NEG	NEG
28	0	NEG	NEG
20	-0.1	NEG	NEG
98	0.122	POS	NEG
12	-0.333	NEG	NEG
22	0.364	POS	NEG
56	0.036	NEG	NEG
8	-0.25	NEG	POS
16	0.25	POS	POS
34	0	NEG	POS

Figure 16: Classification Result in KNIME

From the picture above, it can be seen that the inputted data was successfully matched with the existing dictionary data and classified according to the positive and negative lists. Based on the calculation results, the accuracy value generated based on this classification is 70.2% and is still below the desired standard of 85%.

Improvement on previous literature work:
The limited number of words that can be detected in a text mining process is the main problem that usually arises. The application of text mining algorithms in previous studies was very limited due to the small amount of text obtained from data sources[3]. In this study, this problem was overcome by providing a larger capacity for text to be mined.
Another problem is the lack of a data bank of words used to perform sentiment analysis and if the data bank is considered sufficient there will be another problem that will be faced, namely the inability of the algorithm to find linkages between words [6], the existing algorithm can only detect word by word. To overcome this problem, in this study, researchers combined lexicon based sentiment analysis with the flair NLP framework.
To ensure that the desired results can be achieved, the same steps are carried out with different tools and then compare the results. From the results of this comparison, it is found that there is still a lack of accuracy from the results of the application of the sentiment analysis algorithm that has been carried out and further research is still needed to improve its accuracy. However, the research was able to overcome some of the problems in previous research.

5. CONCLUSION

In this study, we are trying to initiate the automation of the selection of positive and negative words to make it easier for faculty on campus to find out how many negative opinions from students related to lecture activities carried out in that semester and the results can be seen from the application of the methodology used.

The implementation of Lexicon Based Sentiment Analysis using the Flair NLP Framework has succeeded in direct language translation using the Google Translate API and the translated data can be classified using the application used. There is a weakness in the use of this application where it cannot be known how accurate the classification used is.

Because the author is looking for another comparison using the KNIME Workflow, through the workflow that has been created, the classification stages are carried out using Lexicon Based Sentiment Analysis. Based on the implementation of the workflow that has been made, sentiment analysis has been successfully carried out and it can be distinguished which comments are positive and which are negative. Through this workflow, it can also be mapped how accurate the classification results from the algorithm used are.

This study succeeded in mapping students' opinions about the lectures they had undergone, opinions with positive sentiments indicated that there were no problems in the lectures they were undergoing, on the other hand negative sentiments indicated that students had problems. in his lectures. However, the accuracy of mapping and classification using the algorithm is still less than the predetermined standard.

There is still much that can be improved regarding this research. In the future, more data is needed to be tested, in addition to what the previous researchers said, the list of positive and negative words for lexicon based sentiment analysis must continue to be added. In the future, research will be carried out not only for word classification but also for sentence classification.

REFRENCES:

[1] J. Jusmawati, S. Satriawati, and B. M. Sabillah, "Pengaruh Pembelajaran berbasis Daring terhadap Minat Belajar Mahasiswa PGSD Unimerz pada Mata Kuliah Pendidikan Matematika," *J. Kaji. Pendidik. Dasar*, vol. 5, no. 2, pp. 106–111, 2020, [Online]. Available: https://journal.unismuh.ac.id/index.php/jkpd/article/view/3934.

[2] R. Syafrina, Alfiati; Habibah, Sy.; Meuthia Karina, "Hubungan Antara Minat Belajar dengan Hasil Belajar Siswa dalam Mata Pelajaran IPA Pada Kelas V SD Negeri Garot Geuceu Aceh Besar," *J. Ilm. Pendidik. Guru Sekol. Dasar FKIP Unsyiah V*, vol. 2, no. 1, pp. 61–77, 2017.

[3] D. S. Indraloka and B. Santosa, "Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia," *J. Sains dan Seni ITS*, vol. 6, no. 2, pp. 51–56, 2017, doi: 10.12962/j23373520.v6i2.24419.

[4] V. B. Kobayashi, S. T. Mol, H. A. Berkers, G. Kismihók, and D. N. Den Hartog, "Text Mining in Organizational Research," *Organ. Res. Methods*, vol. 21, no. 3, pp. 733–765, 2018, doi: 10.1177/1094428117722619.

[5] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," in *Procedia Computer Science*, 2020, vol. 167, no. 1, pp. 1258–1267, doi: 10.1016/j.procs.2020.03.442.

[6] K. W. Lee, D. Lee, and H. J. Hong, "Text Mining Analysis of tTeacher's Reports on Student Suicide in South Korea," *Eur. Child Adolesc. Psychiatry*, vol. 29, no. 1, pp. 453–465, 2020, doi: 10.1007/s00787-019-01361-1.

[7] M. Fortunatus, P. Anthony, and S. Charters, "Combining textual features to detect cyberbullying in social media posts," in *Procedia Computer Science*, 2020, vol. 176, no. 1, pp. 612–621, doi: 10.1016/j.procs.2020.08.063.

[8] B. Pengembangan dan Pembinaan Bahasa, "Kendala," *KBBI Daring*, 2016. https://kbbi.kemdikbud.go.id/entri/kendala (accessed Nov. 01, 2020).

[9] B. Pengembangan dan Pembinaan Bahasa, "Pembelajaran," *KBBI Daring*, 2016. https://kbbi.kemdikbud.go.id/entri/pembelajaran (accessed Nov. 01, 2020).

[10] H. Hasan, "Kendala yang Dihadapi Guru Dalam Proses Belajar Mengajar Matematika di SD Negeri Gani Kabupaten Aceh Besar," *J. Pesona Dasar*, vol. 1, no. 4, pp. 40–51, 2015, [Online]. Available: http://jurnal.unsyiah.ac.id/PEAR/article/view/7524.

[11] A. Sadikin and A. Hamidah, "Pembelajaran Daring di Tengah Wabah Covid-19," *J. Ilm. Pendidik. Biol.*, vol. 6, no. 2, pp. 214–224, 2020, doi: 10.22437/bio.v6i2.9759.

[12] A. A. Aziz and A. Starkey, "Predicting Supervise Machine Learning Performances for Sentiment Analysis Using Contextual-Based Approaches," *IEEE Access*, vol. 8, pp. 17722–17733, 2020, doi: 10.1109/ACCESS.2019.2958702.

[13] S. S. S. J. Kaur S.K. Sehra, "A Systematic Literature Review of Sentiment Analysis Techniques," *Int. J. Comput. Sci. Eng.*, vol. 5, no. 4, pp. 22–28, 2017, [Online]. Available: https://www.ijcseonline.org/full_paper_view.php?paper_id=1235.

[14] A. Mitra, "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)," *J. Ubiquitous Comput. Commun. Technol.*, vol. 2, no. 3, pp. 145–152, 2020, doi: 10.36548/jucct.2020.3.004.

[15] A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, "FLAIR: An easy-to-use framework for state-of-the-art NLP," *NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Demonstr. Sess.*, pp. 54–59, 2019.

[16] Y. Li, X. Zhou, Y. Sun, and H. Zhang, "Design and implementation of Weibo sentiment analysis based on LDA and dependency parsing," *China Commun.*, vol. 13, no. 11, pp. 91–105, 2016, doi: 10.1109/CC.2016.7781721.

[17] J. W. C. N.-M. H. L.-S. L. H. C. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publication, 2013.

[18] R. Helilintar and I. N. Farida, "Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa," *J. Sains dan Inform.*, vol. 4, no. 2, pp. 80–87, 2018, doi: 10.34128/jsi.v4i2.140.

[19] Haline, "26_Sentiment_Analysis_Lexicon_Based_Approach." https://nodepit.com/workflow/com.knime.hub%2FUsers%2Fhaline%2FPublic%2F26_Sentiment_Analysis_Lexicon_Based_Approach.

[20] B. Liu, *Sentiment Analysis: A Fascinating Problem*. 2012.