S. T. Patil1, Om Chakane2(11910565), Pranali Pardeshi2(11911122), Arshad Patel2(11910269), Pratiksha Patil2(11910542), Prashanth Bijamwar2(11910304)

Department of Computer Engineering, VIT Pune

[[1]](#footnote-1)

**Automated Subjective Answer Checker**

*Abstract* — With the growth in population, and necessity of education, it is proving hard for assessors to check the relevance and accuracy of the answers written by students. In this work we present a new approach of calculating the score of each answer (entered by student) based on the dataset on which the machine is trained. For every answer being entered it was important to reward points based on the usage of words and their importance.

Keywords — Deep Descriptive Answer Scoring, Long Short Term Memory (LSTM), Machine Learning (ML), Recurrent Neural Network (RNN),

# Introduction

One of the significant part of education is the examination which is a measure of students learning ability. After examination, the teachers spend most of their time for evaluating the marks of the students and the evaluation takes bulk usage of human effort, time and cost. An automated subjective answer checker can reduce the efforts during the evaluation. Today, many automated evaluation systems exist and they analyze a piece of text based on semantics, spelling and context. The evaluation of descriptive answers is still an open problem. Major problem among the existing systems is their efficiency. The subjective nature of the answer scripts evaluation corresponds to variations in awarding of grades by different human evaluators. As seen as an unfair method of grading by students. This difficulty of grading answer sheets can be rectified by answer script evaluation tools which grade answer scripts automatically. An automated subjective answer checker must be capable of scoring the answer papers within the range of those awarded by human evaluators. It must be consistent in the way it grades the answer scripts and thus it can save the time and cost of evaluation. Currently there exists many automated essay evaluation systems based on keyword matching, sequence matching and using bag of words model. Nowadays, many researchers focus on the fastest and most accurate area of machine learning, i.e the deep learning. By introducing an automated subjective answer checker using deep learning can improve the efficiency of answer evaluation. Here, a system is introduced to automatically score answer scripts. The main aim of this model is to extract the semantics to efficiently represent the text in answer scripts and develop a model from the key and evaluated answer scripts to grade non evaluated answer scripts using deep learning. The objective of this model is to extract the semantics to efficiently represent the text in answer scripts and develop a model from the key and evaluated answer scripts to grade non evaluated answer scripts using deep neural networks. It is a combination of NLP and machine learning. The learning is done by using the Recurrent Neural Network and LSTM cells. Deep Neural networks are able to capture semantics of text in order to and the similarity between texts. For efficiently represent semantics of the sentences as embedding vectors we use LSTM-Recurrent Neural Networks .The embed-ding vector corresponding to the last word will be the entire representation of the sentence in its semantic form. The fully connected neural network layer automatically learns and predicts the scores for the semantic representation of the answer based on previous knowledge. Significance of the system is that it is useful for valuation of large number of answer scripts and fast valuation. It is a new step in the fields of information retrieval, document similarity, semantic evaluation and essay evaluation. The goal of the system is to replace the traditional human evaluation of the answer sheet that depends on several factor such as time, mindset, presentation style and so on.

# Literature Review

[1] Sheeba Praveen, Published in International Journal of Innovative Research in Computer and Communication Engineering. Vol. 2, Issue 11, November 2014.As observed that these systems contain only multiple choice questions and there was no provision to extend these systems to subjective questions. The paper presents an approach to check the degree of learning of the student/learner, by evaluating their descriptive exam answer sheets. By representing the descriptive answer in the form of graph and comparing it with standard answer are the key steps in our approach. Main drawback of the system will be Non Mathematical subjects only., Less efficiency in similarity matching., Multiple sentence answers are difficult to grade.

[2] Algorithm for Automatic Evaluation of Single Sentence Descriptive Answer International Journal of Inventive Engineering and Sciences (IJIES) ISSN: 23199598, Volume-1, Issue-9, August 2013.The proposed system was a system that would seek to implement an application which will be able to evaluate the subjective answer to a question. It will allot the marks according to the percentage of accuracy present in the answer. This system was a software system in which user will be authenticated by using user login. After the authentication, users will be provided with the questions. The proposed system was designed to evaluate answers for five users providing five different answers. The standard answer is stored in the database with the description meaning and keywords. Then it will evaluate each answer by matching the keywords or the key concepts as well as its synonyms with the standard answer. It will also check the grammar and spellings of the words. After the evaluation, it will grade the answer depending on the correctness of the answer. Here Drawback will be, Not easy to grade multiple sentences. As seen drawbacks of the other systems .In our system we overcome the drawbacks such as manual efforts, consume time, mistakes. Our main aim is that our system will be helpful for various universities and academic.

# Methodology/Experimental

## Materials/Components/Flowchart/Block Diagram/Theory

## Information about LSTM

*Long short-term memory (LSTM)* is an artificial [recurrent neural network](https://en.wikipedia.org/wiki/Recurrent_neural_network)(RNN) architecture[[1]](https://en.wikipedia.org/wiki/Long_short-term_memory#cite_note-lstm1997-1) used in the field of [deep learning](https://en.wikipedia.org/wiki/Deep_learning). LSTM has feedback connections. It can not only process single data points (such as images), but also entire sequences of data (such as speech or video

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three *gates* regulate the flow of information into and out of the cell.

The advantage of an LSTM cell compared to a common recurrent unit is its cell memory unit. The cell vector has the ability to encapsulate the notion of forgetting part of its previously stored memory, as well as to add part of the new information. To illustrate this, one has to [inspect the equations](https://theaisummer.com/understanding-lstm/#lstm-long-short-term-memory-cells) of the cell and the way it processes sequences under the hood.

The LSTM RNN will learn the temporal data from the embedding layer and the embedding vector corresponding to the final glove vector will be the semantic representation of the entire answer. This is given as input to the dropout layer and then to the fully connected neural network layer with a softmax activation function.

## Module of Proposed System

The system comprises of four modules and they are Login module. Information extraction module. Weighting module and Score Generation module.

*LOGIN MODULE*

The login module authenticates both the user & the admin. Once the authentication it done, the user and admin can perform their individual activities.

1. ***Admin Login***: The admin needs to enter the user name and the password for authentication. Once authenticated, the admin can now create the question & store the answer for the same in the database. The admin can also add students, subjects & tests for those subjects. The admin should keep all the keywords present in the answer in capital letters. Thus admin should store the answer with a subject expert helping him to identify the keywords present in the answer. The question will be displayed to the user and the answer stored will be used as the standard answer for comparing with the user's answer.

2. ***User Login***: The user login enables the user to write the answer for the question displayed. The user is asked to enter his login id and test id. If all the credentials are satisfied then the student is redirected to the page when the question & a text box for the answer is displayed. Once the user has completed writing the answer, he/she can submit the answer for evaluation

Fig. 1 shows the accuracy obtained for each epochs for the different LSTM-RNN types. A simple LSTM contains only one LSTM layer whereas in a deep LSTM, many LSTM layers will be there for sentence embedding. LSTM in its core, preserves information from inputs that has already passed through it using the hidden state. Unidirectional LSTM only preserves information of the past because the only inputs it has seen are from the past. Cohen’s kappa is used to measures inter- annotator agreement. It is defined as K = (Po – Pe)/ (1= Pe) (1)

where Po is the empirical probability of agreement and Pe is the expected agreement when both annotators assign labels randomly. Pe is estimated using a per-annotator empirical prior over the class labels. The kappa statistic, which is a number between 1 and 1. The maximum value

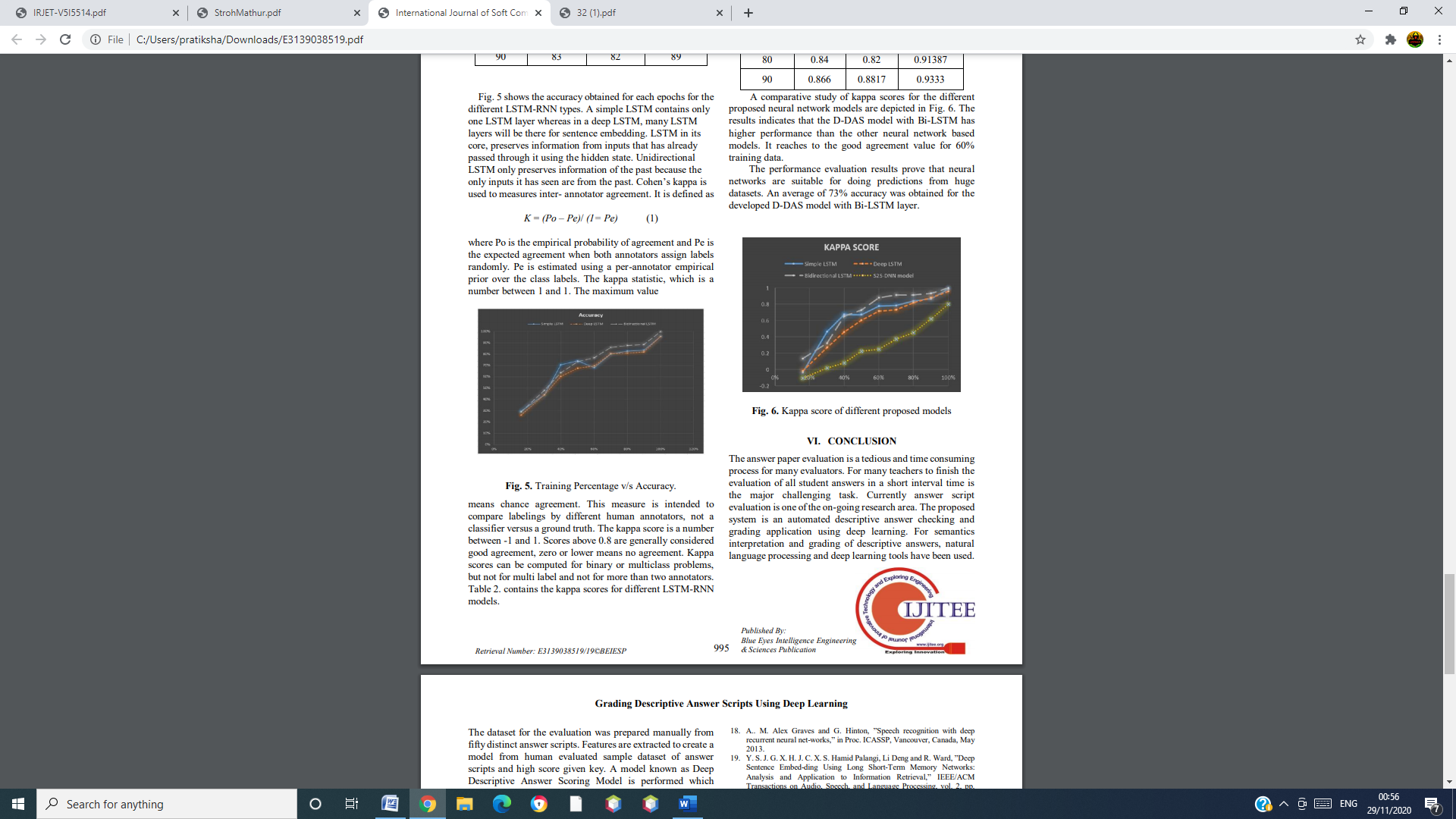


Fig 1 training percentage vs. accuracy

means chance agreement. This measure is intended to compare labeling by different human annotators, not a classifier versus a ground truth. The kappa score is a number between -1 and 1. Scores above 0.8 are generally considered good agreement, zero or lower means no agreement. Kappa scores can be computed for binary or multiclass problems, but not for multi label and not for more than two annotators.

A comparative study of kappa scores for the different proposed neural network models are depicted in Fig. 2

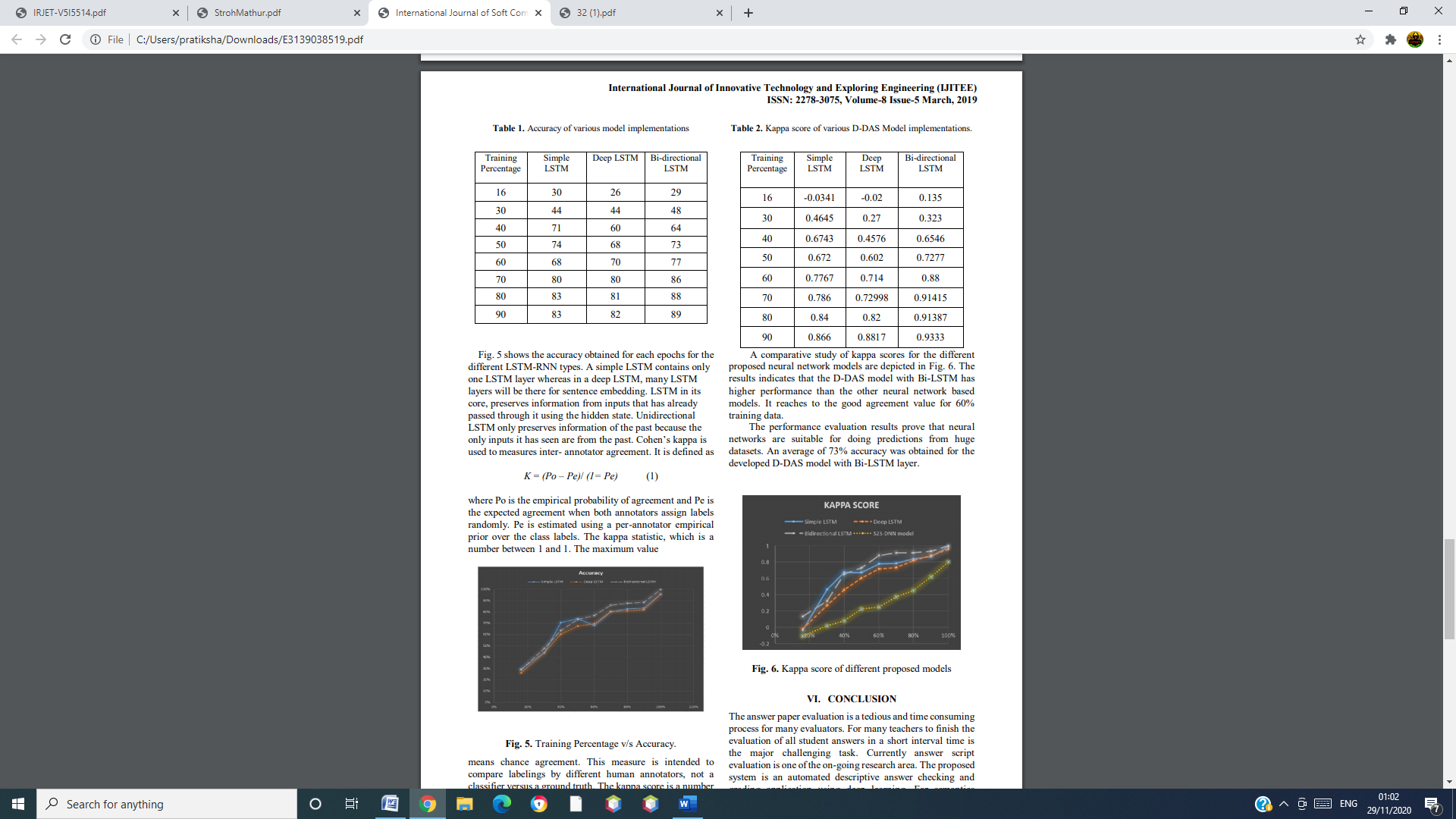


Fig 2 Kappa score of different proposed models

*Homographic Matrix*

Briefly, the homography relates the transformation between two planes. That is we represent a point in one plane with respect to another plane.

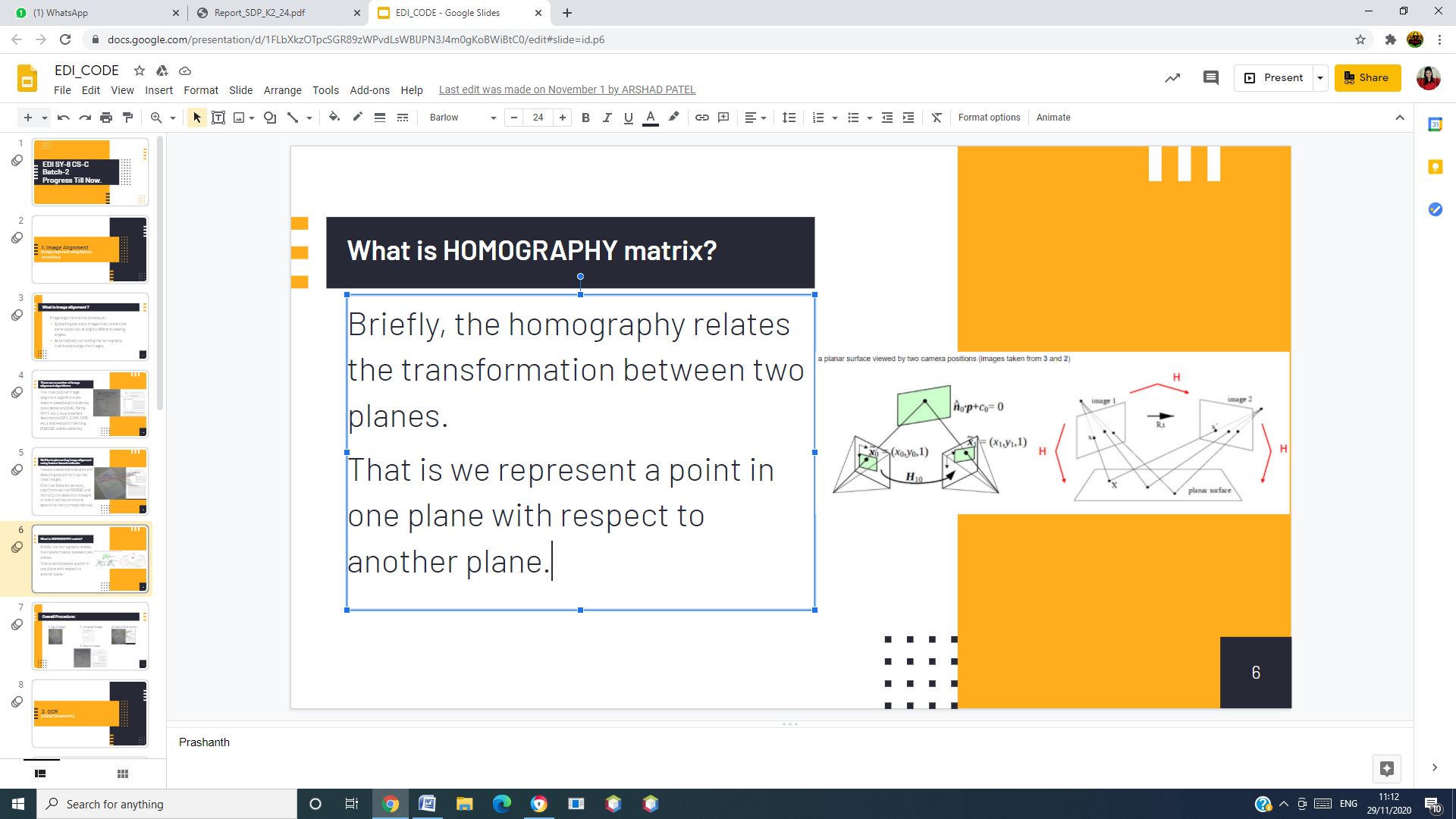
**

Fig 3 Homography Matrix

## Synthesis/Algorithm/Design/Method

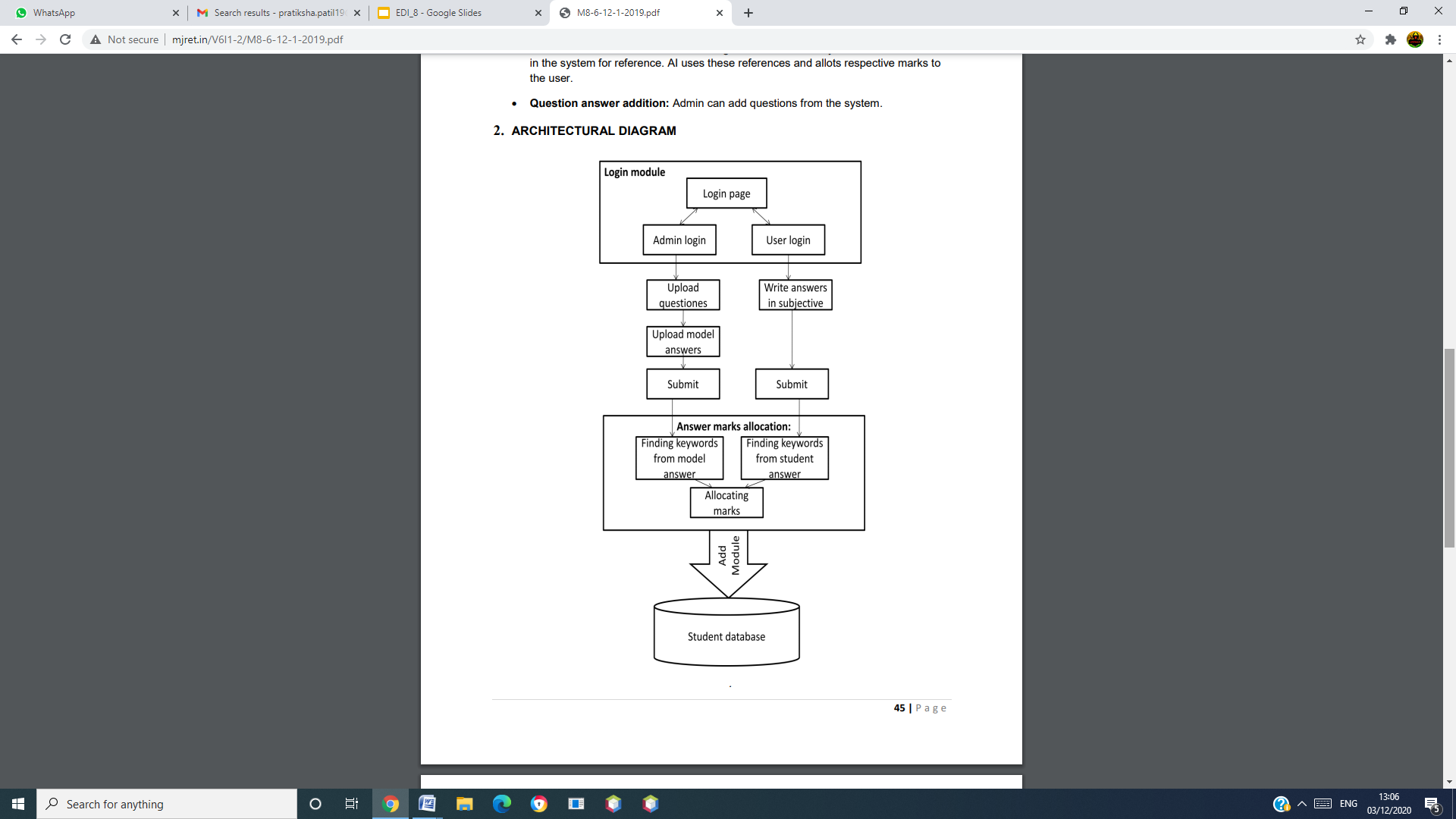


Fig 4 Architecture Diagram

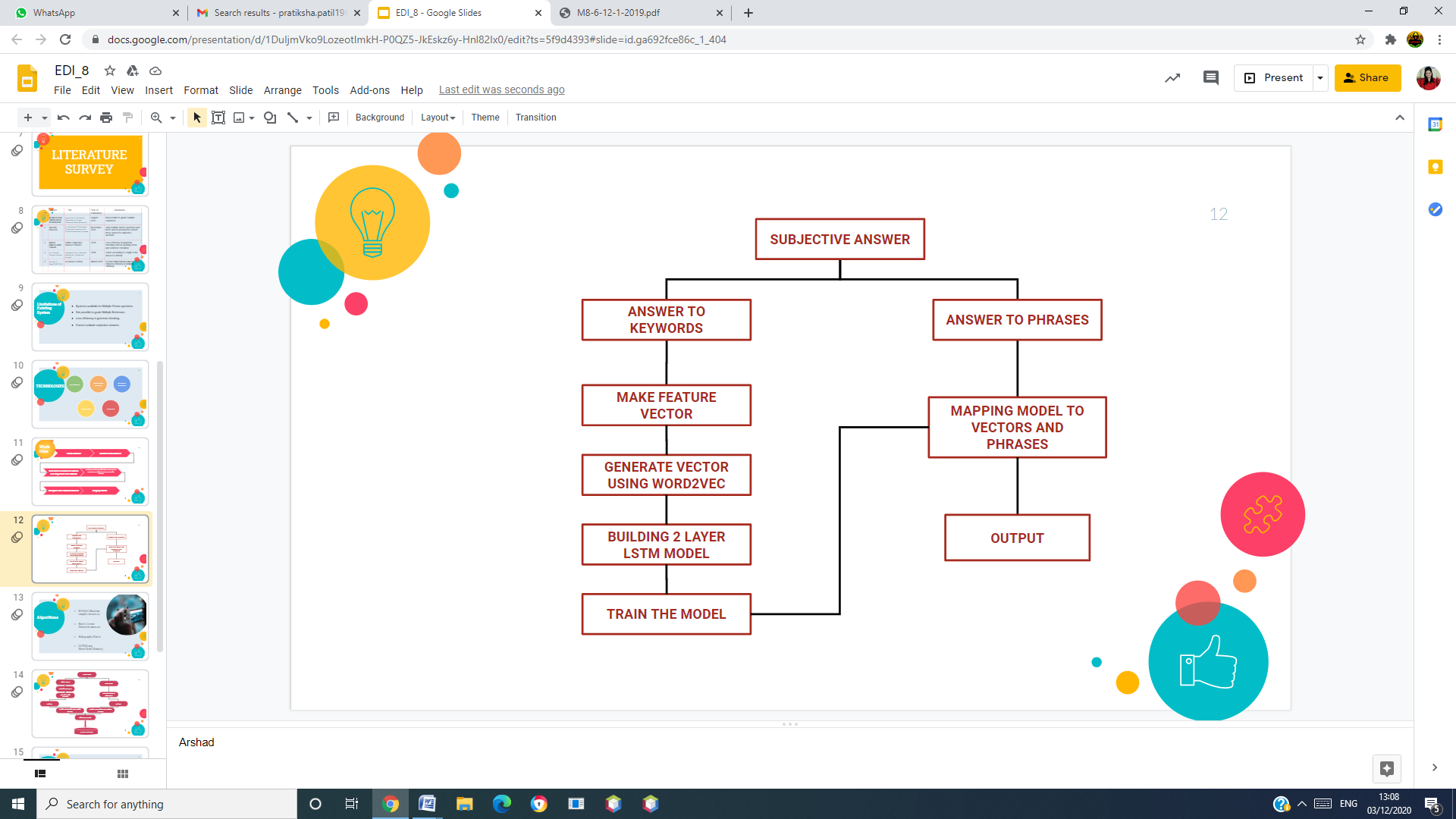


Fig 5. Block Diagram

## Characterization/Pseudo Code/ Testing

## 

Fig 6. Matching Keywords

# Results and Discussions

Web Implementation using Django. Created Database in Pg-Admin. Added your database name in settings.py. Installed psycopg2 database adapter.

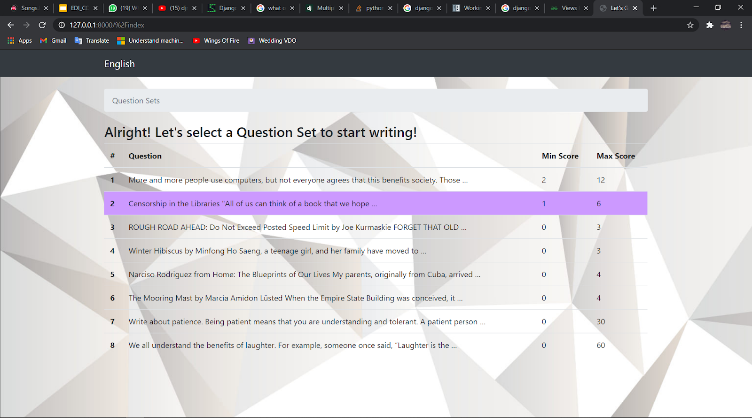


Fig 7.

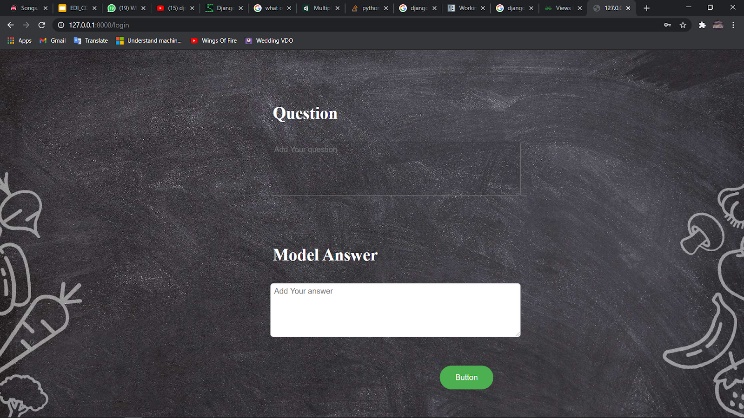


Fig 8. Store the model answer



Fig 9. Guidelines for Teachers

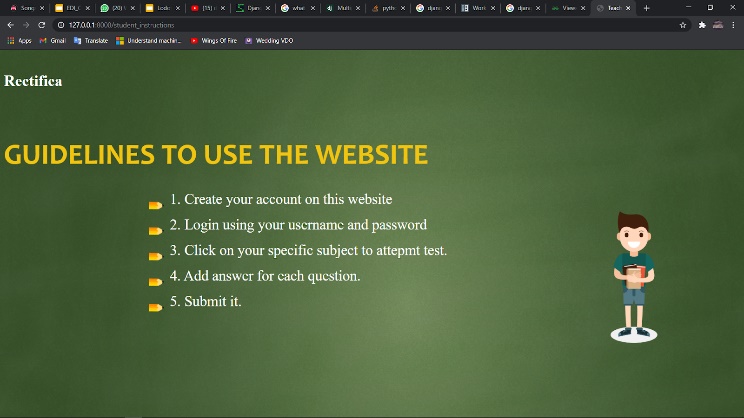


Fig 10. Guidelines to use Website

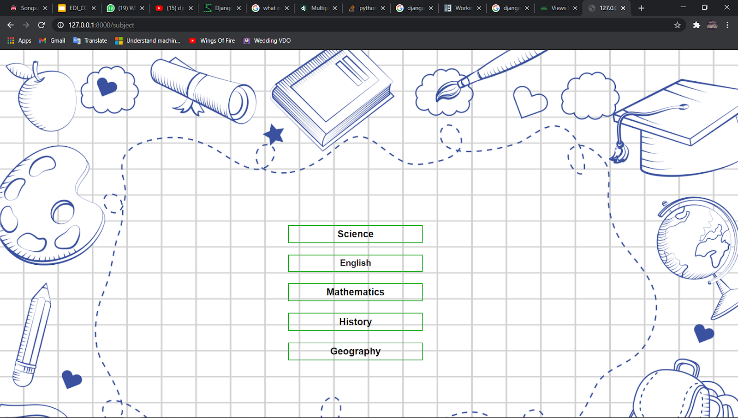


Fig 11. Selection of Subject

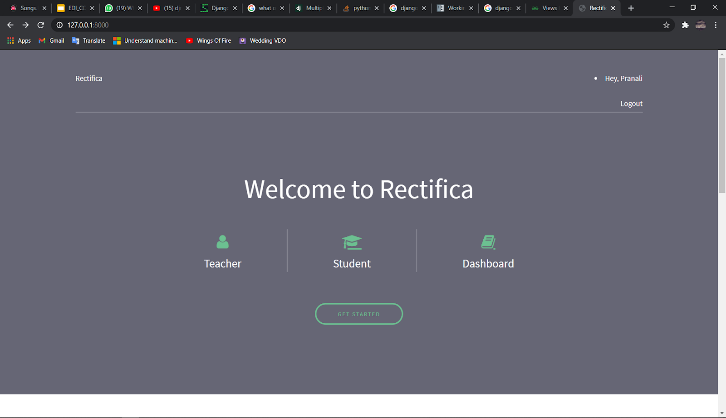


Fig 12. Home Page

Now, here are results from application.

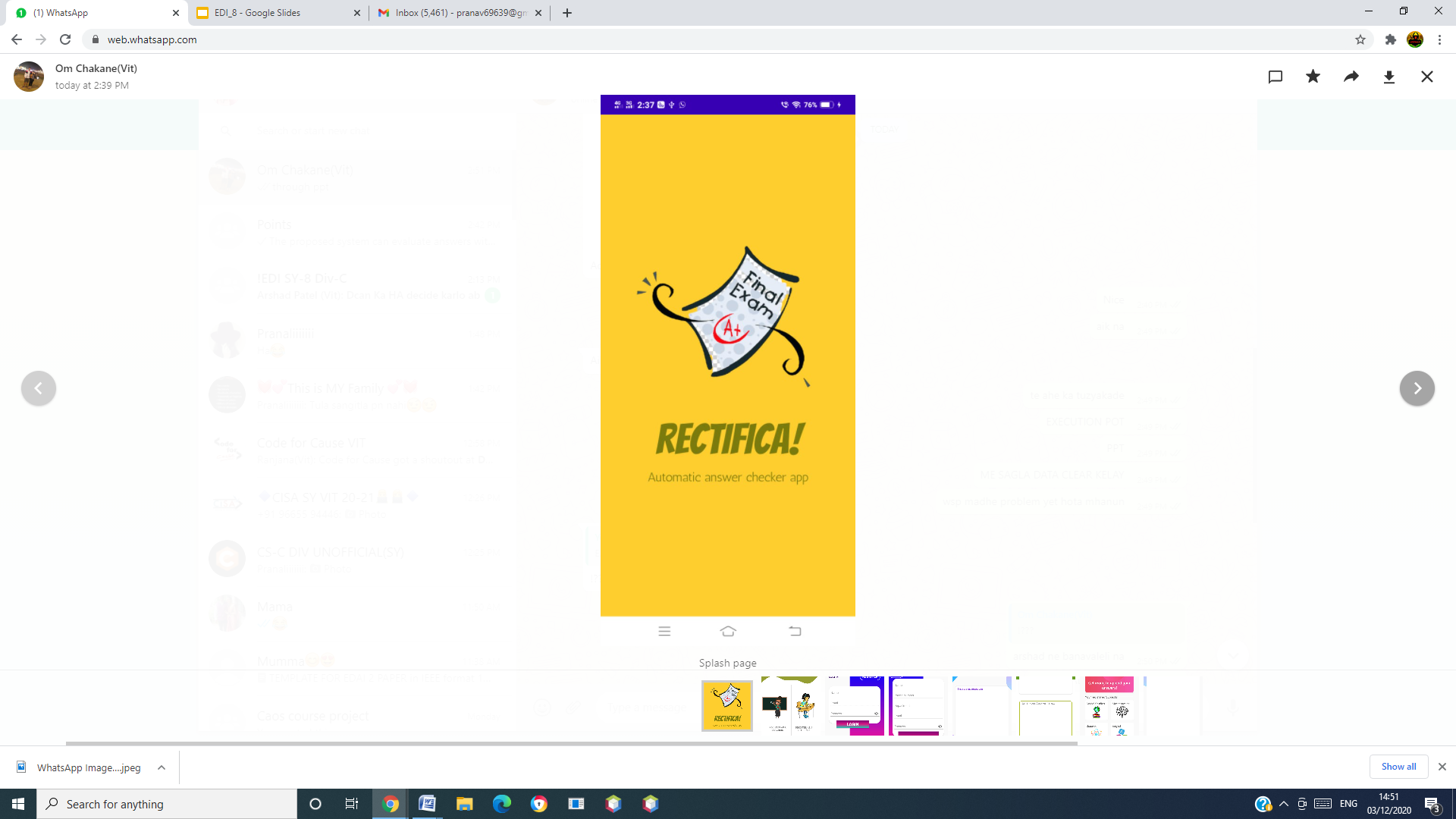


Fig 13 Splash page

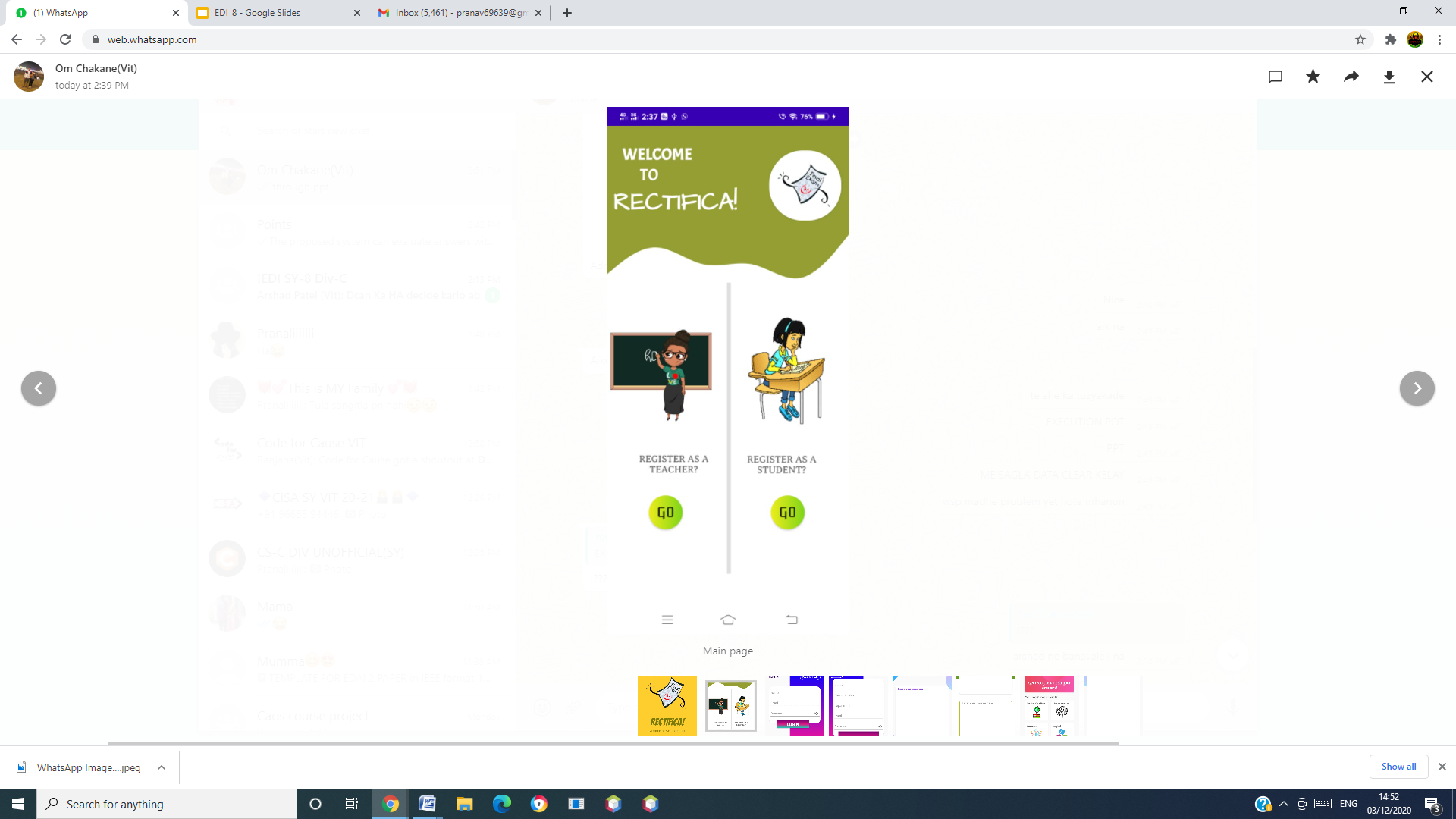


Fig 14 Main page

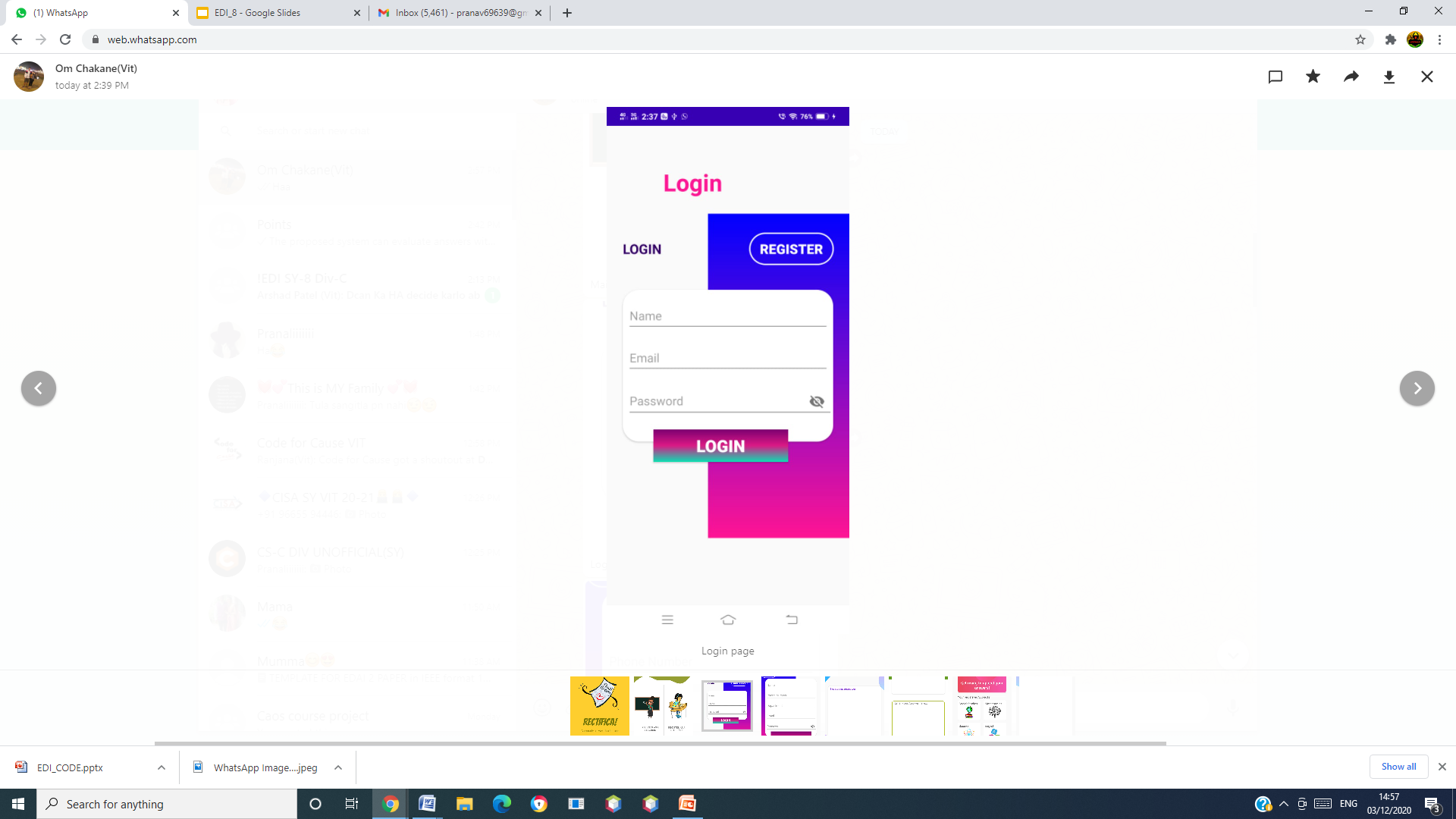


Fig 15 Login page

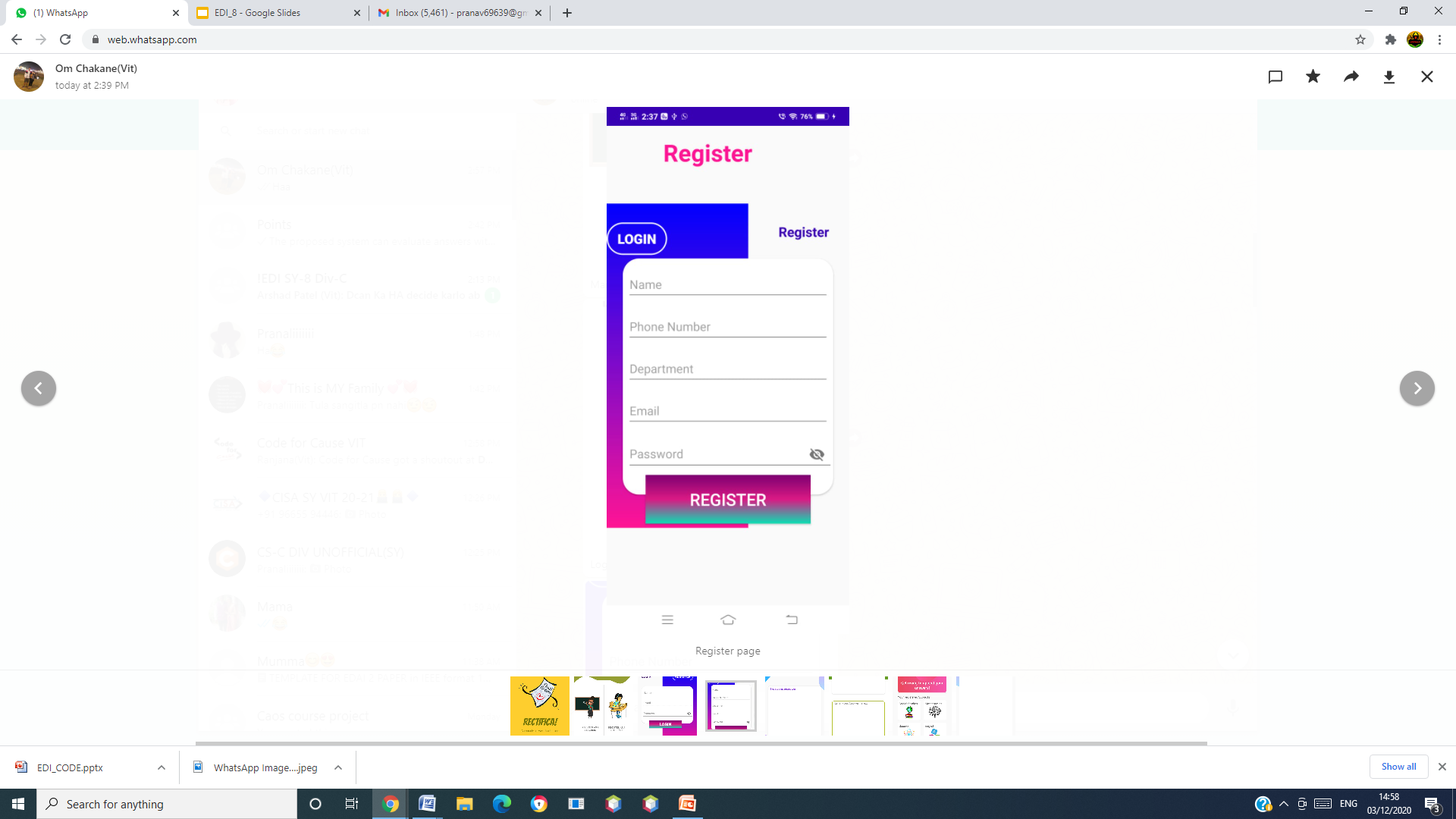


Fig 16 Register page

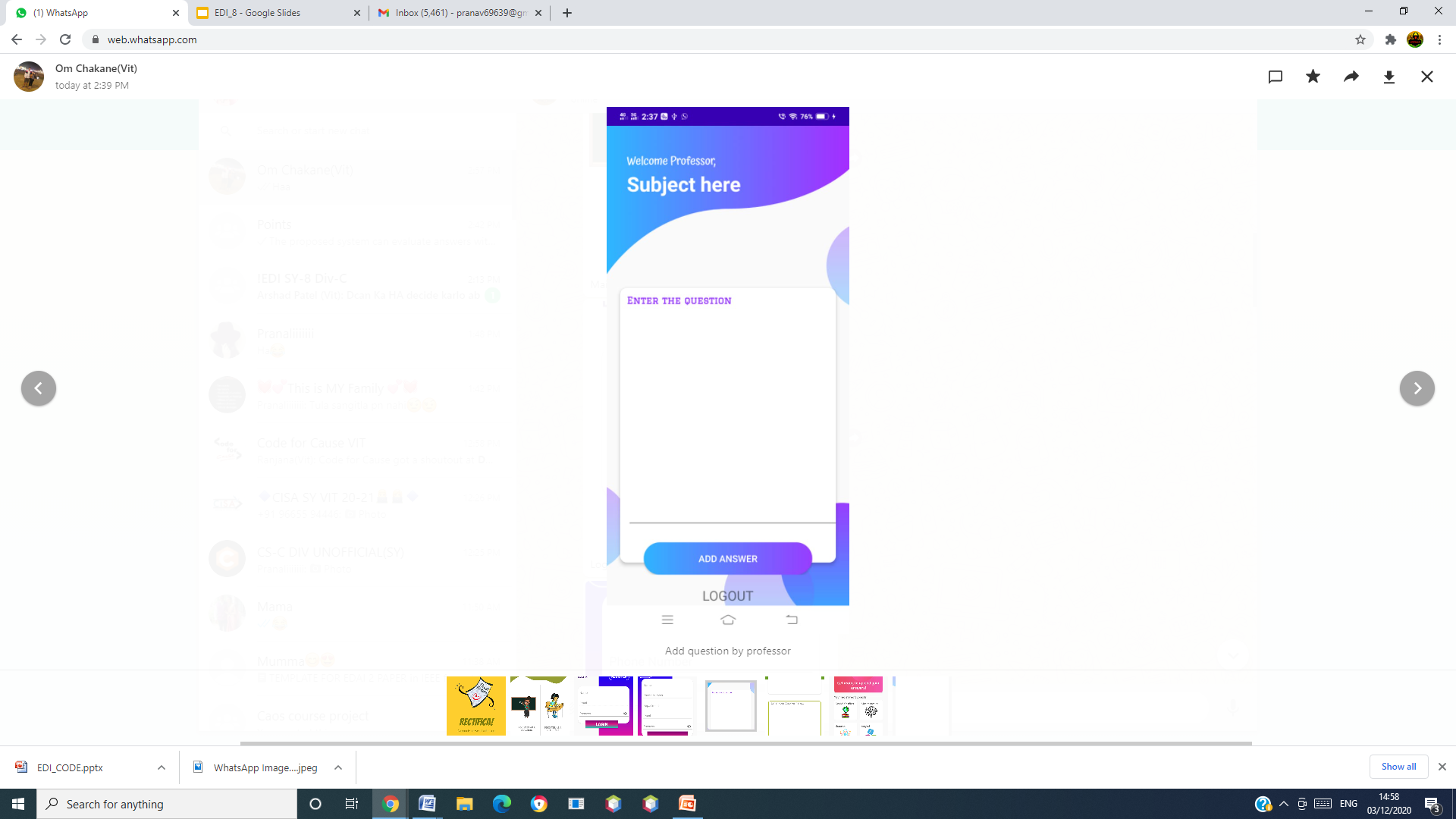


Fig 17 Add question by professor

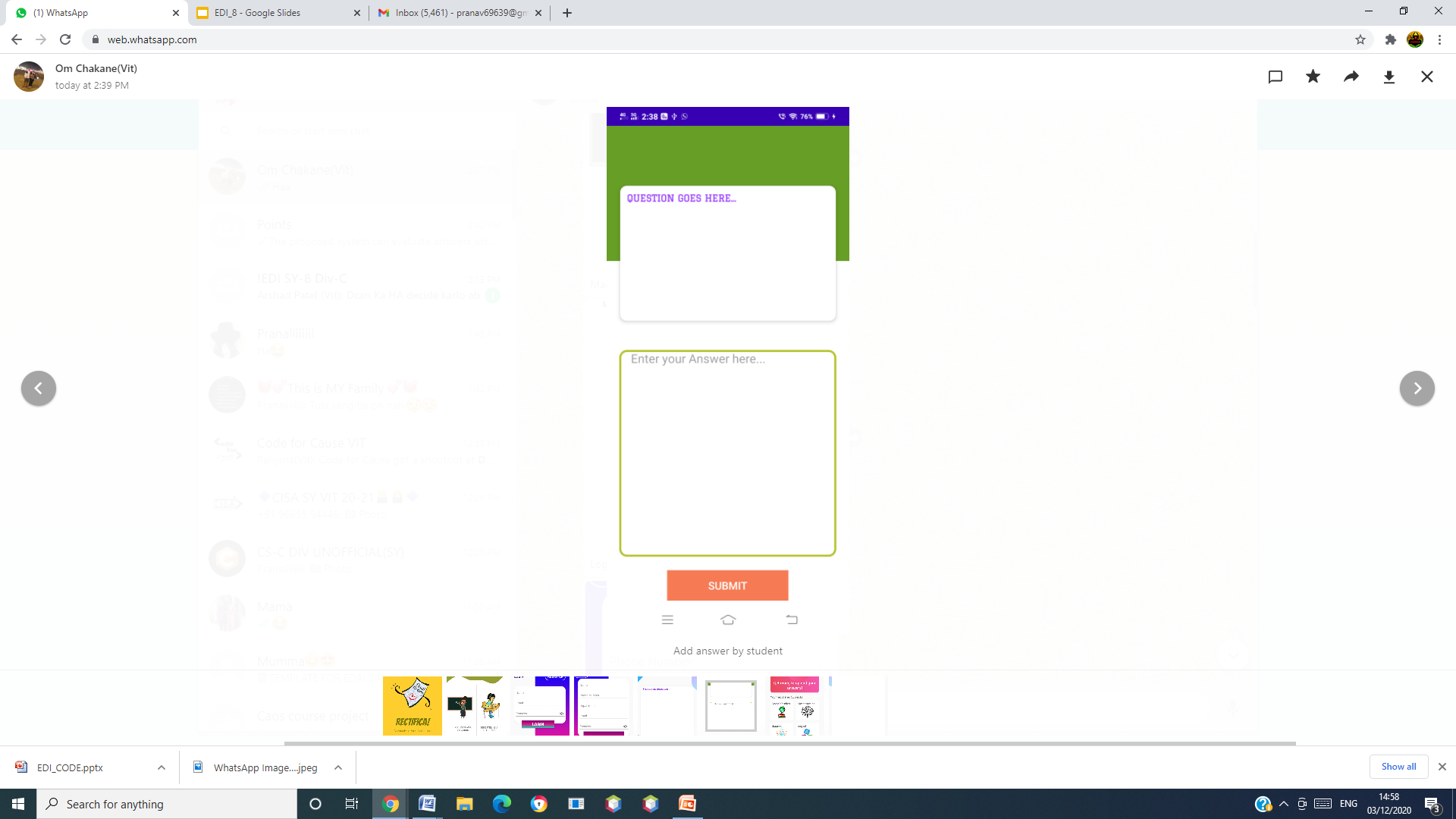


Fig 18 Add answer by student

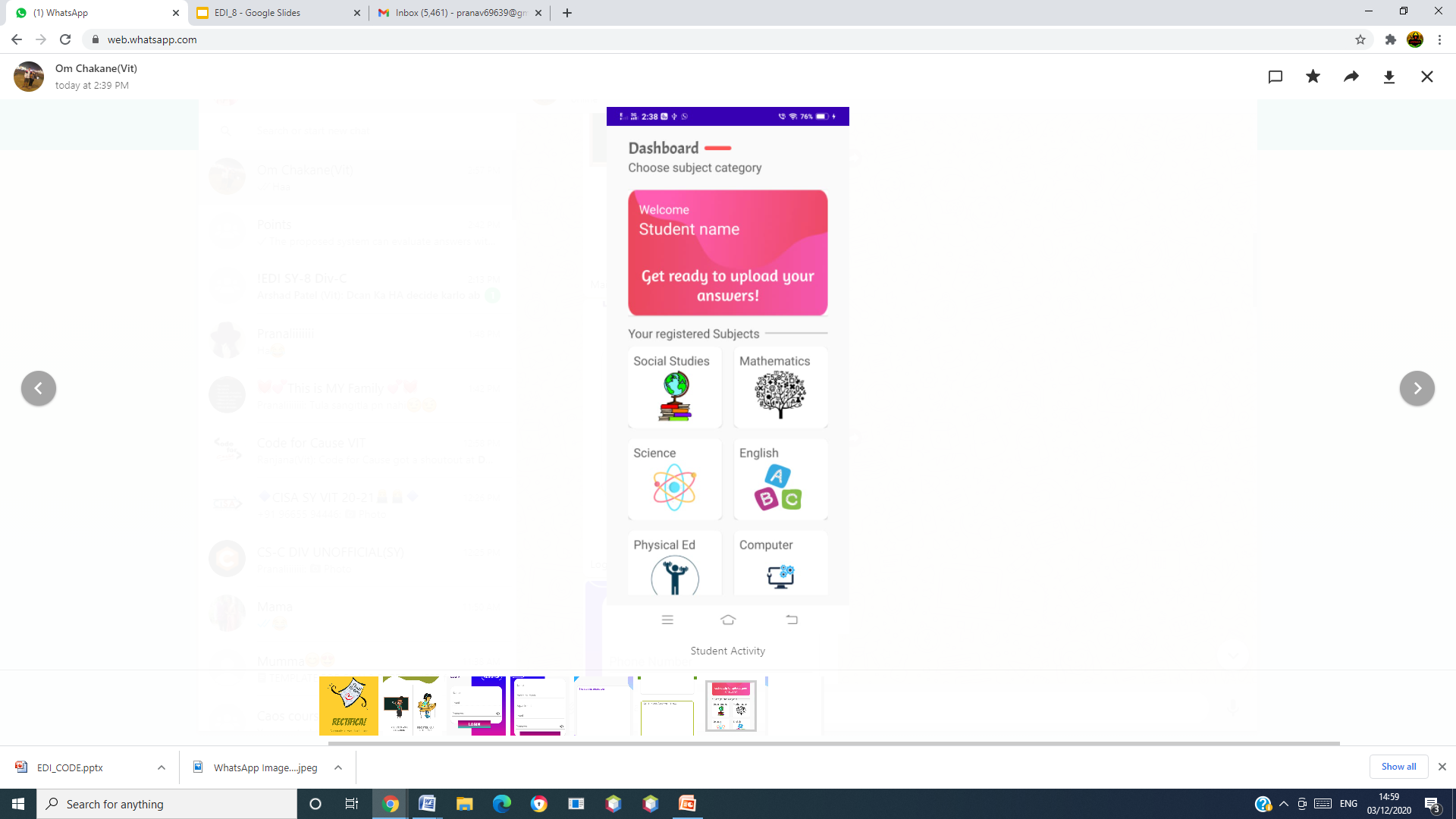


Fig 19 Student Activity

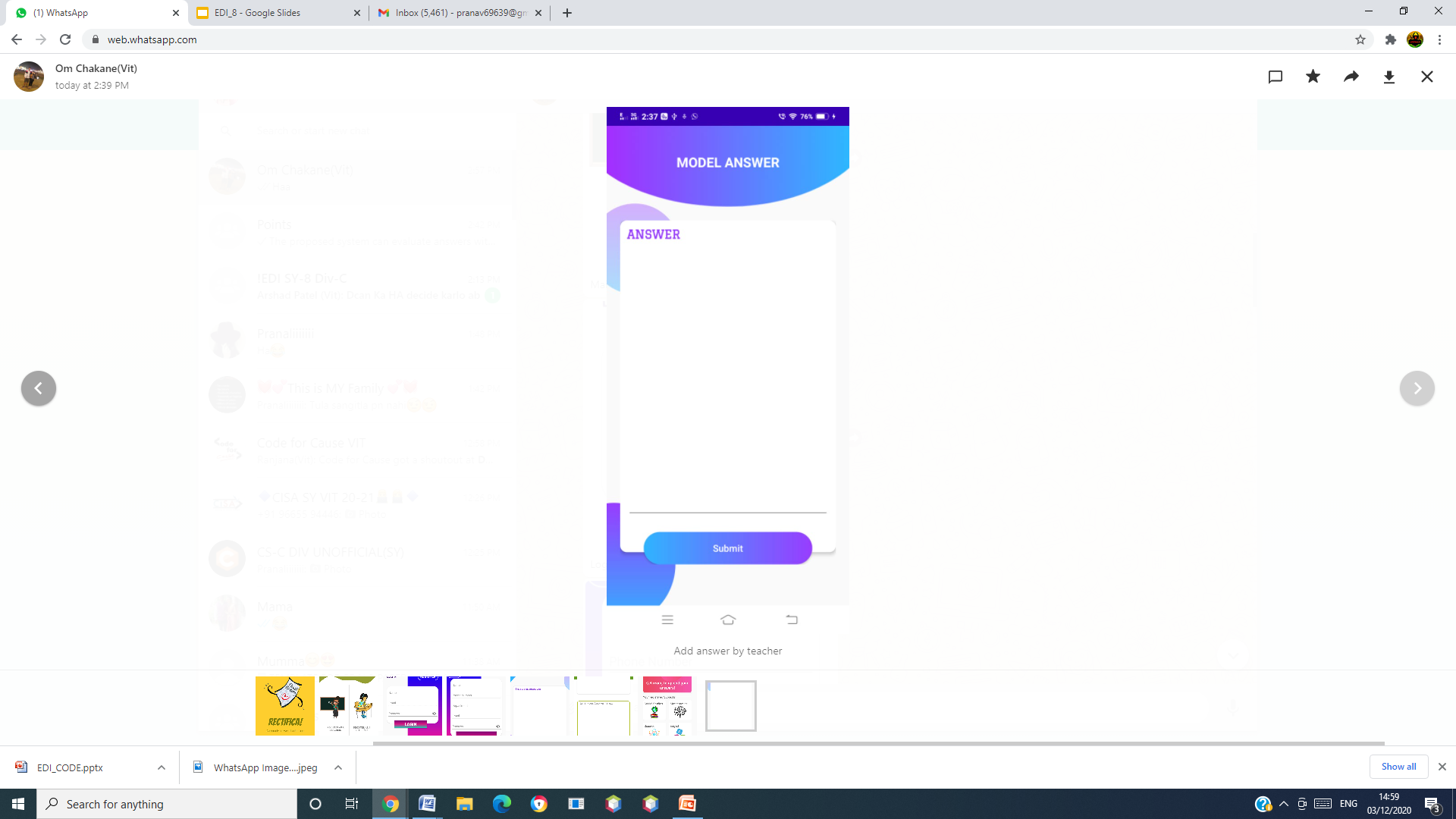


Fig 20 Add answer by teacher

# Limitations

# The proposed system can evaluate answers with English words or paragraph ,only.

# Future Scope

We can further extend it to include not only English words, but also various other languages, images and graphs so that we can span a larger number of variations in answer sheets across various regions.

# Conclusion

Examinations play a very important role in colleges, universities and various other educational institutes Many educational institutes have their examinations conducted online. But these exams only contain multiple choice questions which are providing to be very efficient in testing the student's aptitude ,on the other hand fail to measure the conceptual knowledge a student or learner must possess. Therefore subjective answer must be included in online examinations. The proposed system evaluates the answer based on the keywords. By comparing the standard answer and the student's answer marks is obtained if the student utilizes all the keywords mentioned in the standard answer. Hence the said system could be of great utility to the educators whenever they need to take a quick test for revision purpose, as it saves them the trouble of evaluating the bundle of papers. Also this system totally evaluating any circumstances of business. In future we are planning to evaluate subjective answer with diagrams and mathematical expressions. The current system only evaluates answer written in English. Further it can be extended to evaluate answer written in other languages also.

# Acknowledgment

we would like to thank and express our gratitude to “prof. dr. s. t . patil ” sir for their guidance and support in our project.

# References

1. “Sheeba Praveen.”An Approach to Evaluate subjective Questions for Online Examination System” published in International Journal Of Innovative research in Computers and communication Engineering Vol 2,Issue 11,November 2014. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
2. Gunjal,Mrunal M ,Sayli M Pawar and PrakashJ.Kulkarni,”Evaluation of Subjective answers using GLSA enhanced with contextual synonyms “Published in International Journal on Natural Language processing Computing(INLC) Vol 4.No1,February 2015
3. Ani Thomas, MKKowar & Sanjay Sharma “Intelligent Fuzzy Decision Making For Subjective Answer Evaluation using Utility” published by Emerging Trends in Engineering and Technology 2008 ICETET '08 First International conference on Date 16-18 July 2008
4. P. Sravanthi and D. B. Srinivasu, ”Semantic Similarity Between Sentences,” International Research Journal of Engineering and Technology (IRJET), vol. 2, pp. 156- 161, 2017.
5. J. C. Ming Che Lee and T. C. Hsieh, ”A Grammar-Based Semantic Similarity Algorithm for Natural Language Sentences,” Hindawi Publishing Corporation Scientific World Journal, vol. 2014, pp. 1- 1, 2014.
6. A. Pawar and V. Mago, ”Calculating the similarity between words and sentences using a lexical database and corpus statistics,” IEEE Transactions on Knowledge and data Engineer-ing, February 2018.
7. Jamsheedh C. V, Aby Abahai T and Surekha Mariam Varghese, “A Fair Assessment System For Evaluation And Grading Of Text In Degitized Descriptive Answer Scripts,” RACCCS, vol. 4, pp. 346- 352, 2016.
8. F. B. Hasim Sak, Andrew Senior, ”Long short-term memory recurrent neural network archi-tectures for large scale acoustic modeling,” in Proceedings of the Annual Conference of International Speech Communication Association, 2014.

1. [↑](#footnote-ref-1)