## TRACKING STUDENT ACADEMIC ACTIVITIES AND PREDICTING THE TECHINICAL SKILLS USING LATEST TECHNOLOGIES

A project report submitted in partial fulfillment of the requirements for the award of the degree of

**MASTER OF COMPUTER APPLICATIONS**

By

Sadhu Chitti Tarun Raj

20VV1F0017

Under the Esteemed Guidance of

#### Dr. G. JAYA SUMA, M.Tech, Ph.D.

#### Professor

Department of Information Technology



**DEPARTMENT OF INFORMATION TECHNOLOGY JAWAHARLAL NEHRU TECHNOLOGICAL**

**UNIVERSITY GURAJADA**

**UNIVERSITY COLLEGE OF ENGINEERING VIZIANAGARAM – 535003, A.P.**

**2020- 2022**

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**JNTUGV – UNIVERSITY COLLEGE OF ENGINEERING**

**VIZIANAGARAM**



### CERTIFICATE

This is to certify that the dissertation report entitled **“TRACKING STUDENT ACADEMIC ACTIVITIES AND PREDICTING THE TECHINICAL SKILLS USING LATEST TECHNOLOGIES”** submitted by **SADHU CHITI TARUN RAJ** bearing registration number **20VV1F0017** in partial fulfillment for the award of the degree of **Master of Computer Applications (MCA)** from Jawaharlal Nehru Technological University Gurajada – University College of Engineering Vizianagaram. This bonafide work was carried out by him under my guidance and supervision during year 2020-2022.

The results embodied in this dissertation have not been submitted to any other University or Institute for the award of any degree or diploma.

#### Signature of Project Guide Signature Head of the Department

|  |  |
| --- | --- |
| G. JAYA SUMA, M.Tech, Ph.D, | Dr. B.Tirimula Rao, |
| Professor | Assistant Professor & HOD |
| Dept. of Information Technology | Dept. of Information Technology |

**(Signature of External Examiner)**

### DECLARATION

I, SADHU CHITTI TARUN RAJ (Reg. No: 20VV1F0017) declare that this submission represents my ideas in my own words, and where others’ ideas or words have been included, I have adequately cited and referenced the source. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Signature)

S.Ch.Tarun Raj

(20VV1F0017)

Date:

Place:

### ACKNOWLEDGEMENT

This acknowledgment transcends the reality of formality when I express deep gratitude and respect to all those people behind the screen who inspired and helped us in the completion of of this project work.

I take the privilege to express my heartfelt gratitude to my guide **Dr. G. JAYA SUMA, M.Tech, Ph.D, Professor, Department of Information Technology, JNTUG University College of Engineering Vizianagaram** for her valuable suggestions and constant motivation that greatly helped me in the successful completion of the dissertation. His wholehearted cooperation and the keen interest shown to him at all stages are beyond words of gratitude.

**`**With great pleasure and privilege, I wish to express my heartfelt sense of gratitude and indebtedness to **Dr.B.Tirimula Rao, Assistant Professor, & HOD, Head of Department of Information Technology, JNTUG – UCEV Vizianagaram,** for his supervision.

I extend heartfelt thanks to our principal **Prof. R. Rajeswara Rao** for providing intensive support throughout my dissertation.

I am also thankful to all the Teaching and Non-Teaching staff of the Information Technology Department, JNTUG University College of Engineering Vizianagaram, for their direct and indirect help provided to me in completing the dissertation.

I extend my thanks to my parents and friends for their help and encouragement in the success of my dissertation.

S.Ch.TarunRaj

(20VV1F0017)

**ABSTRACT**

### ABSTRACT

This report introduces the process of creating part of a course management system which is a data-driven website used by the administrator. The major part of the group component and minor part of the marking component is implemented by me and the implementation details will be introduced in the report. The implementation uses a tool called Flask Framework which is an excellent open-source web application framework for complex data-driven website development. The major part of this report will introduce how Python to use to create a database table report, web page user interface, and inner logic to handle user requests by going through the group component implementation process.  In this project, the importance of Python and back-end web development is explained. Using the tools and commands of these languages we can make an interactive and user-friendly interface that can be easily accessed by people. Back-end was earlier known for limited purposes but now when it is explored thoroughly proved to be one of the best languages for website development and highly advanced mobile applications.

**TABLE OF CONTENTS**

**TABLE OF CONTENTS Page.No**

**ABSTRACT**

**LIST OF FIGURES**

1. INTRODUCTION …………………………………………………………….

1.1 Web Frameworks………………………………………………………….

1.1.1 Problem statement ………………………………………………

1.1.2 Flask…………………………………………………………..

1.1.3 Databases ………………………………………………

1.14 Applications ………………………………………………………..

1.2 Introduction to NLP ………………………………………

1.3 Introduction to Machine Learning **………………………………….**

**2. LITERATURE SURVEY ……………………………………………………..**

**3. SOFTWARE AND HARDWARE REQUIREMENTS …………………….**

**4. DATA SET ……………………………………………………………………...**

**5. ARCHITECTURE ……………………………………………………………**

**6. METHODOLOGY …………………………………………………………….**

6.1 NLP Techniques

6.1.1 Data Preprocessing …………………………………………

6.1.2 Tokenization ……………………………………………….

6.1.3 Lower case ………………………………………….

6.1.4 Removing Stop words and Punctuation…………………………….

6.1.5 Stemming ……………………..

6.1.6 Lemmatization ………………………………………………………

6.1.7 Features Extraction-Used TF-IDF Vectorized ……………………….

6.2 Algorithms

6.2.1. Support vector machines …………………………………..

6.2.2 Naïve Bayes …………………………………………………

6.2.3 Logistic Regression ………………………………………

6.2.4 Random Forest Classifier …………………………….

**7. IMPLEMENTATION** ………………

**8. RESULTS AND ANALYSIS ………………………..**

**9. CONCLUSION ……………………………………..**

**LIST OF FIGURES**

# CHAPTER 1 INTRODUCTION

# 

1. **INTRODUCTION**

  The web is the only true object-oriented system”, says Alan Kay, the inventor of object-oriented programming. This is because the web's service-oriented architecture embodies the OO paradigm's early ideas much better than the way it has been implemented in mainstream programming languages. Indeed, the web has become increasingly a system of interconnected services. 

* 1. **Web frameworks:**

A web framework is a code library that makes a developer’s life easier when building reliable, scalable, and maintainable web applications. Web frameworks encapsulate what developers have learned over the past twenty years while programming sites and applications for the web. Frameworks make it easier to reuse code for common HTTP operations and to structure projects so other developers with knowledge of the framework can quickly build and maintain the application.

Frameworks provide functionality in their code or through extensions to perform common operations required to run web applications. These common operations include:

* HTML, CSS, JavaScript, and Bootstrap’s
* URL routing
* Database manipulation
* Security against Cross-site request

**1.1.1 Hyper Text Markup Language (HTML):**

This language provides the format for specifying simple logical structure and links in a hypertext document. As a markup language, special formatting commands are placed in the text describing how the final version should appear. These formatted documents are interpreted by a Web browser which usesthe HTML code to format the page being displayed. Although most professionals use special authoring tools to write HTML documents and to manage sites, developers of e-commerce sites and applications need to know the nitty-gritty detail of HTML, and this is what you will study. HTML has had several versions over the years. "HTML 2.0" was the first standard HTML specification which was published in 1995.

HTML 4.01 was a major version of HTML and it was published in late 1999. Though HTML 4.01 version is widely used but currently we are having HTML 5 version which is an extension to HTML 4.01, and this version was published in 2012.

HTML pages are created by tagging textual information with HTML markup. HTML markup consists of tags, which appear inside angled brackets < and >. An example of an HTML tag is, which causes text to appear in bold. only notes where text should begin to appear in bold, while the tag marks the end of the emboldening. Most HTML tags have a corresponding end tag, which is specified by the name of the tag preceded by the / character.

**1.1.2 CASCADING STYLE SHEETS (CSS):**

There is no format to follow for teaching the aesthetics of style - most people, though, can recognize something that follows a classical design. But some things can be said about the style of a website. For instance, when Web pages belong to the same website, each page should have a consistent look in order to provide familiarity for the user. Style sheets (sometimes referred to as templates) are used in desktop publishing to provide consistency when formatting text. The format applied by the stylesheet could be to indent every first line of a paragraph by 2cm, insert a page break at the end of every chapter, and so on. Naturally, due to multimedia, Web pages not only have to consider text formatting, but also visual and sound presentation, and various multimedia formats in general. Before we continue, let us briefly discuss the advantages and disadvantages of using style sheets.  
Bootstrap  
3.1 Introduction  
One of the problem with basic HTML design is that the webpage may look different in different browser or device(e.g. mobile, tablet and laptop). Therefore, we may need to modify the code according to browser or device. The problem can be easily resolved by using Bootstrap.Bootstrap is a framework which uses HTML, CSS and JavaScript for the web design. It is supported by all the major browsers e.g. Firefox, Opera and Chrome etc. Further, Bootstrap includes several predefined classes for  
easy layouts e.g. dropdown buttons, navigation bar and alerts etc. Lastly, it is responsive in nature i.e. the layout changes automatically according to the device e.g. mobile or laptop etc.  
3.2 Setup  
Bootstrap needs atleast 3 files for its operation which can be downloaded from the Bootstrap website.  
• bootstrap.css (Line 7) : This file contains various CSS for bootstrap.  
• bootstrap.js (Line 16) : This file contains various JavaScript functionalities e.g. dropdown and alerts etc.  
• jQuery.js (Line 17) : This file is the jQuery library which can be downloaded from the ‘jQuery’ website.  
It is required for proper working of ‘bootstrap.js’.

3.2.1 Download and include files  
These files are downloaded and saved inside the ‘asset’ folder. Next, we need to include these files in the HTML document as below,

3.2.2 Add CDN  
Another way to include the files is CDN. In this method, we need not to download the files, but provide the links to these files, as shown in Lines 8, 17 and 19 of below code. Note that, in this case the code will not work in offline mode

### 3.2.3 Check setup Let’s write our first code to check the correct setup of the Bootstrap

3.3 Grid system  
Bootstrap divides the each row into 12 columns. Then following commands can be used to specify the width  
the columns  
• col-lg-4 : It will select 4 columns. Choose any number between 1-12. The ‘lg’ stand for large screen (e.g.  
large computer screen).  
• col-md-5 : ‘md’ = medium screen  
• col-sm-3 : ‘sm’ = small screen  
• col-xs-3 : ‘xs’ = extra small screen  
• col-lg-offset-4 : skip first 4 columns. Simimlary use ‘md’, ‘sm’ and ‘xs’ with offset

JavaScript  
4.1 Introduction  
• JavaScript is a dynamic language which is used for designing the web pages on the client side.  
• It is case sensitive language.  
• It is untyped language i.e. a variable can hold any type of value.  
• // is used for comments.  
• ; i used for line termination.  
• JavaScript code should be added at the end i.e. just above the closing-body-tag.  
• It is better to write the JavaScript code in separate file as shown in next section.  
4.2 First code  
The JavaScript code can be written in the ‘html’ file or in the separate ‘JavaScript file (.js)’ as shown in this  
section,  
4.2.1 JavaScript in HTML file  
In HTML file, the JavaScript codes can be written inside the ‘script’ tag as shown in Lines 11-13 of below code.  
The code will write the message “Hello World from JavaScript!” on the web page. Open the ‘js.html’ in the browser  
to see the message.

4.2.1.1 JavaScript in separate file  
• The JavaScript code is saved in the file ‘my\_javascript.js’ which is located inside the folders ‘asset/js’. Note  
that, no ‘script tag’ is used here.

HTML, CSS, Bootstrap, Javascript and jQuery  
Hello World from JavaScript!  
4.2.1.1 JavaScript in separate file  
• The JavaScript code is saved in the file ‘my\_javascript.js’ which is located inside the folders ‘asset/js’. Note  
that, no ‘script tag’ is used here.  
  
4.3 Keywords, Datatypes, Variables and Operators  
4.3.1 Keywords  
• Below are the reserved keywords in the JavaScript which can not be used as ‘variable’ and ‘function’ names  
etc.

4.3.2 Datatypes  
JavaScript has three types of data,  
• Numbers : 123, 32.32  
• Strings : “Meher”, “Krishna Patel”, “123”  
• Boolean : true, false

4.3.3 Variables  
Variables can be define using keyword ‘var’. Further, JavaScript is untyped language i.e. a variable can hold any  
type of value.  
• In the below HTML, ‘p’ tag with id ‘p\_name’ is defined at Line 10. This id will be used to write some text  
using JavaScript,

jQuery  
5.1 Introduction  
jQuery is the Javascipt library.  
5.1.1 Requirements  
• First create a file ‘jQuery.html’. Next, we need three files, i.e. one CSS (Line 9) and two Javascipt (Lines  
23-24), for using the jQuery in this tutorial. These files are saved inside the folder ‘asset’, which can be  
downloaded from the repository. Lines 9 and 24 are the bootstrap files, whereas Line 23 is the actual  
‘jQuery’ file which is required to use the jQuery.

5.1.2 Add contents  
Next add one jumbotron (Line 15-19) and few labels (Lines 21-26) in the html file as shown in Fig. 5.1.  
1 <!-- jQuery.html -->  
2  
3 <!DOCTYPE html>  
4 <html>  
5 <head>  
6 <title>jQuery</title>  
7  
8 <!-- CSS -->  
9 <link href="asset/css/bootstrap.min.css" rel="stylesheet">  
10 <!-- Add Custom styles below -->  
11  
12 </head>  
13 <body>  
14  
15 <div class="jumbotron">  
16 <div class="col-md-offset-4 container">  
17 <h1><b>jQuery</b></h1>  
18 </div>  
19 </div>  
20  
21 <span class="label label-lg label-default" id="l\_default">Default</span>  
22 <span class="label label-primary" id="l\_primary">Primary</span>  
23 <span class="label label-success" id="l\_success">Success</span>  
24 <span class="label label-info" id="l\_info">Info</span>  
25 <span class="label label-warning" id="l\_warning">Warning</span>  
26 <span class="label label-danger" id="l\_danger">Danger</span>  
27  
28  
29  
30  
31 <!-- Javascript -->  
32 <!-- put jQuery.js before bootstrap.min.js; and then add custom jquery -->  
33 <script src="asset/js/jquery-3.3.1.min.js"></script>  
34 <script src="asset/js/bootstrap.min.js"></script>

</body>  
37 </html>

5.2.3.2 Toggle label based on the buttons  
In previous section, we wrote the code which can toggle only one label. In this section, we will toggle multiple  
labels by writing one-generalized-jQuery

5.3 Selectors  
In this section, we will use following HTML to understand the various combination of the selectors which are  
summarized in Table 5.1,  
<!DOCTYPE html>  
<html>  
<head>  
<title>jQuery Selectors</title>  
</head>  
<body>  
<p>jQuery Selectors</p>  
<p class="foo"> Paragraph with class.</p>  
<p><span>Paragraph with span.</span></p>  
<p id="bar">Paragraph with ID. <span class="foo">Span with class and inside the paragraph-with-id.</  
˓→span></p>  
<div>  
<p my-id="my\_para">Paragraph (with custom id) inside the div</p>  
<p my-id="my\_para2"><span>Span inside the paragraph (with custom id) inside the div</span></p>  
</div>  
<!-- Javascript -->  
<script src="asset/js/jquery-3.3.1.min.js"></script>  
</body>  
</html>

5.3.1 Select elements  
5.3.1.1 Select HTML elements  
HTML elements, e.g. div, p, span and table etc., can be selected by writing them with the quotation mark as  
shown in below codes.  
• Here, ‘$(‘p’) is the command, whereas the lines below it are the outputs.  
• The outputs show the ‘selected elements’; and the total number of selected elements i.e. ‘length: 6’.  
• If we put the cursor to one of the outputs, then it will highlight the corresponding elements in the html

Similarly, we can select ‘div’ elements  
$('div')  
{...}  
0: <div>  
length: 1  
5.3.1.2 Select class  
Class can be selected using ‘.’ operator. In below code, class ‘foo’ is selected.  
$(".foo")  
{...}  
0: <p class="foo">  
1: <span class="foo">  
length: 2  
5.3.1.3 Select ID  
ID can be selected using ‘#’ operator. In the below code, id ‘bar’ is selected.  
$('#bar')  
[...]  
0: <p id="bar">  
length: 1

5.3.1.4 Combining selectors  
We can select a element with specific class name using ‘.’ operator as shown in below code. Here element ‘span’  
with class ‘foo’ is selected,

HTML, CSS, Bootstrap, Javascript and jQuery  
$("span.foo")  
{...}  
0: <span class="foo">  
length: 1  
5.3.1.5 Multiple selectors  
Use ‘,’ to select different types of elements. In the below code, “paragraph with id ‘bar’ ” and “elements with class  
‘foo’ ” are selected ,  
$('p#bar, .foo')  
{...}  
0: <p class="foo">  
1: <p id="bar">  
2: <span class="foo">  
length: 3  
5.3.1.6 Select descendant element  
Select ‘span’ inside the ‘p’ (use space between ‘span’ and ‘p’),  
$('p span')  
{...}  
0: <span>  
1: <span class="foo">  
2: <span>  
length: 3  
5.3.1.7 Select child element  
The ‘>’ is used to select the child of an element.

HTML, CSS, Bootstrap, Javascript and jQuery  
$("span.foo")  
{...}  
0: <span class="foo">  
length: 1  
5.3.1.5 Multiple selectors  
Use ‘,’ to select different types of elements. In the below code, “paragraph with id ‘bar’ ” and “elements with class  
‘foo’ ” are selected ,  
$('p#bar, .foo')  
{...}  
0: <p class="foo">  
1: <p id="bar">  
2: <span class="foo">  
length: 3  
5.3.1.6 Select descendant element  
Select ‘span’ inside the ‘p’ (use space between ‘span’ and ‘p’),  
$('p span')  
{...}  
0: <span>  
1: <span class="foo">  
2: <span>  
length: 3  
5.3.1.7 Select child element  
The ‘>’ is used to select the child of an element.  
$('div>span')  
{...}  
length: 0  
Note: There is one element inside ‘div’ but it’s path is ‘div->p->span’ therefore ‘span’ is a child of ‘p’ (not of  
‘div’). But tgis ‘span’ is the descendant of ‘div’ therefore can be accessed using space,  
$('div span')  
{...}  
0: <span>  
length: 1  
5.3.1.8 Attribute selector  
Custom attributes can be selected as below,  
$('[my-id = my\_para]')  
{...}  
0: <p my-id="my\_para">  
length: 1

$('p[my-id != my\_para]') // select p with 'my-id != my\_para'  
{...}  
0: <p>  
1: <p class="foo">  
2: <p>  
3: <p id="bar">  
4: <p my-id="my\_para2">  
length: 5  
$('p[class = foo]')  
{...}  
0: <p class="foo">  
length: 1  
5.3.2 Filters  
Various filter options (e.g. filter, first, last and has etc.) can be used after ‘:’ as shown below,  
$('p:first') // select first element  
{...}  
0: p  
$('p:last') // last element  
{...}  
0: <p my-id="my\_para2">  
length: 1  
$('p:not(.foo)') // 'p' which does not have class 'foo'  
{...}  
0: <p>  
1: <p>  
2: <p id="bar">  
3: <p my-id="my\_para">  
4: <p my-id="my\_para2">  
length: 5  
$('p:has(span)') // 'p' which has 'span' inside it  
{...}  
0: <p>  
1: <p id="bar">  
2: <p my-id="my\_para2">  
length: 3  
$('p:hidden') // hidden 'p'  
{...}  
length: 0  
$('p:visible') // visible 'p'  
{...}  
0: <p>  
1: <p class="foo">  
2: <p>  
3: <p id="bar">  
4: <p my-id="my\_para">  
5: <p my-id="my\_para2">  
length: 6  
$('p:contains(with class)') // paragraph which contains class inside it

{...}  
0: <p class="foo">  
1: <p id="bar">  
length: 2  
$('p').filter(".foo") // same as p.foo  
{...}  
0: <p class="foo">  
length: 1  
$('p').not(".foo") // p with not class foo  
{...}  
0: <p>  
1: <p>  
2: <p id="bar">  
3: <p my-id="my\_para">  
4: <p my-id="my\_para2">  
length: 5  
$('p').last() // last element of 'p'  
{...}  
0: <p my-id="my\_para2">  
length: 1  
$('p').has('span') // p with span  
{...}  
0: <p>  
1: <p id="bar">  
2: <p my-id="my\_para2">  
length: 3  
$('p').is('.foo') // does 'p' contain 'foo'  
true  
$('p').children() // children of 'p'  
{...}  
0: <span>  
1: <span class="foo">  
2: <span>  
length: 3  
$('p').children(".foo") // child of 'p' with class 'foo'  
{...}  
0: <span class="foo">  
length: 1

### 1.1.3 Python Language Introduction:

[Python](https://www.geeksforgeeks.org/python-programming-language/) is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.Python is a programming language that lets you work quickly and integrate systems more efficiently.  
What is Python REST API Framework  
Python REST API framework is a set of utilities based on werkzeug to easily build Restful API. It keep a clean  
codebase and is easy to configure and extends.  
It does not decide how you want to render your data, or where they lives, or other decisions.  
Instead it give you a good start with an extensible architecture to build your own API.  
Python REST API Framework has not been create for the usual case. Instead it give you some hook to your very  
special ressource provider, your very special view management and your very special way of displaying data.  
Python REST API Framework is fully REST compilant; It implement the common verbs:  
• GET  
• POST  
• UPDATE  
• DELETE  
• HEAD  
It also implement:  
• PAGINATION  
• AUTHENTICATION  
• RATE-LIMIT  
• DATA VALIDATION  
• PARTIAL RESPONSE

Architecture  
Python REST API Framework is base on the MVC pattern. You define some endpoints defining a Ressource, a  
Controller and a View with a set of options to configure them.

Controller  
Manage the way you handle request. Controller create the urls endpoints for you. List, Unique and Autodocumented  
endpoints.  
Controller also manage pagination, formaters, authentication, authorization, rate-limit and allowed method.  
DataStore  
Each method of a Controller call the DataStore to interact with data. The DataStore must be able to retreive data from  
a ressource.  
Each datastore act on a particular type of ressource (database backend, api backend, csv backend etc...). It must be  
able to validate data, create new ressources, update existing ressources, manage filters and pagination.  
Optional configuration option, that can be unique for a particular datastore like Ressource level validation (unique  
together and so), ForeignKey management...  
View  
Views defines how the data must be send to the client. It send a Response object and set the needed headers, mime-type  
and other presentation options like formaters.  
How To use it  
To create as many endpoint as you need. Each endpoints defining a ressource, a controller and a view. Then add them  
to the rest\_api\_framework.controllers.

Ressource  
Ressource define your data. Where are your data ? How can they be accessed ? What they look likes?  
• ressource\_name: will be used to build the url endpoint to your ressource.  
• ressource: where your ressource lies.this argument tell the datastore how they can be accessed. It can be the  
database name and the database table for a SQL datastore or the url endpoint to a distant API for exemple.  
• model: describe how your data look like. Wich field it show, how to validate data and so on.  
• datastore: the type of your data. There is datastore for simple Python list of dict and SQLite datastore.  
They are exemple on how to build your own datastore depending on your needs.  
Controller  
The controller define the way your data should be accessed. Should the results be paginated ? Authenticated ? Rate-  
limited ? Wich it the verbs you can use on the resource ? and so on.  
• list\_verbs: define the verbs you can use on the main endpoint of your ressource. If you dont’ use “POST”,  
a user cannot create new ressources on your datastore.

unique\_verbs: define the verbs you can use on the unique identifier of the ressource. actions depending on  
the verbs follows the REST implementation: PUT to modify an existing ressource, DELETE to delete a  
ressource.  
View  
view define How your ressoources should be rendered to the user. It can be a Json format, XML, or whatever. It can  
also render pagination token, first page, last page, number of objects and other usefull informations for your users.  
• response\_class: the response class you use to render your data.  
To test you application locally, you can add:  
if \_\_name\_\_ == '\_\_main\_\_':  
from werkzeug.serving import run\_simple  
from rest\_api\_framework.controllers import WSGIDispatcher  
app = WSGIDispatcher([ApiApp])  
run\_simple('127.0.0.1', 5000, app, use\_debugger=True, use\_reloader=True)  
then type “python app.py” and your API is up and running  
Options  
Each of this dicts can take an optional parameter: “option”. This parameter is a dict containing all the options you  
want to use with either the datastore, the view or the controller.  
You can learn more about optional parameters in the documentation of each topic : datastore, view, controller  
Using a database  
Instead of using a python dict, you may want to actualy save your data in a database. To do so, you just have to change  
your datastore and define your ressources in a way SQL datastore can understand.  
SQLiteDataStore use sqlite3 as database backend. ressources will be a dict with database name and table name. The  
rest of the configuration is the same as with the PythonListDataStore

**1.1.2 Flask**

Since this bachelor project is about implementing a dashboard for Flask, a bit more information about Flask is provided and details are left out for Django

Flask is called a micro framework because it does not require particular tools or libraries. It has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools. Extensions are updated far more regularly than the core Flask program.

Flask is easy to get started with as a beginner because there is little boilerplate code for getting a simple app up and running

from flask import Flask app = Flask( n a m e )

@app. route(’/’)

def hello world ():

return ’ Hello World!’

if n a m e == ’ m a i n ’: app. run()

With the code snippet above, a simple web-service is created (line 1 and 2). Moreover, an endpoint is connected to the web-service (line 4). The web-service can be started with the code expressed on the last line (line 9). Each Flask web-service contains the parts that are expressed in this code snippet, but are usually more complex[1](#_bookmark7) than this simple example

**Virtual environments**

Use a virtual environment to manage the dependencies for your project, both in development and in production.

What problem does a virtual environment solve? The more Python projects you have, the more likely it is that you need to work with different versions of Python libraries, or even Python itself. Newer versions of libraries for one project can break compatibility in another project.

Virtual environments are independent groups of Python libraries, one for each project. Packages installed for one project will not affect other projects or the operating system’s packages.

Python 3 comes bundled with the [env](https://docs.python.org/3/library/venv.html" \l "module-venv) module to create virtual environments. If you’re using a modern version of Python, you can continue on to the next section.

**1.1.3 Databases**

A database stores application data in an organized way. The application then issues queries to retrieve specific portions as they are needed. The most commonly used databases for web applications are those based on the relational model, also called SQL databases in reference to the Structured Query Language they use. But in recent years document-oriented and key-value databases, informally known together as NoSQL databases, have become popular alternatives

**SQL Databases**

Relational databases store data in tables, which model the different entities in the ap‐ plication’s domain. For example, a database for an order management application will likely have customers, products, and orders tables. A table has a fixed number of columns and a variable number of rows.

The columns define the data attributes of the entity represented by the table. For example, a customer table will have columns such as name, address, phone, and so on. Each row in a table defines an actual data element that consists of values for all the columns. Tables have a special column called the primary key, which holds a unique identifier for each row stored in the table. Tables can also have columns called foreign keys, which reference the primary key of another row from the same or another table. These links between rows are called relationships and are the foundation of the relational database model.

**Python Database Framework**

* **SQLite**

SQLite is a software library that implements a self-contained, serverless, zero configuration, transactional SQL database engine. SQLite is one of the fastest-growing database engines around, but that's growth in terms of popularity, not anything to do with its size. The source code for SQLite is in the public domain.

SQLite is an in-process library that implements a self-contained, serverless, zero configuration, transactional SQL database engine. It is a database, which is zero configured, which means like other databases you do not need to configure it in your system.

SQLite engine is not a standalone process like other databases, you can link it statically or dynamically as per your requirement with your application. SQLite accesses its storage files directly.

**1.3 Introduction to Machine Learning**

Machine Learning is a subset of artificial intelligence that has primarily been used to create algorithms that allow the computer to benefit from data and prior knowledge on its own. Arthur Samuel first demonstrated terminological machine learning in 1959. It can be described as follows.

**Working of Machine Learning:**

"The Machine Learning Model must learn from real data, establish prediction models, and predict outputs while we provide new data," says the author. The amount of data influences the accuracy of the output because the majority of the data is used to build the best model for forecasting the new data point.

Machine learning Algorithms

Building Logical Models

Output

**Training**

**Input past data**

**Learn from data New data**

Fig: 3 Work Flow Of Machine Learning

**Classification Of Machine Learning:**

Machine Learning can be categorized into three types:

1. Supervised learning

2. Unsupervised learning

3. Reinforcement learning

**Supervised Learning:**

Supervised machine learning is a process in which we receive labeled sample data for the machine learning scheme to train it and predict the outcome. The machine generates a model with fewer details so that datasets and studies each dataset. Once the preparation and processing are completed, we will have sample data to validate the model and see if it predicts the same performance. The goal of supervised learning is to convert input data into output data. Simple supervised learning is dependent on supervision and is analogous to a student learning something under the supervision of an instructor.

The goal of supervised learning is to connect input and output data. Supervised learning is dependent on supervision.

Training data(contains both input&out put&pattern in them

Algorithms

Prediction

(final out put)

Classifier

New input data

Fig:4 Supervised Learning

Supervised machine learning is depicted graphically. Original preprocessed data sets containing known variables and targets are divided into training data and test data in supervised learning. (Above) The training data are used to train a learning algorithm in an attempt to develop an accurate predictive model during the training phase. (Center) The test data are then applied to the model to validate it, and the predictive accuracy is evaluated. (Below) After the model has been validated, new data is fed into it in an attempt to make new predictions.

**Advantages of Supervised learning:**

* This type of learning is simple to grasp. It is the most widely used type of learning method. People should begin learning ML by practicing supervised learning.
* The training data is only required to train the model. Because of its size, it takes up a lot of room. However, it is erased from memory once training is completed because it is no longer relevant.
* We'd already know how many classes are in the data.
* After training, the model will determine which specific data needs to be predicted, as all of the data in the collection is unimportant.
* It is extremely useful in solving real-world computational problems.

**Disadvantages of supervised learning:**

* Its performance is limited by the fact that it cannot handle complex ML problems.
* It is unable to generate its own labels. This means that, unlike unsupervised learning, it cannot discover data on its own.
* Any new data we enter must come from one of the given classes. If you enter watermelon data into a collection of apples and oranges, it may classify watermelon as one of these, which is incorrect.
* To train a supervised learning-based model, a good computer with quality processors is required. It necessitates a high level of computation power, which not all PCs may have.

**Unsupervised Learning:**

Unsupervised learning is a mechanism that allows a computer to learn without supervision. Because training data is presented to a computer with data that has not been labeled or classified the algorithm must function without any control over that data. Unsupervised learning attempts to reconstruct input data into new features or a collection of items with matching patterns.

**Reinforcement Learning:**

Strengthening learning is an interpretive-based learning process in which the learning assistant is rewarded for each correct action and fined for each incorrect action. The assistant learns from the feedback and improves its performance as a result. The assistant deals with and examines the world to strengthen learning. The goal of the assistant is to earn more reward points.

**CHAPTER-2**

**LITERATURE SURVEY**

**LITERATURE SURVEY**

The more popular and widely used communication system is the electronic mail or Email communication system. Many organizations throughout the world have been exerting efforts in identifying spam emails. The authors who have discussed the identification of these spam or ham are described further. The filtering techniques are needed for the classification of the emails to be classified as spam/ham. The spam email filtering system is proposed by Selamatand Mohamad in which emails are classified by the use of two different feature selection methods. This considers the Malay and English and rough set theory method and TF-IDF were used to select the features from the email dataset after completion of pre-processing on the data. By presenting the practically best results, the classification process is applied with the machine learning algorithm. Harisinghaneyin proposed another work for the email dataset classification based on machine learning algorithms.

The implementation of algorithms that when they are applied to data for preprocessing present effective results, including the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, k-nearest neighbors (KNN) algorithm, and Naive Bayes. The techniques for optimization have also been adapted in this paper. A method of feedforward neural networks is used by Farisin for the detection of emails as spam and effective optimization of results. The Krill Herd algorithm was used to train the neural network. Training and testing can be performed by dividing the preprocessed dataset into two halves. Then the comparison between the results attained from the neural network of optimized classification and other optimization algorithms such as genetic algorithms and backpropagation is done.

A lot of research and literature studies have been done on email spam detection as well as social media and Twitter signaling spam detection. There is no comprehensive systematic literature review on SMS spam detection because it is a relatively new area of research. Although SMS communication first gained traction in 2000, it gained even more traction after the introduction of Android phones in 2006. SMS spam is becoming more popular among spammers as the number of people using SMS as a mode of communication increases. As a result, SMS spam detection research arose out of necessity, and it began in earnest around 2007. The goal of this review is to gain proper knowledge in the field of spam mail detection.

[S.K Tuteja] (2016) [1] The author has worked with different machine learning algorithms for email classification such as Neural Network (NN), Support Vector Machine (SVM), J48 Decision Tree based classifier, and Naïve Bayes. The dataset used by the author the was Spam Base dataset. Paperwork per work, the author didnmenmeet the ion advantages and disadvantages of any algorithm.

[G. Mujtaba] [L. Shuib] [R. G. Raj] [N. Majeed] [M. A. Al- Garadi] (2017) [2] Proposed the basic three steps which are common in every classification process. The first step is pre-processing in which the given text is converted into tokens and this step is also used for the removal of stop words. The second step is the learning process and, in this feature re,a set is built which is very much necessary for the classification of emails. The last step is a classification of email as ham or spam ban y using an efficient algorithm. Algorithms like support vector machines, logistic regression, regression trees, and random forest are considered for classification. They used the Phishing Corpus dataset and with the help of Bag of words and afeatextractionionn approach classified the email as ham or spam. In his study, they did not mention the different tools for reduction methods for email classification.

[S. Ajaz] [M. T. Nafis] [V. Sharma] (2017) [3] They collected email datasets from the online available websites and used Naïve Bayes for filtering emails. He proposed a hybrid approach using the secure hash method and Naive Bayes to filter email data but could not provide information regarding the misuse of storage resources and network bandwidth. By the siusingecure Hash Algorithm, the email is considered as a message M due to a generated function. The message M is further classified into S and L where L stands for ham email or genuine email and on the other,h stands for spam email.

[V.K Singh] [S. Bhardwaj] (2018) [8] They worked on the solution for combining classification techniques to get better results for spam filtering. The author took the help of data mining and using that collected all the information regarding success, current problems, and previous failures of spam filtering. The method was based on binary classification where 1 was used for Spam emails and 0 was used for Ham emails. They combined the 2 methods that are Machine Learning and Knowledge Engineering for the filtering of emails. The performance of the proposed method was very poor on the combined KNN and SVM algorithms.

**CHAPTER-3**

**SOFTWARE AND HARDWARE**

**REQUIREMENTS**

**Software Requirements:**

Environment : Python,Flask

Text editor : Visual studio code editor,command line interface

Web Technologies: HTML, CSS ,Bootstrap & JQuery

Programming Languages: Python and MySQL.

Operating System : windows 11

Edition : Windows 11 Home Single Language

Version : 21H2

Installed on : ‎12-‎04-‎2022

OS build : 22000.856

Serial number : YLW02F55

Experience : Windows Feature Experience Pack 1000.22000.856.0

**Hardware Requirements:**

Device name: Tarun

Processor: 12th Gen Intel(R) Core(TM) i5-12400 2.50 GHz

Installed RAM : 8.00 GB (7.72 GB usable)

Device ID : CF4B8C39-66B9-441C-B549-34CFAA688F4A

Product ID : 00356-24544-73544-AAOEM

System type : 64-bit operating system, x64-based processor

Pen and touch : No pen or touch input is available for this display

**CHAPTER-4**

**DATA SET**

**DATA SET :**

**Context:**

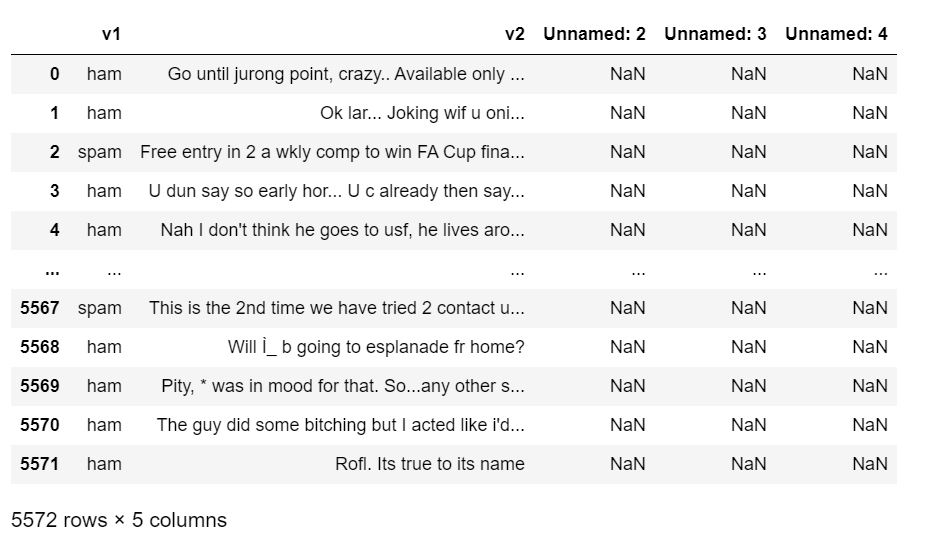
The SMS Spam Collection is a collection of SMS-tagged messages gathered for SMS Spam research. It contains one set of 5,574 SMS messages in English that have been classified as ham (legitimate) or spam.

**Content:**

Each file has one message per line. Each line is made up of two columns: v1, which contains the label (ham or spam), and v2, which contains the raw text.

This corpus was compiled from free or low-cost research sources on the Internet:

* 425 SMS spam messages were manually extracted from the Grumbletext website. This is a UK forum where cell phone users make public claims about SMS spam messages, most of which do not include the spam message itself. Identifying the text of spam messages in claims is a difficult and time-consuming task that required carefully scanning hundreds of web pages.
* A subset of 3,375 SMS randomly selected ham messages from the NUS SMS Corpus (NSC), a dataset of approximately 10,000 legitimate messages collected for research at the National University of Singapore's Department of Computer Science. The messages are mostly from Singaporeans, mostly from students at the University. These messages were gathered from volunteers who were made aware that their contributions would be made public.
* Caroline Tag's Ph.D. thesis yielded a list of 450 SMS ham messages.
* Finally, we've included the SMS Spam Corpus v.0.1 Big. It contains 1,002 ham SMS messages and 322 spam messages.



**CHAPTER-5**

**ARCHITECTURE**

**Architecture:**

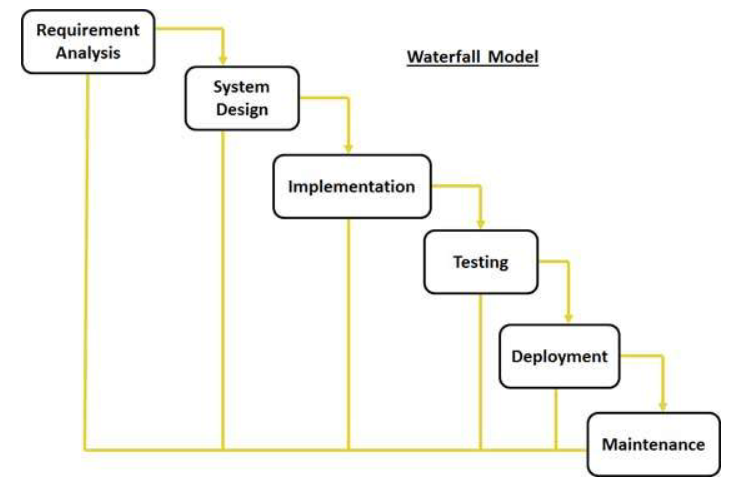
**CHAPTER-6**

**METHODOLOGY**

**WATERFALL MODEL:-**

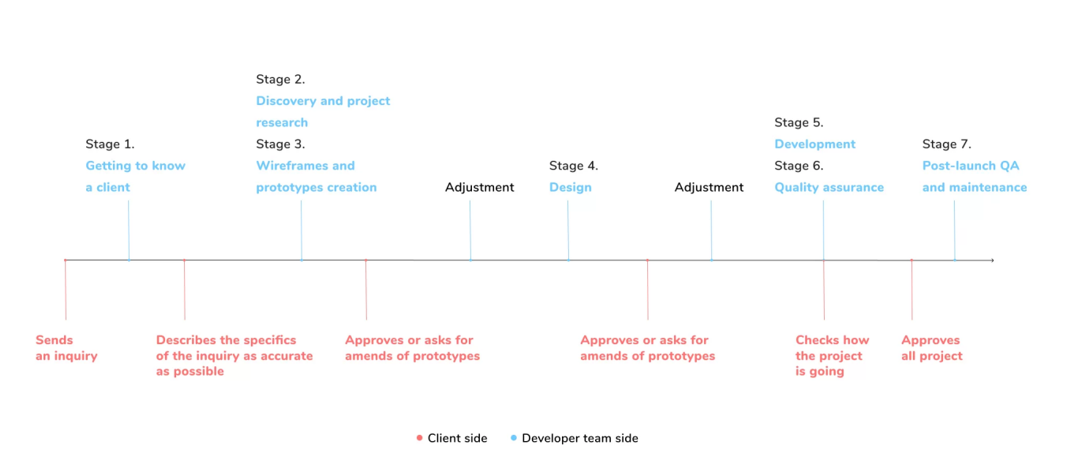
The Waterfall Model was first Process Model to be introduced. It is also referred to as a linear-sequential life cycle model. It is very simple to understand and use. In a waterfall model, each phase must be completed before the next phase can begin and there is no overlapping in the phases.Waterfall model is the earliest SDLC approach that was used for software development .The waterfall Model illustrates the software development process in a linear sequential flow; hence it is also referred to as a linear-sequential life cycle model. This means that any phase in the development process begins only if the previous phase is complete. In waterfall model phases do not overlap.  
**Waterfall Model design:-**

Waterfall approach was first SDLC Model to be used widely in Software Engineering to ensure success of the project. In "The Waterfall" approach, the whole process of software development is divided into separate phases. In Waterfall model, typically, the outcome of one phase acts as the input for the next phase sequentially.



**The sequential phases in Waterfall model are:**  
**Requirement Gathering and analysis:** All possible requirements of the system to be  
developed are captured in this phase and documented in a requirement specification doc.  
**System Design:** The requirement specifications from first phase are studied in this phase  
and system design is prepared. System Design helps in specifying hardware and system  
requirements and also helps in defining overall system architecture.  
**Implementation:** With inputs from system design, the system is first developed in small  
programs called units, which are integrated in the next phase. Each unit is developed and  
tested for its functionality which is referred to as Unit Testing.  
**Integration and Testing:** All the units developed in the implementation phase are  
integrated into a system after testing of each unit. Post integration the entire system is tested for any faults and failures.  
**Deployment of system:** Once the functional and non functional testing is done, the product is deployed in the customer environment or released into the market.  
**Maintenance:** There are some issues which come up in the client environment. To fix those issues patches are released. Also to enhance the product some better versions are released.Maintenance is done to deliver these changes in the customer environment.  
All these phases are cascaded to each other in which progress is seen as flowing steadily  
downwards like a waterfall through the phases. The next phase is started only after the defined set of goals are achieved for previous phase and it is signed off, so the name "Waterfall Model". In this model phases do not overlap.

**Waterfall Model Application**  
Every software developed is different and requires a suitable SDLC approach to be followed based on the internal and external factors. Some situations where the use of Waterfall model is most appropriate are:  
1.Requirements are very well documented, clear and fixed.  
2.Product definition is stable.  
3.Technology is understood and is not dynamic.  
4.There are no ambiguous requirements.  
5.Ample resources with required expertise are available to support the product.  
6.The project is short.  
**Waterfall Model Pros & Cons**  
**Advantage**The advantage of waterfall development is that it allows for departmentalization and control. A schedule can be set with deadlines for each stage of development and a product can proceed through the development process model phases one by one.  
Development moves from concept, through design, implementation, testing, installation,  
troubleshooting, and ends up at operation and maintenance. Each phase of development  
proceeds in strict order.  
**Disadvantage**  
The disadvantage of waterfall development is that it does not allow for much reflection or revision.



\

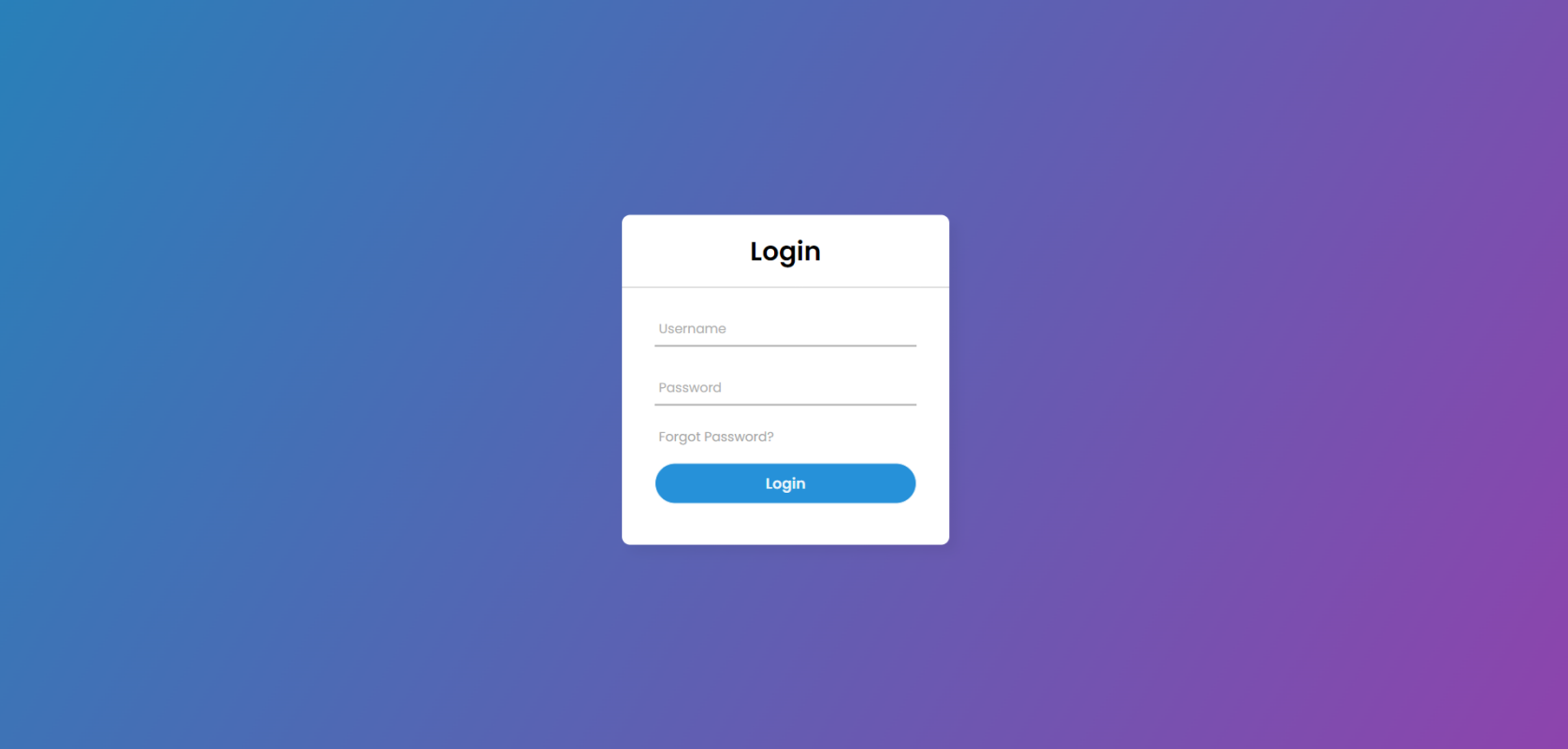
**CHAPTER-8**

**RESULTS AND ANALYSIS**

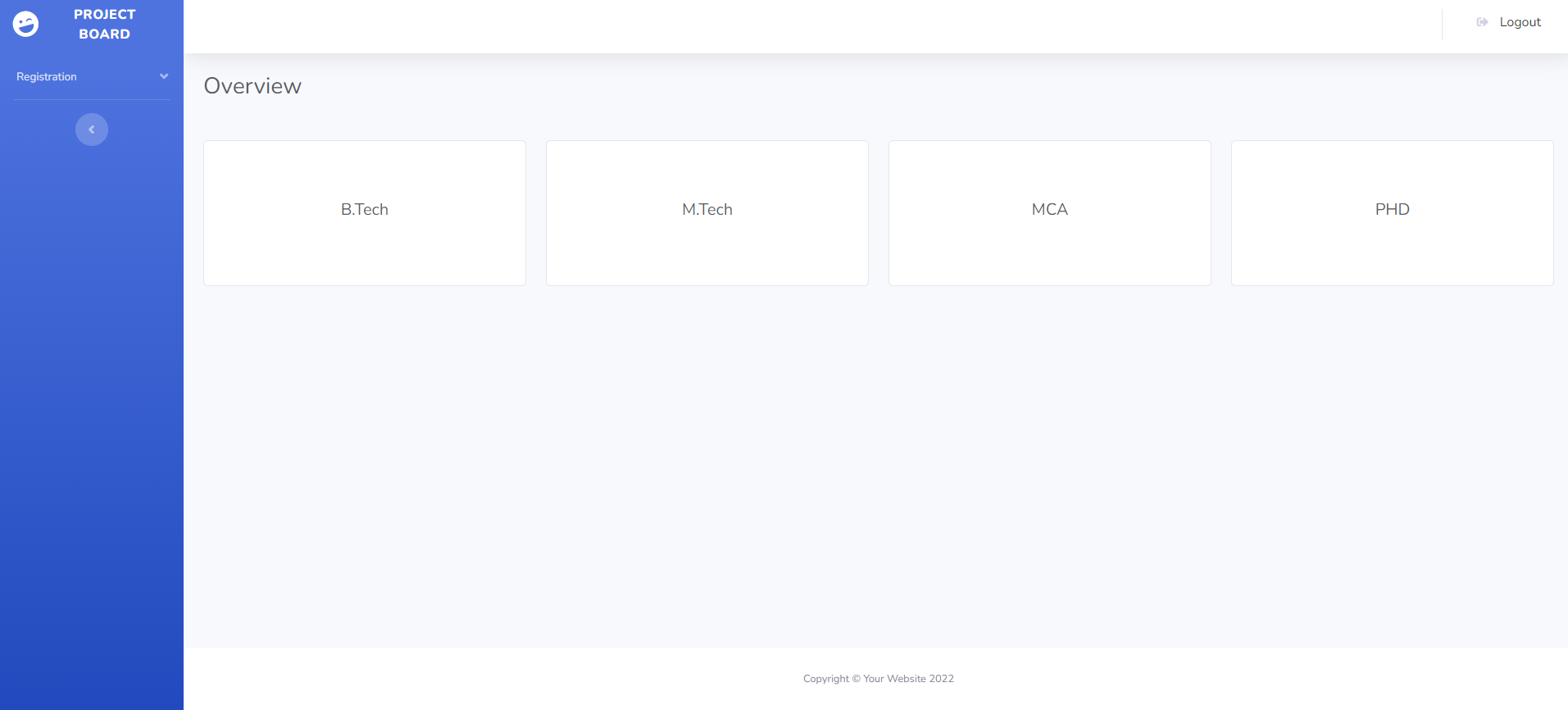
**RESULTS AND ANALYSIS**

**RESULTS**

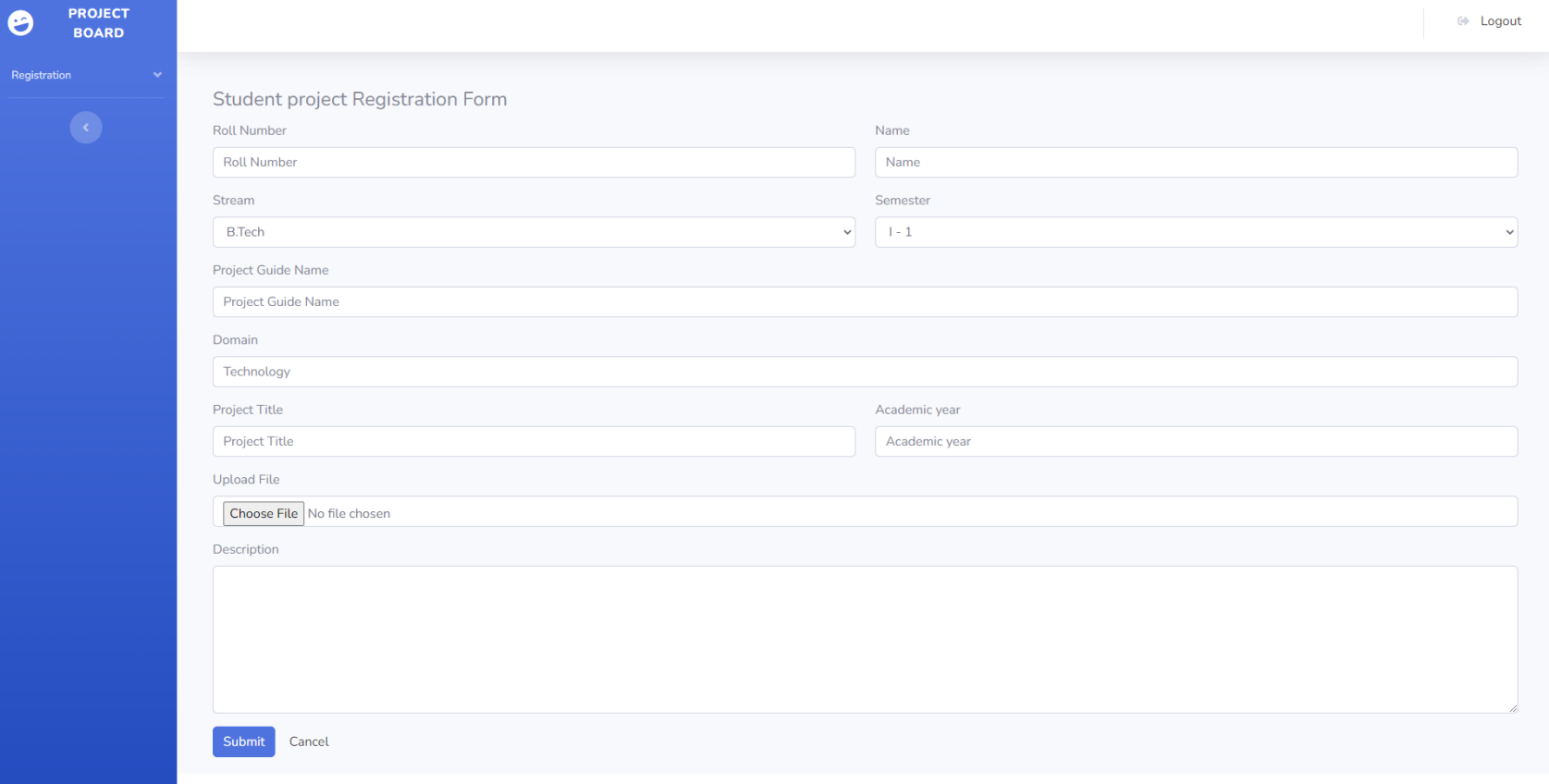
**1)Login page**

****

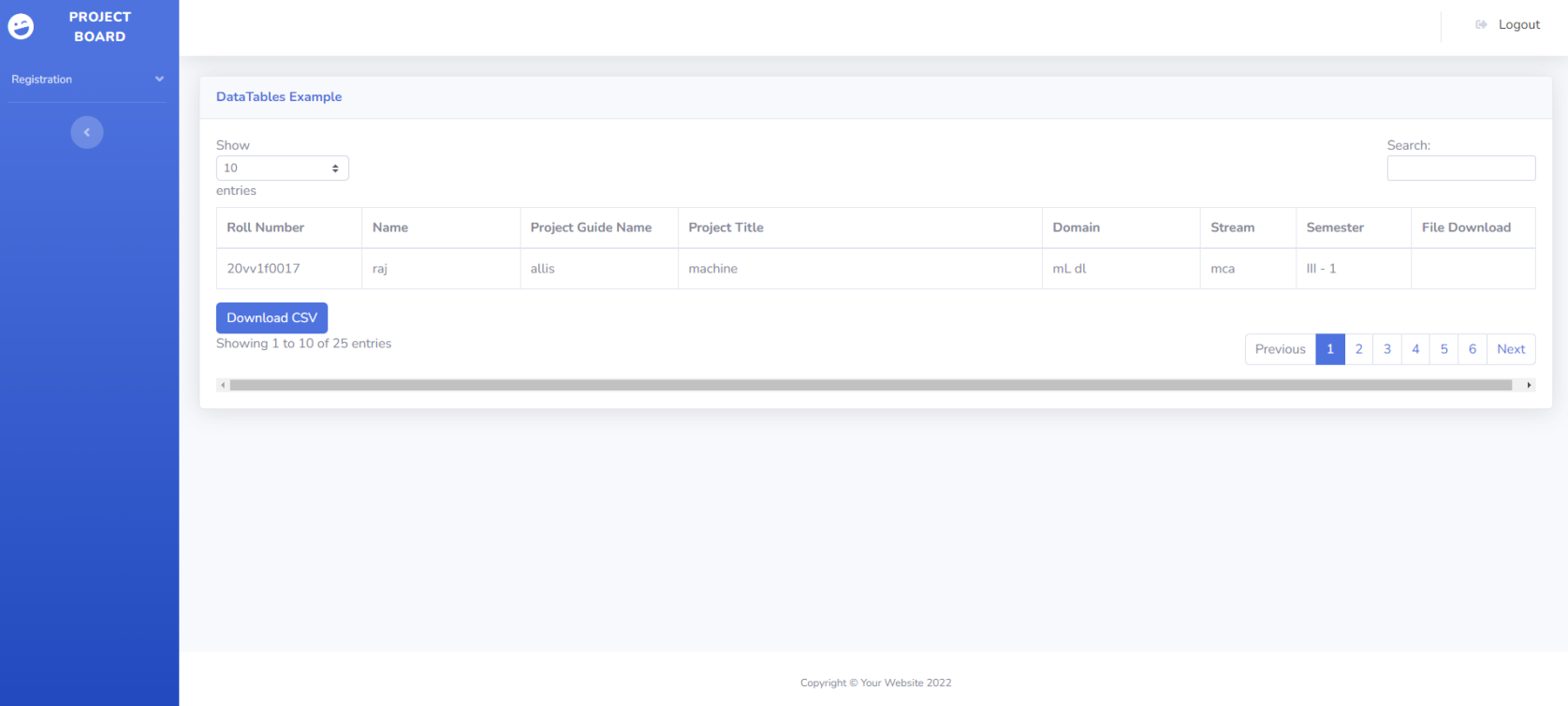
**2)Dash Board**

****

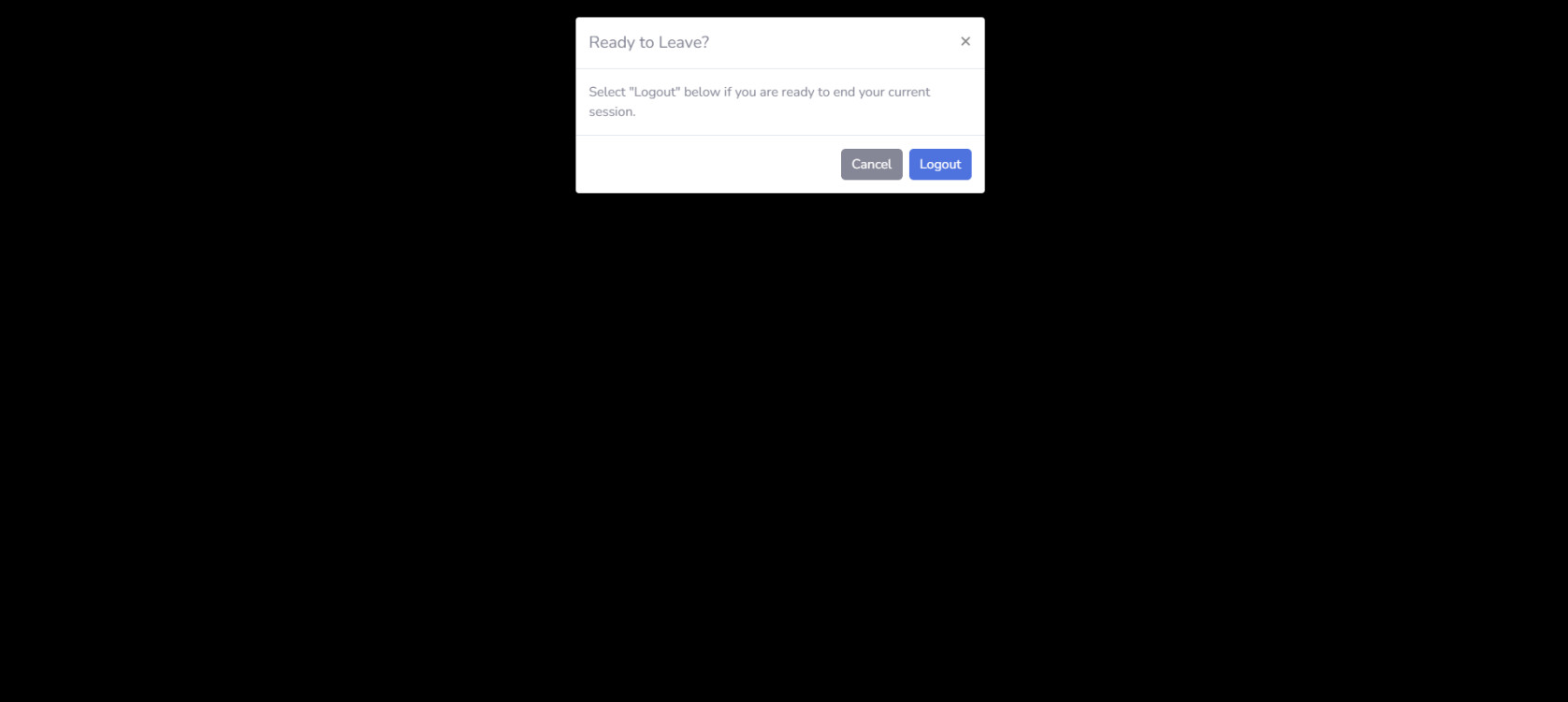
**3)Registration form**



**4)Data Table**



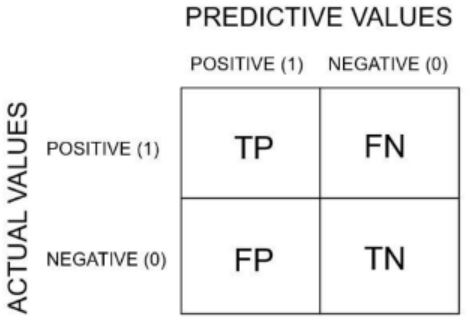
**5)Logout**



These terms are also of extreme importance in Machine Learning. We need them for evaluating ML algorithms or better their results. We will present in this Python Machine Learning four important metrics. These metrics are used to evaluate the results of classifications. The metrics are:

* + - Accuracy
    - Precision
    - Recall
    - F1-Score

We will present each of these measures and go over the advantages and disadvantages of each. Each metric evaluates a classifier's performance in a unique way. The most crucial factor in machine learning will be the metrics.



**Accuracy:**

Since machine learning is of importance to us, accuracy is also utilised as a statistical metric. A classifier's accuracy is measured statistically as the ratio of right predictions (both True Positives (TP) and True Negatives (TN)) to the total of all predictions (including False Positives (FP) and False Negatives (FN)) made by the classifier (FN). Consequently, the equation for calculating binary accuracy.

Accuracy =

**Precision:**

Precision is the ratio of the correctly identified positive cases to all the predicted positive cases, the correctly and the incorrectly case predicted as positive. Precision is the fraction of retrieved documents that are relevant to the query. The formula is:

Precision =

**F1 Score:**

Precision and recall both contribute equally to the F1 score, which can be viewed as a weighted average of these two metrics.

The formula for the F1 score is :

F1 Score =

We have used python to find better accuracy. We have taken the dataset from Spam.csv in kaggle.

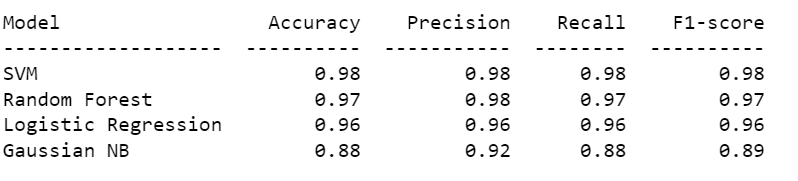


Fig: 15 Output

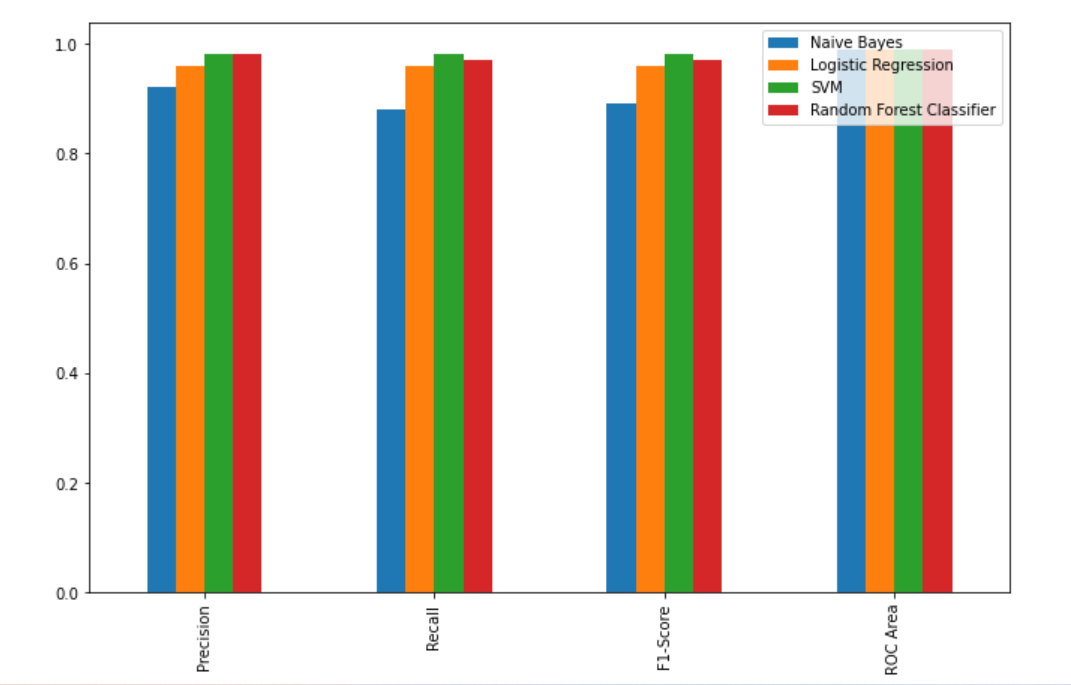


Fig:16 Evaluation metrics Results of bar plot in algorithms

**Accuracy:**

* By using four algorithms i.e., Support vector machines,Random Forest classifier,Logistic Regression and Gaussian Naïve Bayes algorithms.we find the higher accuracy.
* Support Vecor Machines have the heighst accuracy when compared to other algorithms.

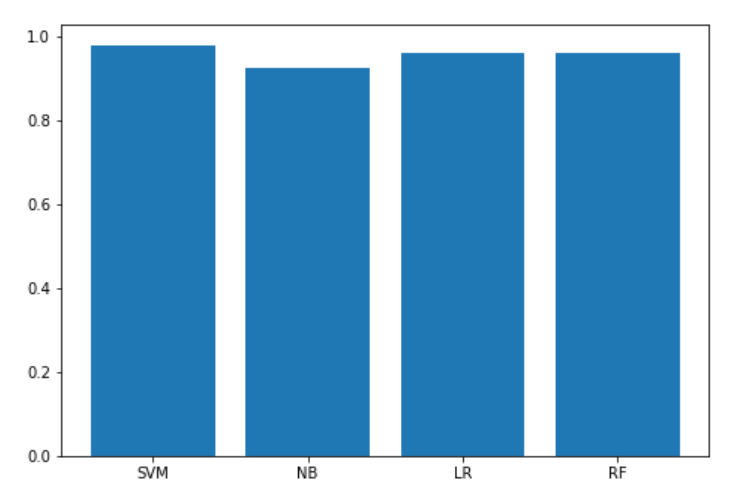


Fig: 17 Accuracy Results of bar plot in algorithms

**CHAPTER-9**

**CONCLUSION AND FUTURE SCOPE**

**CONCLUSION**

Spam email is one of the most demanding and troublesome internet issues in today’s world of communication and technology. is almost impossible to think about email without considering the issue of spam. spammers by generating spam mails are misusing this communication facility and thus affecting organizations and many email user. The machine learning model used by google has now advanced to the point that it can detect and filter out spam and phishing emails with About 98 Percent accuracy by using Support vector machines.