GATES INSTITUTE OF TECHNOLOGY

(Accredited By NAAC)

(MANAGED BY G. RAGHUNATHA REDDY EDUCATIONAL TRUST)
(APPROVED BY A.I.C.T.E. NEW DELHI)
(AFFILIATED TO JNT UNIVERSITY, ANANTAPUR)
N.H-44, Gootyananthapuram (Village), GOOTY, ANANTAPUR (DIST)

A

PROJECT REPORT ON

AN AUTOMATED MACHINE LEARNING LEVERAGED STUDENT RESULT AND SENTIMENT ANALYSIS SYSTEM

Submitted in partial fulfillment of the requirements

for the award of the degree of

Bachelor of Technology

In

Computer Science and Engineering

By

KUMMETHA RAVINDRA REDDY	17F21A0561
PULA SAI GANA SANDEEP	17F21A0568
VANKAMADDI SURESH	17F21A0591
KOVVURU VENU TEJASWINI	17F21A0594

Under the Esteemed Guidance Of

DR. A.V .L .N. SUJITH M.TECH., PH.D

ASSISTANT PROFESSOR & HEAD

Department of Computer Science and Engineering,

GATES Institute of Technology, Gooty

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GATES INSTITUTE OF TECHNOLOGY, GOOTY

CERTIFICATE

This is to certify that the project report entitled **"AN AUTOMATED MACHINE LEARNING LEVERAGED STUDENT RESULT AND SENTIMENT ANALYSIS SYSTEM"** that is being submitted by **Kummetha Ravindra Reddy (17F21A0561), Pula Sai Gana Sandeep (17F21A0568), Vankamaddi Suresh (17F21A0591), Kovvuru Venu Tejaswini (17F21A0594),** in partial fulfillment of the requirements for the award of degree of Bachelor of Technology in **COMPUTER SCIENCE AND ENGINEERING** from GATES institute of Technology is a record of Bonafied work carried out by them under guidance and supervision.

Project Guide:

Dr. A.V .L .N. Sujith M.Tech., Ph.D

Assistant Professor & Head,

Department of CSE,

GATES Institute of Technology,

Gooty – 515401.

Project Guide:

Dr. A.V .L .N. Sujith M.Tech., Ph.D

Assistant Professor & Head,

Department of CSE,

GATES Institute of Technology,

Gooty – 515401.

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

We are very much thankful to our beloved Correspondent **Smt V.K.Padmavathamma garu** for providing the necessaries in completion of the course.

We wish to thank **Sri G Raghunath Reddy** sir our beloved Managing Director & **Smt. V.K. Vani** Director providing us all facilities that are required for our project.

We cordially thank to our Principal **Dr. Bhaskar Patel** M.Tech(IISc),Ph.D(UK),PGCE(UK) for providing the necessaries in completion of this project.

We wish to thank our HOD **Dr. A.V .L .N. Sujith** **M.Tech., Ph.D** for providing us all facilities that are required for completing of our project.

We express our sincere thanks to our project guide **Dr. A.V .L .N. Sujith** **M.Tech., Ph.D** of Computer Science and Engineering department, GATES Institute Of Technology, Gooty, Anantapur, without whose innovative and imaginative approach, regular monitoring and timely advices, this project would never have been realized

We thank our teaching and nonteaching staff of the department of Computer Science & Engineering, GATES Institute of Technology, Gooty, Anantapur

Project Associates:

KUMMETHA RAVINDRA REDDY	**17F21A0561**
PULA SAI GANA SANDEEP	**17F21A0568**
VANKAMADDI SURESH	**17F21A0591**
KOVVURU VENU TEJASWINI	**17F21A0594**

ABSTRACT

The process of analyzing the released results soon after the completion of semester examinations of B-Tech students by the institution is the most essential and substantial extravaganza. Either the class teacher or a HOD or a specific in charge of the institution has to spend more than a week in order to do so traditionally. Certain person is manually managing the things starting from entering the roll number of each and every student into the results portal of the JNTUA official website, gathering the data regarding the results of each student, making the computation of pass percentage of each student, number of backlogs that a particular student has, subject wise pass percentage analysis and so on.
With the rapidly evolving Computer technology and automation methods using artificial intelligence and machine learning technologies, the entire process of results analysis can be facilitated in order to lessen the manpower.

Presently as technology is changing and the volume of automation in day to day life is gradually increasing. The improvisations aim at lessening the manpower, and increase trustworthiness, perfection, and provide less time consumption to endorse the result. Our project proposes and implements a simple and astonishing method based upon software that can be replaced with the conventional implementation and deals with upgrading the traditional approach to the automatic version as follows

- A simple interface asks for the txt file containing list of hall ticket numbers of students of a batch and the link that should be specified in order to access the results.
- Using python web scraping, we are going to fetch the results of each and every student.
- Results are analyzed accordingly.
- Analyzed data should also be emailed to corresponding officials of the institution.

Hence offering decreased human stress and increasing perfection.

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

Automation is a term for technology applications where human input is minimized. This includes business process automation (BPA), IT automation, personal applications such as home automation and more.

TYPES OF AUTOMATION:

Basic Automation: Basic automation takes simple, rudimentary tasks and automates them. This level of automation is about digitizing work by using tools to streamline and centralize routine tasks, such as using a shared messaging system instead of having information in disconnected silos. Business process management (BPM) and robotic process automation (RPA) are types of basic automation.

Process Automation: Process automation manages business processes for uniformity and transparency. It is typically handled by dedicated software and business apps. Using process automation can increase productivity and efficiency within your business. It can also deliver new insights into business challenges and suggest solutions. Process mining and workflow automation are types of process automation.

Integration Automation: Integration automation is where machines can mimic human tasks and repeat the actions once humans define the machine rules. One example is the "digital worker." In recent years, people have defined digital workers as software robots that are trained to work with humans to perform specific tasks. They have a specific set of skills, and they can be "hired" to work on teams.

Artificial Intelligence (AI) Automation: The most complex level of automation is artificial intelligence (AI) automation. The addition of AI means that machines can "learn" and make decisions based on past situations they have encountered and analyzed. For example, in customer service, virtual assistants can reduce costs while empowering both customers and human agents, creating an optimal customer service experience.

AI AND MACHINE LEARNING IN AUTOMATION:

Automation: Automation encompasses everything activities both mundane and business critical. Basic automation is programmed to perform repetitive tasks so humans don't have to.

Artificial Intelligence: AI is programmed with logic and rules to mimic human decision making. AI can be used to detect threats such as changes in user behavior or increased data transfers.

Machine Learning: Machine Learning uses data and experiences to "Learn" without additional programming. It offers more sophisticated and informed insights with each new dataset.

1.1 MOTIVATION

In the last decade, sentiment analysis has been widely applied in many domains, including business, social networks and education. Particularly in the education domain, where dealing with and processing students' opinions is a complicated task due to the nature of the language used by students and the large volume of information, the application of sentiment analysis is growing yet remains challenging. Several literature reviews reveal the state of the application of sentiment analysis in this domain from different perspectives and contexts. However, the body of literature is lacking a review that systematically classifies the research and results of the application of natural language processing (NLP), deep learning (DL), and machine learning (ML) solutions for sentiment analysis in the education domain. In this article, we present the results of a systematic mapping study to structure the published information available. We used a stepwise PRISMA framework to guide the search process and searched for studies conducted between 2015 and 2020 in the electronic research databases of the scientific literature. We identified 92 relevant studies out of 612 that were initially found on the sentiment analysis of students' feedback in learning platform environments. The mapping results showed that, despite the identified challenges, the field is rapidly growing, especially regarding the application of DL, which is the most recent trend. We identified various aspects that need to be considered in order to contribute to the maturity of research and development in the field. Among these aspects, we highlighted the need of having structured datasets, standardized solutions and increased focus on emotional expression and detection.

1.2 OBJECTIVE OF THE PROJECT.

The present education system represents a landscape that is continuously enriched by a massive amount of data that is generated daily in various formats and most often hides useful and valuable information. Finding and extracting the hidden “pearls” from the ocean of educational data constitutes one of the great advantages that sentiment analysis and opinion mining techniques can provide. Sentiments and opinions expressed by students are a valuable source of information not only for analyzing students’ behavior towards a course, topic, or teachers but also for reforming policies and institutions for their improvement. Although both sentiment analysis and opinion mining seem similar, there is a slight difference between the two: the former refers to finding sentiment words and phrases exhibiting emotions, whereas the latter refers to extracting and analyzing people’s opinions for a given entity. For this study, we consider that both techniques are used interchangeably. The sentiment/opinion polarity, which could either be positive, negative, or neutral, represents one’s attitude towards a target entity. Emotions, on the other hand, are one’s feelings expressed regarding a given topic. Since the 1960s, several theories about emotion detection and classification have been developed. The study conducted by Plutchik categorizes emotions into eight categories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust.

Sentiment analysis can be conducted at a word, sentence, or a document level. However, due to the large number of documents, manual handling of sentiments is impractical. Therefore, automatic data processing is needed. Sentiment analysis from the text-based, sentence or document-level corpora is employed using natural language processing (NLP). Most research papers found in the literature published until 2016–2017 employed pure NLP techniques, including lexicon and dictionary-based approaches for sentiment analysis. Few of those papers used conventional machine learning classifiers. Recent years have seen a shift from pure NLP-based approaches to deep learning-based modeling in recognizing and classifying sentiment, and the number of papers published recently on the undertaken topic has increased significantly.

The popularity and importance of students’ feedback have also increased recently, especially in the times of the COVID-19 pandemic, when most educational institutions have transcended traditional face-to-face learning to the online mode. Figure 1 shows the country-wise comparison breakdown of interest over the past six years in the use of sentiment analysis for analyzing students’ attitudes towards teacher assessment.

1.3 PROBLEM DEFINITION

- Results Analysis will help you understand how your students are learning, and how they have performed in each area of the syllabus. You can analyse data at the whole college and individual candidate level, look at your overall results, or focus on how your students have performed in specific areas of the syllabus.
- We now offer more detailed results data to help you improve teaching and learning in your college.
- You can use the information to identify areas of strength and weakness in order to focus your teaching where it's most needed. It can also be used to provide management information for your college.
- You can analyse data at the whole college and individual candidate level, look at your overall results, or focus on how your students have performed in specific areas of the syllabus. The information will help you identify overall trends in your college’s performance and provide constructive feedback for individual students.
- Results Analysis is available through Cambridge International Direct. Exams Officers can give access to teachers at their college who can then view the data through the College Support Hub.
- It is a more important and stressful job which requires a person to be on his toes for more than hours a day per week.
- The deadlines and targets are too specific and hence difficult to catch up day to day professional duty as well as fetching and analyzing the results all the day.
- The traditional approach which we are using right now, requires us to do constant work repeatedly.

2. LITERATURE SURVEY

2.1 INTRODUCTION:

The Student result analysis and Performance report generator is used in generating the performance report of students according to year, branch, section and subject wise which is quite a difficult and time-consuming process in every college. The Student Result Analysis and performance Report Generator helps the teacher to analyze the result and generates its report by just one clicks and it also allows the students to see their academic performance subject, year or semester wise by just only uploading their list of hall tickets and dynamic website link on which results are issued by the university. HOD of the department can see the pass percentage of the student through pictorial repetition which can be categorized subject wise, section wise, overall performance of branch or single student result. It's a web based application which can be accessed from anywhere through a web browser.

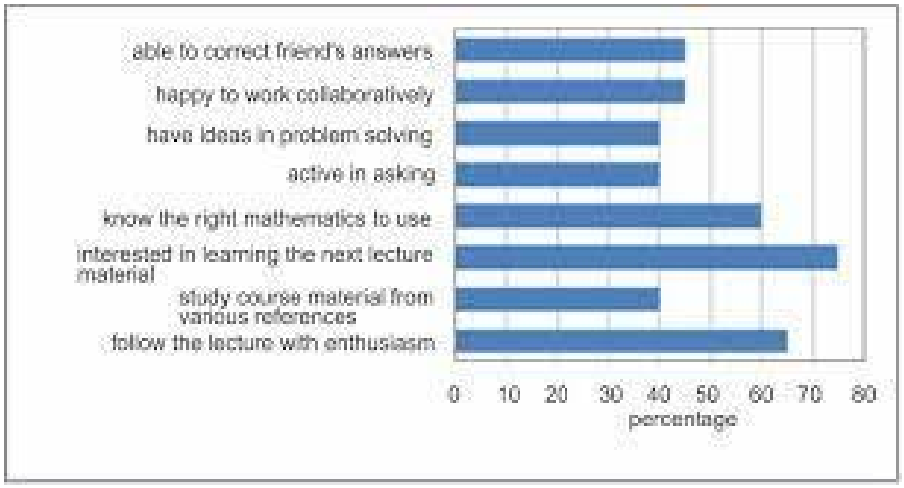

Fig. 2.1.a: Sample analysis of students results

Currently a person is sitting in front of a computer specifically to extract the results of each and every student's hall ticket in the dynamic web page, and copying the results into excel sheets one by one. Just imagine the time consumption of the present scenario would be if we want to collect the entire results of each and every student who are present in the entire college. Yes it is innumerable and exactly beyond our imagination. Our project speaks to the solution to this question. A completely automated results extraction will be a

check to excess time consumption such that there will be no more human involvement needed in students' results analysis.

2.2 EXISTING SYSTEM:

Current system is a bit time consuming as they require some amount of work to be done prior for example creating an excel sheet which contains student information like marks, personal details or enter all these details manually.

- In the existing system, a certain person is manually managing the following things.
- Right now we are not having any type of automated tools to extract results directly from the websites to excel.
- But we are having many tools available for evaluation of further calculations like:

- Individual and subject wise pass & fail percentage of students.

- Branch wise overall percentage.

- Sentiment analysis of students regarding a particular subject.

2.3 DISADVANTAGES OF EXISTING SYSTEM:

- The current system is time consuming.
    - In the current system, some manual work is to be done which makes it complex.
    - In current system user must have some knowledge about creating and managing the particular file which is to be upload to the system for data extraction
    - Some of the current software is not platform independent.
    - This system is a very time-consuming process. And the data can't be accessed accurately.
- If the student wants more information about their courses the admin cant be able to provide that.
- Students need to go manually to get the information.
- This system is complex.

2.4 PROPOSED SYSTEM:

This is the tracking system which is used to keep track of the students overall academic performance and generate its report. In this system, a user must upload the result of the student in the PDF form and then you can see the performance report of the student. System can also compare the student result with other students and the user can see the comparative analysis report of the batch of the student. In the proposed system, the user has to download the result in the PDF form and upload to it in the system and the rest of the things will be done by the system which makes it simple and fast.

2.5 ADVANTAGES OF PROPOSED SYSTEM:

- As the proposed system is a web application so you can access it from anywhere in the world only you must remember your id and password.

- Faculty do not have to prepare the excel sheet of marks of each student, only upload the result excel is automatically created.

- It is very simple as users only have to upload the PDF of the result because the rest of the work is automated.

- No high configuration system required, only required is the supporting browser.

Phase 1: Results Extraction from Dynamic Web Pages

- A Python Program is going to be executed by the user initially by providing following inputs.
- URL link of which University (JNTUA) is going to announce results of individual students.
- A text file (.txt) containing the list of all hall ticket numbers of individual students in each branch (finely formatted by line separation).
- The program iterably retrieves results of each and every student as per the pre initiated file (.txt)
- Finally the entire results are stored in an Excel sheet.

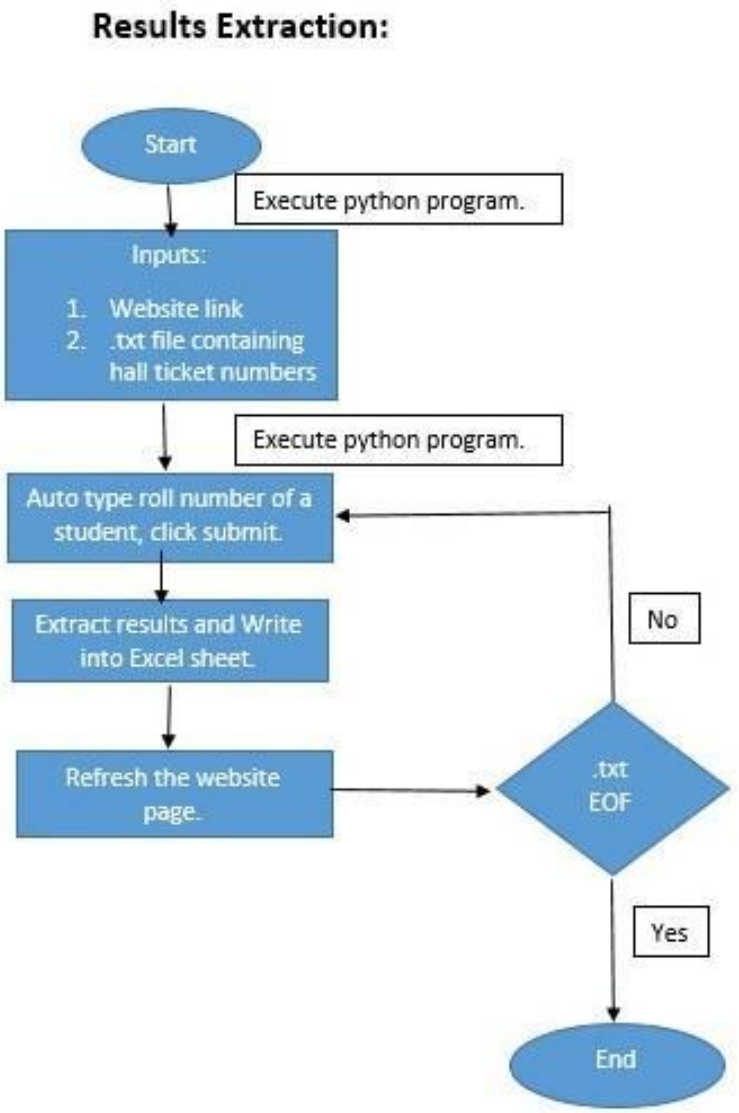

Fig – 2.4.a: Architecture of Results Extraction Phase

Figure shows the interaction between the different modules of the system with the user and shows what they can access in the system. There are two actors in the whole system and what they can access is shown by this simple use case diagram

Phase 2: Sentiment Analysis of the Extracted Results.

Sentiment analysis is a task that focuses on polarity detection and the recognition of emotion toward an entity, which could be an individual, topic, and/or event. In general, the aim of sentiment analysis is to find users' opinions, identify the sentiments they express, and then classify their polarity into positive, negative, and neutral categories.

In the last decade, sentiment analysis has been widely applied in many domains, including business, social networks and education.

Particularly in the education domain, where dealing with and processing students' opinions is a complicated task due to the nature of the language used by students and the large volume of information, the application of sentiment analysis is growing yet remains challenging.

Several literature reviews reveal the state of the application of sentiment analysis in this domain from different perspectives and contexts.

However, the body of literature is lacking a review that systematically classifies the research and results of the application of natural language processing (NLP), deep learning (DL), and machine learning (ML) solutions for sentiment analysis in the education domain.

In general, there are three different levels at which sentiment analysis can be performed: the document level, sentence level, and aspect level.

Sentiment analysis at the document level aims to identify the sentiments of users by analysing the whole document.

Sentence-level analysis is more fine-grained as the goal is to identify the polarity of sentences rather than the entire document.

Aspect-level sentiment analysis focuses on identifying aspects or attributes expressed in reviews and on classifying the opinions of users towards these aspects.

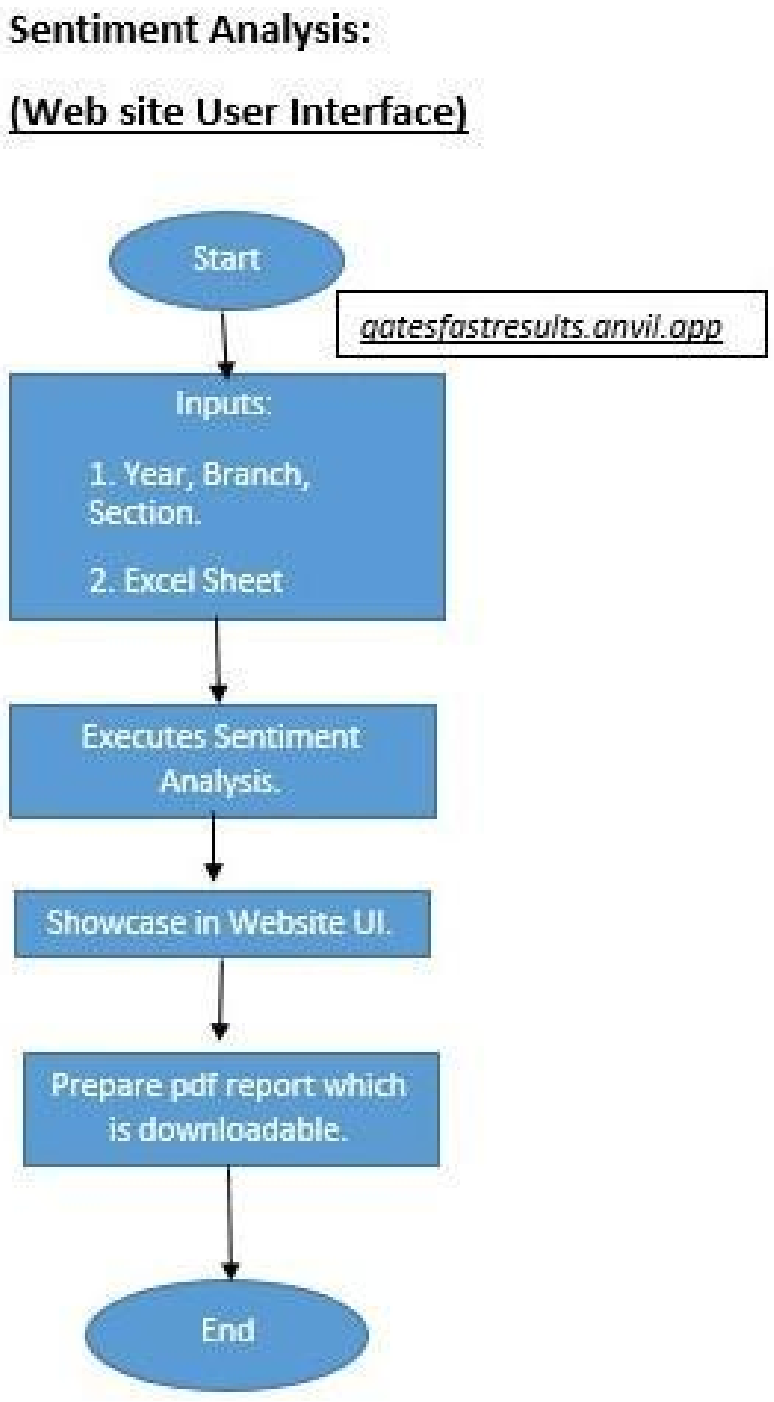

Fig – 2.4.b: Architecture of Sentiment Analysis Phase

3.SENTIMENT ANALYSIS

3.1 INTRODUCTION

The number of papers published recently indicates a growing interest towards the application of NLP/DL/ML solutions for sentiment analysis in the education domain. However, to the best of our knowledge, in order to establish the state of evidence, the body of literature is lacking a review that systematically classifies and categorizes research and results by showing the frequencies and visual summaries of publications, trends, etc. This gap in the body of literature necessitated a systematic mapping of the use of sentiment analysis to study students' feedback. Thus, this article aims to map how this research field is structured by answering research questions through a stepwise framework to conduct systematic reviews. In particular, we formulated multiple research questions that cover general issues regarding investigated aspects in sentiment analysis, models and approaches, trends regarding evaluation metrics, bibliographic sources of publications in the field, and the solutions used, among others.

The main contributions of this study are as follows:

- A systematic map of 92 primary studies based on the PRISMA framework;
- An analysis of the investigated educational entities/aspects and bibliographical and research trends in the field;
- A classification of reviewed papers based on approaches, solutions, and data representation techniques with respect to sentiment analysis in the education domain;
- An overview of the challenges, opportunities, and recommendations of the field for future research exploration.

3.2 OVERVIEW OF SENTIMENT ANALYSIS:

Sentiment analysis is a task that focuses on polarity detection and the recognition of emotion toward an entity, which could be an individual, topic, and/or event. In general, the aim of sentiment analysis is to find users' opinions, identify the sentiments they express, and then classify their polarity into positive, negative, and neutral categories. Sentiment analysis systems use NLP and ML techniques to discover, retrieve, and distil information and opinions from vast amounts of textual information.

In general, there are three different levels at which sentiment analysis can be performed: the document level, sentence level, and aspect level. Sentiment analysis at the document level aims to identify the sentiments of users by analysing the whole document. Sentence-level analysis is more fine-grained as the goal is to identify the polarity of sentences rather than the entire document. Aspect-level sentiment analysis focuses on identifying aspects or attributes expressed in reviews and on classifying the opinions of users towards these aspects.

As can be seen from **Figure 3.2.a**, the general architecture of a generic sentiment analysis system includes three steps. Step 1 represents the input of a corpus of documents into the system in various formats. This is followed by the second step, which is document processing. At this step, the entered documents are converted to text and pre-processed by utilizing different linguistic tools, such as tokenization, stemming, PoS (Part of Speech) tagging, and entity and relation extraction. Here, the system may also use a set of lexicons and linguistic resources. The central component of the system architecture is the document analysis module (step 3) that also makes use of linguistic resources to annotate the pre-processed documents with sentiment annotations. Annotations represent the output of the system—i.e., positive, negative, or neutral—presented using a variety of visualization tools. Depending on the sentiment analysis form, annotations may be attached differently. For document-based sentiment analysis, the annotations may be attached to the entire documents; for sentence-based sentiments, the annotations may be attached to individual sentences; whereas for aspect-based sentiment, they are attached to specific topics or entities.

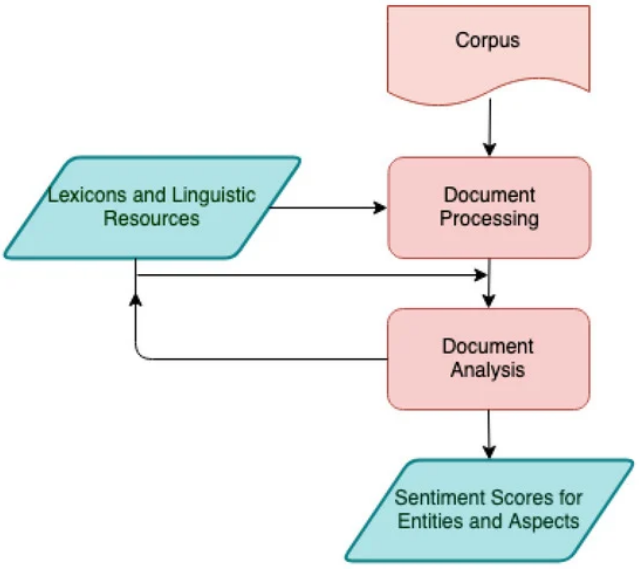

Fig – 3.2.a: The architecture of a generic sentiment analysis system

Sentiment analysis has been widely applied in different application domains, especially in business and social networks, for various purposes. Some well-known sentiment analysis business applications include product and services reviews, financial markets, customer relationship management, and marketing strategies and research, among others. Regarding social networks applications, the most common application of sentiment analysis is to monitor the reputation of a specific brand on Twitter or Facebook and explore the reaction of people given a crisis; e.g., COVID-19. Another important application domain is in politics, where sentiment analysis can be useful for the election campaigns of candidates running for political positions.

Recently, sentiment analysis and opinion mining has also attracted a great deal of research attention in the education domain. In contrast to the above-mentioned fields of business or social networks, which focus on a single stakeholder, the research on sentiment analysis in the education domain considers multiple stakeholders of education including teachers/instructors, students/learners, decision makers, and institutions. Specifically, sentiment analysis is mainly applied to improve teaching, management, and evaluation by analysing learners' attitudes and behaviour towards courses, platforms, institutions, and teachers.

From the learners' perspective, there are a number of papers that have applied sentiment analysis to investigate the correlation of attitude and performance with learners' sentiments as well as the relationship between learners' sentiments and drop-out rates in Massive Open Online Courses (MOOCs). Regarding teachers' perspectives, sentiment analysis has been widely adopted by researchers to examine various teacher-associated aspects expressed in students' reviews or comments in discussion forums. These aspects include teaching pedagogy, behaviour, knowledge, assessment, and experience, to name a few. Sentiment analysis was also used in a number of studies to analyse student's attitudes towards various aspects related to an institution; i.e., tuition fees, financial aid, housing, food, diversity, etc. Regarding courses, aspect-based sentiment analysis systems have been implemented to identify key aspects that play a critical role in determining the effectiveness of a course as discussed in students' reviews and then examine the attitudes and opinions of students towards these aspects. These aspects primarily include course content, course design, the technology used to deliver course content, and assessment, among others.

4. SYSTEM SPECIFICATION

4.1 INTRODUCTION

This part will describe the entire project system specification. In order to make the report professional the desired system specification will be described as well as the recommended hardware component in addition to the software will be used. All the components will be described in next paragraphs.

4.2 HARDWARE REQUIREMENTS:

We tried to make all the hardware components going to be used in this project as simple as possible. Hence the only device going to be used in this project is the computer. This computer must have high specifications in order to handle the system's operation especially when the Automated Python program needs high RAM and a high processor in order to execute the results extraction as quickly as possible.

- RAM : 4GB or Higher
- Processor : Intel i3 or above
- Hard Disk : 500MB or more

4.3 SOFTWARE REQUIREMENTS:

As we discussed earlier, there are two phases involved in the entire project. These two phases will be having two individual programs to be executed separately. First to extract results from dynamic webpage and prepare an excel sheet containing students results, second to perform sentiment analysis on pre extracted excel sheet and to prepare a PDF report.

- Python 3.8 (preferably latest version)
- Selenium (pip install selenium)
- Beautiful soup (pip install bs4 / pip install beautifulsoup4)
- Chromium web-driver (for chrome users)
- Google chrome

Note: The versions of the Google Chrome and Chromium web-driver should be one and the same.

5. SYSTEM DESIGN

The system is designed at the beginning then research has been done about the component that will be used in the project to achieve the task.

5.1 SYSTEM DESIGN AND DEVELOPMENT:

The design of the project will contain software which will handle the entire automated process starting from results extraction to the PDF report preparation. The figure below shows the system design process. In this part a block diagram will be created in order to show the main parts of the project and give a closer look to the component that will be used to create the system. The block diagram below shows the main parts of the project.

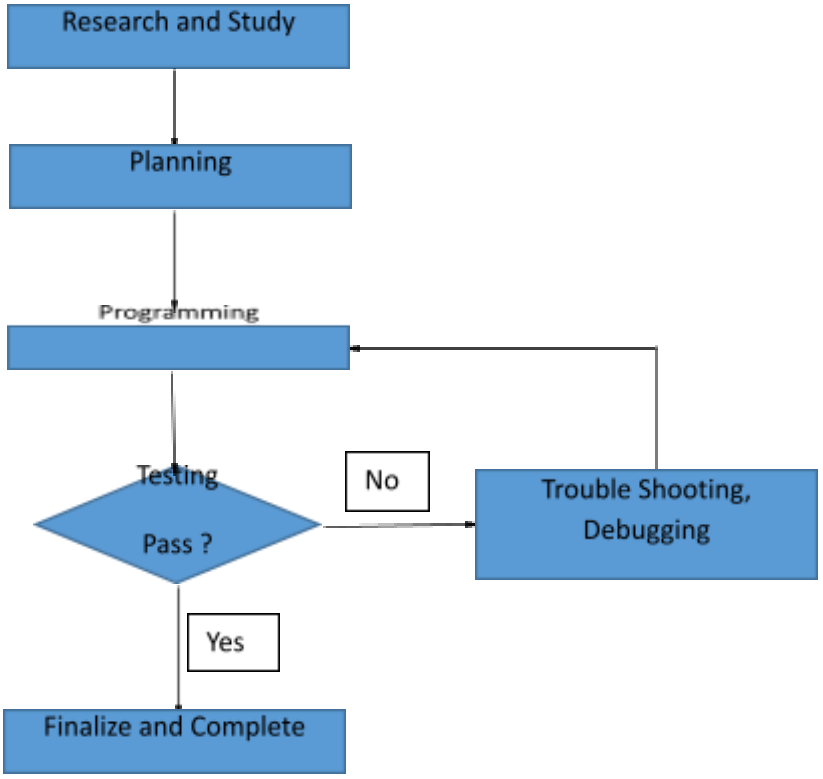

Fig – 5.1.a: Design and development Flow Chart

5.2 AUTOMATED RESULTS EXTRACTION (PHASE 1):

An algorithm for automated results extraction has been developed using Python 3.8 and few installed modules (installed using pip commands) as we discussed earlier. The whole integrated system will work as shown in the figure below.

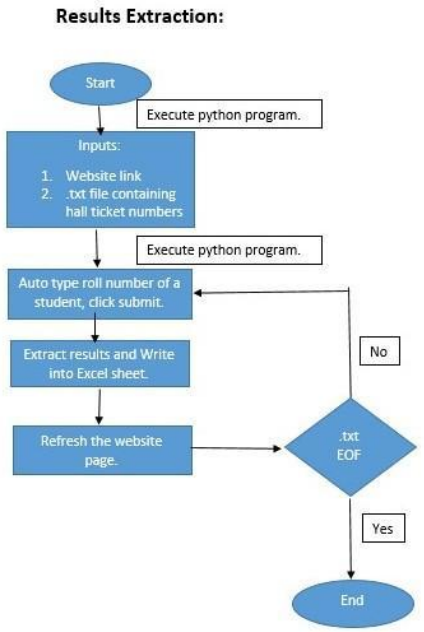

Fig – 5.2.a: Automated Results Extraction Block Diagram

5.3 SENTIMENT ANALYSIS ON EXTRACTED RESULTS (PHASE 2):

A linear regression based machine learning algorithm for sentiment analysis of the results (excel sheet generated by Phase 1) has been developed using Python 3.8 and few installed modules (installed using pip commands) as we discussed earlier. The whole integrated system will work as shown in the figure below.

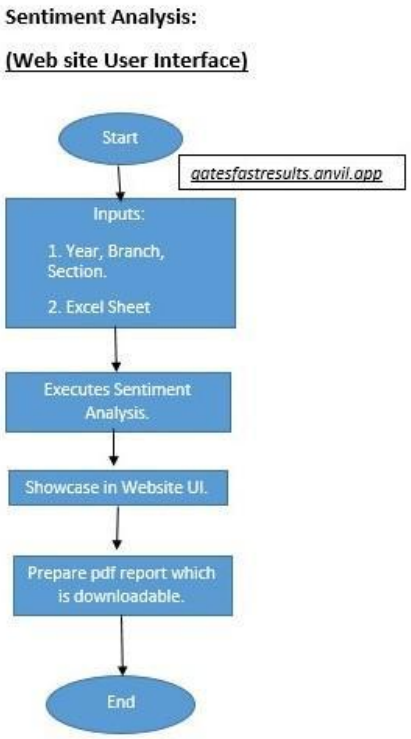

Fig – 5.2.b: Sentiment Analysis block diagram

5.4 IMPORTANT MODULES AND ITS INSTALLATIONS.

- Python 3.8 (Preferably latest version)
  https://www.python.org/
- Google Chrome
  https://www.google.com/intl/en_in/chrome/
- Chromium web-driver (for chrome users)
  https://chromedriver.chromium.org/

Note: The versions of the Google Chrome and Chromium web-driver should be one and the same.

- PIP installations
  https://pypi.org/

- Selenium (pip install selenium)
- Beautiful Soup (pip install bs4 (or) pip install beautifulsoup4)
- Numpy (pip install numpy)
- Pandas (pip install pandas)
- Matplotlib (pip install matplotlib)
- Seaborn (pip install seaborn)
- OS (pip install os (or) pip install os0)
- Dabl (pip install dabl)
-

5.5 DEVELOPMENT SYSTEM TOOLS:

System development tools are a combination of tools and instruments used with the hardware and software modules in order to make the desirable project. The explanation of these tools will be explained below.

5.5.1 PERSONAL COMPUTER:

The computer is considered one of the main devices in the project. The personal computer will work as an interface between the user and the system. The personal computer will operate the program in order to extract the results from the dynamic web page, prepare Excel sheets and apply sentiment analysis in order to prepare a pdf report as a final output.

5.5.2 RESULTS EXTRACTION:

The design of the entire process is divided into two phases of which phase 1 is all about extraction of results from the dynamic web pages and phase 2 is all about performing the sentiment analysis and preparing a PDF report. The results extraction algorithm design is very important since the algorithm will decide the results of each individual student.

In order to extract the results from the dynamic web pages we are going to perform but scrapping upon the dynamic web page such that information displayed in the dynamic web pages will be automatically saved into Excel sheet. Selenium is a portable framework for testing web applications. Selenium provides a playback tool for authoring functional tests without the need to learn a test scripting language (Selenium IDE). It also provides a test domain-specific language (Selenese) to write tests in a number of popular programming languages, including C#, Groovy, Java, Perl, PHP, Python, Ruby

and Scala. The tests can then run against most modern web browsers. Selenium runs on Windows, Linux, and macOS. It is open-source software released under the Apache License 2.0.

5.5.3 SENTIMENT ANALYSIS:

Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

We perform sentiment analysis on the Excel sheet which is previously generated as we discussed earlier and then prepare a report which is of the form of PDF containing 15 to 20 pages such that the report covers in detail analysis of students results it will also help the management of college help students to gain more marks in the examinations by providing suitable training to the students.

6. IMPLEMENTATION AND PROGRAMMING

6.1 INTRODUCTION:

The project will be discussed on the basis of what has been done in the previous chapter. Detailed discussion about automation of the results extraction system is also included in this chapter. In addition the discussion of how to achieve the goals and objectives and do the project very well explained.

Web scraping is applied on the dynamic web pages in order to extract each and every result based on the input given which contents each and every hall ticket number of the students belonging to the particular department in the college. Meanwhile the obstructed results will be further processed into a beautiful clean data set and then saved as an excel sheet.

The reason behind preparation of Excel sheet instead of using database is in excel sheet must be separated to the higher officials of the college management and right now we are not having any kind of other resources which will perform automation basic actions on dynamic web pages in order to extract the results but there are few softwares exists in the real world which is the Excel sheets as in and generate finally design a report based on the Excel sheets. Hence by considering these reasons we are specifically generating Excel sheets by using python programming.

Now by using the Excel sheet which is generated as we discussed in the above paragraph, we are going to perform sentiment analysis by using the concept of linear regression in machine learning. As a part of sentiment analysis the results are analyzed and converted into pictorial representation by comparing each and every student's results in a particular department.

6.2 RESULTS ANALYSIS (PHASE 1):

6.2.1 INPUT FORMAT FOR PHASE 1:

Initially a user has to execute the python executable program (.exe) file which requires two types of inputs. The link of the dynamic web page will be the first input and a text file (.txt file) containing the list of hall ticket numbers of each and every student belonging to a specific department of the college will be the second input.

Note: The roll numbers which are present in the text file must have a separator. In this scenario we are using the line separated line that is each and every roll number in the text file (.txt file) must be started within the new line. If we want to use another operator apart from using the lines operator we have to ensure that the source code must be changed based on the separator we are using. For more understanding, refer split() concept in python language.

https://www.w3schools.com/python/ref_string_split.asp

And that's it! We've got all the inputs which are necessary to execute the results extraction program and generate excel sheets of student's marks.

```
In [53]: from selenium import webdriver
         from time import sleep
         import openpyxl
         workbook = openpyxl.Workbook()
         sheet = workbook.active
         empty=[' ',' ',' ',' ']
         from webdriver_manager.chrome import ChromeDriverManager
         options = webdriver.ChromeOptions()
         options.binary_location = r"C:\Program Files\Google\Chrome\Application\chrome.exe"

         chrome_driver_binary = r"C:\Users\Dell\Desktop\Imp on Desktop\Final Year Project\Src Code\chromedriver.exe"
         driver = webdriver.Chrome(chrome_driver_binary, chrome_options=options)



         # intialisation
         ids=['17F21A0501','17F21A0503','17F21A0504','17F21A0505','17F21A0506','17F21A0507','17F21A0508','17F21A0509','17F21A0510','17F21A0511','1
         7F21A0512','17F21A0513','17F21A0514','17F21A0515','17F21A0516','17F21A0517','17F21A0518','17F21A0519','17F21A0520','17F21A0521','17F21A05
         22','17F21A0523','17F21A0524','17F21A0525','17F21A0526','17F21A0527','17F21A0528','17F21A0529','17F21A0530','17F21A0531','17F21A0532','17
         F21A0533','17F21A0534','17F21A0536','17F21A0537','17F21A0538','17F21A0539','17F21A0540','17F21A0541','17F21A0542','17F21A0543','17F21A054
         4','17F21A0545','17F21A0546','17F21A0547','17F21A0548','17F21A0549','17F21A0550','17F21A0551','17F21A0552','17F21A0553','17F21A0554','17F
         21A0555','17F21A0556','17F21A0557','17F21A0558','17F21A0559','17F21A0560','17F21A0561','17F21A0562','17F21A0564','17F21A0565','17F21A056
         6','17F21A0568','17F21A0569','17F21A0570','17F21A0571','17F21A0573','17F21A0575','17F21A0576','17F21A0577','17F21A0578','17F21A0579','17F
         21A0580','17F21A0581','17F21A0582','17F21A0583','17F21A0584','17F21A0586','17F21A0587','17F21A0588','17F21A0589','17F21A0590','17F21A059
         2','17F21A0593','17F21A0594','17F21A0595','17F21A0596','17F21A0597','17F21A0598','17F21A0599','17F21A05A1','17F21A05A2','17F21A05A3','17F
         21A05A4','17F21A05A5','17F21A05A6','17F21A05A7','17F21A05A8','17F21A05A9','17F21A05B0','17F21A05B1','17F21A05B2','17F21A05B3','17F21A05B
         4','17F21A05B5','18F21A0501','18F21A0502']

         # id
         url='https://jntuaresults.ac.in/view-results-56736399.html'
         # code
         #driver=webdriver.Chrome(ChromeDriverManager().install())
         #driver = webdriver.Chrome()
```

Fig - 6.2.1.a: Figure showing the sample pre defined inputs in program

6.2.2 EXECUTION OF PHASE 1:

```
        name+=o+" "
    pen.append(name)
    name = []

    pen.extend(totals)                    # Suject Marks
    pen.append(sum(totals))
    pen.append((sum(totals)/(len(subjects)*100))*100)      # Percentage
    pen.append(pcount)
    pen.append(fcount)
    if fcount == 0:
        pen.append("Yes")
    else:
        pen.append("No")

    sheet.append(pen)
    workbook.save("MyResults.xlsx")

print("Work sheet done!\n\n\n")
```

```
C:\Users\Dell\Anaconda3\lib\site-packages\ipykernel_launcher.py:12: DeprecationWarning: use options instead of chrome_options
  if sys.path[0] == '':
evaluation checked for 17F21A0501
evaluation checked for 17F21A0503
evaluation checked for 17F21A0504
evaluation checked for 17F21A0505
evaluation checked for 17F21A0506
evaluation checked for 17F21A0507
evaluation checked for 17F21A0508
evaluation checked for 17F21A0509
evaluation checked for 17F21A0510
evaluation checked for 17F21A0511
evaluation checked for 17F21A0512
evaluation checked for 17F21A0513
evaluation checked for 17F21A0514
evaluation checked for 17F21A0515
evaluation checked for 17F21A0516
```

Fig - 6.2.2.a: Execution of the Program.

While coming to the execution part, by taking each and every roll number text file as an input the executable program will pass the roll number as keys to the dynamic web page where the results are declared by the University board of examinations, and the results are fetched from the website accordingly, allowed whether processing of data and preparing the Excel sheet as shown in the figure 6.2.2.a.

6.2.3 OUTPUT FORMAT FOR PHASE 1:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	RollNo	Name	COMPILER	DATAWAF	DESIGNPA	DESIGN&A	WEB&INTI	ARTIFICIA	WEB&INTI	DATAWAF	COMPREH	ADVANCE	Total	Percentag	PassCount	failCount	AllClear
2	17F21A0501	SHAIK ABDUL ABBAS	58	67	72	64	51	61	97	93	94	27	684	68.4	10	0	Yes
3	17F21A0503	AIMANA ZAHERA	46	51	55	54	40	55	84	92	86	27	590	59	10	0	Yes
4	17F21A0504	DUDEKULA AJAD	49	62	55	58	42	61	91	91	92	26	627	62.7	10	0	Yes
5	17F21A0505	GOLLA ANUSHA	56	57	64	61	47	61	92	92	90	27	647	64.7	10	0	Yes
6	17F21A0506	DUDEKULA ANWAR	48	62	72	63	49	62	94	93	94	29	666	66.6	10	0	Yes
7	17F21A0507	CHEKKA ARAVIND	51	65	55	51	41	58	91	93	92	29	626	62.6	10	0	Yes
8	17F21A0508	SYEDA AYESHA SIDDIQUA	33	56	50	52	30	27	92	93	91	28	552	55.2	7	3	No
9	17F21A0509	VALLIPI BHAVYA	61	70	73	64	54	67	96	97	92	29	703	70.3	10	0	Yes
10	17F21A0510	CHITRA BHAVYA	43	66	67	59	44	61	92	96	93	29	650	65	10	0	Yes
11	17F21A0511	NAGARI CHAITHANYA KESAVA	43	59	51	58	43	58	90	93	89	27	611	61.1	10	0	Yes
12	17F21A0512	KAPPADI CHANDAN REDDY	26	41	32	54	25	45	91	88	75	29	506	50.6	7	3	No
13	17F21A0513	SILAR CHANDINI	62	66	82	66	63	63	98	98	98	28	724	72.4	10	0	Yes
14	17F21A0514	CHITRA CHANDRASEKHAR REDDY	28	41	61	51	30	45	90	90	92	26	554	55.4	8	2	No
15	17F21A0515	KUMMARA DEEPTHI	49	64	75	60	49	75	96	92	85	26	671	67.1	10	0	Yes
16	17F21A0516	KOTTILA DILEEP	29	28	46	18	24	48	76	81	88	25	463	46.3	6	4	No
17	17F21A0517	UNDABANDA DILEEP KUMAR	42	45	58	30	31	49	89	90	87	27	548	54.8	8	2	No
18	17F21A0518	TUGGALI DINESH BABU	52	60	60	49	45	58	91	94	90	28	627	62.7	10	0	Yes
19	17F21A0519	RABBANUGUNDU DINESH KUMAR	46	54	65	51	45	54	91	92	94	28	620	62	10	0	Yes
20	17F21A0520	DONGALA HARICHANDANA	45	53	72	52	47	54	94	94	91	30	632	63.2	10	0	Yes
21	17F21A0521	BOGGULA HARIKRISHNA	47	50	70	49	45	55	90	91	92	29	618	61.8	10	0	Yes
22	17F21A0522	GOLLA HARIKRISHNA	55	69	76	60	46	55	92	96	92	28	669	66.9	10	0	Yes
23	17F21A0523	BHOGATI HARINATH REDDY	41	37	52	47	34	30	92	92	85	29	539	53.9	7	3	No
24	17F21A0524	GONGATI HARINI	56	56	76	68	61	67	99	98	97	29	707	70.7	10	0	Yes
25	17F21A0525	THOTA HARSHAVARDHAN	48	51	61	55	35	47	90	94	90	28	599	59.9	9	1	No

Fig - 6.2.3.a: Excel sheet with finely processed data

The figure 6.2.3.a shows the final representation of the Excel sheet generated soon after the completion of phase 1. This Excel sheet can be considered as the final output of phase 1 of our project.

6.3 SENTIMENT ANALYSIS (PHASE 2):

Initially a user has to execute the python executable program (.exe) file which requires only one input that is Excel sheet which is previously generated in the first phase of the project. And that's it a PDF report of 15 to 20 pages will be generated as the final output of the Phase 2.

Mainly we are using the linear regression concept in machine learning.

6.3.1 WHAT IS MACHINE LEARNING?

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and by the use of data. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

A subset of machine learning is closely related to computational statistics, which focuses on making predictions using computers; but not all machine learning is statistical learning. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a related field of study, focusing on exploratory data analysis through unsupervised learning. In its application across business problems, machine learning is also referred to as predictive analytics.

6.3.2 WHAT IS REGRESSION?

Regression is a statistical method used in finance, investing, and other disciplines that attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).

$$Y = f(X_i, B) + e_i$$

Y = dependent variable

f = function

X_i = independent variable

B = unknown parameters

e_i = error terms

6.3.3 LINEAR REGRESSION:

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering and the number of independent variables being used.

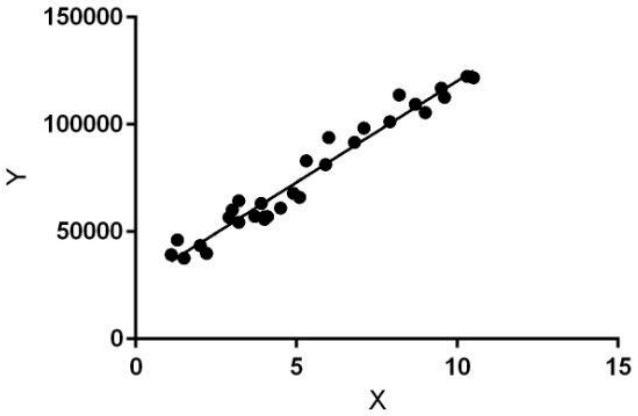

Fig 6.3.3.a: Linear regression plot

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.

In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best fit line for our model.

$$Y = \theta 1 + \theta 2.X$$

x: input training data (univariate – one input variable(parameter))

y: labels to data (supervised learning)

When training the model – it fits the best line to predict the value of y for a given value of x. The model gets the best regression fit line by finding the best $\theta 1$ *and* $\theta 2$ *values.*

$\theta 1$*: intercept*

$\theta 2$*: coefficient of x*

6.3.4 PICTORIAL REPRESENTATION:

As a result of the sentiment analysis, we will be able to analyze the results of the students and represent the results in pictorial representation as shown in the figures below. By using these pictorial representations, we will be preparing the automated PDF report of analyzed results.

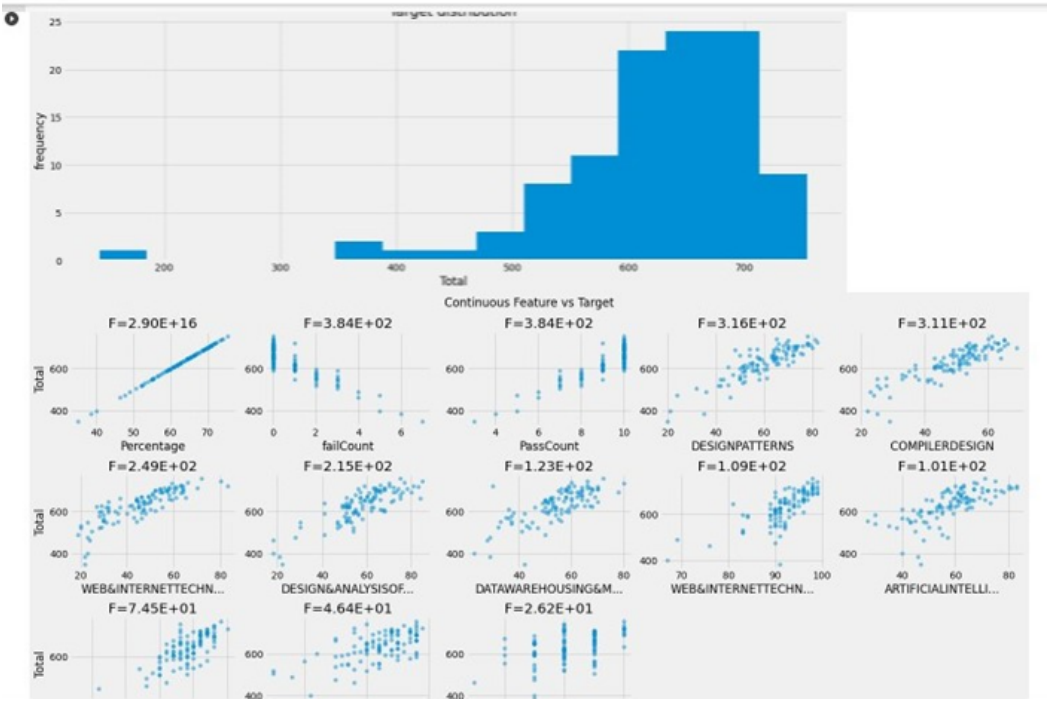

Fig 6.3.4.a: Total vs Frequency

Figure 6.3.4.a shows the Total vs Frequency plot at the top of the graph that is showing the ranges of Total marks obtained by the students grouped by categorical classification. Also the small graphs which are shown in the above figure, represents the subject wise classification of total marks obtained by the students. Such that based on the subject wise performance of the students observed, faculty or respective department's staff can able to concentrate on those subjects. Hence results are improved in further attempts or in further batches.

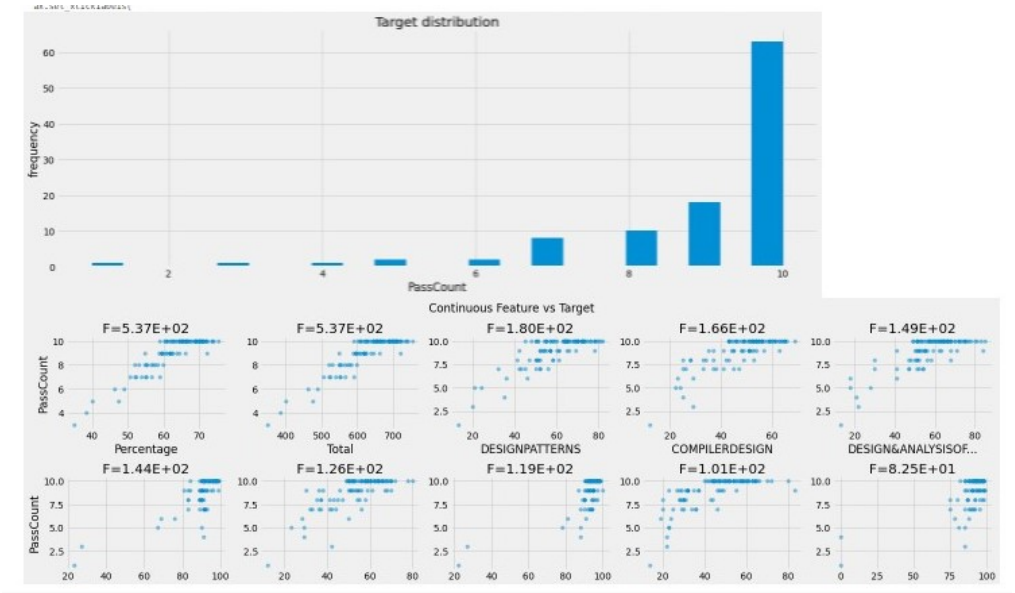

Fig 6.3.4.b: Passcount vs Frequency

Figure 6.3.4.b shows the Pass count vs Frequency plot at the top of the graph that is showing the ranges of Total Pass count obtained by the students grouped by categorical classification. Also the small graphs which are shown in the above figure, represents the subject wise classification of total Pass Count obtained by the students. Such that based on the subject wise performance of the students observed, faculty or respective department's staff can able to concentrate on those subjects. Hence results are improved in further attempts or in further batches.

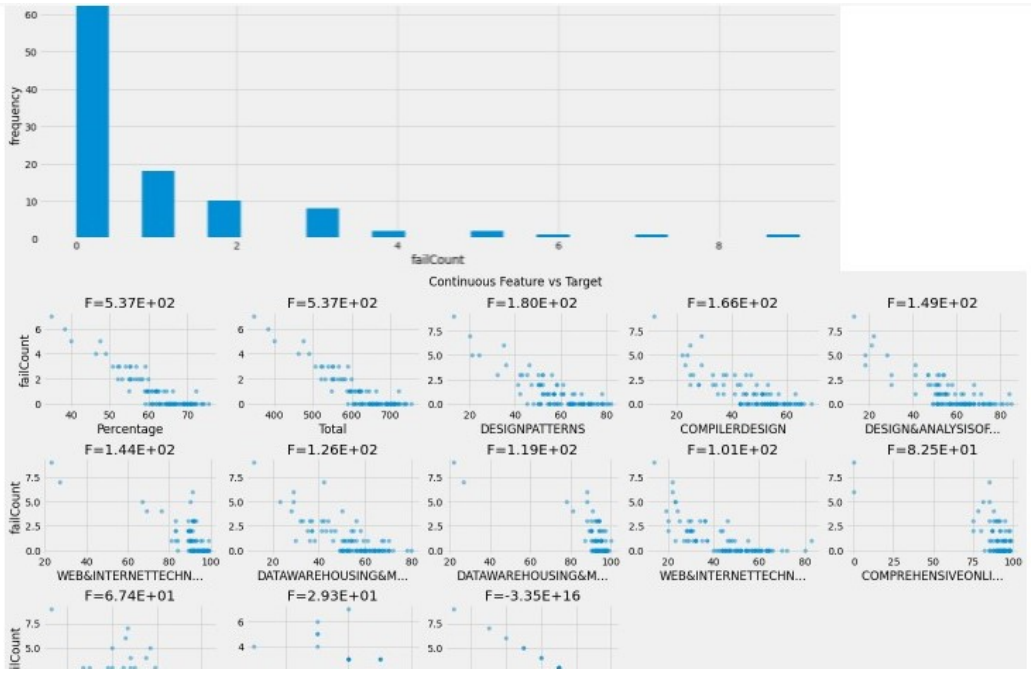

Fig 6.3.4.c: Fail count vs Frequency

Figure 6.3.4.c shows the Fail count vs Frequency plot at the top of the graph that is showing the ranges of all students whose performance was below average and failed in respective subjects obtained by the students grouped by categorical classification. Also the small graphs which are shown in the above figure, represents the subject wise classification of fail marks obtained by the students. Such that based on the subject wise performance of the students observed, faculty or respective department's staff are able to concentrate on those subjects. Hence results are improved in further attempts or in further batches.

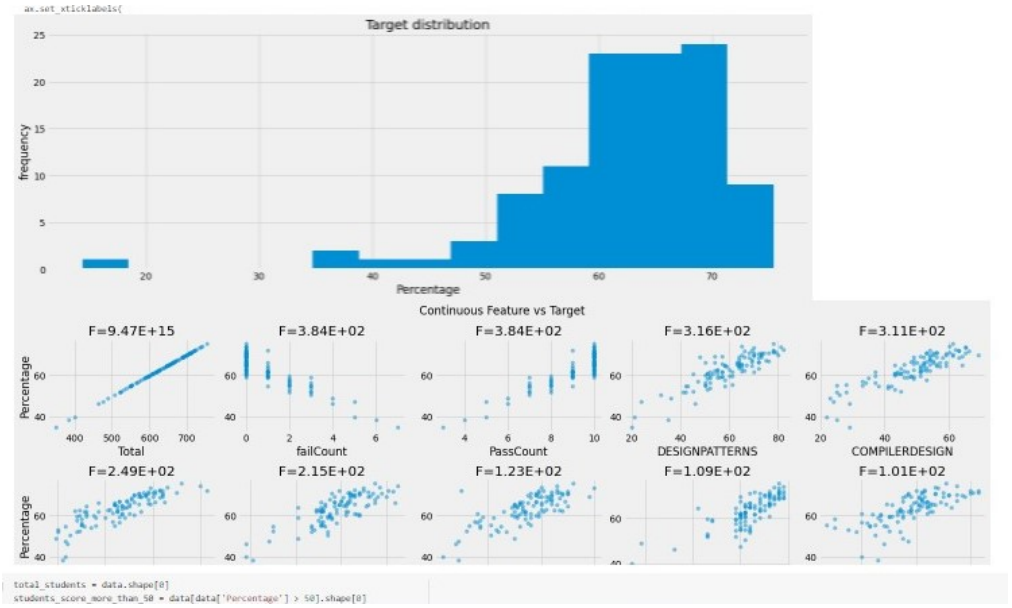

Fig 6.3.4.d: Percentages vs Frequency

Figure 6.3.4.d shows the Percentage range vs Frequency plot at the top of the graph that is showing the ranges of all students whose performance was below average and failed in respective subjects obtained by the students grouped by categorical classification percentage wise. Also the small graphs which are shown in the above figure, represents the subject wise classification of the percentages obtained by the students. Such that based on the subject wise performance of the students observed, faculty or respective department's staff are able to concentrate on those subjects. Hence results are improved in further attempts or in further batches.

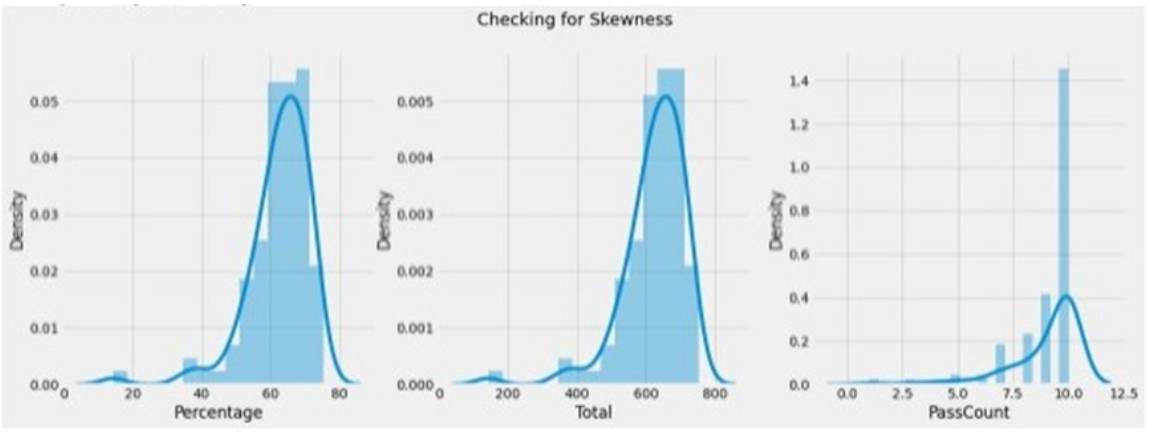

Fig 6.3.4.e: Percentages vs Total vs Pass Count

Figure 6.3.4.e shows the Percentage range vs Frequency vs Pass Count plot at the top of the graph that is showing the ranges of all students whose performance was below average and failed in respective results obtained by the students grouped by categorical classification percentage wise. Also the small graphs which are shown in the above figure, represents the subject wise classification of the percentages obtained by the students. Such that based on the subject wise performance of the students observed, faculty or respective department's staff are able to concentrate on those subjects. Hence results are improved in further attempts or in further batches.

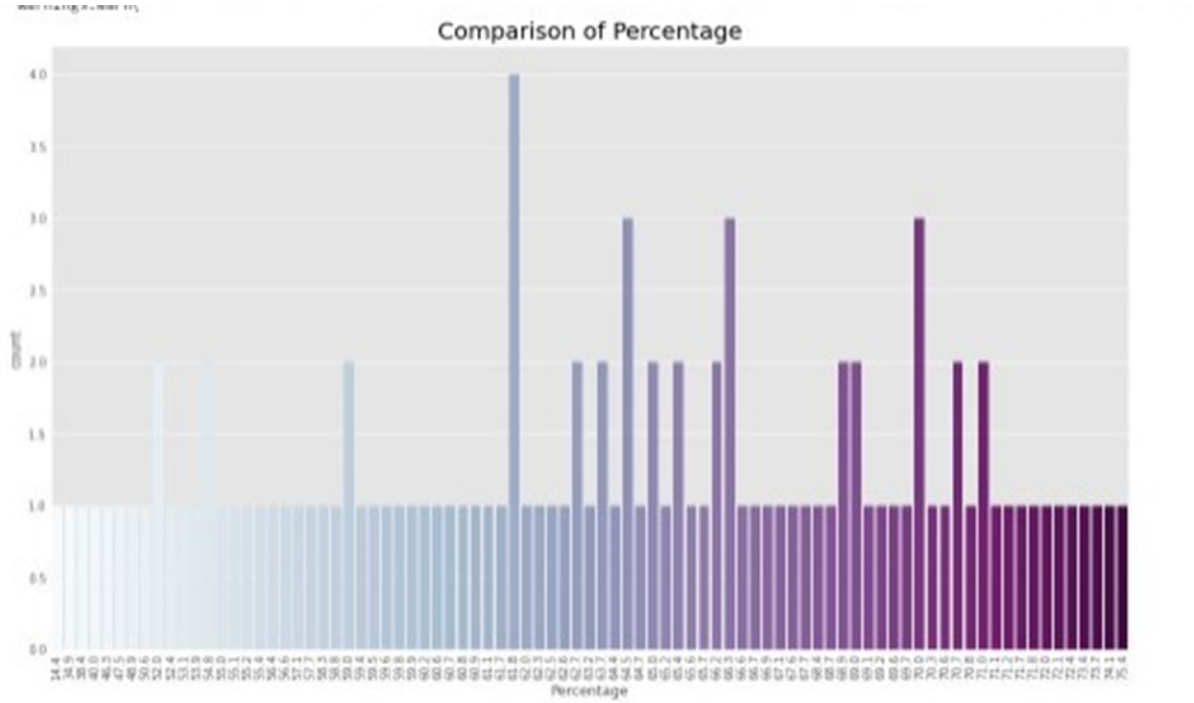

Fig 6.3.4.f: Individual Percentages comparison

Figure 6.3.4.f shows the individual Percentage range plot at the top of the graph that is showing the ranges of all students whose performance was below average and failed in respective subjects obtained by the students grouped by categorical classification percentage wise. Such that all student's results are shown in a single plot. We can easily point out the ranges of percentages, and their holding student's roll number. We can also highlight the class topper of the results.

Hence results are improved in further attempts or in further batches.

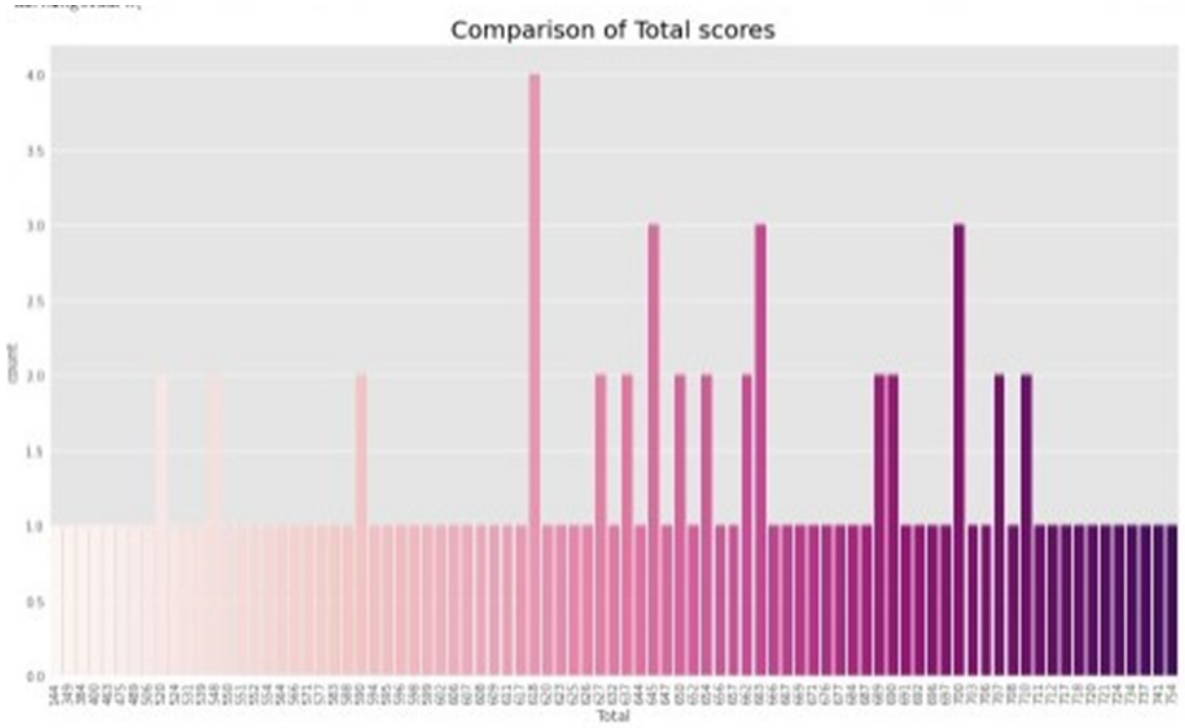

Fig 6.3.4.g: Total scores comparison

Figure 6.3.4.g shows the individual Total range plot at the top of the graph that is showing the ranges of all students whose performance was below average and failed in respective subjects obtained by the students grouped by categorical classification Total marks wise. Such that all student's results are shown in a single plot. We can easily point out the ranges of Total marks, and their holding student's roll number. We can also highlight the class topper of the results.

Hence results are improved in further attempts or in further batches.

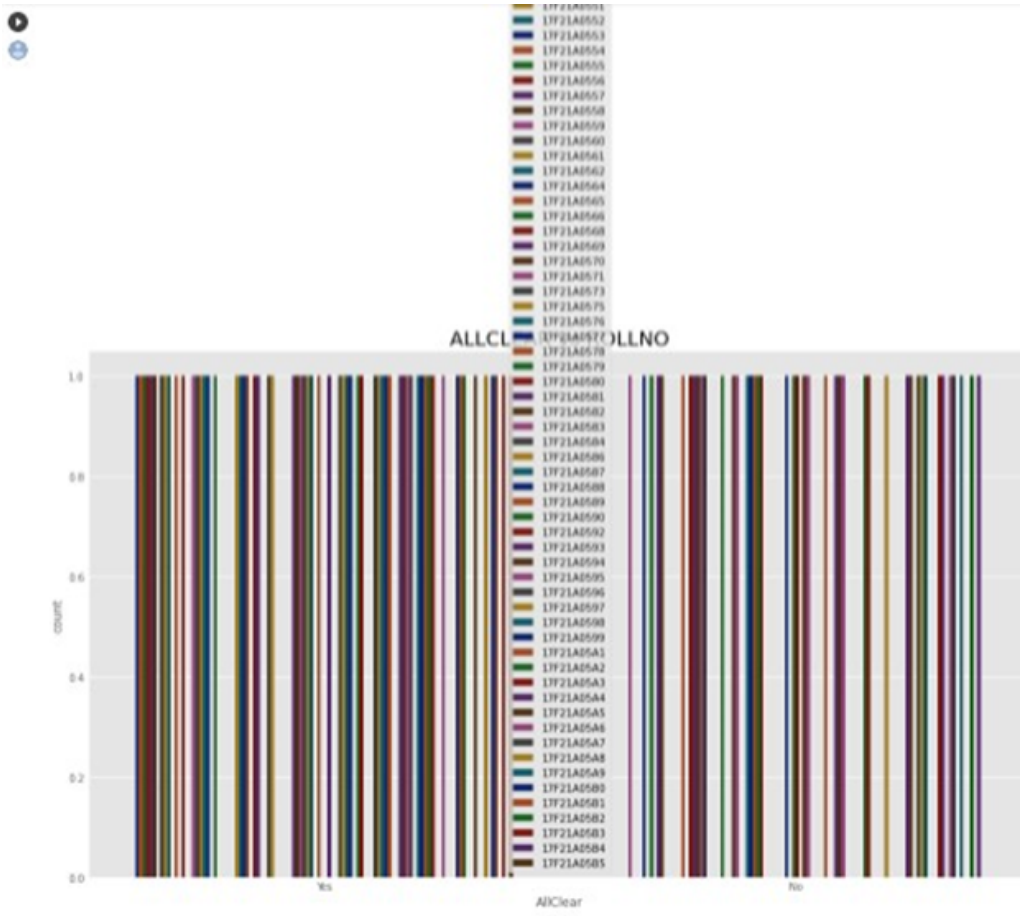

Fig 6.3.4.h: Allclear Yes vs Allclear No

Figure 6.3.4.h shows the individual All Clear Yes or No range plot at the top of the graph that is showing the ranges of all students whose performance was below average and failed in respective subjects obtained by the students grouped by categorical classification All clear wise. Such that all student's results are shown in a single plot. We can easily point out the ranges of All Cleared students, and their holding student's roll number. We can also highlight the class topper of the results. Hence results are improved in further attempts or in further batches.

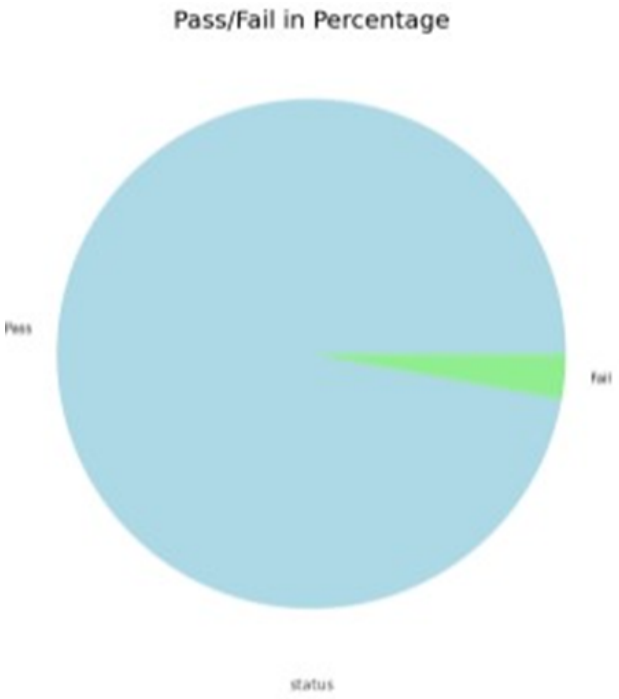

Fig 6.3.4.i: Pass vs Fail in percentage

7 TESTING AND VALIDATION

7.1 TESTING:

In general, algorithm have been made to extract accurate results from dynamic web pages. When we speak about testing the program which is used for results extraction from the dynamic web pages, the following testing is performed against the program in Phase 1 of our project.

7.1.1 FUNCTIONAL TESTING:

Functionality testing is performed to verify whether the product/application meets the intended specifications and the functional requirements mentioned in the documentation. Functional tests are written from a user's perspective.

Status: while we are extracting the results from the dynamic web pages, we are getting the exact match results of each and every student.

7.1.2 STRUCTURAL TESTING:

In Structural testing, a white box testing approach is taken with focus on the internal mechanism of a system or component.

Status: able to observe in the Excel sheet that the results are finely cleaned such that not even a single cell has a null value or mismatched value.

7.1.3 INTEGRATION TESTING:

Integration testing is the activity of software testing in which individual software modules are combined and tested as a group. Testing in which software components or hardware components or both are combined and tested to evaluate the interaction between them.

Status: At the initial stage, we wrote two programs. One is to extract results and create a data set, another is to convert the data set into a clear Excel sheet. When we merged both and converted into a single program, we got the same output as we got before doing the same thing.

7.1.4 UNIT TESTING:

Testing of individual software components or groups of related components Testing conducted to evaluate whether systems or components pass data and control correctly to one another

Code(Program)

Module

Function

Sub-Program

7.1.4.a RE-TESTING: Retesting means executing the same test case after fixing the bug to ensure the bug fixing.

Status: We had an eagle eye upon each and every source line of code, and tried to solve bugs raised here and there. After huge efforts, we had a final program executed with no errors.

7.1.5 STRESS TESTING:

- Stress Tests determine the load under which a system fails, and how it fails.
- Testing conducted to evaluate a system or component at or beyond the limits of its specified requirements to determine the load under which it fails and how.
- A graceful degradation under load leading to non-catastrophic failure is the desired result. Often Stress Testing is performed using the same process as Performance Testing but employing a very high level of simulated load.

Status: We had tried so many attempts to achieve a final executable program which runs under stress when we give a huge set of input (say 1000 hall ticket numbers in a row). As a result of our efforts, our program will executes seamlessly unless and until the computer has a proper and stable internet connection.

7.2 PERFORMANCE AND GUIDELINES:

- ➔ The results are perfectly extracted from the dynamic web pages such that no issues are raised over mismatching of marks.
- ➔ Unless and until the computer is provided with a seamless power supply and stable internet connection, the program will perform well.
- ➔ There is no need for physical inspection of a person while the entire process is going on.
- ➔ Mostly auto saved files.
- ➔ We prefer to disable google chrome auto update feature such that no mismatch of google chrome and chromium web driver raises.
- ➔ There is no such restriction applied that only google chrome is the only browser which is supported by the project. All the things we need to change in the source code is to replace the necessary lines related to the chromium and download the web browser accordingly. Say…
    - ◆ Chromium web driver for Google Chrome.
        - https://chromedriver.chromium.org/
    - ◆ Firefox web driver for Mozilla Firefox.
        - https://www.browserstack.com/guide/run-selenium-tests-using-firefox-driver
    - ◆ Safari web driver for Apple Safari
        - https://developer.apple.com/documentation/webkit/testing_with_webdriver_in_safari
- ➔ .

8. CONCLUSION

In the last decade, sentiment analysis enabled by NLP, machine learning, and deep learning techniques has also been attracting the attention of researchers in the educational domain in order to examine students' attitudes, opinions, and behavior towards numerous teaching aspects. In this context, we provided an analysis of the related literature by applying a systematic mapping study method. Specifically, in this mapping study, we selected 92 relevant papers and analyzed them with respect to different dimensions such as the investigated entities/aspects on the education domain, the most frequently used bibliographical sources, the research trends and patterns, what tools were utilized, and the most common data representation techniques used for sentiment analysis.

We have shown an overall increasing trend of publications investigating this topic throughout the studied years. In particular, there was a significant growth of articles published during the year 2020, where the DL techniques were mostly represented.

The mapping of the included articles showed that there is a diversity of interest from researchers on issues such as the approaches/techniques and solutions applied to develop sentiment analysis systems, evaluation metrics to assess the performance of the systems, and the variety of datasets with respect to their size and format.

- A Solely Automated Results Analysis System project, we are going to make an attempt to effectively introduce the concept of automation using python in order to analyse results of semester examinations.
- We then explain the concept of how proposed methodology is going to affect the present practice.
- We describe the proposed system and explain the features implemented by our proposed system.
- We also give a brief overview of technologies used during the development of our proposed system.
- This project can be further refined and extended by introducing new and more innovative features.

In light of the findings highlighted by the body of knowledge, we have identified a variety of challenges regarding the application of sentiment analysis to examine students' feedback. Consequently, recommendations and future directions to address these challenges have been provided. We believe that this study's results will inspire future research and development in sentiment analysis applications to further understand students' feedback in an educational setting.

In future work, our plan is to further deepen the analysis that we performed in this mapping study by conducting systematic literature reviews (SLRs), as also suggested by

9. REFERENCES

- Plutchik, R. The Nature of Emotions. *Am. Sci.* 2001, *89*, 344–350. [Google Scholar] [CrossRef]
- Cambria, E.; Schuller, B.; Xia, Y.; Havasi, C. New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intell. Syst.* 2013, *28*, 15–21. [Google Scholar] [CrossRef]
- Feldman, R. Techniques and Applications for Sentiment Analysis. *Commun. ACM* 2013, *56*, 82–89. [Google Scholar] [CrossRef]
- Yang, L.; Li, Y.; Wang, J.; Sherratt, R.S. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access* 2020, *8*, 23522–23530. [Google Scholar] [CrossRef]
- Carosia, A.; Coelho, G.P.; Silva, A. Analyzing the Brazilian financial market through Portuguese sentiment analysis in social media. *Appl. Artif. Intell.* 2020, *34*, 1–19. [Google Scholar] [CrossRef]
- Capuano, N.; Greco, L.; Ritrovato, P.; Vento, M. Sentiment analysis for customer relationship management: An incremental learning approach. *Appl. Intell.* 2020, *50*, 1–14. [Google Scholar] [CrossRef]
- Python tutorial: https://www.studytonight.com/python/
- Article on automation: https://automatetheboringstuff.com/