

# CONTEXT ANALYSIS-BASED AUTOMATION OF STUDENT DEBARRING SYSTEM

V. Kumuda Valli, Sauradip Sengupta

Under the guidance of

Dr. Siddharth Swarup Rautaray, Dr. Manjusha Pandey

**Abstract-** Students must document their leave of absence for not attending classes for a specific period where they are required to provide appropriate and genuine reasoning for the same. These applications need to be passed through an automated system that determines whether they are genuine and marks their attendance accordingly.

The disadvantage of the current automated system is that it only considers the content typed by the student as a reason for not attending the class and not the context of it. This method of examining student responses might overlook a genuine reason why a student was not attending the class, and marking them absent might eventually lead to them getting debarred from attending examinations. Thus, it does not give a clear idea of the reasoning as to why the said student is not attending class.

Text analysis is a machine learning technique used to automatically derive insights from unstructured data. This research proposes using text analytics to contextualize the student debarring system. To achieve this, we are employing the use of the bag of words module and the NLTK library present in python to develop a custom dictionary of keywords about valid reasons for attending classes.

**Index Terms-** Data analysis, Context orientation, Examination system, Automation

## I. INTRODUCTION

Our current education system gives heavy importance to attendance as it is a measure that not only evaluates the student's academic performance but his overall discipline and regularity. But maintaining a record of all the attending students became quite tedious for the professors, and sometimes the support staff is known to make errors. Thus, having an automated system for attendance facilitates the professors greatly.

In the existing system, the student gets a '1' or a positive value for all the classes he/she has attended and a '0' or a null value for the classes they didn't. Therefore, the percentage can be calculated by taking the sum of the positive values, dividing it by the total number of classes held, and multiplying the quotient by 100. Based on this percentage, a student might be debarred from giving their examinations, if this percentage falls too low.

But this existing system, though automated still has major disadvantages. It only considers whether the student was present or not, but doesn't examine the reason behind it. There might be a multitude of justifiable reasons why the student was not present in this class, like any health issue, or an emergency family situation. Or in other cases, the class might not have been held at all, due to a natural disaster, or

the professor falling ill. Despite the myriad of reasons, this system would simply award the student a '0' or a null value, affecting their attendance percentage and leading to them getting debarred.

Text analytics is an automated process that translates large volumes of unstructured text into quantitative data to analyze and, map trends, and form patterns. It is a machine learning technique that when combined with data visualization tools can help us understand the data better thus making better decisions Bag of words is an NLP technique of text modeling and a method of feature extraction with text which describes the occurrence of words within a document. The model gains its name as it disregards the order of words in which they occur, and the structure of the document, along with grammatical details. It is only concerned about where the known words occur, not how. The Natural Language Toolkit (NLTK) is a platform used for building Python programs that work with human language data for application in statistical NLP. It contains text-processing libraries for tokenization, parsing, classification, stemming, tagging, and semantic reasoning. It also includes graphical demonstrations and sample data sets.

We propose to make use of text analysis to contextualize the data and employ the Bag of Words module and use the NLTK library in Python.

## **II. STATEMENT OF THE PROBLEM AND OBJECTIVE**

The proposed work is to make a context-aware examination debarring system based on a data analysis algorithm that would help in a context-based automated debarring system.

## **III. BACKGROUND STUDY AND TECHNOLOGY GAPS IDENTIFIED**

Some previous papers published in this field is

Year of Publication	Title	Proposition	Advantages	Disadvantages
2019	Making the student voice count: using qualitative student feedback to enhance the student experience [1]	Outlines how qualitative data can be used to gain insights into the experience of different students including online and on-campus students. [2] It employs a conceptual and semantic analysis of the text. The program examines textual material for the frequency and co-occurrence of concepts that cluster together in textual data and then generates	Concepts are presented as grey dots. Groups of semantically related concepts are indicated as themes which are shown by circles. Similar concepts gather in the maps according to their semantic relationship to other concepts, indicating that adjacent concepts are related. Leximancer software used here in this study, adds objectivity to the findings and increases the reliability of findings.	1) To prevent anonymity, plural and singular words were grouped for the algorithm to work but the student in concern of this research paper can be absent for multiple classes at once. 2) Coder reliability can be an issue during manual coding processes and lead to human interpretation biases or inconsistencies due to preconceived ideas by human coders.

		visual concept maps to present how concepts relate to one another		
2020	Text Classification of College Attendance[3]	Explores the natural language characteristics of college-going choice among high school students[4]	Indicates that there are some distinct, recognizable topics in the corpus, showing graphically this homogeneity of topics across essays.	Should have posed more open-ended questions
2020	Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs [5]	This paper is a framework to automatically analyze the opinions of students expressed in reviews.[6]	This paper is a framework to automatically analyze the opinions of students expressed in reviews.	This paper is a framework to automatically analyze the opinions of students expressed in reviews.
2021	Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study [7]	Provided an analysis of the related literature by applying a systematic mapping study method [8]	Identified a variety of challenges regarding the application of sentiment analysis to examine students' feedback. Recommendations and future directions to address these challenges have been provided.	Systematic Literature Reviews (SLR) have not been conducted.
2021	Sentiment Analysis and Topic Modeling on Tweets about Online Education during COVID- [9]	This study investigates the effectiveness of online education by analyzing the sentiments of its stakeholders' using social media data.[10]	Topic modeling through LSA suggests that the uncertainty of opening date institutions is among the most concerning topics for students. Additionally, the lack of technical skills and network challenges in rural areas are major concerns for the students.	The dataset used in this study has been obtained by the Twitter API using the keywords related to the topic, thus making the dataset extremely sensitive even to minor changes in the keywords.
2021	Student feedback sentiment analysis: A review[11]	Extracts data from students' comments and performs [12]	Extracts features from the text entered in the form of comments and take into account, the	1) The model only seems to be performing well on a small dataset

		sentiment analysis on them using various methods like the likerat scale, word cloud, and sentiment score associated with each comment	emotions associated with each feedback which satisfies the main goal of our project	Multiple class assignments. Here the comments are rated as good bad or neutral whereas our model needs to be strictly categorical (yes or no)
2021	Capturing Student Feedback and Emotions in Large Computing Courses: A Sentiment Analysis Approach[13]	Student emotions are analyzed via sentiment analysis to provide teachers with an appropriate feedback mechanism[14]	Sentiment analysis is performed to make accurate visualization of emotions which were then passed through a sentiment analysis function thus quantifying each emotion. Data analysis was performed to summarize and make predictions of how students found certain classes by the frequency of words under each emotion	Deals with a wide array of emotions for text analysis to work. In the context of this research study, the emotions to be mapped with attendance could even be singular which could render this model inefficient
2021	Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study[15]	Employs NLP and deep learning methods to classify research papers and classified them into 3 broad categories based on aspect sentiment analysis conducted on opinions stated by students based on a questionnaire conducted.[16]	It systematically categorizes the research, frequencies, and visual summaries of publications, trends of certain words, etc. This gave an overview of the body of knowledge and the research area and identify the number of publications and the type of research and results.	The grouping in this method excludes certain types of papers. In the context of this paper, however, they require every possible reason stated by an absentee to make accurate predictions
2022	Online engagement and performance on formative	Recommends using a variety of educational technologies to	Leveraging engagement through self-regulated learning and self-assessment by using a	Does not extend SDT and engagement theory to TEL education

	assessments mediate the relationship between attendance and course performance[17]	pave multiple pathways to academic success [18]	variety of technologies is recommended to promote performance in higher education	
2022	Evaluating the construct validity of text embeddings with application to survey questions[19]	View the quality evaluation of text embedding models as a measurement validity problem[20]	Is grounded in measurement theory, providing a clear framework about what needs to be tested, and applied to many substantive research questions in computational social science.	1) Focuses only on the validity of text embeddings. 2) The synthetic survey question data set and the ESS data do not cover nearly most survey question types

#### IV. PROPOSED MODEL / TOOL.

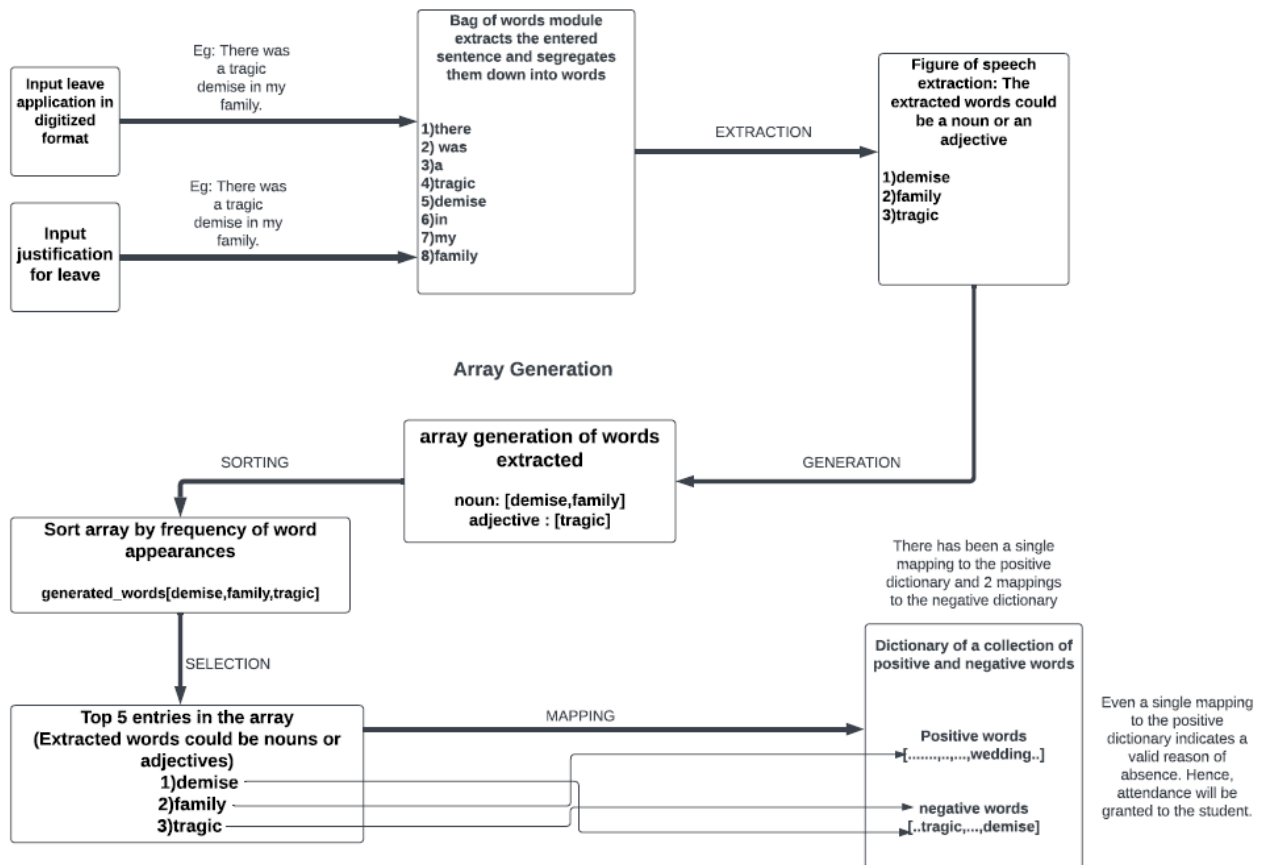


Figure 1: Proposed Model

## V. IMPLEMENTATION AND RESULTS

For the implementation of the said project we have devised an approach wherein:

The leave application written by the concerned student will be entered into our automation system where the nouns and adjectives present in the letter will be segregated. Based on the words obtained, we will construct a list where we employ the use of a bag of words to populate it with the occurrences of nouns and adjectives extracted based on their frequencies i.e. the number of times they have occurred within the letter after which we prioritize the top 5 occurrences of the extracted words (based on the frequency of occurrence) which are then mapped to a dictionary consisting of positive and negative words where the positive dictionary consists of words indicating possible reasons of absence and the negative dictionary consists of words indicating the reasons we are omitting to grant attendance to the student. After mapping the extracted words to their respective presence in our created dictionaries of possible positive and negative reasons we devise an analysis. The number of mappings to the positive and negative dictionaries is counted. In case the number of positive dictionary mappings is more than negative dictionary mapping, we derive a conclusion that the concerned student has not expressed a valid reason for absence through their application. On the contrary, if the number of negative dictionary mappings is more, we conclude that the reason for absence expressed by the concerned student is valid. In case the number of mappings on either dictionary is equal, we present this as a neutral case and throw an error accordingly.

Take an example,

*Respected sir,*

*I am down with a fever and flu because of this I will not be able to come to the office for at least 5 days. As per my family doctor, it is best that I take rest and recover properly before resuming work. I have asked Sauradip to check on my clients and will try to periodically check my email if you need anything urgent. Please grant me leave for the aforementioned period. If you need additional information, please let me know.*

*Yours Sincerely,*

*Rajesh*

If this input is given to our program, the following output would be generated

```
no of words that got mapped with +ve dictionary: 1
no of words that got mapped with -ve dictionary: 4
```

## DEBARRING SYSTEM

```
ex_percentage_of_attendance = 60

if flag == True:
    ex_percentage_of_attendance +=10

if ex_percentage_of_attendance >= 75:
    print(f"RESULT : NOT DEBARRED")
else:
    print("RESULT = DEBARRED")
```

✓ 0.3s

RESULT = DEBARRED

The number of negative words mapped in the letter is more than the number of positive words, so he has been awarded attendance for that class. But his prior attendance was 60%, and even after he got a hike of 10% due to his letter, he still fell short of the required 75%. Therefore he has been debarred from the examination.

Further implementation and a refined code will be presented at the conference

## VI. CONCLUSION

We started with a system where the students would get debarred from giving their attendance based on content alone. In the existing system, only their absence would be noted without the contextual reason behind the text. This might lead to misclassification in some cases. Our proposal addresses this problem, by introducing a context-based system that would greatly reduce the number of cases to be manually analyzed by the authorities. Our future scope for this proposal is to devise a solution to expand our dictionary of words and tackle edge cases in a more efficient manner.

## VII. REFERENCES

- [1] SHAH, MAHSOOD & PABEL, ANJA. (2019). MAKING THE STUDENT VOICE COUNT: USING QUALITATIVE STUDENT FEEDBACK TO ENHANCE THE STUDENT EXPERIENCE. JOURNAL OF APPLIED RESEARCH IN HIGHER EDUCATION. AHEAD-OF-PRINT. 10.1108/JARHE-02-2019-0030.
- [2] ABD-ELRAHMAN, A., ANDREU, M. AND ABBOTT, T. (2010), "USING TEXT DATA MINING TECHNIQUES FOR UNDERSTANDING FREESTYLE QUESTION ANSWERS IN COURSE EVALUATION FORMS", RESEARCH IN HIGHER EDUCATION JOURNAL, VOL. 9, PP. 12-23
- [2] ABD-ELRAHMAN, A., ANDREU, M. AND ABBOTT, T. (2010), "USING TEXT DATA MINING TECHNIQUES FOR UNDERSTANDING FREESTYLE QUESTION ANSWERS IN COURSE EVALUATION FORMS", RESEARCH IN HIGHER EDUCATION JOURNAL, VOL. 9, PP. 12-23
- [3] Z. KASTRATI, A. S. IMRAN AND A. KURTI, "WEAKLY SUPERVISED FRAMEWORK FOR ASPECT-BASED SENTIMENT ANALYSIS ON STUDENTS' REVIEWS OF MOOCs," IN IEEE ACCESS, VOL. 8, PP. 106799-106810, 2020, DOI: 10.1109/ACCESS.2020.300073
- [3] Z. KASTRATI, A. S. IMRAN AND A. KURTI, "WEAKLY SUPERVISED FRAMEWORK FOR ASPECT-BASED SENTIMENT ANALYSIS ON STUDENTS' REVIEWS OF MOOCs," IN IEEE ACCESS, VOL. 8, PP. 106799-106810, 2020, DOI: 10.1109/ACCESS.2020.3000739.
- [4] KASTRATI, Z.; DALIPI, F.; IMRAN, A.S.; PIREVA NUCI, K.; WANI, M.A. SENTIMENT ANALYSIS OF STUDENTS' FEEDBACK WITH NLP AND DEEP LEARNING: A SYSTEMATIC MAPPING STUDY. APPL. SCI. 2021, 11, 3986. [HTTPS://DOI.ORG/10.3390/APP11093986](https://doi.org/10.3390/app11093986)
- [5] PLUTCHIK, R. THE NATURE OF EMOTIONS. AM. SCI. 2001, 89, 344–350.
- [5] PLUTCHIK, R. THE NATURE OF EMOTIONS. AM. SCI. 2001, 89, 344–350
- [6] MUJAHID, M.; LEE, E.; RUSTAM, F.; WASHINGTON, P.B.; ULLAH, S.; RESHI, A.A.; ASHRAF, I. SENTIMENT ANALYSIS AND TOPIC MODELING ON TWEETS ABOUT ONLINE EDUCATION DURING COVID-19. APPL. SCI. 2021, 11, 8438. [HTTPS://DOI.ORG/10.3390/APP11188438](https://doi.org/10.3390/app11188438)
- [7] D.D. DSOUZA, D.P. DEEPIKA, E.M. NAYAK, N.D. ADESH SENTIMENTAL ANALYSIS OF STUDENT FEEDBACK USING MACHINE LEARNING TECHNIQUES  
INT. J. RECENT TECHNOL. ENG., 8 (1S4) (2019), PP. 986-991
- [8] ABDESSALEM W. K. B., AMDOUNI S. (2011). E-RECRUITING SUPPORT SYSTEM BASED ON TEXT MINING METHODS. INTERNATIONAL JOURNAL OF KNOWLEDGE AND LEARNING, 7(3), 220–232.  
<https://doi.org/10.1504/IJKL.2011.044542>
- [9] AGGARWAL C. C., ZHAI C. (2012). A SURVEY OF TEXT CLASSIFICATION ALGORITHMS. IN AGGARWAL C. C., ZHAI C. (EDS.), MINING TEXT DATA (PP. 163–222). NEW YORK, NY: SPRINGER.  
[HTTPS://DOI.ORG/10.1007/978-1-4614-3223-4\\_6](https://doi.org/10.1007/978-1-4614-3223-4_6)
- [10] N. ALTRABSHEH, M. COCEA, S. FALLAHKHAIR  
LEARNING SENTIMENT FROM STUDENTS' FEEDBACK FOR REAL-TIME INTERVENTIONS IN CLASSROOMS  
A. BOUCHACHIA (ED.), ADAPTIVE AND INTELLIGENT SYSTEMS. ICAIS 2014. LECTURE NOTES IN COMPUTER SCIENCE, SPRINGER, CHAM (2014)  
[HTTPS://DOI.ORG/10.1007/978-3-319-11298-5\\_5](https://doi.org/10.1007/978-3-319-11298-5_5)



[11]SARIS WE, OBERSKI DL, REVILLA M, ZAVALA-ROJAS D, LILLEOJA L, GALLHOFER IN, GRUNER T (2011) THE DEVELOPMENT OF THE PROGRAM SQP 2.0 FOR THE PREDICTION OF THE QUALITY OF SURVEY QUESTIONS

[12] PAOLA ADINOLFI, ERNESTO D'AVANZO, MILTIADIS D LYTRAS, ISABEL NOVO-CORTI, AND JOSE PICATOSTE. 2016. SENTIMENT ANALYSIS TO EVALUATE TEACHING PERFORMANCE. INTERNATIONAL JOURNAL OF KNOWLEDGE SOCIETY RESEARCH (IJKSR), VOL. 7, 4 (2016), 86--107

[13] NABEELA ALTRABSHEH, MIHAELA COCEA, AND SANAZ FALLAHKHAIR. 2014A. LEARNING SENTIMENT FROM STUDENTS' FEEDBACK FOR REAL-TIME INTERVENTIONS IN CLASSROOMS. IN INTERNATIONAL CONFERENCE ON ADAPTIVE AND INTELLIGENT SYSTEMS. SPRINGER, 40--49.

[14] KASTRATI, Z.; DALIPI, F.; IMRAN, A.S.; PIREVA NUCI, K.; WANI, M.A. SENTIMENT ANALYSIS OF STUDENTS' FEEDBACK WITH NLP AND DEEP LEARNING: A SYSTEMATIC MAPPING STUDY. APPL. SCI. 2021, 11, 3986. [HTTPS://DOI.ORG/10.3390/APP11093986](https://doi.org/10.3390/app11093986)

[14] KASTRATI Z, DALIPI F, IMRAN AS, PIREVA NUCI K, WANI MA. SENTIMENT ANALYSIS OF STUDENTS' FEEDBACK WITH NLP AND DEEP LEARNING: A SYSTEMATIC MAPPING STUDY. APPLIED SCIENCES. 2021; 11(9):3986. [HTTPS://DOI.ORG/10.3390/APP11093986](https://doi.org/10.3390/app11093986)

[15] LU, C., CUTUMISU, M. ONLINE ENGAGEMENT AND PERFORMANCE ON FORMATIVE ASSESSMENTS MEDIATE THE RELATIONSHIP BETWEEN ATTENDANCE AND COURSE PERFORMANCE. INT J EDUC TECHNOL HIGH EDUC 19, 2 (2022). [HTTPS://DOI.ORG/10.1186/S41239-021-00307-5](https://doi.org/10.1186/s41239-021-00307-5)

[15] LU, C., CUTUMISU, M. ONLINE ENGAGEMENT AND PERFORMANCE ON FORMATIVE ASSESSMENTS MEDIATE THE RELATIONSHIP BETWEEN ATTENDANCE AND COURSE PERFORMANCE. INT J EDUC TECHNOL HIGH EDUC 19, 2 (2022). [HTTPS://DOI.ORG/10.1186/S41239-021-00307-5](https://doi.org/10.1186/s41239-021-00307-5)

[16]ARUN R., SURESH V., MADHAVAN C. E. V., MURTHY M. N. N. (2010). ON FINDING THE NATURAL NUMBER OF TOPICS WITH LATENT DIRICHLET ALLOCATION: SOME OBSERVATIONS. IN ZAKI M. J., YU J. X., RAVINDRAN B., PUDI V. (EDS.), ADVANCES IN KNOWLEDGE DISCOVERY AND DATA MINING (PP. 391–402). BERLIN, GERMANY: SPRINGER. [HTTPS://DOI.ORG/10.1007/978-3-642-13657-3\\_43](https://doi.org/10.1007/978-3-642-13657-3_43)

[17] EPJ DATA SCI. 11, 39 (2022)

[HTTPS://DOI.ORG/10.1140/EPJDS/S13688-022-00353-7](https://doi.org/10.1140/EPJDS/s13688-022-00353-7)

[17] EPJ DATA SCI. 11, 39 (2022)

[HTTPS://DOI.ORG/10.1140/EPJDS/S13688-022-00353-7](https://doi.org/10.1140/EPJDS/s13688-022-00353-7)