10-2018

SUFAT: An analytics tool for gaining insights from student feedback comments

Siddhant PYASI
Singapore Management University, siddhantp.2015@smu.edu.sg

Swapna GOTTIPATI
Singapore Management University, SWAPNAG@smu.edu.sg

Venky SHANKARARAMAN
Singapore Management University, venks@smu.edu.sg

DOI: https://doi.org/10.1109/FIE.2018.8658457

SUFAT - An Analytics Tool for Gaining Insights from Student Feedback Comments

Siddhant Pyasi, Swapna Gottipati, Venky Shankararaman
School of Information Systems
Singapore Management University
Singapore
siddhantp.2015,swapnag,venky@smu.edu.sg

Abstract— **Teacher evaluation is a vital element in improving student learning outcomes. Course and instructor feedback given by students, provides insights that can help improve student learning outcomes and teaching quality. Teaching and course evaluation systems help to collect quantitative and qualitative feedback from students. Since manually analysing the qualitative feedback is painstaking and a tedious process, usually, only the quantitative feedback is often used for evaluating the course and the instructor. However, useful knowledge is hidden in the qualitative comments, in the form of sentiments and suggestions that can provide valuable insights to help plan improvements in the course content and delivery. In order to efficiently gather, analyse and provide deeper insights from student feedback by topics, we developed a user-friendly application, StUdent Feedback Analysis Tool (SUFAT). The tool is an independent desktop application that can be installed and used by any non-technical user. The tool takes an excel sheet with comments as an input and generates an excel sheet with visual reports of summaries that include sentiments and suggestions as an output. The tool can benefit instructors to quickly analyse the termly qualitative feedback and take appropriate actions. We intend to release the tool for public use so that instructors can download and install the tool on their computer system.**

Keywords— Student feedback, teaching evaluations, insights, sentiments, suggestions, text analytics, tool.

I. INTRODUCTION

Evaluation of teaching has several purposes, including collecting feedback for teaching improvement, developing a portfolio for job applications, or gathering data as part of personnel decisions, such as reappointment or promotion and tenure[1, 2, 3]. Hence, evaluation enhances student learning and simultaneously encourages teachers to maintain their professional growth. It is concerned with gathering evidence from a range of sources that inform on teacher's performance and using this to support improvements in teaching practice. One major source is the student feedback, through teaching evaluation systems that are implemented in educational institutions.

Students provide feedback related to teaching, content and learning in two distinct forms namely, quantitative (numeric) ratings for survey questions on Likert scale and qualitative comments (text). The teaching component refers to aspects such as instructors' interaction, delivery style, ability to motivate students, out of class support, etc. The content refers to aspects related to course details such as concepts covered, lecture notes, labs, exams, projects, etc. The learning refers to aspects related to student learning experience such as understanding concepts, developing skills, applying skills acquired, value gained from the course, etc.

An instructor usually analyses the quantitative feedback to the survey questions and accordingly acts upon to improve the course. However, the quantitative scores do not give sufficient insights as to what changes should the instructor implement in order to address low scores in a specific component. In addition to using this quantitative feedback, if the qualitative comments are also analysed, the hidden insights from the feedback aids instructor in improving the learning and teaching process by addressing the low scores.

The conceptual framework for student feedback analysis proposed by [4] is a starting point for the community of stakeholders to consider how qualitative and quantitative feedback can help in making informed decisions with respect to teaching, learning, and curriculum improvements. The framework consists of four major components; Text Analytics Model, Data Processing, Extraction, and Summarization. These components enable to processes the student's textual comments and generate visual outputs for the users of the evaluation system to improve the teaching and learning process. The benefits of the framework is twofold. Firstly, it supports academic managers in faculty related decisions such as faculty recruitment, award nominations and as well as personality development decisions. Secondly, it supports the course designers or the instructors to identify the gaps in the course content and course delivery process and improve the teaching process.

Based on the framework presented in [4], in this paper, we describe the design and implementation of an application we have implemented, StUdent Feedback Analysis Tool (SUFAT). This tool uses text mining and visualization techniques and aggregates the textual comments into sentiments and suggestions categorised by topics, that is, aspects which most students talk about. Additionally, in this paper, we have extended the work presented in [4] by providing the detailed system architecture, model enhancements using topic modelling, and describing an end to end tool for the faculty to analyse the qualitative feedback comments and gain valuable insights. Using SUFAT, the instructor can use the additional insights to amend the course with specific focus on issues of

978-1-5386-1174-6/18/$31.00 ©2018 IEEE

concern. In other words, the instructor will be able to make informed decisions by analysing the textual comments. For example, if students provided a low score to the topic related to "course labs, project and assignment" and then voice their sentiments and suggestions in the comment section, the instructor can combine both these feedback in order to gain a better understanding of gaps and what needs to be improved.

SUFAT enables a deeper analysis of textual comments and provides user-friendly reports through the use of topic modelling, sentiment analysis, natural language processing (NLP) and visualization techniques. Sentiment analysis aims at classifying the data into positive or negative polarities using supervised methods or unsupervised methods. NLP techniques helps to automate the processing of human language [6, 8]. Such processing helps in the subsequent tasks of classification and clustering of comments into suggestive texts, and aid in increasing the accuracy of sentiment analysis. Topic modelling techniques are capable of identifying the hidden topics of interest in textual documents [5]. The topics are part of discussions and studying the sentiment or suggestions by the key topics in the dataset provides categorised analysis in a summarised format [9]. The goal of visualization techniques is to provide user-friendly summaries of the sentiments and suggestions by topics of interest that are extracted from student comments in a visual representation that supports search, comparison, and analysis.

The main contribution of our work is the innovative application of sentiment analysis models, natural language processing (NLP) techniques, Latent Dirichlet Allocation (LDA) model and visualization techniques in the education domain. SUFAT empowers the instructors with additional insights that can be gleamed from students' qualitative feedback and helps continually improve the student learning experience through the provision of three capabilities; (1) Provision of functions and visualizations for analysing student qualitative feedback; (2) User-friendly reports that supports any non-IT instructor to extract, analyse and visualize sentiments and suggestions from the students comments; (3) Easy installation of the tool as an integrated executable file that works on windows operating system.

The paper will be structured as follows. Section II will be devoted to literature review of related work. Section III describes a pedagogical scenario in which we developed the solution. Section IV describes the concept architecture of the solution model. We describe the implementation details of the tool and its features in Section V. In section VI, we present the results of evaluation of the tool and its limitations. We conclude in Section VII suggesting some interesting future directions of our work.

II. RALATED WORK

In this section, we focus on works related to textual content mining and sentiment mining relevant to student feedback analysis.

Textual Content Mining: Text mining and natural language processing techniques are useful for opinion mining research. Opinion extraction aims at automatically finding attitudes or opinions about specific targets, such as named entities, consumer products or public events [6, 7]. An opinion without its target being identified is of limited use [10]. Early adoptions of topic models for educational data include the work of Haruechaiyasak and Damrongrat [11], who recommended articles from Wikipedia by calculating the similarity measures among topic distributions of the articles. Kuang et al. provided resource recommendation for users in an e-learning system based on contents and user log activities by applying LDA models [10].

Ming and colleagues applied hierarchical LDA models to predict the grades of students [12]. Zhang et al. applied LDA model to online discussions of four Chinese classrooms to extract topics and display the temporal profiles of the topics [13]. Sherin used LDA models to extract fragments (categories) of ideas from student interviews [14]. Southavilay et al. used LDA models to mine cloud data from Google Docs to gain insights on how learners' collaborative activities, ideas and concepts are developed during the process of writing [15]. Wong et al. used probabilistic and LDA models to analyze discussion forum data and present it in user-friendly visualization charts [15]. The above studies point to the promising potential of LDA models to capture conceptual topics in education datasets.

Sentiment Analysis in Education Data Mining: Sentiment classification aims at classifying the data into positive or negative polarities [8] using supervised methods or unsupervised methods. Early works of sentiment analysis of student feedback using data mining approach include Altrabsheh et al. who devised a system to analyze sentiments in real time to provide real-time intervention in the classroom [16]. They combined support vector machines and naïve bayes methods. Rashid et al. used generalized sequential pattern mining and association rule mining to analyze opinion words from student feedback [17]. Gamon et al. took another approach for analyzing sentiments in free flowing text – as is with student feedback as well – by building a system, Pulse, that brought together algorithms that clustered topics and classified sentiments with intuitive visualization to allow a deeper analysis of customer feedback and sentiment on special topics [18]. Ila et al. combined clustering and sentiment classification models to extract the topics and the sentiments from the student's feedback [19]. The limitation with clustering model is that each comment is only assigned to a single cluster and the topics coherence quality is low.

In our solution, we propose LDA models to address this limitation and the comments are assigned to more than one cluster. At the same time, we also use the classification based approach for the sentiment mining task. We also incorporate suggestion extraction models, and provide an integrated solution for teaching evaluation analytics. In the next section, we describe the pedagogical scenario based on which we designed the solution.

Table 1: Sample textual comments from teaching evaluations by students. Italic, underlined and bold shows multiple topics in a single comment provided by the student

Comment	Topics	Sentiments	Suggestion?
Don't understand why Tableau was brought in when it was already covered in other modules. Additionally, too much focus was on Sharepoint when we could have used more time to explore other softwares such as LifeRay.	*Course_Value_Use_Challenge* Course_Value_Use_Challenge	*Negative* Negative	
Prof is very helpful,friendly and approachable. She is able to answer questions that the students raised in class or during consultations and able to provide constructive feedback and improvements on the projects as well.	*Faculty_Feedback* Faculty_Feedback	*Positive* Positive	
Prof is very friendly and helpful and kind. profs handles the q&a well throughout the class	*Faculty_Approachable_Fairness* Faculty_Feedback	*Positive* Positive	
The instructor is highly encouraging and approachable. She makes it a point to keep lessons engaging and interesting.	*Faculty_Feedback* Faculty_Interaction_Engagement	*Positive* Positive	
The course gives a good exposure to enterprise web solutions. However, I feel the course is too reliant on Sharepoint as Sharepoint is not a perfect software as it has many flaws. It would be good to get exposure to other types of web portals to have better exposure.	*Course_Value_Use_Challenge* Course_Value_Use_Challenge Course_Value_Use_Challenge	*Positive* Negative Negative	 Yes
Should provide more time for teams to decide on chosen company.	Course_Value_Use_Challenge	Negative	Yes
Need clearer definitions on grading for ICE and other assessments	Course_Value_Use_Challenge	Negative	Yes

III. Pedagogical Scenario

The model and results presented in this paper are based on the teaching evaluation feedback process followed at the Singapore Management University. At the end of each term, student feedback is gathered through a questionnaire which includes questions relating to students' perceptions of the instructor and their learning experience in the course. The teaching evaluation system uses an integrated web application, which is an environment that integrates courses, faculty and students. The questions are adapted and developed from the literature on measuring tertiary teaching and learning. Students provide both quantitative and qualitative feedback on both course and instructor aspects. Both the summarized quantitative data by questions as well as compilation of random qualitative comments in raw form are made available to the respective instructors as individual reports. The comments are not specific to any questions in the survey form and therefore, the categorization by topics of interest is not provided in the current reports.

This feedback is then passed to the faculty for deeper analysis and to discover the major concerns of the students about the instructor, course content and delivery. Sample student evaluations are shown in the Table I.

Table I shows two key types of feedback; sentiments and suggestions. We observe that the feedback is focussed on the specific topics relevant to the course or faculty. At times, each comment might have multiple topics relevant to the course. For example, the comment, "*The instructor is highly encouraging and approachable.* She makes it a point to keep lessons engaging and interesting" is based on two topics; "Faculty feedback" and "Faculty interaction engagement". The sentiments on both topics are positive. Occasionally, the students also provide suggestions. For example, the comment, "Should provide more time for teams to decide on chosen company" is a suggestion on the topic, "Course value, use and challenge".

To manually extract the detailed insights from the raw data, instructor is required to spend more effort and time before he or she can plan the improvements in the course design and development. Therefore, the tool we developed, SUFAT, aids the instructor to extract these insights automatically and efficiently. The tool takes students' comments in an excel spreadsheet as an input and generates the summarised insights in the form of user-friendly visuals in an excel sheet. The instructor can use this visualization to analyse the students' concerns and plan the action items to improve the course content and delivery.

IV. Conceptual Archtecture of SUFAT

In this section, we first present the conceptual architecture of the tool and then describe each component. Most architectural diagrams depict the general relationship with the data and algorithms. In this paper, we focus on the detailed architecture with more precise information on each component of the architecture. Figure 1 depicts the conceptual architecture of SUFAT. The tool implementation uses most of the components described in this architecture. However, in order to enable an independent installation of the tool on the instructor machine and for ease of use, some components are modified.

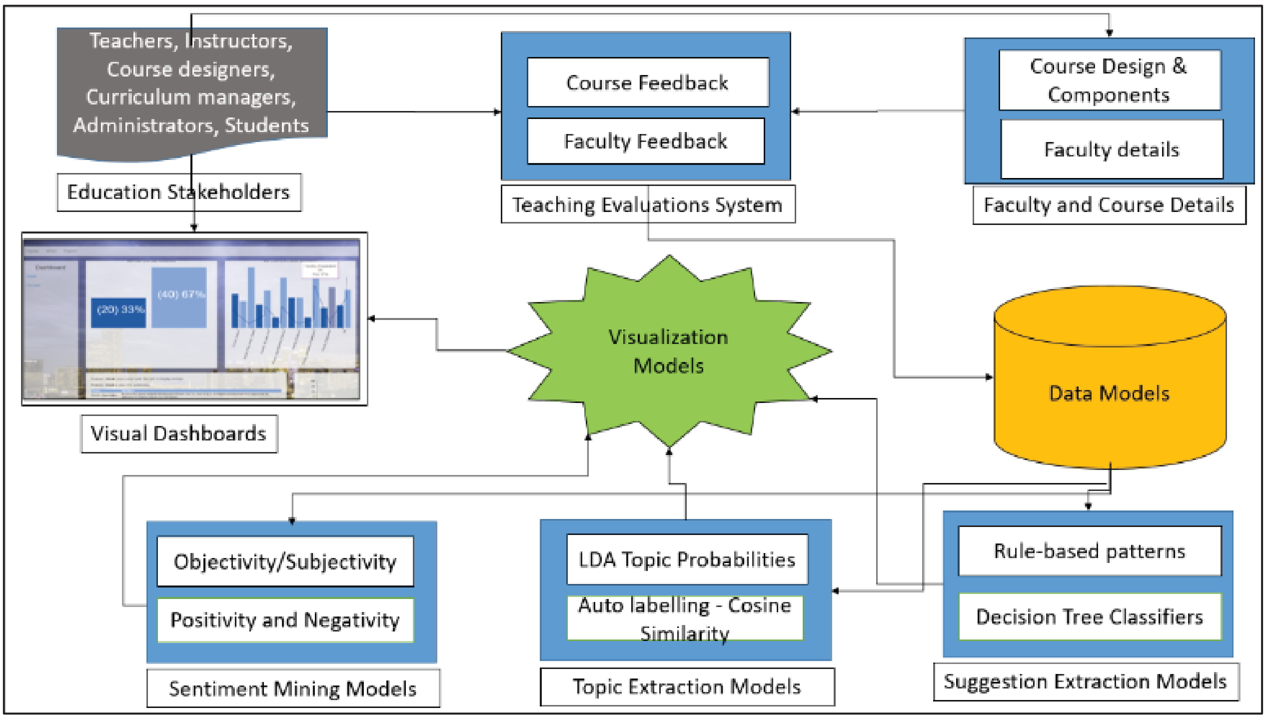


Figure 1: Conceptual architecture overview of SUFAT

The tool implementation will be described in Section V. We first describe the details of each component.

A. Education Stakeholders

Instructors and students are key stakeholders of the teaching and learning process. Instructors are part of a university's pedagogy model. They follow university's education model, and play a key role in supporting the success of students' education and development. Students are core integral part of this model as they participate in the learning activities, use resources, achieve learning outcomes and represent the outcomes of the pedagogy model.

B. Course Design and Delivery

Universities define their pedagogy models and develop guidelines for the course content design and development [22]. These are general guidelines and as the course outcomes are very distinguished, each course might consist of different components to achieve the learning outcomes of the course. For example, a course such as "Introduction to Programing" in computer science school might have components such as labs, projects, lecture slides, exercises, quizzes and tests. Whereas, a course such as "Financial Markets" from business school might have components such as case studies, assignment, research, projects, quizzes and tests. [22]. Therefore, the course details are useful inputs that help to extract the relevant topics from the feedback.

C. Teaching Evaluation System

Teaching evaluation system (TES) is an integral part of all universities to enable performance based compensations to instructors and to ensure high quality teaching and learning. TES is usually internally developed by the university and used to collect the students' evaluations in an online format. Mostly, they are the combinations of quantitative and qualitative data. Responses are usually anonymous and collated along with other details such as faculty profile, course details and student profile, and stored in a database. TES database serves as input to the data models.

D. Data Models

Data models vary for quantitative and qualitative data sets. They also vary according to the data mining, text mining and visualization algorithms. The data models are expected to support the continuous streaming of the data to enable the latest feedback analysis reports are presented to the instructor. The data is usually stored in a raw format, to provide at a glance, the actual feedback given by the students. Simultaneously, the data is also processed, cleaned and stored as processed data models to serve as inputs to the text mining and NLP algorithms.

E. Topic Extraction Models

Topic extraction models are a collection of algorithms that can infer topics from sets of documents. Latent Dirichlet Allocation (LDA) topics model is based on probabilities that associate a topic with a distribution over a set of words [5]. These topics represent the hidden aspects in the comments. In our solution, we use supervise LDA models to enable more accurate topic extraction, auto labelling and auto numbering of the topics for the given set of comments. Standard LDA models are unsupervised and labelling of the topic for each comment is

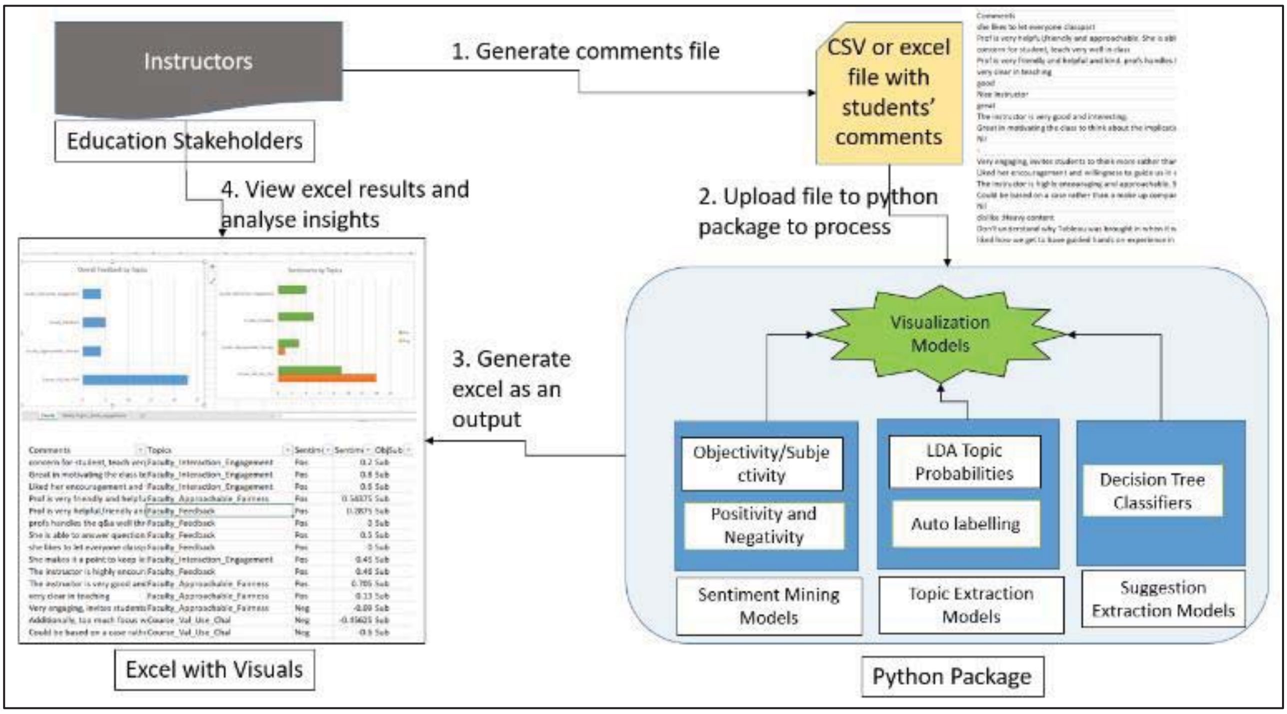


Figure 2: SUFAT Tool implementation and features

a laborious task performed by a human. Therefore, we took a semi-supervised LDA approach to enable auto labelling of the topics generated by the LDA algorithm.

F. Sentiment Mining Models

Student comments are mapped to sentiments; negative sentiments and positive sentiments [6, 20, 21]. Sentiment models are polarity computing algorithms. These algorithms look at the objectivity and the subjectivity of each sentence. Based on the polarity score, the tokenized sentences will be classified as either positive or negative sentiment [6]. Text blob algorithm makes use of Naïve Bayes classifier trained on the movie reviews dataset. In our preliminary experiments, we observed that the basic Textblob tool has limitations with contrasting conjunctions [21]. For example, it is not effective when dealing with comments such as "The instructor is kind but does not have sufficient domain knowledge". Therefore, we further modified Textblob to consider contrasting conjunctions such as "although", "despite" etc.

G. Suggestion Extraction Models

Suggestions refer to comments which provide actionable feedback to the decision makers; instructors and academic managers. For example, "The course needs to focus on the code as much as the business side" is a suggestion from the student feedback on the course content whereas, the comment, "sounding a little more upbeat may help with the class's energy level" is a suggestion for the instructor. We adopt a vector space representation of a document where each comment is evaluated as a document term frequency [23]. We implemented decision tree C5.0 text classification method to extract suggestions from a given set of comments.

H. Visualization Models

Visual models are the data models for the visual dashboards. They provide the format and labels needed for generating the reports and charts useful for representing the insights from the students' comments in a user-friendly manner. The visual models provide the features to create interactive reports and enable deeper analysis on the feedback provided by the students.

I. Visual Dashboards

The goal of visualization dashboard is to provide user-friendly summaries of the insights obtained from students' comments. The design goal is to ensure a user-friendly report that supports comparison and analysis. A graphical representation of the text using a word cloud, which is a popular approach, is adopted to provide a quick view. In our tool, we also designed reports for comment insights, sentiment insights and suggestion insight categorised by the topics. In order to ensure easy installation and enhanced usability for non-technical users, we use a desktop excel sheet with the visualizations as a final output to the user.

V. Features and Implementation Of the Tool

In this section, we describe implementation details, features of the tool, and the usability of the tool. Figure 2 depicts the features of the tool and the process flow to generate the summaries.

A. Implementation Details

The tool is developed using python language. The program is invoked with an input file, csv or excel spreadsheet that contains a list of all students' comments. Python package consists of three key components. The first component of the python program identifies the topics. Figure 3 shows the code snippet topic extraction that is based on genism LDA models [24].

```
fixed_dict = []
for key, elem in seeded_dict.items():
#       print(seeded_dict)
    fixed_dict.append(elem.split())
print(fixed_dict)
#  print(doc_clean) #resultof the cleaning

#built the dictionary
dictionary = corpora.Dictionary(doc_clean) #texts
print("#######dictionary#######")
#  print(dictionary)
#bag of words
corpus = [dictionary.doc2bow(text) for text in doc_clean]

#convert list of documents into document term matrix
doc_term_matrix = [dictionary.doc2bow(doc) for doc in doc_clean]

k_topics = 3
# Creating the object for LDA model using gensim library
Lda = gensim.models.ldamodel.LdaModel

# Running and Trainign LDA model on the document term matrix.
ldamodel = Lda(doc_term_matrix, num_topics=k_topics, id2word = dictionary, passes=20)

print(ldamodel.print_topics(num_topics=5, num_words=50))

#create empty dictionary
ldamodel_Showtopics = {}
```

Figure 3; Code snippet of the LDA models from Jupyter UI

The second component is the sentiment extraction model. We use Textblob for sentiment extraction [21]. Each word is scored and the cumulative score of the sentence will indicate whether the sentiment is positive (if greater than or equal to zero) or negative (if less than zero). This is a generic model and in our preliminary analysis, the accuracy on our dataset was low. Therefore, we created a reverse list to improve the sentiment scores of certain words specific to education domain. Figure 4 shows the code snippet of sentiment model.

```
j = senti(temp)
checksum1 = 0
checksum2 = 0
checksum3 = 0
checksum4 = 0
for p in rev_list:
    checksum1 = checksum1 + len(re.findall('\\b'+p+'\\b', temp, flags=re.IGNORECASE))
if checksum1 > 0:
    q = j*-1    #change the polarity if found words from rev list
else:
    for pl in dip_list:
        checksum2 = checksum2 + len(re.findall('\\b'+pl+'\\b', temp, flags=re.IGNORECASE))
    if checksum2 > 0:
        r = j-0.6
    else:
        r = j
    for pl in spe_neg_list:
        checksum3 = checksum3 + len(re.findall('\\b'+pl+'\\b', temp, flags=re.IGNORECASE))
    if checksum3 > 0:
        s = r-0.3
    else:
        s = r
    for pl in spe_pos_list:
        checksum4 = checksum4 + len(re.findall('\\b'+pl+'\\b', temp, flags=re.IGNORECASE))
    if checksum4 > 0:
        q = s+0.3
    else:
        q = s
```

Figure 4; Code snippet of the Sentiment model from Jupyter UI

The third component is suggestion extraction, which is based on decision trees. Decision trees are supervised classification algorithms and therefore require training datasets. C5.0 decision tree algorithm is used. One key approach we take here is that the stopwords are not removed in the processing stage as they are crucial in identifying the suggestions. The training set for suggestions is created with human judgement and stored in the database. Figure 5 shows the code snippet for loading the training dataset from the database.

```
operations = [
    migrations.CreateModel(
    name='TrainSet',
    fields=[
            ('id', models.AutoField(auto_created=True, primary_key=
            ('comment', models.TextField(blank=True, db_column='Com
            ('sentiment', models.TextField(blank=True, db_column=Su
        ],
        options={
            'db_table': 'train_set',
            'managed': False,
        },
      ),
  ]
```

Figure 5: Code snippet for loading the suggestion training dataset and creating the model.

The trained model is created as a pre-processing step. Once trained, the model can automatically identify selected comments as suggestions.

B. Features

The tool provides two main features for the users. Firstly, the feature of uploading the students' feedback in excel or csv format. The tool is developed as a desktop application so that the instructor can install and use it on their individual machines. The tool can also be converted into a web application with some minor modifications. The key disadvantage with the desktop application is that excel reporting is not interactive, Secondly, the insights from the data are currently presented in the fixed form of charts and tables. Unlike a web application reporting feature, where the users can provide various inputs for search and comparison, the excel sheet is limited to the fixed reporting.

In this subsection, we discuss the reporting features in detail. Recall that the main goal of the tool is to provide insights from comments that are useful in making decisions relevant to improvement in teaching and learning process. Therefore, the tool generates summarised reports in an excel spreadsheet for the users to quickly digest the insights and then focus on areas of concern. For this paper, we used the report based on data from an undergraduate course in Information Systems program, "Enterprise Web Solutions".

1) Overall feedback report

The first report generated is the overall statistics of the comments. The comments which are insignificant such as "Nil", "good" etc., are removed in the cleaning process. The statistics are shown for each topic of interest related to the course and the faculty. Figure 6 depicts the overall feedback report.

Figure 6 shows that topic, "Course_Project_Assignment" is of high interest to the students. Faculty can focus on this topic for analyzing whether the sentiment on this is positive or negative. Accordingly, the course improvements can be made by analysing the suggestions.

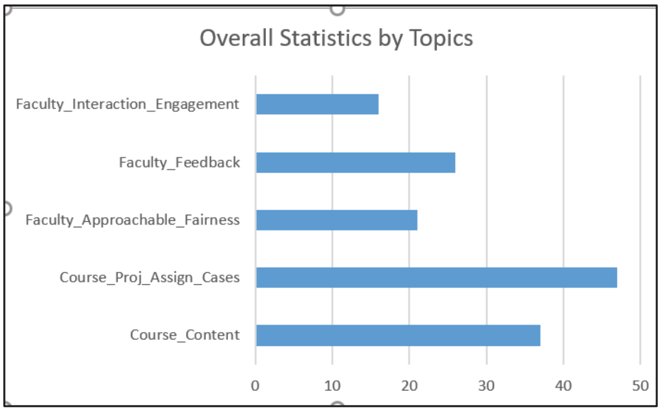


Figure 6: Overall feedback statistics for each topic in the comments

2) Sentiments report

The second report shows the overall sentiment statistics. The statistics shown depict the positive and negative sentiments. Figure 7 shows the sentiments report doe each topic.

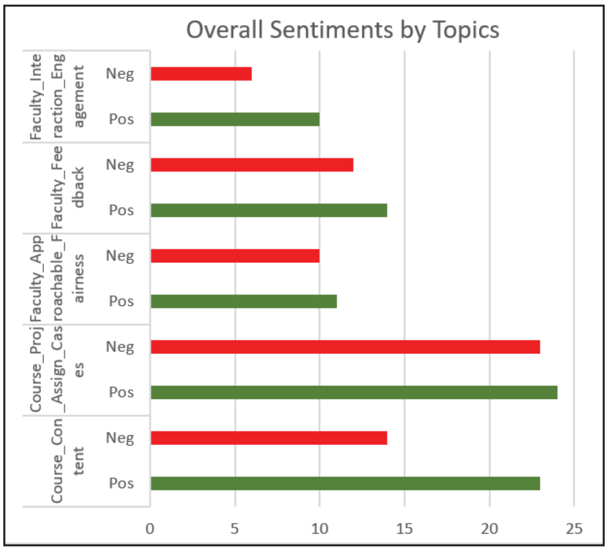
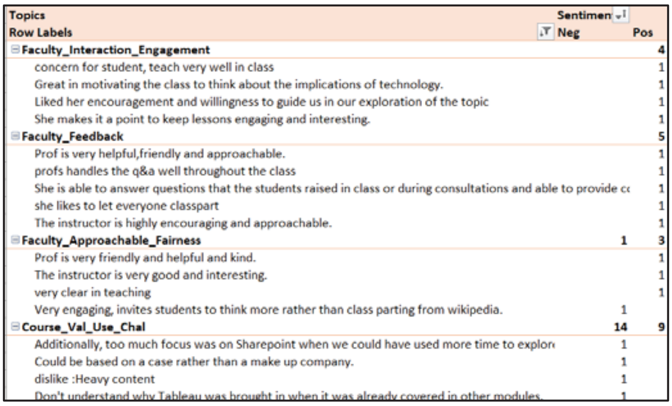


Figure 7: Sentiment statistics by topics

From Figure 7, we observe that the faculty sentiments are mostly positive and course sentiments is a mix of positive and negative. The faculty can now focus more on those aspects where the sentiment is mostly negative. The faculty can retain the positive aspects of the course and work on improving the negative aspects of the course. To further conduct deeper analysis, the next report provides additional details of sentiments.

3) Detailed sentiments report

To conduct deeper analysis on the comments, the faculty can refer to the detailed sentiment report. This is a tabular report that depicts the comments by the topics and sentiment types. Figure 8 shows the detailed summary report.



Topics / Row Labels	Sentiment Neg	Pos
Faculty_Interaction_Engagement		**4**
concern for student, teach very well in class		1
Great in motivating the class to think about the implications of technology.		1
Liked her encouragement and willingness to guide us in our exploration of the topic		1
She makes it a point to keep lessons engaging and interesting.		1
Faculty_Feedback		**5**
Prof is very helpful,friendly and approachable.		1
profs handles the q&a well throughout the class		1
She is able to answer questions that the students raised in class or during consultations and able to provide c		1
she likes to let everyone classpart		1
The instructor is highly encouraging and approachable.		1
Faculty_Approachable_Fairness	**1**	**3**
Prof is very friendly and helpful and kind.		1
The instructor is very good and interesting.		1
very clear in teaching		1
Very engaging, invites students to think more rather than class parting from wikipedia.	1	
Course_Val_Use_Chal	**14**	**9**
Additionally, too much focus was on Sharepoint when we could have used more time to explor	1	
Could be based on a case rather than a make up company.	1	
dislike :Heavy content	1	
Don't understand why Tableau was brought in when it was already covered in other modules.	1	

Figure 8: Detailed sentiments report

4) Suggestions report

Recall that students tend to provide suggestions for improvements. Figure 9 shows the overall suggestions report in a tabular form by topics. The faculty can focus on the topic that received more number of suggestions.

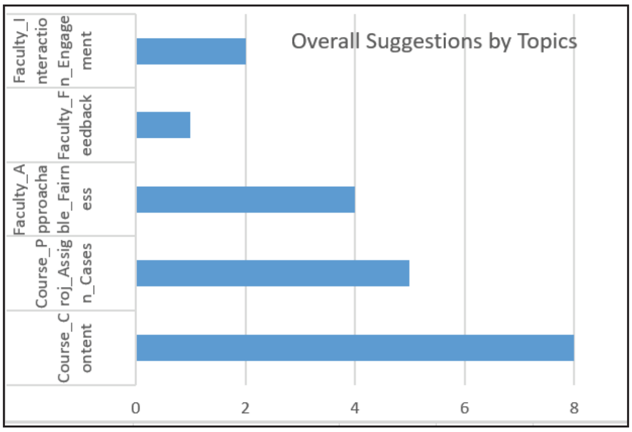
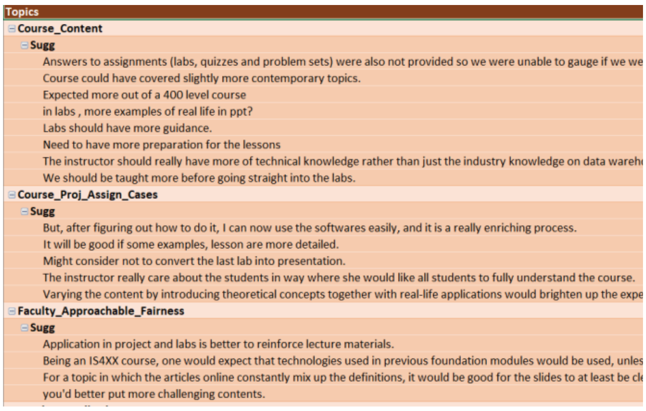


Figure 9: Suggestions report by topics

5) Detailed suggestions report

To conduct deeper analysis of the suggestions, the faculty can refer to the detailed suggestions report. Figure 10 shows the detailed suggestions report.



Topics
Course_Content
Sugg
Answers to assignments (labs, quizzes and problem sets) were also not provided so we were unable to gauge if we we
Course could have covered slightly more contemporary topics.
Expected more out of a 400 level course
in labs , more examples of real life in ppt?
Labs should have more guidance.
Need to have more preparation for the lessons
The instructor should really have more of technical knowledge rather than just the industry knowledge on data wareh
We should be taught more before going straight into the labs.
Course_Proj_Assign_Cases
Sugg
But, after figuring out how to do it, I can now use the softwares easily, and it is a really enriching process.
It will be good if some examples, lesson are more detailed.
Might consider not to convert the last lab into presentation.
The instructor really care about the students in way where she would like all students to fully understand the course.
Varying the content by introducing theoretical concepts together with real-life applications would brighten up the expe
Faculty_Approachable_Fairness
Sugg
Application in project and labs is better to reinforce lecture materials.
Being an IS4XX course, one would expect that technologies used in previous foundation modules would be used, unles
For a topic in which the articles online constantly mix up the definitions, it would be good for the slides to at least be cle
you'd better put more challenging contents.

Figure 10: Detailed suggestions report

This is a tabular report that depicts the comments that are suggestions, organized by the topics and sentiment types.

6) Excel based table of insights

Figure 11 shows the overall summarized insights table which includes the comments together with the topics of concerns, sentiments, sentiment scores and suggestions.

Comments	Topics	Sentiment	Sentime	Suggestion
Prof is making beast effort to reply to students' feedback and	Faculty_Feedback	Pos	0	Oth
this course can be considered a whole new course in SIS as th	Faculty_Interaction_En	Neg	-0.455	Oth
I feel that unclear instruction in labs helps me more in learnir	Faculty_Interaction_En	Neg	-0.35	Sugg
The hands on experience gained in this class definitely helps	Faculty_Feedback	Pos	0	Oth
On the side note, problem sets are not really useful and sch l	Faculty_Feedback	Neg	-0.638	Oth
instructor is overseas but still try to manage the course	Faculty_Approachable_	Pos	0	Oth
learn how to manage data warehouse by hand-on exercise ir	Faculty_Feedback	Pos	0	Oth
you'd better put more challenging contents.	Faculty_Interaction_En	Neg	-0.1	Sugg
you'd better put more challenging contents.	Faculty_Approachable_	Neg	-0.1	Sugg
Due to unforeseen circumstances, it is unfair to feedback.	Faculty_Approachable_	Neg	-0.313	Oth
But definitely the course in this case is certainty affected and	Faculty_Feedback	Pos	0	Oth
She's not here most of the time, but I enjoyed sitting in her fe	Faculty_Approachable_	Neg	-0.4	Oth
Despite her external commitments, she showed dedication tc	Faculty_Feedback	Pos	0	Oth
I heard Datawarehousing is a lot harder this semester, but I'r	Faculty_Feedback	Neg	-0.033	Oth
Application in project and labs is better to reinforce lecture m	Faculty_Approachable_	Pos	0.5	Sugg
The work load is a challenge but manageable.	Faculty_Feedback	Pos	0	Oth
The instructor has been kind and helpful.	Faculty_Approachable_	Pos	0.6	Oth
She has always gone out of her way to help students despite	Faculty_Interaction_En	Pos	0	Oth
The course has taught me some interesting stuff about data v	Faculty_Interaction_En	Neg	-0.35	Oth
instructor tried to make the course as interactive as possible	Faculty_Approachable_	Pos	0	Oth

Figure 11: Overall insights of all comments

C. Process Flow

In this sub-section, we describe the three steps needed to use the tool shown in Figure 2. Step 1, the users collects the qualitative feedback in an excel sheet or csv file. Step 2, execute the python file. Step 3, select the file location as input, and once the code is executed, an output file is generated with the reports explained in the previous sub-section. The output file is located in the same folder as the input file.

VI. Evaluation of the Tool

We evaluated the tool using statistical approaches popular in text mining such as accuracy, precision and F-scores. We first present the statistical evaluations. We then discuss the limitations of the tool based on the performance of the algorithms used in the development of the tool and its usability.

A. Statistical evaluations

For sentiment classification, we used Polarity Analyser and Textblob. Textblob has better performance that Polarity Analyser; Recall is 96.17%, Precision is 67.47% and F-score of 79.30%.We tested various classification models for suggestion extraction; Generalized Linear Models (GLM), Support Vector Machine (SVM), Conditional Inference Tree (CTREE) and Decision Tree C5.0. C5.0 has better performance than other models; recall is 80.2%, precision is 77.5% and F-Score is 78.1%. For topics, we compared LDA models with k-means clustering with cosine similarity scores. LDA models were capable of providing multiple topics and more relevant topics for the comments whereas cosine clusters assigned a single topic to the comment.

B. Discussions and limitations

The tool has two main limitations. Firstly, limitations related to the techniques used. Secondly, limitations related to the usability of the tool. Text mining techniques such as classification and topic models have limitations in terms of performance. In our statistical analysis, we observed that sentiments and suggestion extraction has better performance but certain comments with bad grammatical structure can be wrongly classified. The topic models generate multiple topics for the comment and we need to extract the most relevant topics for a given comment. We used cosine similarity to map the topics to the seeded topics and hence generate the coherent topics with the automated labels. This approach has its limitations in the performance. However, the cosine similarity, between comments and seeded topics, has lower performance than LDA topic models.

We also observed that the suggestion extraction classifier has performance limitations. In our experiments, we observed that stopwords play an important role in suggestion detection. However, the accuracy can be further improved by studying the grammatical structure. Usability issues depend on the technology background of the user. Though the conceptual architecture supports web based application for generating interactive reporting, academicians may not have the necessary background to set up the web environment. To handle this issue, we implemented a desktop application based on python. From our experience, once the users follow the installation instructions, they will be able to use the three-step process of uploading the data file and analysing the output file of insights. Therefore, there is a need for clear user guide with instructions for installation and program execution.

VII. Conclusions

In this paper, we presented a student feedback analysis tool, SUFAT. The tool is useful in extracting insights from the qualitative feedback comments provided by students. The textual comments contain sentiments and suggestions given by the students. The tool extracts the comments, sentiments and suggestions based on the topic the student is commenting on. The tool is user-friendly and we hope faculty will benefit from the tool and gain insights from textual feedback comments by the students, leading to enhancements of the teaching and learning process. We intend to release the tool and the user guide of SUFAT.

Acknowledgment

This research was supported by the Singapore Ministry of Education, Tertiary Education Research Fund under the research grant reference number MOE2016-2-TR44. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore Ministry of Education. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centers in Singapore Funding Initiative

REFERENCES

[1] Leckey, J., & Neill, N. (2001). Quantifying quality: the importance of student feedback. Quality in Higher Education, 7(1), 19-32.

[2] Donovan, J., Mader, C. E., & Shinsky, J. (2010). Constructive student feedback: online vs. traditional course evaluations. Journal of Interactive Online Learning

[3] Scott, G., Grebennikov, L., & Shah, M. (2008). Using qualitative data to prove and improve quality in Australian higher education. Evidence based Decision Making: Scholarship and Practice, 97.

[4] S. Gottipati, V. Shankararaman and S. Gan, "A conceptual framework for analyzing students' feedback," 2017 IEEE Frontiers in Education Conference (FIE), Indianapolis, IN, USA, 2017, pp. 1-8. doi:10.1109/FIE.2017.8190703

[5] Blei, D. M., Ng A. Y., and Jordan M. I. 2003. Latent Dirichlet allocation. Journal of Machine Learning Research, 3 (March, 2003), 993-1022.

[6] B. Liu. Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, 2nd ed, 2010.

[7] Jiang, Long, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. In Proceedings (ACL-2011).

[8] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. In EMNLP '02: Proceedings of the 2002 pages 79–86, July 2002.

[9] Ramanand, J., Bhavsar, K., Pedanekar, N. (2010) Wishful thinking: finding suggestions and 'buy' wishes from product reviews. In: Proceedings of the NAACL HLT 2010 Workshop pp. 54–61

[10] Wei Kuang, Nianlong Luo, and Zilei Sun. 2011. Resource recommendation based on topic model for educational system. Information Technology and Artificial Intelligence Conference (ITAIC), 2011 2:370–374, Aug

[11] Choochart Haruechaiyasak and Chaianun Damrongrat. 2008. Article recommendation based on a topic model for wikipedia selection for schools. 5362:339–342

[12] Ming, N., & Ming, V. (2012). Predicting student outcomes from unstructured data. Proceedings of the 20th Conference on User Modeling, Adaptation, and Personalization.

[13] Zhang, Y., Law, N., Li, Y., and Huang, R. (2012). Automatic extraction of interpretable topics from online discourse. In International Conference of the Learning Sciences (ICLS) 2012, Volume 1, (p.443-450).

[14] Sherin, B. (2012). Computing student science conceptions with Latent Dirichlet Allocation. Paper presented at the 10th International Conference of the Learning Sciences.

[15] Vilaythong Southavilay, Kalina Yacef, Peter Reimann, and Rafael A. Calvo. 2013. Analysis of collaborative writing processes using revision maps and probabilistic topic models. In Proceedings (LAK '13), NY, USA, 38-47.

[16] Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Learning sentiment from students' feedback for real-time interventions in classrooms. In Adaptive and Intelligent Systems (pp. 40-49). Springer.

[17] Rashid, A., Asif, S., Butt, N. A., & Ashraf, I. (2013). Feature Level Opinion Mining of Educational Student Feedback Data using Sequential Pattern Mining and Association Rule Mining. International Journal of Computer Applications,81(10), 31-38.

[18] Gamon, M., Aue, A., Corston-Oliver, S., & Ringger, E. (2005). Pulse: Mining customer opinions from free text. In Advances in Intelligent Data Analysis VI(pp. 121-132). Springer Berlin Heidelberg.

[19] Nitin, Gokran Ila; Shankararaman, Venky; Gottipati Swapna. (2015) Analyzing Educational Comments for Topics and Sentiments: A Text Analytics Approach. (2015). Frontiers in Education Conference 45th, October 2015.

[20] Polarity Analyser from https://github.com/sayonetech/text-polarity-analyser/blob/master/polarity_analyser.py

[21] Loria, S. (2014), Textblob: Simplified Text Processing. http://textblob.readthedocs.org/en/dev

[22] Brinthaupt, Thomas M.; Clayton, Maria A.; Draude, Barbara J.; Calahan, Paula T. (2014)How Should I Offer This Course? The Course Delivery Decision Model (CDDM). Journal of Online Learning & Teaching . Jun2014, Vol. 10 Issue 2, p326-336. 11p.

[23] Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: vector space models of semantics. J. Artif. Int. Res. 37, 1 (January 2010), 141-188.

[24] Rehr uv rek, Petr Sojka Software Framework for Topic Modelling with Large Corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. (22 May 2010), pp. 45-50