SITS/Computer Engineering/Projects/UG/2019-20/B1

A Project Report on

AUTOMATED ESSAY GRADING

BY

Mr. VINAY SANGA	Exam Seat No. B150574344
Mr. ABHISHEK JAGTAP	Exam Seat No. B150574260
Mr. SHREYANSH PATIL	Exam Seat No. B150574295
Mr. PRASHANT RAUT	Exam Seat No. B150574312

Guide
Dr. G. S. Navale

Sinhgad Institutes

Department of Computer Engineering
Sinhgad Institute of Technology and Science
Pune-411 041

[2019-20]

Sinhgad Institutes

C E R T I F I C A T E

This is to certify that *Mr. VINAY SANGA Exam No. B150574344*, *Mr. ABHISHEK JAGTAP Exam No. B150574260*, *Mr. SHREYANSH PATIL Exam No. B15057255*, *Mr. PRASHANT RAUT Exam No. B150574312* have successfully completed the Project Stage-II entitled *Automated Essay Grading* under my supervision, in the partial fulfillment of Bachelor of Computer Engineering of Savitribai Phule Pune University.

Date :
Place :

Name of Guide **Dr. G. S. Navale**
Guide **Head of Department**

External Examiner **Dr. R. S. Prasad**
- **Principal**
- **Sinhgad Institute of Technology and Science, Pune 41**

ACKNOWLEDGEMENT

We take this opportunity with great pleasure to express our deep sense of gratitude towards our guide *Prof. Dr. G. S. Navale* for her valuable guidance and incessant encouragement and co-operation extended to us during this project work.

We are also thankful to *Prof. Dr. G. S. Navale* Head, Computer Engineering Department, for providing all departmental facilities for this work.

We would also like to thank *Prof. Dr. R. S. Prasad* Principal, Sinhgad Institute of Technology and Science for his unflinching help, support and cooperation during the project work.

- Vinay Sanga

- Abhishek Jagtap

- Shreyansh Patil

- Prashant Raut

Abstract

In the task of essay grading we assign a grade to an essay in the most human-like manner possible. This problem can be tackled by building statistical models which try and predict how a human would evaluate an essay. In this project, the proposal is to use Long Short Term Memory (LSTM) neural networks to create an essay grader to undertake this daunting task. The grader will give an essay a grade which would be expected from a human grader. The user will be able to see his score instantaneously which is an improvement as compared to human graders. There are some very challenging tasks such as recognizing the context of the essay. Also, the vocabulary and various versions of English will have to be considered while evaluating the essay. The grader will be able to grade an essay without any bias and high accuracy.

Keywords: Essay Grading, Deep Learning, LSTM, Computerised Testing.

List of Abbreviations

The next list describes several abbreviations that will be later used within the body of the document

DL **Deep Learning**

FR **Functional Requirement**

$KLOC$ **Kilo Lines Of Code**

LSA **Latent Semantic Analysis**

$LSTM$ **Long Short Term Memory**

$LSTM - RNN$ **Long Short Term Memory Recurrent Neural Networks**

ML **Machine Learning**

NLP **Natural Language Processing**

RNN **Recurrent Neural Network**

$SDLC$ **Software Development Lifecycle**

SRS **Software Requirement Specifications**

SVD **Singular Vector Decomposition**

SVM **Support Vector Machines**

UML **Unified Modeling Language**

List of Figures

List of Tables

Contents

Chapter 1

INTRODUCTION

Essays are a tool for testing the students' fluency, vocabulary and grammatical correctness in a language. They are also useful to test one's creativity, originality and articulateness. The highly subjective and diverse nature of an individual in writing an essay makes it difficult to grade the essay uniformly across many human graders. In addition to this there are various other biasing factors in grading an essay. There has been research on automatic essay grading since the 1960s. The first systems based their grading on the surface information from essays. These systems were successful though they failed to capture aspects like grammatical correctness and language fluency. Much research has been conducted in the field most notably by Educational Testing Service (ETS). Clubbing this together with the resurgence in new technologies such as neural networks, deep neural networks, there is a whole new world of possibilities due to their capacity of modeling complex patterns in data. These methods do not depend on feature engineering so they are really useful for solving problems in an end-to-end fashion. With this intuition this project aims to the relevant knowledge in the field of education and try to create an essay grader which can make quality education more accessible. The work also explores methods of improving the quality and usability of the system.

1.1 Overview

The preliminaries suggest that the DL and NLP are having a strong capability to implement the Essay Grader accurately. The system will be able to solve some of the oldest pedagogical issues and also reduce the burden on human grader as it will be seen in the upcoming section that the project is expected to outperform some of existing system in practice. This work will also be able to set some benchmark and can even be use as one to be

compared with. The system will be useful to educators, researchers, etc. more of which will be discussed in upcoming sections.

1.2 Motivation

The human graders unknowingly tend to grade an essay biasing towards the individual subject matter presented. Another major drawback is the time required to grade essays can be significantly high.

The present technologies present an excellent opportunity to automate tedious tasks such as essay grading. Deep Learning being affordable to use now is a major push towards the reliance and devising the system.

This project will help in maintaining an unbiased and fair approach towards evaluating the written essay for competitive exams, tests, etc. which is in turn beneficial in multiple ways.

1.3 Problem Defintion and Objectives

To implement an essay grading system to assess essays of length 150-600 words using LSTM-RNN Neural Networks

The specific objectives of the proposed work are:

1. To revise the required concepts.
2. To pre-process the data using NLP.
3. To train an LSTM neural network using Deep Learning.
4. To test and validate the model using existing approaches.
5. To prepare a front end for the project viz. a Web Page.

1.4 Project Scope and Limitations

The project scope encompasses the educational domain and can be used from lower levels to the higher levels of the education cycles.

The lowest implementation would be to use the grader to grade the essays of the primary school students. The level of implementation actually depends on the dataset contents. The quality of the essays dictates the level at which the system can be implemented. The next level of the implementation would be for secondary school to pre-university schools. In this case, the system will be more useful than the previous level as it us more practical to use the system in this case.

The higher level of implementation are competitive and entrance exams for graduate courses. This level is the optimal level of implementation and the system will be able to deliver the most in this scenario.

Apart from this, the system applicability can be extended to evaluating or grading of answers for the theoretical papers. While it is a bit enticing to build a grader for this purpose, the process is intimidating and much more restrictive than grading an essay.

1.5 Methodologies of Problem solving

1. Data Preprocessing:
The essay dataset is preprocessed by removing the stop words, stemming, lemmatizing and then they are formed into a tsv file.

2. Word2Vec Represetation:
The essay is then transformed into word2vec vectors such that they can be used to train the model.

3. Model Building:
In this process, we train the LSTM model and simultaneously tune the hyper parmeters so that the highest obtainable accuracy is achieved.

4. Model Testing:
We test the model on different available test sets and finalize the model to be pickled and used in the web app.

This concludes the first chapter of the report. All the initial details regarding the project were discussed. The idea and theme of the project is now clear and the further proceedings are scheduled successfully. The budgeting is done and the effort required to implement the project is known. In the next chapter, there will be a discussion about the literature survey and the related details of the project.

Chapter 2

LITERATURE SURVEY

The previous chapter tells about the description of this project. The problem statement gives a brief idea about the project and the objectives gives a step-wise execution process of the project. In this chapter, some concrete basis supporting and extending the applicability of the project will be reviewed with the hope of providing better insights into the foundations of the project.

2.1 Literature Survey

Equivocating the statements above this section will discuss the findings done during the literature review of the topic.

Alex Adamson et. al. [1] have implemented an essay grader on the Hewlett Foundation dataset using different ML techniques such as SVM, Latent Semantic Analysis. Each essay was graded by at least two humans. It uses the Quadratic Weighted Kappa as closeness measure. The general pipeline involves extracting features from the raw essays, and iteratively training and using k-folds cross-validation on the model on selected essay sets in order to optimize hyperparameters. It is shown that for essays of intermediate writing level and given enough human graded training examples for a writing prompt, we can automate the grading process for that prompt with fairly good accuracy. However, the classifier was built for this task with limited success.

Derrick Higgins et. al. elaborated Support vector machine (SVM) for classification and regression [2]. SVM technique relies on kernel functions. There were multiple goals in this work. The authors wanted to introduce a concept of essay coherence comprising multiple aspects, and investigate what linguistic features drive each aspect in student essay writing. They have worked with writing experts to develop a comprehensive protocol that details

how coherence in writing can be evaluated, either manually or automatically. Using this protocol, human annotators labeled a corpus of student essays, using the coherence dimensions. However, this approach is relatively new and though better than previous ones is valid only theoretically.

The LSA approach to detect major research topics and themes of a multidisciplinary field was applied by Gang Kou et. al. [3]. The LSA analysis can be summarized in three main steps. The first step is to set up a term-document matrix in which each row stands for a key word or term and each column stands for a document or context in which the key word appears. An entry in the matrix is the frequency of a key word in the corresponding document. The second step is to transform the term frequencies in a term-document matrix using various weighting schemes. The third step is to perform SVD on the matrix to reduce the dimensionality, which is the key feature of the LSA method. The limitations are - a. It is not possible to use LSA on an unestablished field of research; b. Only English journals are considered in this approach.

Kaveh et. al discusses about various kinds of approaches available for constructing a Machine Learning model. They are also using the dataset mentioned in the first paper. They have selected LSTM model [4] as it works the best among all the available ML models. They have discussed their approach as well as the work done by other prominent researchers. Their model does not require feature engineering as compared to some other models. The model performs 5.6% better than the baseline performance. There is still quite a scope in improving the model as per their views.

2.2 Summary

This section gives the important findings obtained during reviews as seen in Section 2.1 and are depicted in Table 2.1

Table 2.1: Literature Review

Ref. No.	**Highlights**	**Observations**
1	• Essays can be graded without the help of the teacher. • Different models are compared and best amongst them is used. • The grades are equivalent to human graders.	• The size of essay which can be graded is limited. • Good quality training examples are required as presently the accuracy is limited to the present set.
2	• Hybrid approach is used to classify different coherence dimensions with a high- or low-quality rank. • In each of the experiments with less fold-cross validations results are reported for large set of essays.	• Choosing appropriate kernel function is difficult and complex task. • In case of using high dimension kernel it generates too many support vectors which reduces training speed.

Ref. No.	Highlights	Observations
3	• Doesn't require feature engineering. • One of the best implementation of the grading system.	• Training time is more as it is a Deep Learning model.
4	• LSA is capable of assuring decent results. It works well on dataset with diverse topics. • LSA can handle Synonymy problems to some extent. • It is faster as compared to other dimensionality reduction models.	• LSA depends on identifying frequent word usage patterns from a collection of text, it is difficult to capture a research area if it is not well established. • The research abstracts collected in this analysis include only English language journals.

From the surveys, it is clear that the project is quite feasible and quite a lot of work has been done in the field related to this project. Also, there are some attributes such as accuracy, scalability, performance, etc. which can be further improved with the use of current technologies. More about the current technologies and the implementation of the project will be discussed in the next chapter.

Chapter 3

SOFTWARE REQUIREMENTS SPECIFICATION

In the last chapter, the basis of implementing this project were discussed. Following that, in this chapter Software Requirement Specification document will be elaborated. In the subsequent sections, there will be a discussion regarding the technical requirements, methodology to be used in SDLC, etc.

3.1 Assumptions and Dependencies

Assumptions –

1. **The essay writer agrees to the limit of the word specified.**
2. **The system is used only for educational purposes.**
3. **The dataset provided is a meaningful dataset so that we abide by 'Garbage in Garbage out' phenomenon in Machine Learning.**
4. **The knowledge provided is static in nature.**

Dependencies –

- **Python 3.7.3**
- **NLTK 3.4.5**
- **Numpy 1.17**
- **Pandas 1.0.2**
- **Gensim 3.7.3**
- **TensorFlow 2.0.0**

- **Keras 2.3.1**
- **Flask 1.1.2**

3.2 Functional Requirements

These are the functions which the system performs and the user can use them on their device.

3.2.1 System Feature 1

The system should display the list of gradable essay topics to the users. These essays will be passed further to the grading system.

3.2.2 System Feature 2

The system should get grade by passing the essay which has been written by user to the grading system which is present at the backend or web server.

3.2.3 System Feature 3

The grading system should be able to grade the essay passed to it by the web server. The predicted grade will be return back to user through same channel by which it was received.

3.3 Nonfunctional Requirements

This section describes the non functional requirements of the project which includes performance, safety, security, etc.

3.3.1 Performance Requirements

The system must be interactive and the delays involved must be less. So in every action-response of the system, there are no immediate delays. In case of opening links, uploading and submitting the essays the delay is below 2 seconds. The result is obtained almost instantly because the grader is preloaded in the memory while opening the website. Also when connecting to the server the delay is based editing on the distance of the 2 systems and the configuration between them so there is high probability that there will be or not a successful connection in less than 20 seconds for sake of good communication.

3.3.2 Safety Requirements

Information transmission should be securely transmitted to server without any changes in information. The essay data and the score is sent to the server as post data and therefore is highly secure.

3.3.3 Security Requirements

There is no need of user login and hence the authentication is not required which saves us from auth safety procedures.

3.3.4 Software Quality Attributes

1. Security
This is achieved by send the essay via POST method and neglecting the need of login.

2. Reliability
The system will uses Flask micro-framework which is reliable.

3. Maintainability
The separation of the modules allows easy maintainability.

4. Portability
The web app uses Bootstrap which is a responsive framework so it can be run both on Desktop browser and Mobile browsers without any fuss.

5. Resource Utilization
The resource requirements are similar to a standard website as most of the work is done at the backend.

3.4 System Requirements

1. **Web Browser – Chrome, Firefox, Microsoft Edge.**
2. **An internet connection with min speed 2 Mbps.**

3.5 Analysis Models: SDLC Model to be applied

CRISP-DM stands for cross industry standard process for data mining. It is a comprehensive data mining methodology and process model that provides anyone from novices to data mining experts with a complete blueprint for conducting a data mining project. CRISP-DM breaks down the life cycle of a data mining project into six phases. These 6 high-level phases of CRISP-DM are still a good description for the analytics process. The architecture is shown in Fig. 3.1

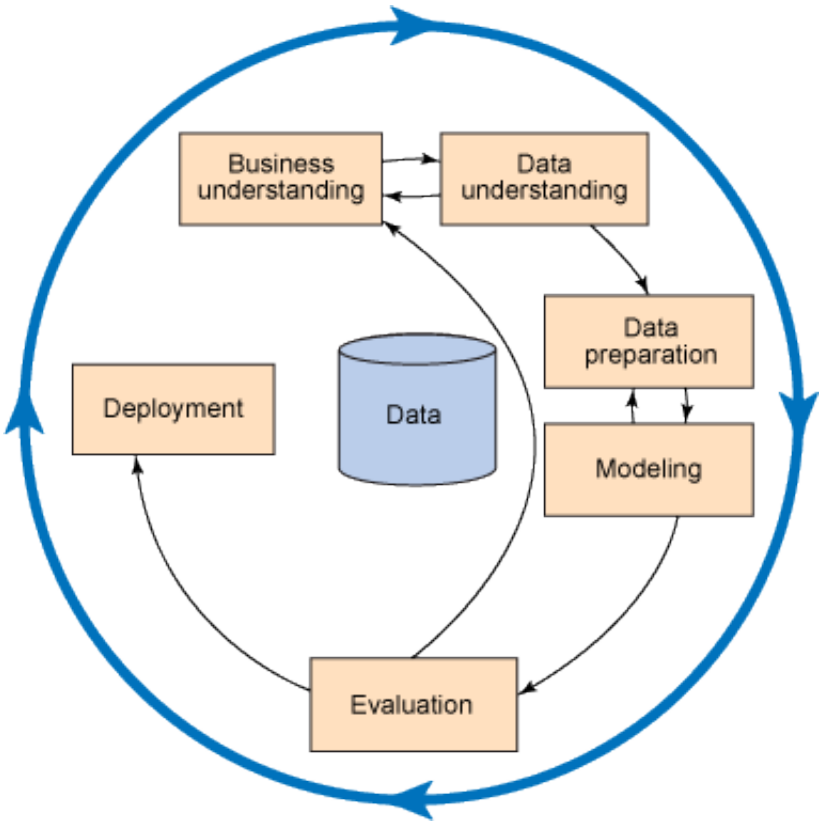


Figure 3.1: CRISP-DM Model[5]

1. **Business Understanding:-**

   First and foremost step in Analytics project is understanding client's business to formulate problem statement. Once we define problem statement then we can drive data accordingly. Also Focus should be on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the below objectives

   - **What the client really wants to accomplish?**
   - **Uncover important factors including the constraints, competitive objectives**

2. **Data Understanding:-**

   An initial data collection is present and subsequent activities in order to get familiar with the data, to identify data quality problems, to discover initial level of insights into the data or to detect interesting subsets to form hypotheses for hidden information are done

   - **Describing the data**
   - **Exploring the data**
   - **Verifying the data quality**

3. **Data Preparation:-**

   It covers all activities to construct the final dataset from the initial raw data. Converting raw data into analytical dataset is very important. Quality of cleaned data will impact on model performance. Data preparation tasks are likely to be performed multiple times and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools.

   - **Cleaning the raw data**
   - **Usage of Sampling Techniques**

4. **Modelling:-**

   In this phase, various modeling techniques are selected and applied and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often necessary.

5. **Evaluation:-**

   Thorough evaluation of the model and review of the steps executed to construct the model is done to check the business objectives. A key objective is to determine if there is some important business issue that has not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

3.6 System Implementation Plan

It is divided into parts as follows -
Part 1: In this part, the concepts and technologies related to the implementation will be learnt. This part is the learning part of the phase 2 and is will take around 8 weeks. Objective 1 will be completed in this part.
Part 2: In this part, the data processing will take place and will take around 5 weeks to complete. This part covers objective 2 of the project.
Part 3: Actual DL Model implementation and testing will be undertaken in this part. This part is the most laborious part of the whole project and will take around 8 weeks to be accomplished. The objectives 3 and 4 will be achieved in this part.
Part 4: This part is the final part of the phase 2. This phase relates to all the housekeeping of the project. The finalization of the report for phase 2 will be done in this part. The system will be fully implementation after this part.

In this chapter, SRS was elaborated and various requirements of the project were studied. In the following chapter, there will be discussion about the implementation architecture and related details using various visual tools.

Chapter 4

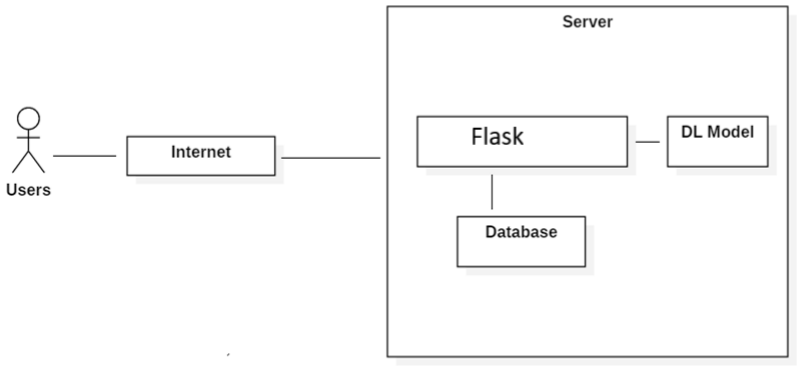
SYSTEM DESIGN

This chapter will discuss the design specifications of the project. Various diagrams are drawn to make the idea of the project clearer to understand. As the project is focussed more on the implementation of DL model and less on other software parts, there are very few UML diagram to be shown and discussed about.

4.1 System Architecture

The architecture of the system is the most fundamental approach to show the overall structure. The grading system is implemented as a Client - Server Architecture. The essay writers(user) are the clients. The model and other implementations as usual reside at the server end. There are not much components in either of the side because the input to output task is simple pipeline to-and-fro the DL Model residing at server end. The architecture is shown in Fig. 4.1.

LSTMs are a special kind of RNN, capable of learning long-term dependencies. They were introduced by Hochreiter & Schmidhuber (1997). They work tremendously well on a large variety of problems, and are now widely used. LSTMs are explicitly designed to avoid the long-term dependency problem, which is present in standard RNN.
All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.It is shown in Fig. 4.2
LSTMs also have this chain like structure(see Fig. 4.3), but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.



System Architecture

Figure 4.1: System Architecture

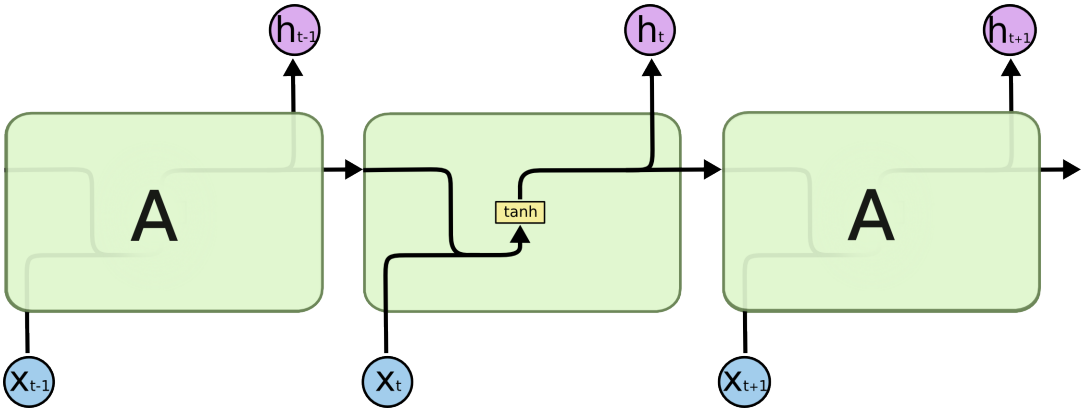


Figure 4.2: Standard RNNs [6]

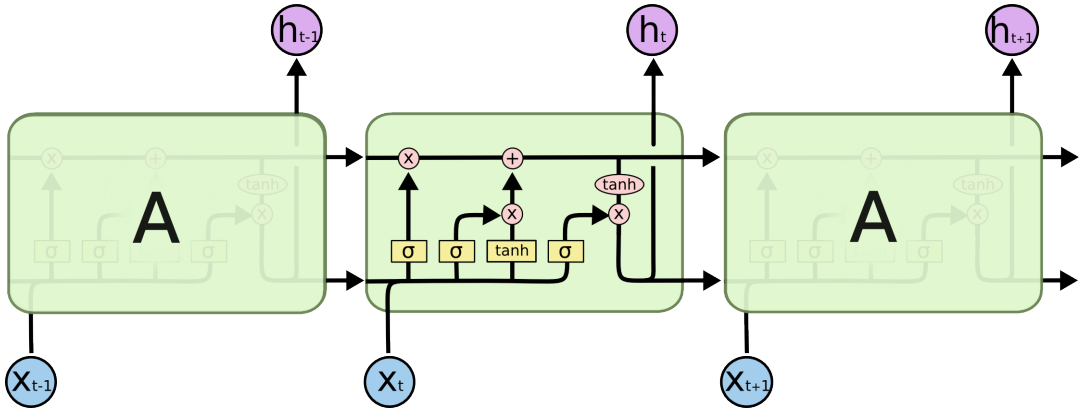
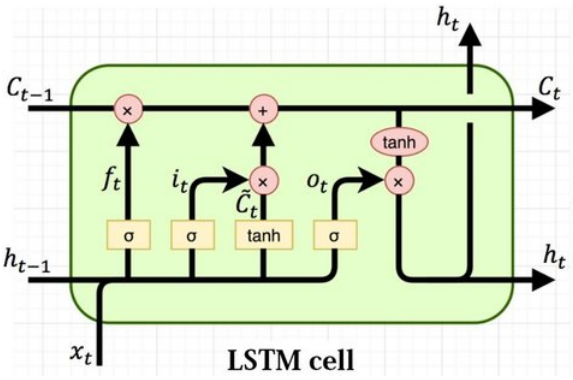


Figure 4.3: LSTMs [6]

4.2 Mathematical Model

Below is the mathemical formulation used inside the neural network to calulate the weights and biases. The gate equations are shown in 4.4.



$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right)$$
$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right)$$
$$o_t = \sigma\left(x_t U^o + h_{t-1} W^o\right)$$
$$\tilde{C}_t = \tanh\left(x_t U^g + h_{t-1} W^g\right)$$
$$C_t = \sigma\left(f_t * C_{t-1} + i_t * \tilde{C}_t\right)$$
$$h_t = \tanh(C_t) * o_t$$

Figure 4.4: Mathematical model diagram

4.3 Entity Relationship Diagrams

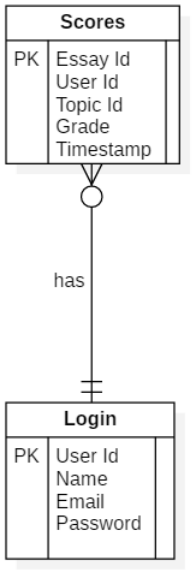
The ER diagram shows the relationship between various entities stored in the database. The database will only be used to store the login data and the scores given by the grader for analytical purposes. Due to this reason, the use of database in the system is minimal and hence the entities in the database are only a couple. The main emphasis of the grading system lies in the DL modeling part which doesn't need to deal with the databases once the training of the model is done. The ER diagram as illustrated in Fig. 4.5 is fairly simple with only two entities named as Login and Scores.

4.4 UML Diagrams

A UML diagram is a diagram based on the UML used to represent a system along with its main actors, roles, actions, artifacts, activities, etc. in order to better understand, alter, maintain, or document information about the system. The UML diagrams pertaining to the system are shown and briefly discussed in the upcoming sections.

4.4.1 Activity Diagram

The activity diagram depicts the flow of control from one activity to another. Fig. 4.6 shows the array of activities such as displaying the essay topics,



ER Diagram

Figure 4.5: ER Diagram

getting the essay from the writers, grading the essay, etc.

4.4.2 Deployment Diagram

Deployment Diagram specifies how the system will be implemented physically. The system has a simple client and server deployment as depicted in Fig. 4.3.2.1. The client will have to use web browser to requests to the system. The actual DL model will be on the web server and will be the one grading the essay submitted by the user via Internet.

4.4.3 Sequence Diagram

Sequence diagram simply depicts interaction between objects in a sequential order i.e. the order in which these interactions take place. The objects into the consideration are User and Scoring Engine as shown in Fig. 4.8.

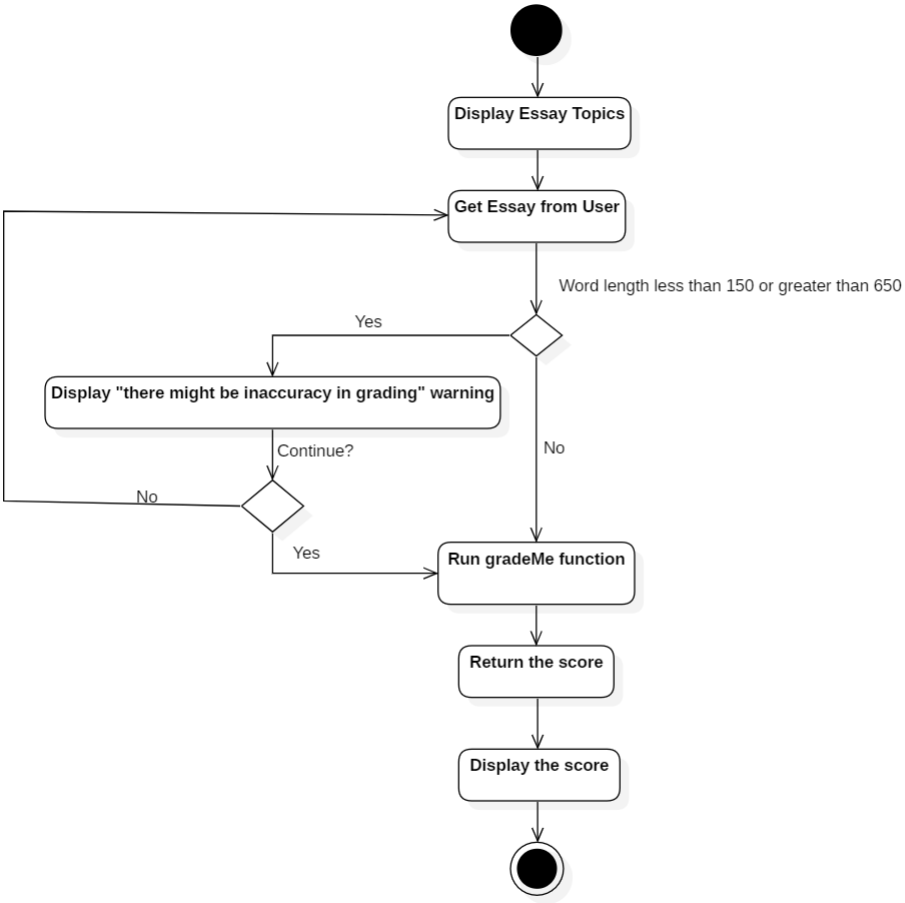


Figure 4.6: Activity Diagram

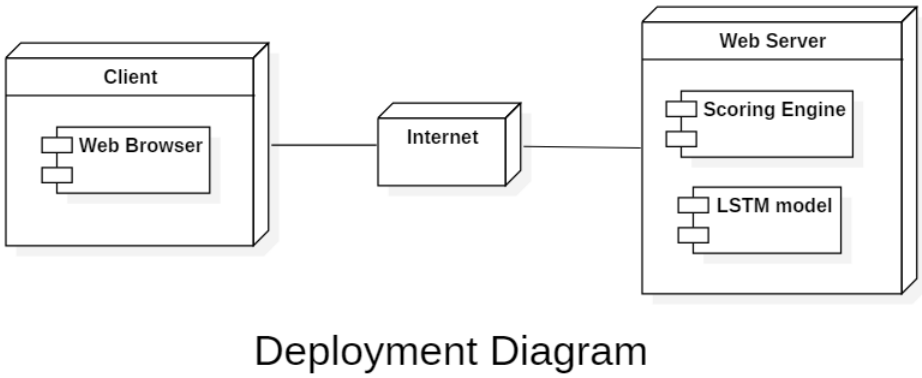


Figure 4.7: Deployment Diagram

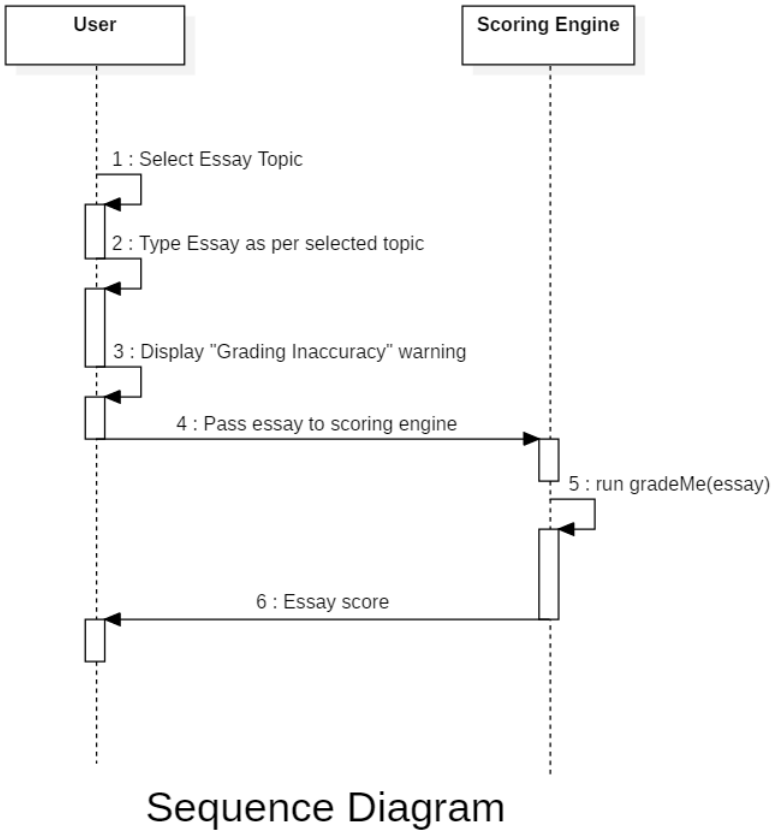


Figure 4.8: Sequence Diagram

This concludes the visual representation on the system. Various UML diagrams were used to instill the structure of the project. It is now almost fully possible to go for implementation of the project but there are some pros and cons which needs to be discussed in the next chapter.

Chapter 5

PROJECT PLAN

A project plan is a formal document designed to guide the control and execution of a project. A project plan is the key to a successful project and is the most important document that needs to be created when starting any business project. The following chapter contains necessary points to be considered for the project plan.

5.1 Project Estimate

The project estimate is prior calculation of the cost, resources and other needs. This is done after the requirements of the project are finalized between both the parties. The estimated cost include the cost of hosting the servers which can be range from INR 5000 - INR 10,000 based on the providers and the type of service used.

5.1.1 Reconciled Estimates

The budget of a project is calculation or estimation of all the efforts and costs required to implement the project. For this project, the budget has been calculated by using CoCoMo model. The basic CoCoMo model was used in Organic mode as the project is small and doesn't have too many complex budgeting factors.

The basic CoCoMo equations are -

$$E = a_b(KLOC)^{b_b} \tag{5.1}$$

$$D = c_b(E)^{d_b} \tag{5.2}$$

$$SS = E/D \tag{5.3}$$

Estimated size of the project = 5 KLOC
So, using equations 5.1 & 5.2, we get

$E = 2.4(5)^{1.05} = 13.01PM$
$D = 2.5(13.01)^{0.38} = 6.63M$
$SS = 13.01/6.63 = 1.96P$

Here, E is Effort (measured in Person Months)
D is Deployment Time (measured in Months)
SS is Staff Size (units is Persons)
Hence, Total Effort required is 13 person months(approx.) yielding a Development Time of 6.63 months and a Staff Size of 2 persons.
As, the team size is 4 persons, the development time of 6.63 months can be speeded up and calculated as follows:

Persons	D
2	**1/6.63**
4	**1/x**

So, $x = 2 * 6.63/4 = 3.3$ **Hence, the project will require 3 month (approx.) to complete (theoretically).**

5.2 Risk Management

Risk management is defined as the process of identifying, assessing and controlling threats to a project. Risk management is a process that seeks to reduce the uncertainties of an action taken through planning, organizing and controlling of both human and financial capital.It is the responsibility of project manager to go through potential threats. The project manager can identify the risks and accordingly control them with the help of other stake holders.

5.2.1 Risk Identification

The risks identified in the project are very few because of high cohesiveness and low coupling involved. Also, open-source technologies are used throughout and no external APIs are used which makes the project self-dependent and hence the probability of error is less.

5.2.2 Risk Analysis

Risk analysis is handling the system response in the risk conditions which are identified. If the system response is incorrect then the developers work on it and try fixing out the identified risks. Risk analysis can affect on project objectives which can work incorrectly in the risk cases.The risk analysis assigns the severity to the risks and developers start working accordingly.

1.High severity: Catastrophic

2.Medium severity: Critical

3.Low severity: Marginal

5.2.3 Overview of Risk Mitigation, Monitoring, Management

There are few risks which were encountered while developing the project, these risk have low probability but high impact. The most important thing is their probability. These errors are not catastrophic as they do not occur regularly. The risks are described in the tables 5.1 and 5.2 and are self-explanatory.

Table 5.1: Risk 1

Risk ID	**1**
Risk Description	Website Crash
Category	Technical
Source	Internet
Probability	Low
Impact	High
Strategy	Wait for a while
Risk status	Not Occured

Table 5.2: Risk 2

Risk ID	**2**
Risk Description	Grade inappropriate
Category	Technical
Source	Internet
Probability	Low
Impact	High
Strategy	Re-evaluation of Model
Risk status	Occurred

5.3 Project Schedule

The project schedule is a set of activities which covers the development of all functionalities in the project. It comes with the start and end date of each and every activity. The project schedule includes every single detail of the project, such as who will be completing each task, the deliverables that will be produced, the goals and objectives the project will achieve and the amount of time it will take to complete the project.

The schedule is shown is shown in Fig. 5.1.

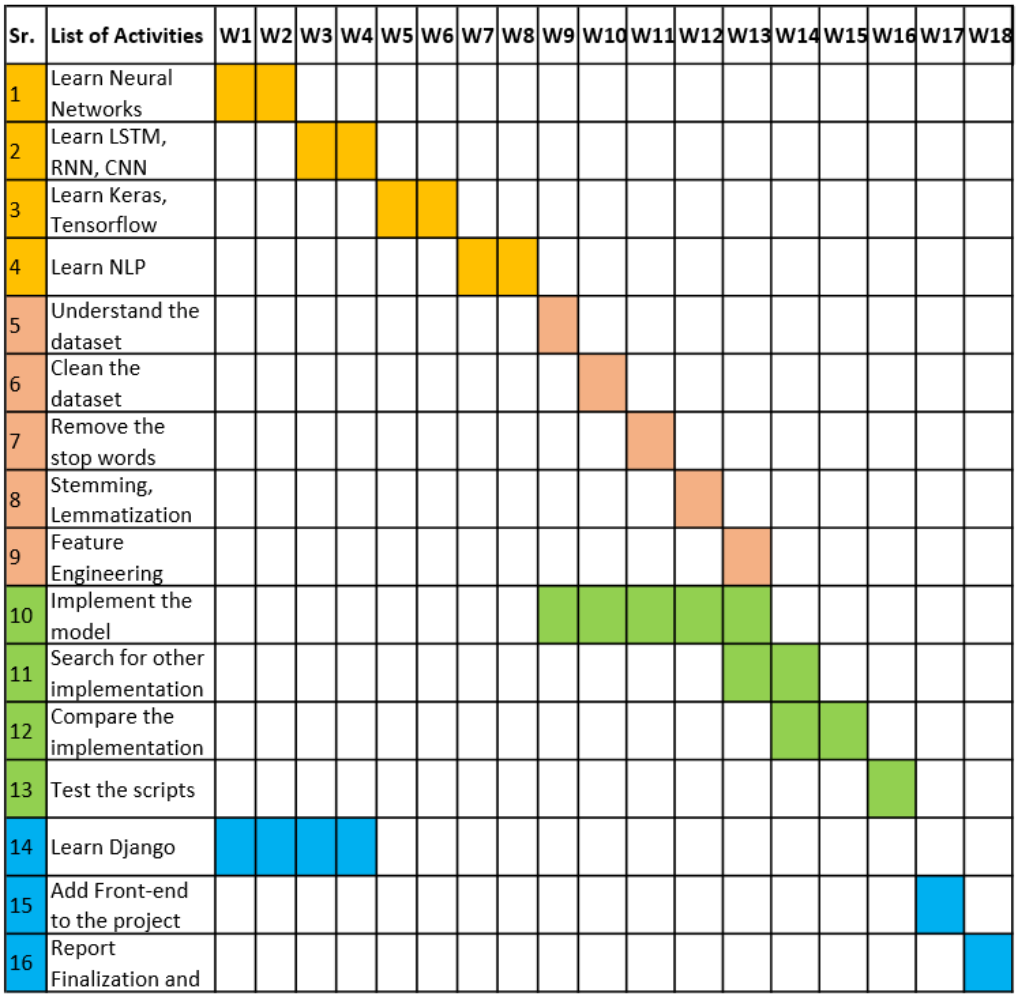


Figure 5.1: Schedule and Implementation Chart

5.3.1 Project Task Set

1. **To revise the required concepts. Learn Neural Networks**
    - **Learn LSTM, RNN, CNN**

- **Learn Keras, Tensorflow**
- **Learn NLP**

2. **To pre-process the data using NLP.**
   - **Understand and Clean the dataset**
   - **Remove the stop words Stemming, Lemmatization**
3. **To train and test LSTM neural network using Deep Learning.**
   - **Implement the model**
   - **Search for other implementations**
   - **Compare the implementations**
   - **Test the scripts**
4. **To prepare a front end for the project.**
   - **Learn Django**
   - **Add Front-end to the project**
   - **Documentation and Paper**

5.4 Team Organization

According to Savitribai Phule Pune University rules, the team was supposed to be of 4 students from Final year of Computer Engineering. The team was formed in June 2019.

5.4.1 Team Structure

Dr. G.S. Navale - Project Guide
Vinay Sanga - Model Preparation and Planning
Abhishek Jagtap, Shreyansh Patil - Front End
Abhishek Jagtap, Shreyansh Patil, Prashant Raut - Back End
All the members contributed equally in the project and all the tasks received appropriate attention from all the members starting from designing, implementation and testing.
In this chapter, we took a glance at the project plan including project estimates along with Risk Management steps. In the next chapter, we will take a look at the implementation of the project and the tools used in making the system.

Chapter 6

PROJECT IMPLEMENTATION

In this chapter, we will see the details regarding the project implementation. We will take a look at the tools and technologies used, algorithm for the project.

6.1 Overview of Project Modules

The web app starts with a page showing the steps to follow in order to use the website. The next module is essay writer module which supports writing and editing of the essays. The final module is a grader module which is responsible for grading the essay. These modules are explained below -

1. The web site :
The website welcomes the new user. This webite helps the new user to navigate between different pages through links . The result of the grader is also displayed here.

2. Essay writer module:
This module helps the essay writer to write the essay or upload the text file of essay.This module further helps to send the essay to grader module.

3. Grader module:
This is the main LSTM network which is responsible for grading the essay. It is a stateful LSTM which grades the human essays. The grader module preprocesses the essay, grades the essays and return the essay score.

6.2 Tools and Technologies Used

1. Flask

Flask is a web framework that provides with tools, libraries and technologies that allow you to build a web application.Flask is a lightweight web frame of Python. It provides the user with libraries, modules and tools to help build web applications.

2. Pandas and Numpy

Pandas is a high-level data manipulation tool. It is built on the Numpy package and its key data structure is called the DataFrame. DataFrames allow you to store and manipulate tabular data in rows of observations and columns of variables.

NumPy is a Python package which stands for 'Numerical Python'. It is the core library for scientific computing, which contains a powerful n-dimensional array object, provide tools for integrating C, C++ etc. It is also useful in linear algebra, random number capability etc.

3. Gensim

Gensim is an open-source library for unsupervised topic modeling and natural language processing, using modern statistical machine learning. It is designed to handle large text collections using data streaming and incremental online algorithms, which differentiates it from most other machine learning software packages that target only in-memory processing. It includes streamed parallelized implementations of fastText, word2vec and doc2vec algorithms as well as latent semantic analysis (LSA, LSI, SVD), non-negative matrix factorization (NMF), latent Dirichlet allocation (LDA), tf-idf and random projections.

4. Tensorflow - Keras

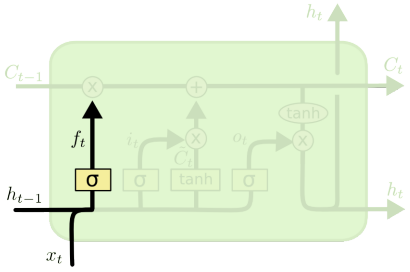
TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google. It lets users build and train ML models easily using intuitive high-level APIs like Keras.

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML.It focuses on being user-friendly, modular, and extensible.Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a

host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code. In addition to standard neural networks, Keras has support for convolutional and recurrent neural networks. It supports other common utility layers like dropout, batch normalization, and pooling. It also allows use of distributed training of deep-learning models on clusters of Graphics processing units (GPU) and tensor processing units (TPU) principally in conjunction with CUDA.

6.3 Algorithm Details

The algorithm used is LSTM and its steps are described as follows:
1. The first step in our LSTM is to decide what information we're going to throw away from the cell state. This decision is made by a sigmoid layer called the "forget gate layer." It looks at h_{t-1} and x_t, and outputs a number between 0 and 1 for each number in the cell state C_{t-1}. A 1 represents "completely keep this" while a 0 represents "completely get rid of this." It is shown is fig. 6.1

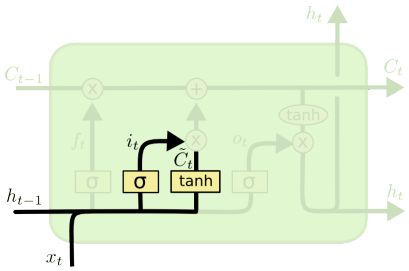


$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] \; + \; b_f\right)$$

Figure 6.1: Forget Gate Layer

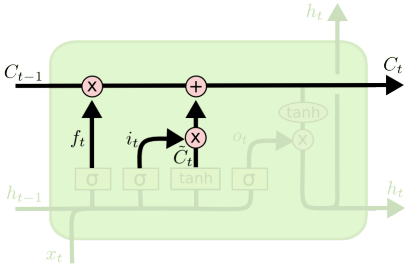
2. The next step is to decide what new information we're going to store in the cell state. This has two parts. First, a sigmoid layer called the "input gate layer" as shown in fig. 6.2 decides which values we'll update. Next, a tanh layer creates a vector of new candidate values, $\tilde{C}_t$, that could be added to the state. In the next step, we'll combine these two to create an update to the state.

3. It's now time to update the old cell state, C_{t-1}, into the new cell state C_t. The previous steps already decided what to do, we just need to do it. We multiply the old state by ft, forgetting the things we decided to forget earlier. Then we add $i_t \times \tilde{C}_t$. This is the new candidate values(see fig. 6.3).



$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] \; + \; b_i\right)$$
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] \; + \; b_C)$$

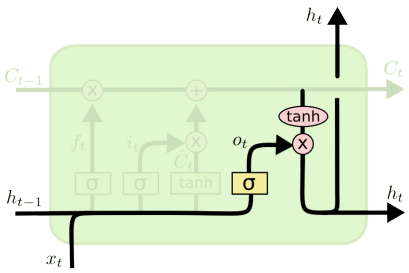
Figure 6.2: Input Gate Layer



$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Figure 6.3: Update Old State

4. Finally, we need to decide what we're going to output. This output will be based on our cell state, but will be a filtered version. First, we run a sigmoid layer which decides what parts of the cell state we're going to output(see fig. 6.4). Then, we put the cell state through tanh (to push the values to be between -1 and 1) and multiply it by the output of the sigmoid gate, so that we only output the parts we decided to.



$$o_t = \sigma\left(W_o \; [\, h_{t-1}, x_t] \; + \; b_o\right)$$
$$h_t = o_t * \tanh\left(C_t\right)$$

Figure 6.4: Final Output

Chapter 7

SOFTWARE TESTING

Software testing is process of checking whether the actual results match the expected results. It checks whether the developed software met the specified requirements and identifies any defect in the software in order to produce a quality product. Testing is required for an effective performance of software application. In this project, only two types of testings were used which are described following.

7.1 Type of Testing

Following types of testing is carried out for automated essay grader during testing phase.

7.1.1 Unit Testing

Unit testing is type of testing where individual unit or component of software is tested. In automated essay grading application the three units 1.essay graders 2.essay writer and 3.hyperlink in all web pages were tested. The three units didn't produce any bugs when tested and functioned as expected.

7.1.2 Integration Testing

It is type of testing where individual components are combined and tested as a group. All the modules were integrated and tested. A bug was found in which the user could not go back to the pages because of cache mismatch. The issue was fixed and all components work properly when integrated.

7.2 Test cases and Test Results

The results obtained after various testing procedures are depicted in table 7.1, 7.2, 7.3, 7.4 respectively. The tables are self-explanatory and hence no further explanation of the tables in detail is needed.

7.2.1 Test Case ID 1

Summary	The local file can be uploaded into the esaay textbox.
Prerequisites	browser , computer.
Test Procedure	Home page click , recent scores page click
Expected Result	When we click on any hyperlink , a new page should open.
Actual Result	All the hyperlinks are working properly.
Status	Pass

Table 7.1: Test case ID 1

7.2.2 Test Case ID 2

Summary	when we click on upload file all text get inserted in textbox. .
Prerequisites	Any file such as text file
Test Procedure	text can be copied from file to textbox.
Expected Result	The local file can be uploaded into the esaay textbox.
Actual Result	The local file can be uploaded into the esaay textbox.
Status	Pass

Table 7.2: Test case ID 2

7.2.3 Test Case ID 3

Summary	When GradeMe Button is clicked the essay scores should be displayed
Prerequisites	Essay with word length between 150 to 600.
Test Procedure	Upload the file containing essay and check the scores.
Expected Result	Proper scores of essay.
Actual Result	Proper scores of essay.
Status	Pass

Table 7.3: Test case ID 3

7.2.4 Test Case ID 4

Summary	The web pages should be properly displayed.
Prerequisites	Latest version of browser.
Test Procedure	Check by resizing the browser window.
Expected Result	All the web pages and UI elements should properly displayed.
Actual Result	All the UI elements and web pages are loading both on mobile as well as desktop browser properly.
Status	Pass

Table 7.4: Test case ID 4

In this chapter, we looked at the testing process and the results obtained after performing various types of testing for the application. In next chapter, we will take a look at the results obtained from the system.

Chapter 8

RESULTS

This section illustrates the generated output by the essay grading system on providing a proper set of input as and when required.

8.1 Outcomes

In the model evaluation, we built about 20 models and chose the best among them. All models were built on the same training and validation sets. The only difference was the selection of hyperparameters in each models. The evaluation criterion used for evaluation is Kappa score. The best model which showed the highest accuracy has been shown in 8.1

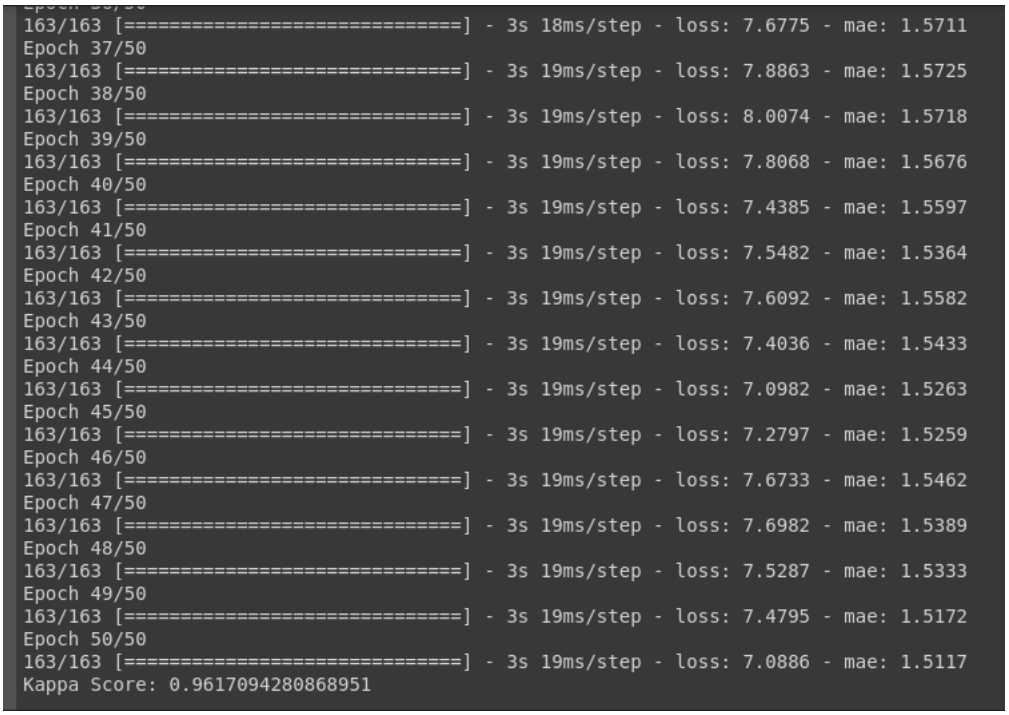


Figure 8.1: Accuracy Score

8.2 Screen Shots

The screenshots from the website running on Google Chrome are shown below. As we can see in fig. 8.2, we are given a essay topic to select from. Based on the topic we are given a text box where we can write the essay or upload a text file which can be seen in fig. 8.3. The final score based on the essay topic selected is shown in fig. 8.4.

In this chapter, we saw some screenshots related to the application. We also discussed how we evaluated the results to find the best model out of all.

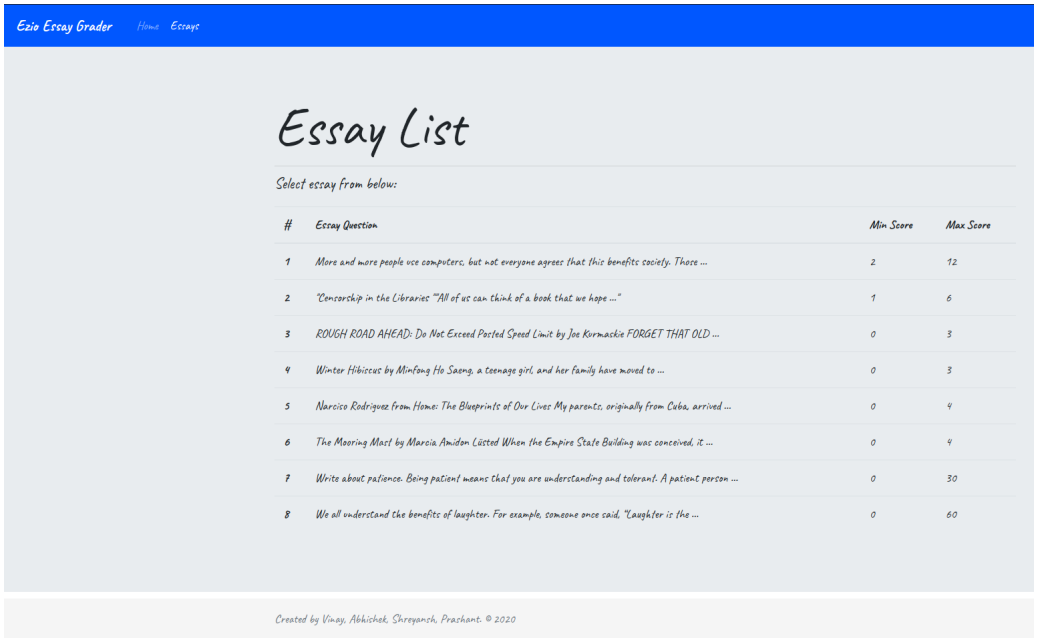


Figure 8.2: Essay topic selection

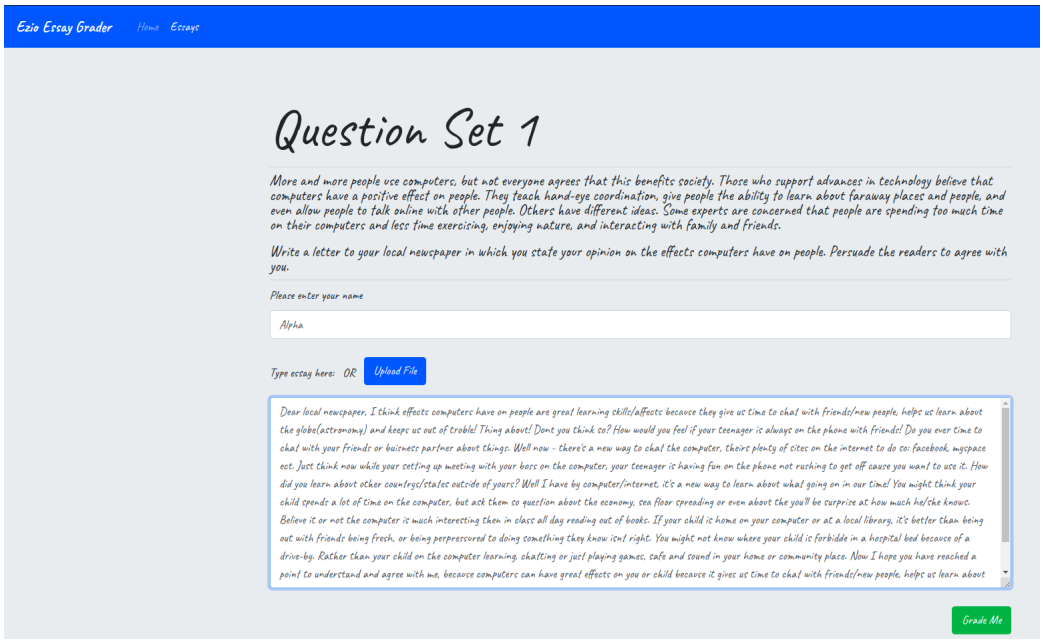


Figure 8.3: Essay writer

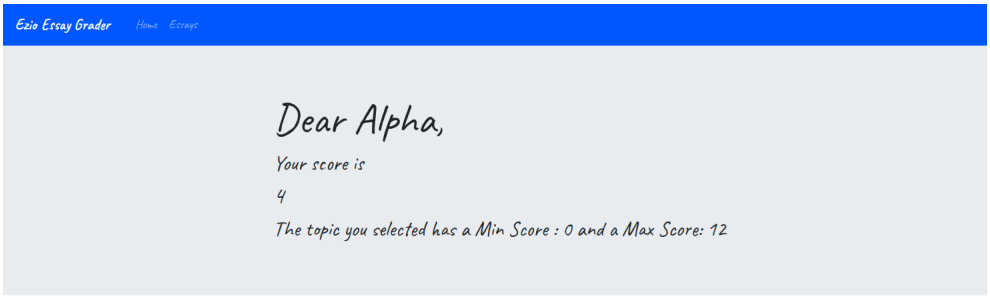


Figure 8.4: Essay Score

Chapter 9

CONCLUSIONS

The essay grading task is a laborious task which was automated with the help of Deep Learning. Various models were studied and the suitable model was used in the research. It was found that the LSTM networks were a perfect fit as discussed. The defined objectives were met and the results have been discussed in the previous chapter.

The literature review has yielded the insights into some of the useful approaches to solve this problem. Also it showed the dependence of the model on the amount of data.

The obtained accuracy is 96.2 % using a 3-folds cross validation. The grader can be used reliably. It was tested on extreme scenarios as well, and the result obtained is satisfactory.

9.1 Future Work

The essay grader can be used by teachers for grading student essays. It can also be used by testing agencies for test of English writing to take the burden off the human graders. The concept can be extended to other languages too if the dataset is available. Lastly, the accuracy of the model can still be increased if more data is feeded to the LSTM network while training. More sophicated models can be developed using transfer learning at the cost of using more resources required to train the model.

9.2 Applications

1. The Automated Essay Grading application can be used in education field to grade the essays of students of school and colleges.
2. The application can used in exams like GRE or TOEFL to grade the essay

with high accuracy.

3. It can be used to categories people with respect to their writing skills.

Appendix A

Problem statement feasibility and completeness analysis

This project is in the domain of Deep Learning. The problems in the Deep Learning domain depend on various factors such as data size, hyper-parameters, etc. The hyperparameters are more deciding than the data size in consideration. So, given proper tuning of the hyperparameters, the Deep Learning problems are approachable in polynomial time and hence are known to be P type problems. The majority of the time is spent in preprocessing of the dataset, followed by training of the models. The training time can be reduced by using GPU-enabled DL libraries.

Hence, it can be concluded that the project model falls in P type.

Appendix B

Appendix: Publications

Automated Essay Grading

Vinay Sanga, Shreyansh Patil, Abhishek Jagtap, Prashant Raut, Geeta S Navale

Department of Computer Engineering, Sinhgad Institute of Technology & Science, Savitribai Phule Pune University, Pune

vinayklind@gmail.com
shreyanshpatil98@gmail.com
abhishekjagtap88@gmail.com
rautprashant902@gmail.com
gsnavale_sits@sinhgad.edu

Abstract

In the task of essay grading a grade is assigned to an essay in the most human-like manner possible. Authors tackle this problem by building statistical models which try and predict how a human would evaluate an essay. In this paper, we propose to use Long Short Term Memory (LSTM) neural networks to create an essay grader to undertake this daunting task. The grader will give an essay a grade which would be expected from a human grader. The user will be able to see his score instantaneously which is an improvement as compared to human graders. There are some very challenging tasks such as recognizing the context of the essay. Also, the vocabulary and various versions of English will have to be considered while evaluating the essay. The grader will be able to grade an essay without any bias and high accuracy.

Keywords— Essay Grading, Deep Learning, LSTM, Computerised Testing, Text Grading.

1. INTRODUCTION

Essays are a tool for testing the students' fluency, vocabulary and grammatical correctness in a language. They are also useful to test one's creativity, originality and articulateness. The highly subjective and diverse nature of an individual in writing an essay makes it difficult to grade the essay uniformly across many human graders. In addition to this there are various other biasing factors in grading an essay. There has been research on automatic essay grading since the 1960s. The first systems based their grading on the surface information from essays. These systems were successful though they failed to capture aspects like grammatical correctness and language fluency. Much research has been conducted in the field most notably by Educational Testing Service (ETS). Clubbing this together with the resurgence in new technologies such as neural networks, deep neural networks, there is a whole new world of possibilities due to their capacity of modelling complex patterns in data. These methods do not depend on feature engineering so they are really useful for solving problems in an end-to-end fashion. With this intuition this project aims to the relevant knowledge in the field of education and try to create an essay grader which can make quality education more accessible. The work also explores methods of improving the quality and usability of the system.

I. MOTIVATION

The human graders unknowingly tend to grade an essay biasing towards the individual subject matter presented. Another major drawback is the time required to grade essays can be significantly high. The present technologies present an excellent opportunity to automate tedious tasks such as essay grading. Availability of powerful Deep Learning libraries is a major push towards the reliance and devising the system. This project will help in maintaining an unbiased and fair approach towards evaluating the written essay for competitive exams, tests, etc. which is in turn beneficial in multiple ways.

II. RELATED WORK

Alex Adamson et. al. [1] have implemented an essay grader on the Hewlett Foundation dataset using different ML techniques such as SVM, Latent Semantic Analysis. Each essay was graded by at least two humans. It uses the Quadratic Weighted Kappa as closeness measure. The general pipeline involves extracting features from the raw essays, and iteratively training and using k- folds cross-validation on the model on selected essay sets in order to optimize hyperparameters. It is shown that for essays of intermediate writing level and given enough human graded training examples for a writing prompt, they can automate the grading process for that prompt with fairly good accuracy. However, the classifier was built for this task with limited success.

Derrick Higgins et. al. elaborated Support vector machine (SVM) for classification and regression [2]. SVM technique relies on kernel functions. There were multiple goals in this work. The authors wanted to introduce a concept of essay coherence comprising multiple aspects, and investigate what linguistic features drive each aspect in student essay writing. They have worked with writing experts to develop a comprehensive protocol that details how coherence in writing can be evaluated, either manually or automatically. Using this protocol, human annotators labeled a corpus of student essays, using the coherence dimensions. However, this approach is relatively new and though better than previous ones is valid only theoretically.

The LSA approach to detect major research topics and themes of a multidisciplinary field was applied by Gang Kou et. al. [3]. The LSA analysis can be summarized in three main steps. The first step is to set up a term-document matrix in which each row stands for a key word or term and each column stands for a document or context in which the key word appears. An entry in the matrix is the frequency of a key word in the corresponding document. The second step is to transform the term frequencies in a term-document matrix using various weighting schemes. The third step is to perform SVD on the matrix to reduce the dimensionality, which is the key feature of the LSA method.The limitations are - a. It is not possible to use LSA on an unestablished field of research; b. Only English journals are considered in this approach.

Kaveh et. al discusses about various kinds of approaches available for constructing a Machine Learning model. They are also using the dataset mentioned in the first paper. They have selected LSTM model [4] as it works the best among all the available ML models. They have discussed their approach as well as the work done by other prominent researchers. Their model does not require feature engineering as compared to some other models. The model performs 5.6% better than the baseline performance. There is still quite a scope in improving the model as per their views.

In their paper, Hongbo Chen et. Al [5] argue that the current AES systems can be further improved by taking into accountthe agreement between human and machine raters. They have proposed a listwise learning to rank approach to automated essay scoring (AES) by directly incorporating the human-machine agreement into the loss function.

Furthermore, evidences supporting the use of Deep Learning are shown by Ronan Collbert et. Al[6]. They have shown how both multitask learning and semi-supervised learning improve the generalization of the shared tasks, resulting in state of the-art performance. They have used Part-Of-Speech Tagging (POS), Chunking, Named Entity Recognition (NER), Semantic Role Labeling (SRL), Language Models and Semantically Related Words (“Synonyms”) similar to previously seen approaches.

In an another approach[7], Peter Phandi et. Al proposed a novel domain adaptation technique based on Bayesian linear ridge regression and domain adaptation. Domain adaptation is the task of adapting knowledge learned in a source domain to a target domain.

TABLE I
LITERATURE SURVEY

Ref No.	Highlights	Observations
1.	• Essays can be graded without the help of the teacher. • Different models are compared and best amongst them is used. • The grades are equivalent to human graders.	• The size of essay which can be graded is limited. • Good quality training examples are required as presently the accuracy is limited to the present set.
2.	• Hybrid approach is used to classify different coherence dimensions with a high- or low-quality rank. • In each of the experiments with less fold-cross validations results are reported for large set of essays.	• Choosing appropriate kernel function is difficult and complex task. • In case of using high dimension kernel it generates too many support vectors which reduces training speed.
3.	• Doesn't require feature engineering. • One of the best implementation of the grading system.	• Training time is more as it is a Deep Learning model.
4.	• LSA is capable of assuring decent results. It works well on dataset with diverse topics. • LSA can handle Synonymy problems to some extent. • It is faster as compared to other dimensionality reduction models.	• LSA depends on identifying frequent word usage patterns from a collection of text, it is difficult to capture a research area if it is not well established. • The research abstracts collected in this analysis include only English language journals.
5.	• A deep neural network model that is capable of representing local conceptual and usage information by using LSTM.	• ATS(automated text scoring) takes more time to generate text score using deep neural network.
6.	• A model for scoring essay dimension of argument strength which is important aspect of argument essays this model gives argument strength. • This model gives argument strength of essays which convince most reader.	• This model is only detecting the scores of argument essays automatically but not gives scores of all types of essays.
7.	• A general deep NN architecture for NLP. • This architecture is extremely fast enabling to take advantage of huge databases.	• When training the SRL task jointly with their language, model achieved state of-the-art performance in SRL without any explicit syntactic features.
8.	• Domain adaptation can achieve better results compared to using just the small number of target domain data or just using a large amount of data from a different domain.	• The effectiveness of using domain adaptation when only have a small number of target domain essays is discussed.

	• This research will help reduce the amount of annotation work needed to be done by human graders to introduce a new prompt.	

III. PROPOSED SYSTEM

The grading system is implemented as a Client – Server Architecture. The essay writers are the clients who can access the app through their web browsers. The model and other implementations as usual reside at the server end. The Essay will be passed to the Server for grading and then will be submitted for grading. After grading, the grade will be shown to the user on the screen. The representation of the system is shown in Figure no. 1

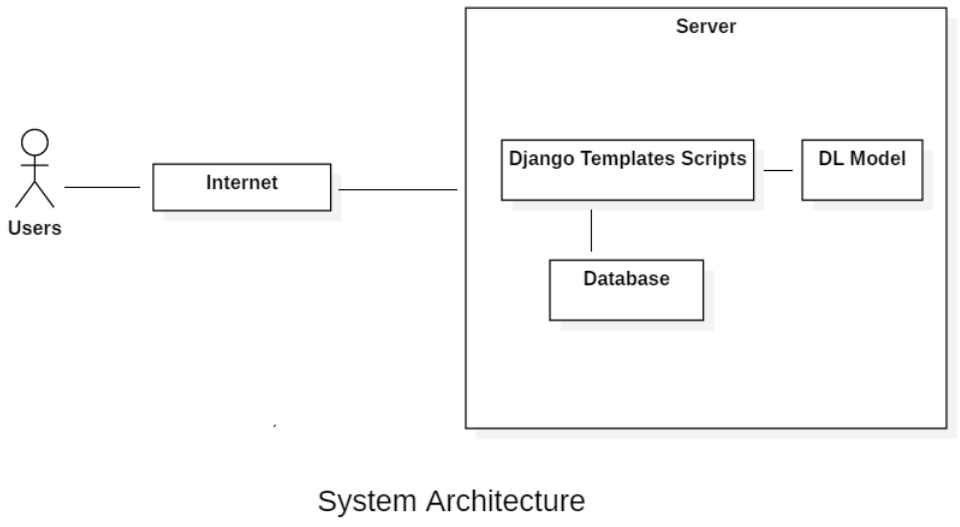


Figure no. 1: System Architecture

The communication between the LSTM predictor and the Backend script is the most important part. If the libraries are not compatible, it can cause issues due to the threading involved in the scripts.

IV. RESULTS

The essay grader achieved an accuracy of 96.2% using a 3-folds cross validation. The model was trained and tested on both Google Colab as well as Local PC. The accuracy fluctuated between 96.17% -98.7% based on the random vector encodings but mostly it lingered around 96.2%. The obtained accuracy can be seen from Figure no. 2

```
163/163 [==============================] - 3s 18ms/step - loss: 7.6775 - mae: 1.5711
Epoch 37/50
163/163 [==============================] - 3s 19ms/step - loss: 7.8863 - mae: 1.5725
Epoch 38/50
163/163 [==============================] - 3s 19ms/step - loss: 8.0074 - mae: 1.5718
Epoch 39/50
163/163 [==============================] - 3s 19ms/step - loss: 7.8068 - mae: 1.5676
Epoch 40/50
163/163 [==============================] - 3s 19ms/step - loss: 7.4385 - mae: 1.5597
Epoch 41/50
163/163 [==============================] - 3s 19ms/step - loss: 7.5482 - mae: 1.5364
Epoch 42/50
163/163 [==============================] - 3s 19ms/step - loss: 7.6092 - mae: 1.5582
Epoch 43/50
163/163 [==============================] - 3s 19ms/step - loss: 7.4036 - mae: 1.5433
Epoch 44/50
163/163 [==============================] - 3s 19ms/step - loss: 7.0982 - mae: 1.5263
Epoch 45/50
163/163 [==============================] - 3s 19ms/step - loss: 7.2797 - mae: 1.5259
Epoch 46/50
163/163 [==============================] - 3s 19ms/step - loss: 7.6733 - mae: 1.5462
Epoch 47/50
163/163 [==============================] - 3s 19ms/step - loss: 7.6982 - mae: 1.5389
Epoch 48/50
163/163 [==============================] - 3s 19ms/step - loss: 7.5287 - mae: 1.5333
Epoch 49/50
163/163 [==============================] - 3s 19ms/step - loss: 7.4795 - mae: 1.5172
Epoch 50/50
163/163 [==============================] - 3s 19ms/step - loss: 7.0886 - mae: 1.5117
Kappa Score: 0.9617094280868951
```

Figure no. 2: Accuracy measure in Kappa Score

The essay topic selection window and the grader score is displayed in figures 3, 4 & 5.

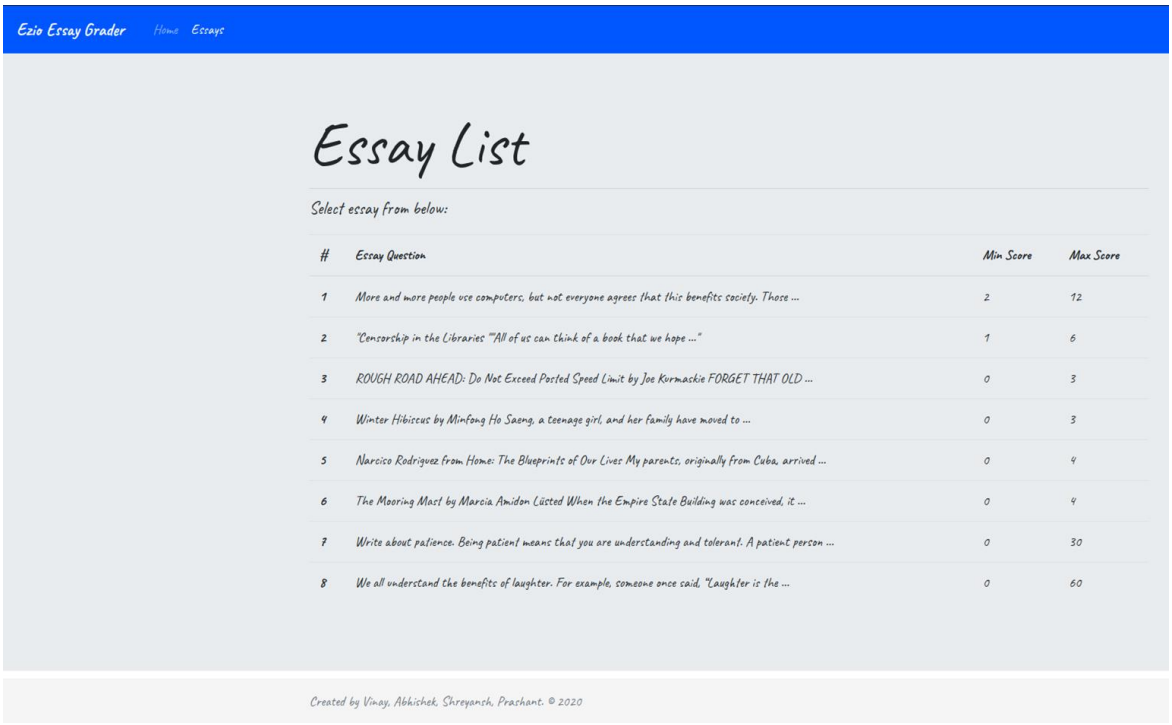


Figure no. 3: List of Essay Topics displayed

Figure no. 4: The text box for writing the essay

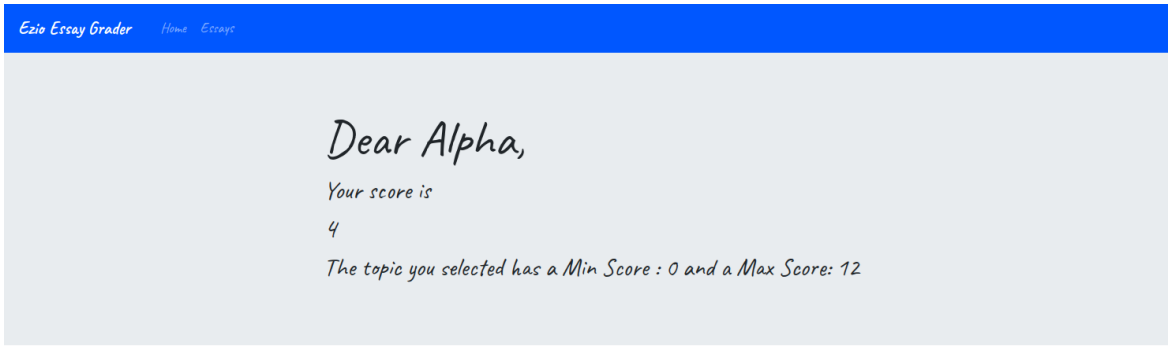


Figure no. 5: Score given by the grader

V. CONCLUSIONS

The essay grading task is a laborious task which can be automated with the help of Deep Learning. The grader which is proposed in the paper has an accuracy of 96.17% which is the highest that has been observed until now. The grader is very good in differentiating between ambiguous sentence formations and the new vector representation is very good in implementing the NLP translations.

VI. FUTURE SCOPE

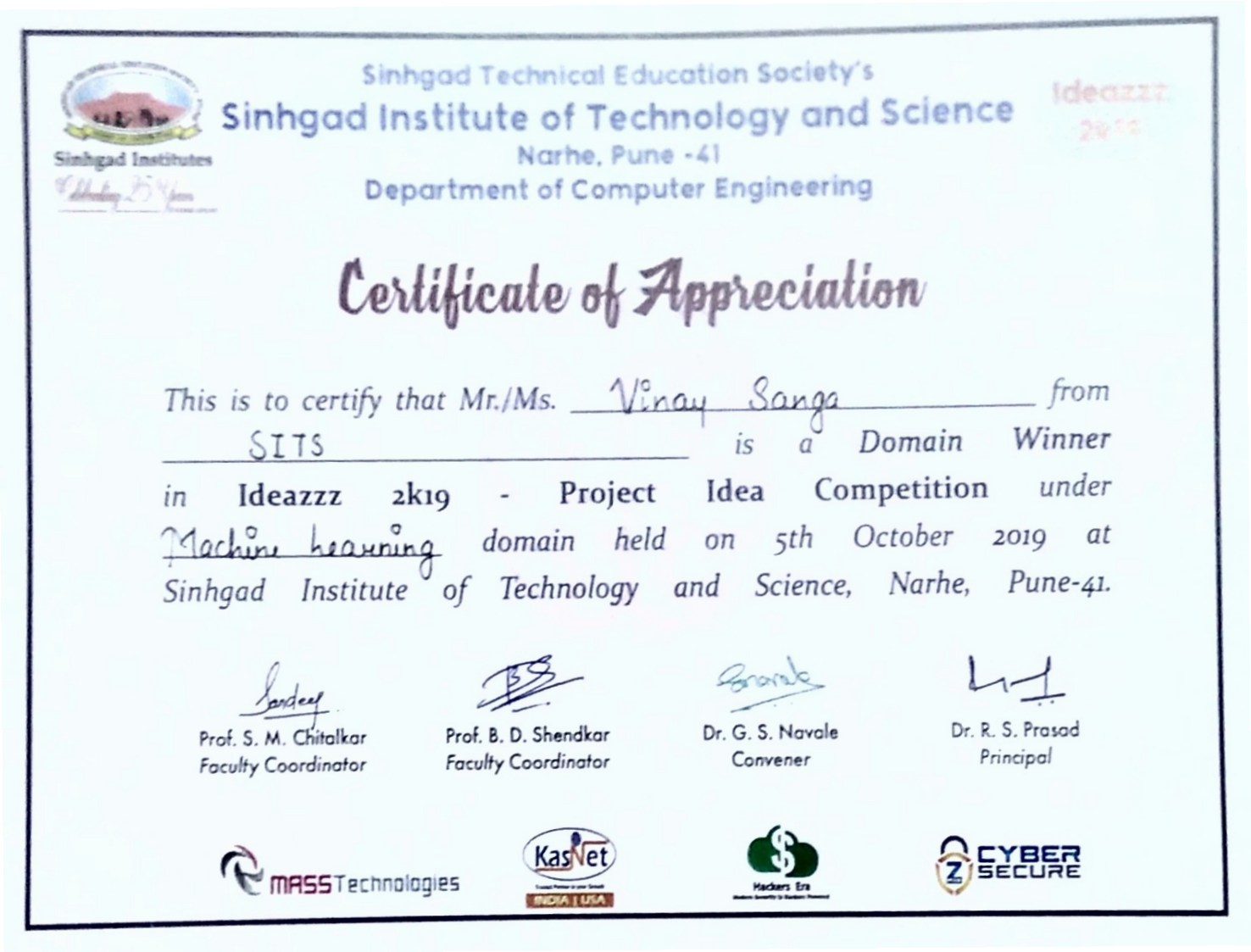
The essay grader can be used by teachers for grading student essays. It can also be used by testing agencies for test of English writing to take the burden off the human graders. The concept can be extended to other languages too if the dataset is available. Lastly, the accuracy of the model can still be increased if more data is feeded to the LSTM network while training. More sophicated models can be developed using transfer learning at the cost of using more resources required to train the model.

REFERENCES

1. Adamson Alex, Andrew Lamb, and Ralph Ma, Automated Essay Grading. 2014
2. Higgins Derrick, Jill Burstein, Daniel Marcu, and Claudia Gentile, Evaluating Multiple Aspects of Coherence in Student Essays. In HLT-NAACL, pp. 185-192. 2004.
3. Gang Kou, Yi Peng, An Application of Latent Semantic Analysis for Text Categorization In International Journal of Computers, Communications & Control (IJCCC) 10(3):357 April 2015
4. Kaveh Taghipour, Hwee Tou Ng, A Neural Approach to Automated Essay Scoring In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, November 2016
5. Isaac Persing and Vincent Ng , Modeling Argument Strength in Student Essays , 2015.
6. Dimitrios Alikaniotis , Helen Yannakoudakis and Marek Rei , Automatic Text Scoring Using Neural Networks , 16 Jun 2016.
7. Ronan Collobert and Jason Weston NJ 08540, A Unified Architecture for Natural Language Processing,2011.
8. Peter Phandi1 , Kian Ming A. Chai2 and Hwee Tou Ng1 , Flexible Domain Adaptation for Automated Essay Scoring Using Correlated Linear Regression , 17-21 September 2015.

Appendix C

Appendix: Certificates



Sinhgad Technical Education Society's

Sinhgad Institute of Technology and Science

Narhe, Pune -41

Department of Computer Engineering

Ideazzz 2k19

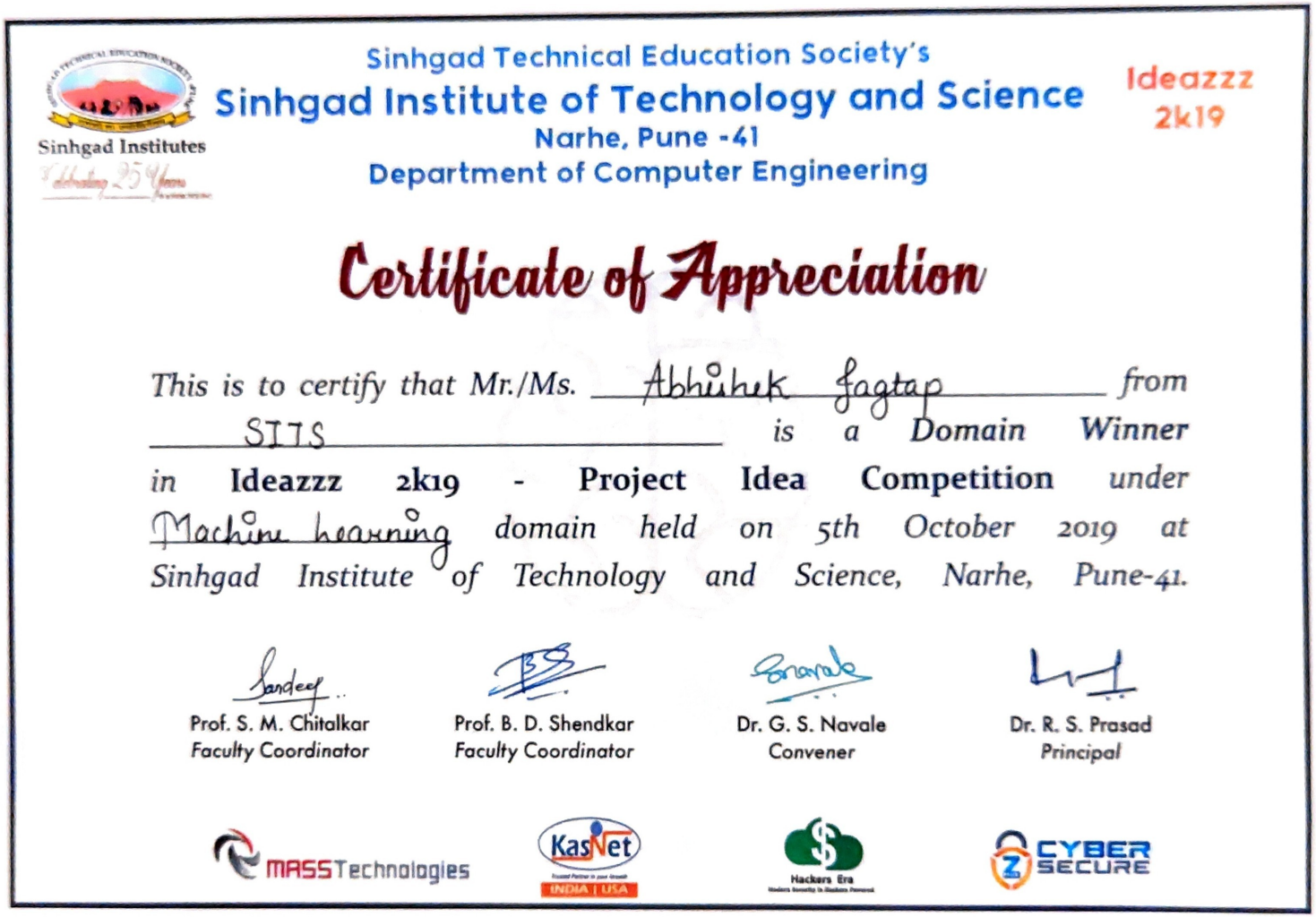
Sinhgad Institutes

Certificate of Appreciation

This is to certify that Mr./Ms. Vinay Sanga *from* SITS *is a Domain Winner in* **Ideazzz 2k19 - Project Idea Competition** *under* Machine learning *domain held on 5th October 2019 at Sinhgad Institute of Technology and Science, Narhe, Pune-41.*

Prof. S. M. Chitalkar	Prof. B. D. Shendkar	Dr. G. S. Navale	Dr. R. S. Prasad
Faculty Coordinator	Faculty Coordinator	Convener	Principal

MASS Technologies KasNet INDIA | USA Hackers Era CYBER SECURE



Sinhgad Institutes

Sinhgad Technical Education Society's

Sinhgad Institute of Technology and Science

Narhe, Pune -41

Department of Computer Engineering

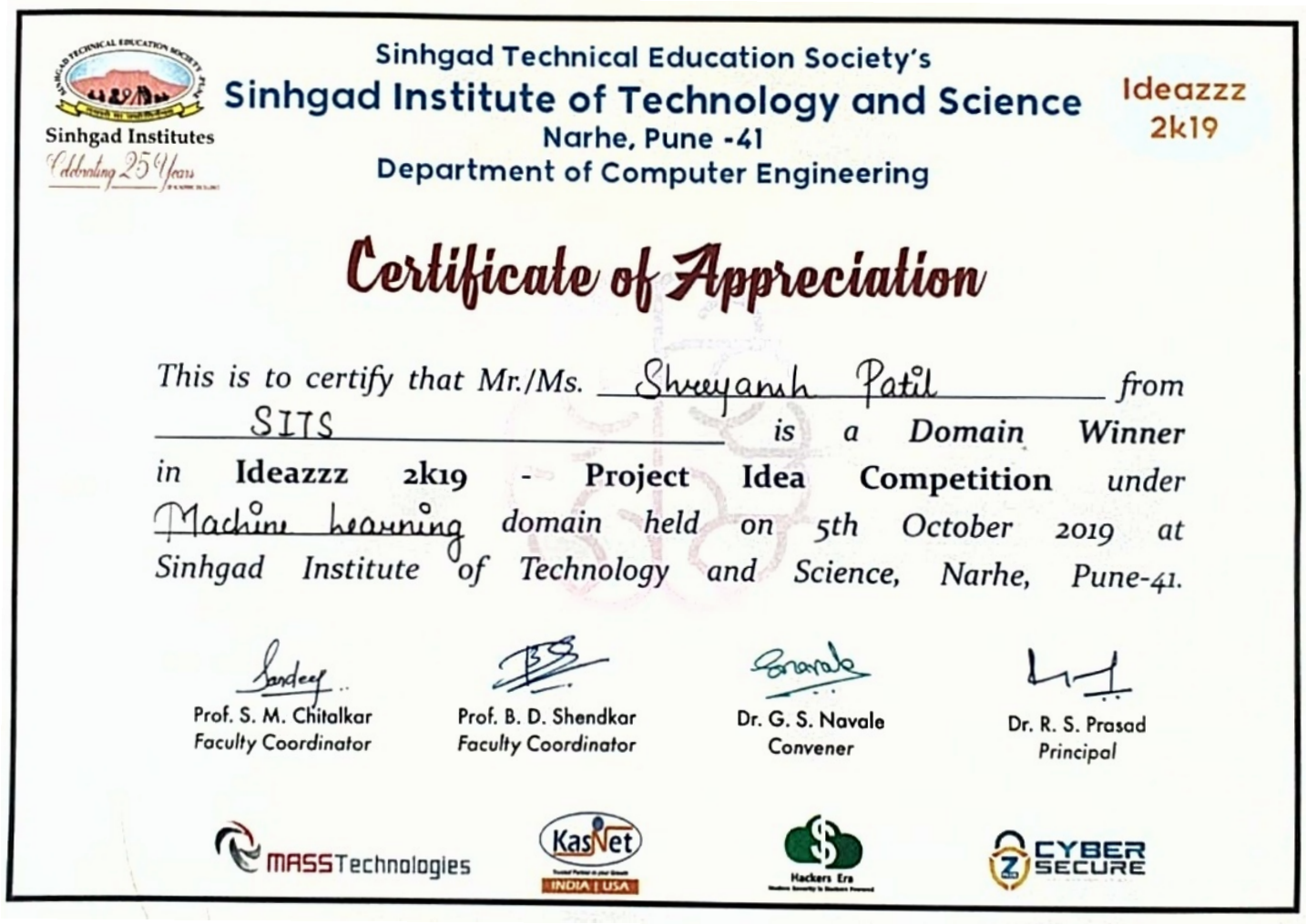
Ideazzz 2k19

Certificate of Appreciation

This is to certify that Mr./Ms. Abhishek Jagtap *from* SITS *is a* **Domain Winner** *in* **Ideazzz 2k19 - Project Idea Competition** *under* Machine learning *domain held on 5th October 2019 at Sinhgad Institute of Technology and Science, Narhe, Pune-41.*

Prof. S. M. Chitalkar	Prof. B. D. Shendkar	Dr. G. S. Navale	Dr. R. S. Prasad
Faculty Coordinator	*Faculty Coordinator*	Convener	*Principal*

MASSTechnologies KasNet INDIA | USA Hackers Era CYBER SECURE



Sinhgad Technical Education Society's

Sinhgad Institute of Technology and Science

Narhe, Pune -41

Department of Computer Engineering

Sinhgad Institutes

Celebrating 25 Years

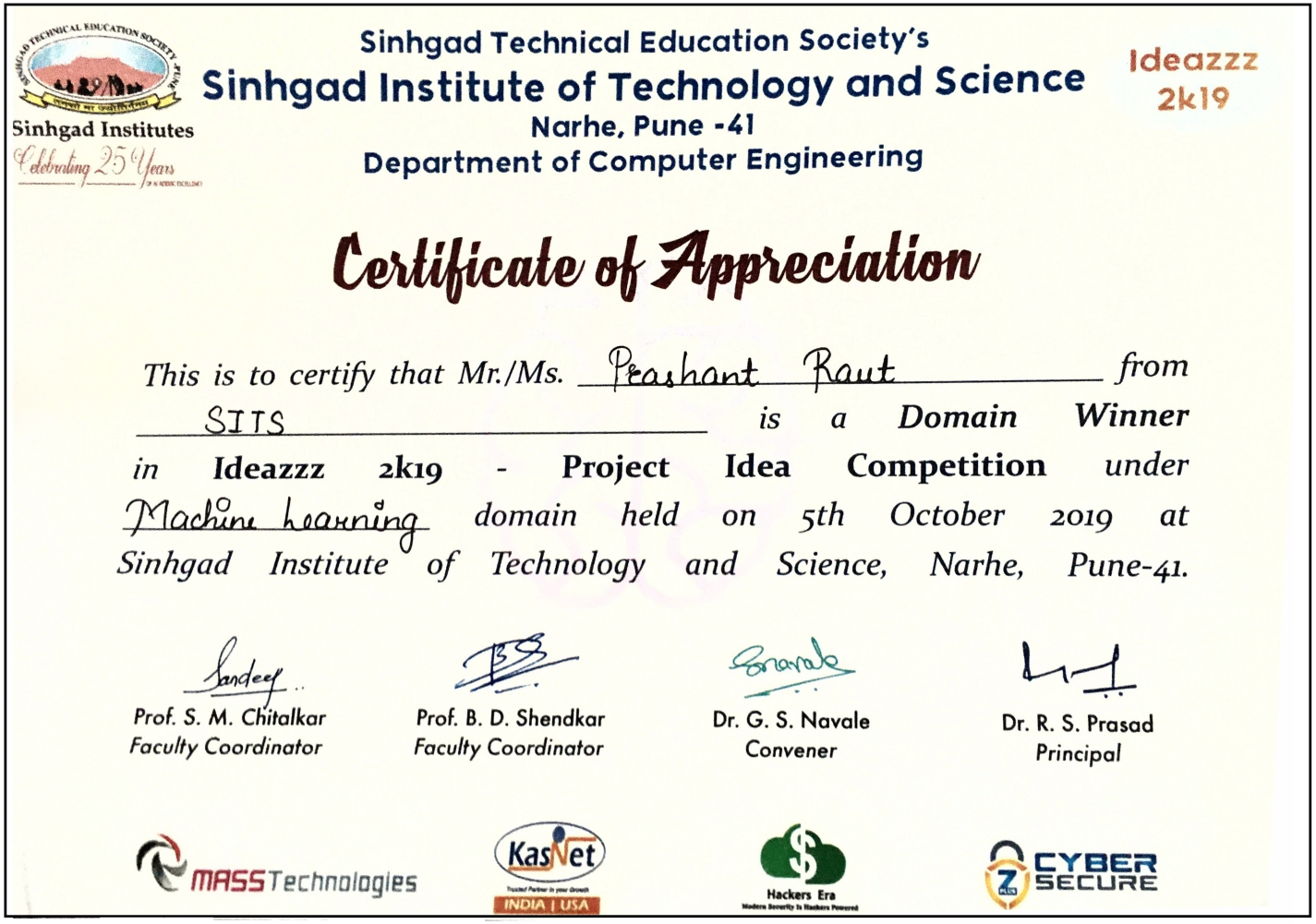
Ideazzz 2k19

Certificate of Appreciation

This is to certify that Mr./Ms. Shreyansh Patil *from* SITS *is a* **Domain Winner** *in* **Ideazzz 2k19 - Project Idea Competition** *under* Machine learning *domain held on 5th October 2019 at Sinhgad Institute of Technology and Science, Narhe, Pune-41.*

Prof. S. M. Chitalkar	Prof. B. D. Shendkar	Dr. G. S. Navale	Dr. R. S. Prasad
Faculty Coordinator	*Faculty Coordinator*	Convener	*Principal*

MASSTechnologies KasNet INDIA | USA Hackers Era CYBER SECURE



Sinhgad Technical Education Society's

Sinhgad Institute of Technology and Science

Narhe, Pune -41

Department of Computer Engineering

Ideazzz 2k19

Sinhgad Institutes

Celebrating 25 Years

Certificate of Appreciation

This is to certify that Mr./Ms. Prashant Raut *from* SITS *is a* ***Domain Winner*** *in* **Ideazzz 2k19 - Project Idea Competition** *under* Machine learning *domain held on 5th October 2019 at Sinhgad Institute of Technology and Science, Narhe, Pune-41.*

Prof. S. M. Chitalkar	Prof. B. D. Shendkar	Dr. G. S. Navale	Dr. R. S. Prasad
Faculty Coordinator	*Faculty Coordinator*	Convener	Principal

MASS Technologies

KasNet INDIA | USA

Hackers Era

CYBER SECURE

SINHGAD INSTITUTE OF
TECHNOLOGY AND SCIENCE, PUNE

Sinhgad Institutes

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr./Mr./Ms. Vinay Sanga

of Sinhgad Institute of Technology and Science *has participated / presented a paper entitled*

Automated Essay Grading *at*

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

YASHIKA PUBLICATIONS

Scopus®

Sinhgad Institutes

SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE, PUNE

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr./Mr./Ms. Abhishek Jagtap

of Sinhgad Institute of Technology and Science *has participated / presented a paper entitled*

Automated Essay Grading *at*

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

YASHIKA PUBLICATIONS Scopus®

Sinhgad Institutes

SINHGAD INSTITUTE OF
TECHNOLOGY AND SCIENCE, PUNE

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr./Mr./Ms. Shreyansh Patil

of Sinhgad Institute of Technology and Science *has participated / presented a paper entitled*

Automated Essay Grading *at*

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

YASHIKA PUBLICATIONS Scopus®

Sinhgad Institutes

SINHGAD INSTITUTE OF TECHNOLOGY AND SCIENCE, PUNE

Presents

International Conference on Future Intelligence in Science & Technology

31st May 2020

Certificate

This is to certify that Prof./Dr./Mr./Ms. Prashant Raut

of Sinhgad Institute of Technology and Science *has participated / presented a paper entitled*

Automated Essay Grading *at*

"International Conference on Future Intelligence in Science & Technology " held on 31st May 2020 at Sinhgad Institute of Technology and Science, Pune

Prof. S.A. Kulkurni
Organizing Secretory, SITS FIST 2020

Dr. R. S. Prasad
General Chair, SITS FIST 2020

YASHIKA PUBLICATIONS Scopus®

Appendix D

Appendix: Plagiarism Report

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	995	Date	July 19,2020
Characters	6996	Exclude Url	

6%	94%	3	45
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

This concludes the first chapter of the report. All the initial details re-garding the project were discussed. The idea and theme of the project isnow clear and the further proceedings are scheduled successfully. The bud-geting is done and the effort required to implement the project is known. Inthe next chapter, there will be a discussion about the literature survey andthe related details of the project.3 Automated Essay GradingChapter 2LITERATURE SURVEYThe previous chapter tells about the description of this project. The problemstatement gives a brief idea about the project and the objectives gives a step-wise execution process of the project. In this chapter, some concrete basissupporting and extending the applicability of the project will be reviewedwith the hope of providing better insights into the foundations of the project.2.1 Literature SurveyEquivocating the statements above this section will discuss the findings doneduring the literature review of the topic.Alex Adamson et. al. [alex] have implemented an essay grader on theHewlett Foundation dataset using different ML techniques such as SVM,Latent Semantic Analysis. Each essay was graded by at least two humans.It uses the Quadratic Weighted Kappa as closeness measure. The generalpipeline involves extracting features from the raw essays, and iterativelytraining and using k-folds cross-validation on the model on selected essaysets in order to optimize hyperparameters. It is shown that for essays ofintermediate writing level and given enough human graded training examplesfor a writing prompt, we can automate the grading process for that promptwith fairly good accuracy. However, the classifier was built for this task withlimited success.Derrick Higgins et. al. elaborated Support vector machine (SVM) forclassification and regression [derrick]. SVM technique relies on kernel func-tions. There were multiple goals in this work. The authors wanted to intro-duce a concept of essay coherence comprising multiple aspects, and investi-gate what linguistic features drive each aspect in student essay writing. Theyhave worked with writing experts to develop a comprehensive protocol thatSITS, B. E. (Computer) 2015 Course, Project Stage II, 2019-204 details how coherence in writing can be evaluated, either manually or auto-matically. Using this protocol, human annotators labeled a corpus of studentessays, using the coherence dimensions. However, this approach is relativelynew and though better than previous ones is valid only theoretically.The LSA approach to detect major research topics and themes of a multi-disciplinary field was applied by Gang Kou et. al. [gang]. The LSA analysiscan be summarized in three main steps. The first step is to set up a term-document matrix in which each row stands for a key word or term and eachcolumn stands for a document or context in which the key word appears.An entry in the matrix is the frequency of a key word in the correspond-ing document. The second step is to transform the term frequencies in aterm-document matrix using various weighting schemes. The third step isto perform SVD on the matrix to reduce the dimensionality, which is thekey feature of the LSA method. The limitations are - a. It is not possible touse LSA on an unestablished field of research; b. Only English journals areconsidered in this approach.Kaveh et. al discusses about various kinds of approaches available forconstructing a Machine Learning model. They are also using the datasetmentioned in the first paper. They have selected LSTM model [kaveh] asit works the best among all the available ML models. They have discussedtheir approach as well as the work done by other prominent researchers.Their model does not require feature engineering as compared to some othermodels. The model performs 5.6% better than the baseline performance.There is still quite a scope in improving the model as per their views.2.2 SummaryThis section gives the important findings obtained during reviews as seen inSection 2.1 and are depicted in Table 2.15 Table 2.1: Literature

ReviewRef.No.HighlightsObservations1•Essays can be graded with-out the help of the teacher.•Different models are com-pared and best amongstthem is used.•The grades are equivalent tohuman graders.•The size of essay which canbe graded is limited.•Good quality training ex-amples are required aspresently the accuracy islimited to the present set.2•Hybrid approach is used toclassify different coherencedimensions with a high- orlow-quality rank.•In each of the experimentswith less fold-cross valida-tions results are reported forlarge set of essays.•Choosing appropriate ker-nel function is difficult andcomplex

cross valida-tions results are reported forlarge set of essays.•Choosing appropriate kel-nel function is difficult andcomplex task.•In case of using high dimen-sion kernel it generates toomany support vectors whichreduces training speed.6 Ref.No.HighlightsObservations3•Doesn't require feature engi-neering.•One of the best implementa-tion of the grading system.•Training time is more as it isa Deep Learning model.4•LSA is capable of assuringdecent results. It works wellon dataset with diverse top-ics.•LSA can handle Synonymyproblems to some extent.•It is faster as compared toother dimensionality reduc-tion models.•LSA depends on identifyingfrequent word usage patternsfrom a collection of text,it is difficult to capture aresearch area if it is not wellestablished.•The research abstracts col-lected in this analysis in-clude only English languagejournals.From the surveys, it is clear that the project is quite feasible and quite alot of work has been done in the field related to this project. Also, there aresome attributes such as accuracy, scalability, performance, etc. which canbe further improved with the use of current technologies. More about thecurrent technologies and the implementation of the project will be discussedin the next chapter. Chapter 3SOFTWARE REQUIREMENTSSPECIFICATIONIn the last chapter, the basis of implementing this project were discussed.Following that, in this chapter Software Requirement Specification documentwill be elaborated. In the subsequent sections, there will be a discussionregarding the technical requirements, methodology to be used in SDLC, etc.3.1 Assumptions and DependenciesAssumptions –1. The essay writer agrees to the limit of the word specified.2. The system is used only for educational purposes.3. The dataset provided is a meaningful dataset so that we abide by'Garbage in Garbage out' phenomenon in Machine Learning.4. The knowledge provided is static in nature.Dependencies – •Python 3.7.3•NLTK 3.4.5•Numpy 1.17•Pandas 1.0.2•Gensim 3.7.3•TensorFlow 2.0.0SITS, B. E. (Computer) 2015 Course, Project Stage II, 2019-208 •Keras 2.3.1•Flask 1.1.23.2 Functional RequirementsThese are the functions which the system performs and the user can usethem on their device.3.2.1 System Feature 1The system should display the list of gradable essay topics to the users.These essays will be passed further to the grading system

Sources	Similarity
The Literature Review | A Complete Step-by-Step Guide a literature review is a survey of scholarly knowledge on a topic.across these studies, there is consistent evidence that body image issues are influenced not by social media usage in general, but by engagement with the visual and interactive aspects of these platforms. https://www.scribbr.com/dissertation/literature-review/	10%
Automated Essay Grading | 2 Models Our general pipeline involved extracting features from the raw essays, and iteratively training and using k-folds cross-validation on our model on selected essay sets in order to optimize hyperparameters. 2.1 Support Vector Regression. For each essay set, we featurized the essays... http://cs229.stanford.edu/proj2014/Alex Adamson, Andrew Lamb, Ralph Ma, Automated Essay Grading.pdf	5%
New Test for Computers: Grading Essays at College Level - The New... https://www.nytimes.com/2013/04/05/science/new-test-for-computers-grading-essays-at-college-level.html	4%

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	885	Date	July 19,2020
Characters	5991	Exclude Url	

9%	91%	4	43
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

3.2.2 System Feature 2The system should get grade by passing the essay which has been written byuser to the grading system which is present at the backend or web server.3.2.3 System Feature 3The grading system should be able to grade the essay passed to it by theweb server. The predicted grade will be return back to user through samechannel by which it was received.3.3 Nonfunctional RequirementsThis section describes the non functional requirements of the project whichincludes performance, safety, security, etc.3.3.1 Performance RequirementsThe system must be interactive and the delays involved must be less. So inevery action-response of the system, there are no immediate delays. In caseof opening links, uploading and submitting the essays the delay is below2 seconds. The result is obtained almost instantly because the grader ispreloaded in the memory while opening the website. Also when connectingto the server the delay is based editing on the distance of the 2 systems andthe configuration between them so there is high probability that there willbe or not a successful connection in less than 20 seconds for sake of goodcommunication.9 3.3.2 Safety RequirementsInformation transmission should be securely transmitted to server withoutany changes in information. The essay data and the score is sent to theserver as post data and therefore is highly secure.3.3.3 Security RequirementsThere is no need of user login and hence the authentication is not requiredwhich saves us from auth safety procedures.3.3.4 Software Quality Attributes1. SecurityThis is achieved by send the essay via POST method and neglecting the needof login.2. ReliabilityThe system will uses Flask micro-framework which is reliable.3. MaintainabilityThe separation of the modules allows easy maintainability.4. PortabilityThe web app uses Bootstrap which is a responsive framework so it can berun both on Desktop browser and Mobile browsers without any fuss.5. Resource UtilizationThe resource requirements are similar to a standard website as most of thework is done at the backend.3.4 System Requirements1. Web Browser – Chrome, Firefox, Microsoft Edge.2. An internet connection with min speed 2 Mbps.10 3.5 Analysis Models: SDLC Model to be appliedCRISP-DM stands for cross industry standard process for data mining. It isa comprehensive data mining methodology and process model that providesanyone from novices to data mining experts with a complete blueprint forconducting a data mining project. CRISP-DM breaks down the life cycle ofa data mining project into six phases. These 6 high-level phases of CRISP-DM are still a good description for the analytics process. The architectureis shown in Fig. 3.1Figure 3.1: CRISP-DM Model11 1. Business Understanding:-First and foremost step in Analytics project is understanding client'sbusiness to formulate problem statement. Once we define problemstatement then we can drive data accordingly. Also Focus should be onunderstanding the project objectives and requirements from a businessperspective, then converting this knowledge into a data mining prob-lem definition and a preliminary plan designed to achieve the belowobjectives•What the client really wants to accomplish?•Uncover important factors including the constraints, competitiveobjectives2. Data Understanding:-An initial data collection is present and subsequent activities in orderto get familiar with the data, to identify data quality problems, todiscover initial level of insights into the data or to detect interestingsubsets to form hypotheses for hidden information are done•Describing the data•Exploring the data•Verifying the data quality3. Data Preparation:-It covers all activities to construct the final dataset from the initial rawdata. Converting raw data into analytical dataset is very important.Quality of cleaned data will impact on model performance. Data prepa-ration tasks are likely to be performed multiple times and not in anyprescribed order. Tasks include table, record and attribute selection aswell as transformation and cleaning of data for modeling tools.•Cleaning the raw data•Usage of Sampling

Techniques4. Modelling:-In this phase, various modeling techniques are selected and appliedand their parameters are calibrated to optimal values. Typically, thereare several techniques for the same data mining problem type. Sometechniques have specific requirements on the form of data. Therefore,stepping back to the data preparation phase is often necessary.12 5. Evaluation:-Thorough evaluation of the model and review of the steps executedto construct the model is done to check the business objectives. Akey objective is to determine if there is some important business issuethat has not been sufficiently

business objectives. Akey objective is to determine if there is some important business issuethat has not been sufficiently considered. At the end of this phase, adecision on the use of the data mining results should be reached.3.6 System Implementation PlanIt is divided into parts as follows -Part 1: In this part, the concepts and technologies related to the implemen-tation will be learnt. This part is the learning part of the phase 2 and is willtake around 8 weeks. Objective 1 will be completed in this part.Part 2: In this part, the data processing will take place and will take around5 weeks to complete. This part covers objective 2 of the project.Part 3: Actual DL Model implementation and testing will be undertakenin this part. This part is the most laborious part of the whole project andwill take around 8 weeks to be accomplished. The objectives 3 and 4 will beachieved in this part.Part 4: This part is the final part of the phase 2. This phase relates to all thehousekeeping of the project. The finalization of the report for phase 2 willbe done in this part. The system will be fully implementation after this part.In this chapter, SRS was elaborated and various requirements of theproject were studied. In the following chapter, there will be discussionabout the implementation architecture and related details using various vi-sual tools.

Sources	Similarity
CCNA2 v6.0 Chapter 9 Exam Answers 2019 - Passed Full Score 100% increases the performance of packet transmission to the internet. enables the easy deployment of applications that require end-to-end traceability.ccna 2 rse 6.0 chapter 9 exam answers 2018 2019 03. dynamic nat with a pool of two public ip addresses. pat using an external interface. https://www.ccna7.com/ccna2-v6-0/ccna2-v6-0-chapter-9-exam-answer-2017/	5%
Cross-Industry process for data mining | by mayank aggarwal | Medium It is a comprehensive data mining methodology and process model that provides anyone from novices to data mining experts with a complete blueprint for conducting a data mining project. CRISP-DM breaks down the life cycle of a data mining project into six phases. These 6 high-level... https://medium.com/@thecodingcookie/cross-industry-process-for-data-mining-286c407132d0	7%

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	1000	Date	July 19,2020
Characters	6632	Exclude Url	

0%	100%	0	43
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

Chapter 4SYSTEM DESIGNThis chapter will discuss the design specifications of the project. Variousdiagrams are drawn to make the idea of the project clearer to understand.As the project is focussed more on the implementation of DL model and lesson other software parts, there are very few UML diagram to be shown anddiscussed about.4.1 System ArchitectureThe architecture of the system is the most fundamental approach to showthe overall structure. The grading system is implemented as a Client - ServerArchitecture. The essay writers(user) are the clients. The model and otherimplementations as usual reside at the server end. There are not muchcomponents in either of the side because the input to output task is simplepipeline to-and-fro the DL Model residing at server end. The architectureis shown in Fig. 4.1.LSTMs are a special kind of RNN, capable of learning long-term depen-dencies. They were introduced by Hochreiter & Schmidhuber (1997). Theywork tremendously well on a large variety of problems, and are now widelyused. LSTMs are explicitly designed to avoid the long-term dependencyproblem, which is present in standard RNN.All recurrent neural networks have the form of a chain of repeating modulesof neural network. In standard RNNs, this repeating module will have avery simple structure, such as a single tanh layer.It is shown in Fig. 4.2LSTMs also have this chain like structure(see Fig. 4.3), but the repeatingmodule has a different structure. Instead of having a single neural networklayer, there are four, interacting in a very special way.SITS, B. E. (Computer) 2015 Course, Project Stage II, 2019-2014 Figure 4.1: System ArchitectureFigure 4.2: Standard RNNs [lstm]Figure 4.3: LSTMs [lstm]15 4.2 Mathematical ModelBelow is the mathemical formulation used inside the neural network to calu-late the weights and biases. The gate equations are shown in 4.4.Figure 4.4: Mathematical model diagram4.3 Entity Relationship DiagramsThe ER diagram shows the relationship between various entities stored inthe database. The database will only be used to store the login data andthe scores given by the grader for analytical purposes. Due to this reason,the use of database in the system is minimal and hence the entities in thedatabase are only a couple. The main emphasis of the grading system lies inthe DL modeling part which doesn't need to deal with the databases oncethe training of the model is done. The ER diagram as illustrated in Fig. 4.5is fairly simple with only two entities named as Login and Scores.4.4 UML DiagramsA UML diagram is a diagram based on the UML used to represent a systemalong with its main actors, roles, actions, artifacts, activities, etc. in orderto better understand, alter, maintain, or document information about thesystem. The UML diagrams pertaining to the system are shown and brieflydiscussed in the upcoming sections.4.4.1 Activity DiagramThe activity diagram depicts the flow of control from one activity to another.Fig. 4.6 shows the array of activities such as displaying the essay topics,16 Figure 4.5: ER Diagramgetting the essay from the writers, grading the essay, etc.4.4.2 Deployment DiagramDeployment Diagram specifies how the system will be implemented phys-ically. The system has a simple client and server deployment as depictedin Fig. 4.3.2.1. The client will have to use web browser to requests to thesystem. The actual DL model will be on the web server and will be the onegrading the essay submitted by the user via Internet.4.4.3 Sequence DiagramSequence diagram simply depicts interaction between objects in a sequentialorder i.e. the order in which these interactions take place. The objects intothe consideration are User and Scoring Engine as shown in Fig. 4.8.17 Figure 4.6: Activity DiagramFigure 4.7: Deployment Diagram18 Figure 4.8: Sequence Diagram19 This concludes the visual representation on the system. Various UMLdiagrams were used to instill the structure of the project. It is now almostfully possible to go for implementation of the project but there are somepros and cons which needs to be discussed in the next chapter. Chapter

5PROJECT PLANA project plan is a formal document designed to guide the control and ex-ecution of a project. A project plan is the key to a successful project andis the most important document that needs to be created when startingany business project. The following chapter contains necessary points to beconsidered for the project plan.5.1 Project EstimateThe project estimate is prior calculation of the cost, resources and otherneeds. This is done after the requirements of the project are finalizedbetween both the parties. The estimated cost include the cost of hostingthe servers which can be range from INR

finalizedbetween both the parties. The estimated cost include the cost of hostingthe servers which can be range from INR 5000 - INR 10,000 based on theproviders and the type of service used.5.1.1 Reconciled EstimatesThe budget of a project is calculation or estimation of all the efforts andcosts required to implement the project. For this project, the budget hasbeen calculated by using CoCoMo model. The basic CoCoMo model wasused in Organic mode as the project is small and doesn't have too manycomplex budgeting factors.The basic CoCoMo equations are -E=ab(KLOC)bb(5.1)D=cb(E)db(5.2)SS=E/D(5.3)SITS, B. E. (Computer) 2015 Course, Project Stage II, 2019-2021 Estimated size of the project = 5 KLOCSo, using equations 5.1 & 5.2, we getE= 2.4(5)1.05= 13.01P MD= 2.5(13.01)0.38= 6.63MSS= 13.01/6.63 = 1.96PHere, E is Effort (measured in Person Months)D is Deployment Time (measured in Months)SS is Staff Size (units is Persons)Hence, Total Effort required is 13 person months(approx.) yielding a Devel-opment Time of 6.63 months and a Staff Size of 2 persons.As, the team size is 4 persons, the development time of 6.63 months can bespeeded up and calculated as follows:PersonsD21/6.6341/xSo,x= 2□6.63/4 = 3.3Hence, the project will require 3 month (approx.) tocomplete (theoretically).5.2 Risk ManagementRisk management is defined as the process of identifying, assessing and con-trolling threats to a project. Risk management is a process that seeks toreduce the uncertainties of an action taken through planning, organizingand controlling of both human and financial capital.It is the responsibility ofproject manager to go through potential threats. The project manager canidentify the risks and accordingly control them with the help of other stakeholders.5.2.1 Risk IdentificationThe risks identified in the project are very few because of high cohesivenessand low coupling involved. Also, open-source technologies are used through-out and no external APIs are used which makes the project self-dependent.

Sources	Similarity

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	980	Date	July 19,2020
Characters	6798	Exclude Url	

11%	89%	5	40
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

Chapter 6PROJECT IMPLEMENTATIONIn this chapter, we will see the details regarding the project implementa-tion. We will take a look at the tools and technologies used, algorithm forthe project.6.1 Overview of Project ModulesThe web app starts with a page showing the steps to follow in order to usethe website. The next module is essay writer module which supports writ-ing and editing of the essays. The final module is a grader module which isresponsible for grading the essay. These modules are explained below -1. The web site :The website welcomes the new user. This webite helps the new user to nav-igate between different pages through links . The result of the grader is alsodisplayed here.2. Essay writer module:This module helps the essay writer to write the essay or upload the text fileof essay.This module further helps to send the essay to grader module.3. Grader module:This is the main LSTM network which is responsible for grading the essay.It is a stateful LSTM which grades the human essays. The grader modulepreprocesses the essay, grades the essays and return the essay score.SITS, B. E. (Computer) 2015 Course, Project Stage II, 2019-2026 6.2 Tools and Technologies Used1. FlaskFlask is a web framework that provides with tools, libraries and technologiesthat allow you to build a web application.Flask is a lightweight web frame ofPython. It provides the user with libraries, modules and tools to help buildweb applications.2. Pandas and NumpyPandas is a high-level data manipulation tool. It is built on the Numpypackage and its key data structure is called the DataFrame. DataFramesallow you to store and manipulate tabular data in rows of observations andcolumns of variables.NumPy is a Python package which stands for 'Numerical Python'. It is thecore library for scientific computing, which contains a powerful n-dimensionalarray object, provide tools for integrating C, C++ etc. It is also useful inlinear algebra, random number capability etc.3. GensimGensim is an open-source library for unsupervised topic modeling and natu-ral language processing, using modern statistical machine learning. It is de-signed to handle large text collections using data streaming and incrementalonline algorithms, which differentiates it from most other machine learn-ing software packages that target only in-memory processing. It includesstreamed parallelized implementations of fastText, word2vec and doc2vec al-gorithms as well as latent semantic analysis (LSA, LSI, SVD), non-negativematrix factorization (NMF), latent Dirichlet allocation (LDA), tf-idf andrandom projections.4. Tensorflow - KerasTensorFlow is a free and open-source software library for dataflow and differ-entiable programming across a range of tasks. It is a symbolic math library,and is also used for machine learning applications such as neural networks.It is used for both research and production at Google. It lets users buildand train ML models easily using intuitive high-level APIs like Keras.Keras is an open-source neural-network library written in Python. It is capa-ble of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano,or PlaidML.It focuses on being user-friendly, modular, and extensible.Kerascontains numerous implementations of commonly used neural-network build-ing blocks such as layers, objectives, activation functions, optimizers, and a27 host of tools to make working with image and text data easier to simplifythe coding necessary for writing deep neural network code. In addition tostandard neural networks, Keras has support for convolutional and recur-rent neural networks. It supports other common utility layers like dropout,batch normalization, and pooling. It also allows use of distributed trainingof deep-learning models on clusters of Graphics processing units (GPU) andtensor processing units (TPU) principally in conjunction with CUDA.6.3 Algorithm DetailsThe algorithm used in the working of the web app is given below.1. Start2. Visit the Website.3. Click on essays.4. Select Essay Topic.5. Type essay as per selected topic or upload the essay as text file.6. Click on Grade me button.7. Essay Score is displayed.8. Stop Chapter 7SOFTWARE TESTINGSoftware testing is process of checking whether the actual results match theexpected results. It checks whether the developed software met the specifiedrequirements and identifies any defect in the software in order to produce aquality product. Testing is required for an effective performance of softwareapplication. In this project, only two types of testings were used which aredescribed following.7.1 Type of TestingFollowing types of testing is carried out for automated essay grader duringtesting phase 7.1.1 Unit TestingUnit testing is type of testing where individual

carried out for automated essay grader duringtesting phase.7.1.1 Unit TestingUnit testing is type of testing where individual unit or component of softwareis tested. In automated essay grading application the three units 1.essaygraders 2.essay writer and 3.hyperlink in all web pages were tested. Thethree units didn't produce any bugs when tested and functioned as expected.7.1.2 Integration TestingIt is type of testing where individual components are combined and testedas a group. All the modules were integrated and tested. A bug was found inwhich the user could not go back to the pages because of cache mismatch.The issue was fixed and all components work properly when integrated.SITS, B. E. (Computer) 2015 Course, Project Stage II, 2019-2029 7.2 Test cases and Test Results7.2.1 Test Case ID 1SummaryThe local file can be uploaded into the esaay textbox.Prerequisitesbrowser , computer.Test ProcedureHome page click , recent scores page clickExpected ResultWhen we click on any hyperlink , a new page shouldopen.Actual ResultAll the hyperlinks are working properly.StatusPassTable 7.1: Test case ID 17.2.2 Test Case ID 2Summarywhen we click on upload file all text get inserted intextbox. .PrerequisitesAny file such as text fileTest Proceduretext can be copied from file to textbox.Expected ResultThe local file can be uploaded into the esaay textbox.Actual ResultThe local file can be uploaded into the esaay textbox.StatusPassTable 7.2: Test case ID 230 7.2.3 Test Case ID 3SummaryWhen GradeMe Button is clicked the essay scoresshould be displayedPrerequisitesEssay with word length between 150 to 600.Test ProcedureUpload the file containing essay and check the scores.Expected ResultProper scores of essay.Actual ResultProper scores of essay.StatusPassTable 7.3: Test case ID 37.2.4 Test Case ID 4SummaryThe web pages should be properly displayed.PrerequisitesLatest version of browser.Test ProcedureCheck by resizing the browser window.Expected ResultAll the web pages and UI elements should properlydisplayed.Actual ResultAll the UI elements and web pages are loading both onmobile as well as desktop browser properly.StatusPassTable

Sources	Similarity
64 Natural language processing interview questions and answers Gensim is a production-ready open-source library for unsupervised topic modeling and natural language processing, using modern statistical machine learning. Gensim is implemented in Python and Cython for top performance and scalability. https://onlinecoursetutorials.com/interview-questions/natural-language-processing-interview-questions-and-answers/	4%
TensorFlow - Wikipedia https://en.wikipedia.org/wiki/TensorFlow	4%
AI as a Service?. Digital Infrastructure and... | Towards Data Science it is a symbolic math library, and is also used for machine learning applications such as neural networks.part of this is the cloud machine learning engine, a managed service that lets developers and data scientists build and run machine learning models in production. https://towardsdatascience.com/ai-as-a-service-b465ddc0c7e0	4%
Keras - Wikipedia it supports other common utility layers like dropout, batch normalization, and pooling.[11]. keras allows users to productize deep models on smartphones (ios and android), on the web, or on the java virtual machine.[3] it also allows use of distributed training of deep-learning models on clusters of... https://en.wikipedia.org/wiki/Keras	4%
What are the 7 Principles of Software Testing | Edureka It checks whether the developed software met the specified requirements and identifies any defect in the software in order to produce a quality product.In order to overcome this Pesticide Paradox, you need to review and revise the set of test cases regularly. https://www.edureka.co/blog/seven-principles-of-software-testing/	3%

SmallSEOTools

PLAGIARISM SCAN REPORT

Words	491	Date	July 19,2020
Characters	3269	Exclude Url	

0%	100%	0	24
Plagiarism	Unique	Plagiarized Sentences	Unique Sentences

Content Checked For Plagiarism

7.2.3 Test Case ID 3SummaryWhen GradeMe Button is clicked the essay scoresshould be displayedPrerequisitesEssay with word length between 150 to 600.Test ProcedureUpload the file containing essay and check the scores.Expected ResultProper scores of essay.Actual ResultProper scores of essay.StatusPassTable 7.3: Test case ID 37.2.4 Test Case ID 4SummaryThe web pages should be properly displayed.PrerequisitesLatest version of browser.Test ProcedureCheck by resizing the browser window.Expected ResultAll the web pages and UI elements should properlydisplayed.Actual ResultAll the UI elements and web pages are loading both onmobile as well as desktop browser properly.StatusPassTable 7.4: Test case ID 4In this chapter, we looked at the testing process and the results obtainedafter performing various types of testing for the application. In next chapter,we will take a look at the results obtained from the system.31 Automated Essay GradingChapter 8RESULTSThis section illustrates the generated output by the essay grading system onproviding a proper set of input as and when required.8.1 OutcomesIn the model evaluation, we built about 20 models and chose the best amongthem. All models were built on the same training and validation sets. Theonly difference was the selection of hyperparameters in each models. Theevaluation criterion used for evaluation is Kappa score. The best modelwhich showed the highest accuracy has been shown in 8.1SITS, B. E. (Computer) 2015 Course, Project Stage II, 2019-2032 Figure 8.1: Accuracy Score8.2 Screen ShotsThe screenshots from the website running on Google Chrome are shownbelow-In this chapter, we saw some screenshots related to the application. Wealso discussed how we evaluated the results to find the best model out ofall.33 Figure 8.2: Essay topic selectionFigure 8.3: Essay writer34 Figure 8.4: Essay Score35 Automated Essay GradingChapter 9CONCLUSIONSThe essay grading task is a laborious task which can be automated with thehelp of Deep Learning. Various researchers have put their best efforts tosolve this issue. The literature review has yielded the insights into some ofthe useful approaches to solve this problem. One particular Deep Learningtechnique known as LSTM, is highly capable in case of sequential data. Thismakes it the perfect fit for implementing an Essay Grader. Other methodsalso exist but are not capable of performing the best in sequential data. Thetarget accuracy for the project is to score more than 95% so that the gradercan be used reliably.9.1 Future WorkThe essay grader can be used by teachers for grading student essays. It canalso be used by testing agencies for test of English writing to take the burdenoff the human graders. The concept can be extended to other languages tooif the dataset is available. Lastly, the accuracy of the model can still beincreased if more data is feeded to the LSTM network while training. Moresophicated models can be developed using transfer learning at the cost ofusing more resources required to train the model.9.2 Applications1. The Automated Essay Grading application can be used in education fieldto grade the essays of students of school and colleges.2. The application can used in exams like GRE or TOEFL to grade the essaywith high accuracy.

Sources	Similarity

Bibliography

[1] **Adamson Alex, Andrew Lamb, and Ralph Ma, *Automated Essay Grading.* 2014**

[2] **Higgins Derrick, Jill Burstein, Daniel Marcu, and Claudia Gentile, *Evaluating Multiple Aspects of Coherence in Student Essays.* In HLT-NAACL, pp. 185-192. 2004.**

[3] **Gang Kou, Yi Peng, *An Application of Latent Semantic Analysis for Text Categorization* In International Journal of Computers, Communications & Control (IJCCC) 10(3):357 April 2015**

[4] **Kaveh Taghipour, Hwee Tou Ng, *A Neural Approach to Automated Essay Scoring* In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, November 2016**

[5] **The Crisp DM Methodology, *http://www.pamanalytics.com/downloads/The%20CRISP-DM%20Methodology.pdf***

[6] **Christopher Olah, *Understanding LSTM Networks* https://colah.github.io/posts/2015-08-Understanding-LSTMs/**