**TASK-PROGESS REPORT**

**IMPLEMENTING WEB APPLICATION**

**CREATE IBM DB2 AND CONNECT WITH PYTHON**

|  |  |
| --- | --- |
| Date | 15 November 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**STEP 1: Import the ibm\_db Python library:**

!pip install --force-reinstall ibm\_db==3.1.0 ibm\_db\_sa==0.3.7   
import ibm\_db

**STEP 2: Identify the database connection credentials:**

dsn\_hostname = "2d46b6b4-cbf6-40eb-bbce- 6251e6ba0300.bs2io90l08kqb1od8lcg.databases.appdomain.cloud"

dsn\_uid = "vjd29721"     
dsn\_pwd = "6TTgx8MRBzT45o3q"   
dsn\_driver = "{IBM DB2 ODBC DRIVER}"   
dsn\_database = "BLUDB" # e.g. "BLUDB"   
dsn\_port = "32328"       # e.g. "32733"   
dsn\_protocol = "TCPIP"# i.e. "TCPIP"   
dsn\_security = "SSL" #i.e. "SSL"

**STEP 3:Create the DB2 database connection:**

dsn = (

"DRIVER={0};"

"DATABASE={1};"

"HOSTNAME={2};"

"PORT={3};"

"PROTOCOL={4};"

"UID={5};"

"PWD={6};"

"SECURITY={7};").format(dsn\_driver, dsn\_database, dsn\_hostname, dsn\_port, dsn\_protocol, dsn\_uid, dsn\_pwd,dsn\_security) print(dsn)

**Now establish the connection to the database:**

conn = ibm\_db.connect(dsn, "", "")

print ("Connected to database: ", dsn\_database, "as user: ",dsn\_uid, "on host: ",dsn\_hostname)

**except:**

print ("Unable to connect:", ibm\_db.conn\_errormsg() )

server = ibm\_db.server\_info(conn)

print("DBMS\_NAME:",server.DBMS\_NAME)

print ("DBMS\_VER:", server.DBMS\_VER)

 print ("DB\_NAME:", server.DB\_NAME)

client = ibm\_db.client\_info(conn)

print("DRIVER\_NAME:",client.DRIVER\_NAME)

print("DRIVER\_VER:",client.DRIVER\_VER)

print("DATA\_SOURCE\_NAME:",client.DATA\_SOURCE\_NAME)

print("DRIVER\_ODBC\_VER:",client.DRIVER\_ODBC\_VER)

print ("ODBC\_VER:", client.ODBC\_VER)

print ("ODBC\_SQL\_CONFORMANCE: ", client.ODBC\_SQL\_CONFORMANCE)

print ("APPL\_CODEPAGE: ", client.APPL\_CODEPAGE)   
print ("CONN\_CODEPAGE:", client.CONN\_CODEPAGE)

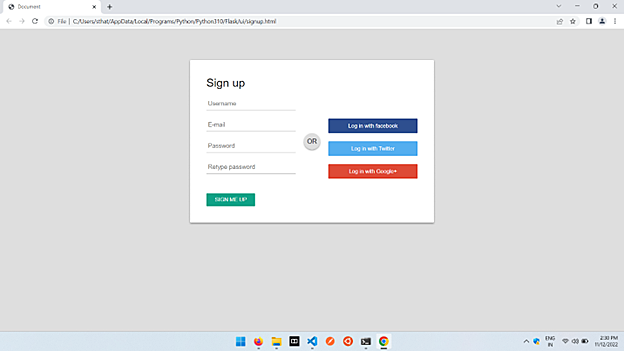
**STEP 4: Close the Connection:**

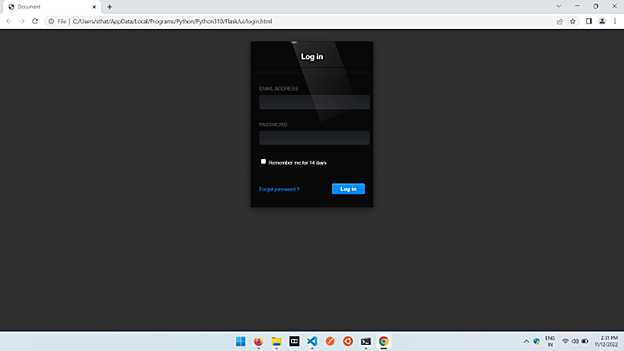
ibm\_db.close(conn)

**IMPLEMENTING WEB APPLICATION**

**Create UI to Interact with Application**

|  |  |
| --- | --- |
| Date | 12 November 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Step 1: Sign up**   


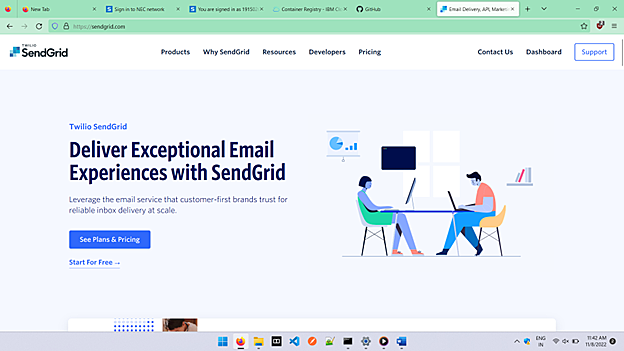
**Step 2: Login Page**  


**SETITNG UP APPLICATION ENVIRONMENT**

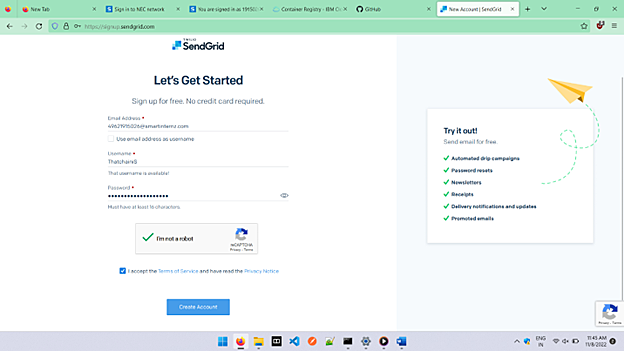
**Create an Account in Send Grid**

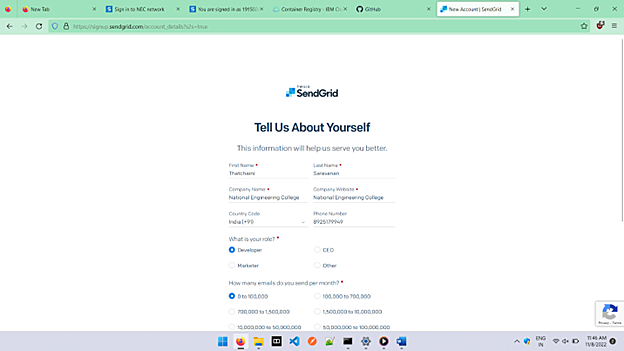
|  |  |
| --- | --- |
| Date | 08 November 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Step 1: Navigate to** [**https://sendgrid.com**](https://sendgrid.com)



**Step 2: Click on start for free and register yourselves by entering required details and click on create account**





**Output:**

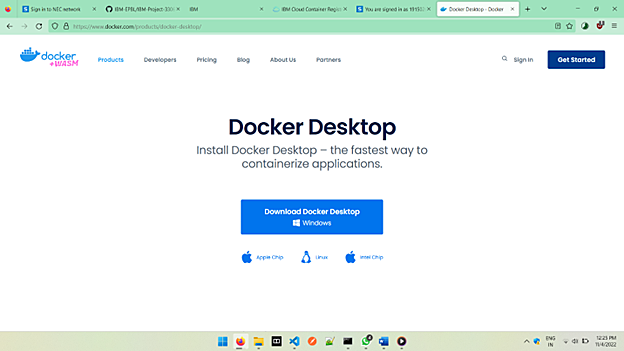
**SendGrid account created successfully.**

**SETITNG UP APPLICATION ENVIRONMENT**

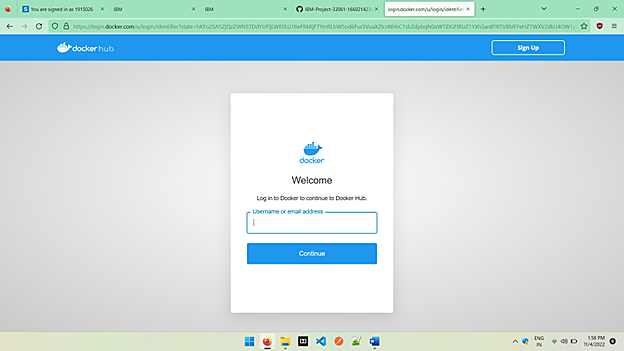
**Docker CLI Installation**

|  |  |
| --- | --- |
| Date | 07 November 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Step 1: Download Docker from docker.com and install it by running the docker Desktop installer.exe file**



**Step 2: Go to hub.docker.com register and create an account and login with the same.**

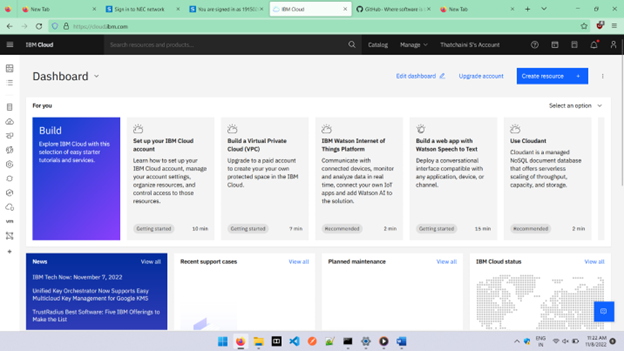


**SETITNG UP APPLICATION ENVIRONMENT**

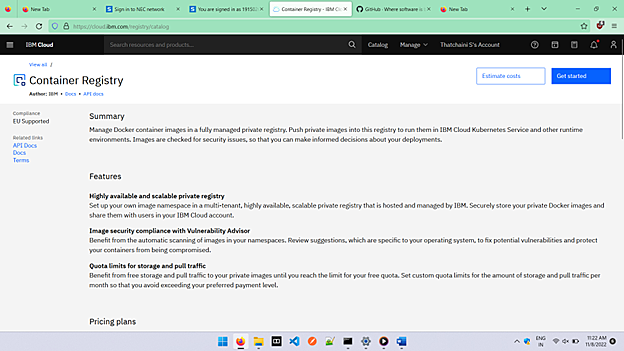
**Install IBM Cloud CLI**

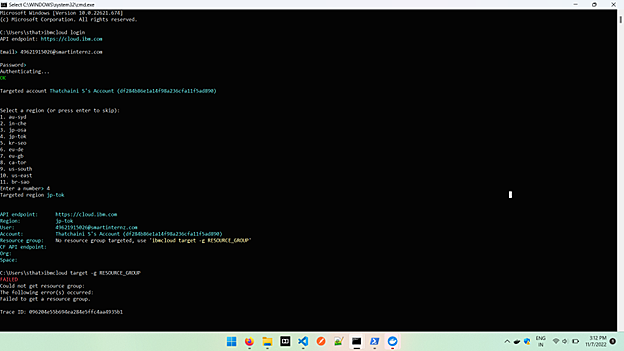
|  |  |
| --- | --- |
| Date | 04 November 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

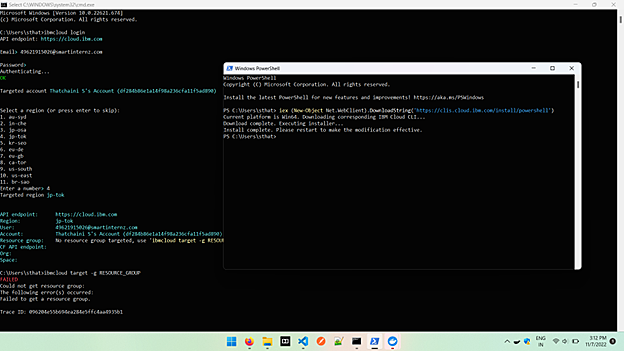
**Step 1:** Navigate to cloud.ibm.com

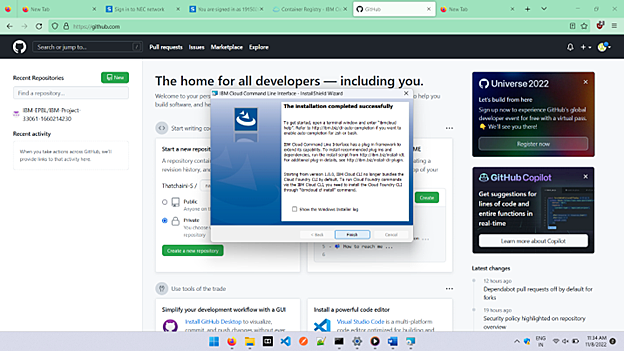


Step 2: Search on Container registry and give get started and follow the steps to install IBM Cloud CLI









**SETITNG UP APPLICATION ENVIRONMENT**

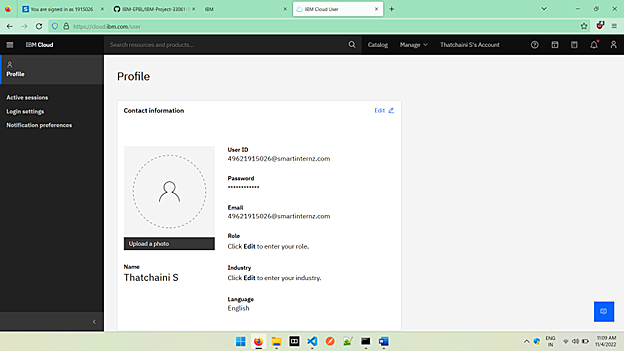
**CREATE IBM Cloud Account**

|  |  |
| --- | --- |
| Date | 04 November 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Step 1: Go to the official website** [**http://cloud.ibm.com/**](http://cloud.ibm.com/)

**Step 2: Sign up using your credentials**

**IBM Cloud Profile:**

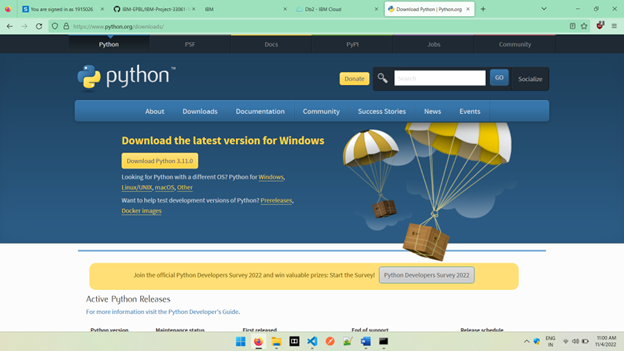


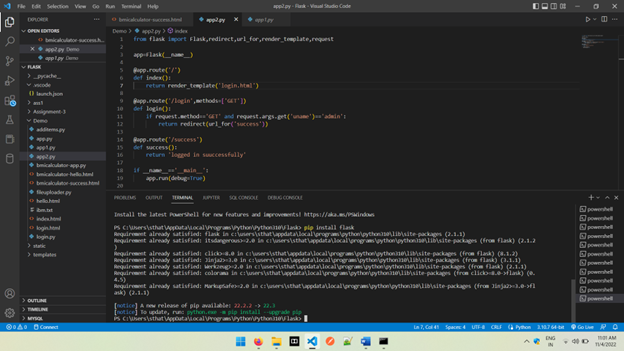
**SETITNG UP APPLICATION ENVIRONMENT**

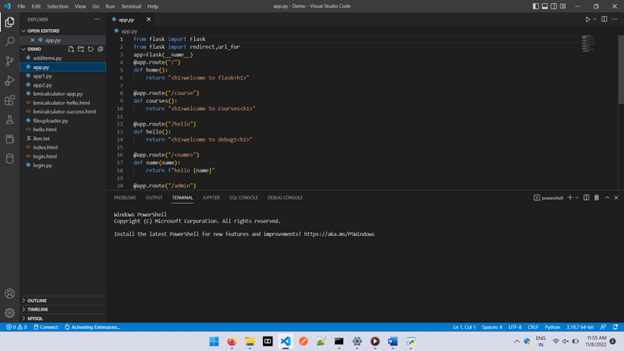
**CREATE FLASK PROJECT**

|  |  |
| --- | --- |
| Date | 04 November 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Step 1: Install Python latest version from python.org**



**Step 2: Install flask using the command pip install flask**



**Project Planning Phase Sprint DeliveryPlan**

|  |  |
| --- | --- |
| Date | 30 October 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Product Backlog,Sprint Schedule, and Estimation (4 Marks)**

    Use the below template to createproduct backlog and sprint schedule.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Functional**  **Requirements (Epic)** | **User Story Number** | **User Story / Task** | **Story Points** | **Priorit y** | **Team Members** |
| S-1 | User Panel | USN-1 | The user will access the website and view theproducts it provides after registering in. | 20 | High | Thatchaini S  Muthulakshmi A  Venkatraman S  Pradeep Rajadurai W |
| S-2 | Admin panel | USN-2 | The administrator's task is to look over the stock database andmonitor on everything  that people are buying. | 20 | High | Thatchaini S  Muthulakshmi A  Venkatraman S  Pradeep Rajadurai W |
| S-3 | Chat Bot | USN-3 | The user can directly talk to Chatbot regarding the products. Get the recommendations based on information provided  by the user. | 20 | High | Thatchaini S  Muthulakshmi A  Venkatraman S  Pradeep Rajadurai W |
| S-4 | final delivery | USN-4 | Container of applications usingdocker kubernetes and deployment the application.  Create the documentation andfinal submit the application | 20 | High | Thatchaini S  Muthulakshmi A  Venkatraman S  Pradeep Rajadurai W |

**Project Tracker, Velocity& Burndown Chart:(4 Marks)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sprint** | **Total Story**  **Points** | **Duration** | **Sprint Start Date** | **Sprint**  **End-Date(Planned)** | **Story Points Completed**  **(as on planned date)** | **Sprint ReleaseDate(actual)** |
| S-1 | 20 | 6 Days | 24 Oct2022 | 29 Oct 2022 |  | 29 Oct2022 |
| S-2 | 20 | 6 Days | 31 Oct2022 | 05 Nov 2022 |  | 05 Nov 2022 |
| S-3 | 20 | 6 Days | 07 Nov 2022 | 12 Nov 2022 |  | 12 Nov 2022 |
| S-4 | 20 | 6 Days | 14 Nov 2022 | 19 Nov 2022 |  | 19 Nov 2022 |

**Velocity:**

Imagine we have a 10-daysprint duration, and the velocityof the team is 20

(Points per sprint). Let’s calculate the team’s average velocity (AV) per iteration unit (story points per day)

|  |  |
| --- | --- |
|  |  |
|  |  |

**PROJECT PLANNING PHASE PREPARE MILESTONE AND ACTIVITY LIST**

|  |  |
| --- | --- |
| Date | 30 October 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Remaining Task:**

|  |  |  |
| --- | --- | --- |
| **MILESTONES** | **ACTIVITY** | **DESCRIPTION** |
| Project development phase | Delivery of Sprint-1,2, 3, 4 | To develop the code and submit the develop the code aftercompletion of testing |
| Implementing web application | Create UI to interact with the application | Create UI   * registration page * login page * view products page * add products page |
|  | Create IBM DB2 and connect with thePython | Create an IBM DB2 in the IBM cloud and connectit to Python. |
| Integrating send grid service | SendGrid integration with the Python | The SendGrid services must be integrated in order for the application to send emails. |
| Developing a chat bot | Building a chat bot and integrate with  The application | Build the chat botand  integrate it to the flask application |
| Deployment of app in IBM cloud | Containerize theapp | Create a docker imageof the application in addition to  push it to the IBM container registry |
|  | Upload imageto IBM container registry | Upload the image to IBM container registry |
|  | Deploy in in Kubernetes cluster | Once the image is uploaded to IBM container registry deploy the image toward IBM kubernetes cluster |

**Completed Tasks:**

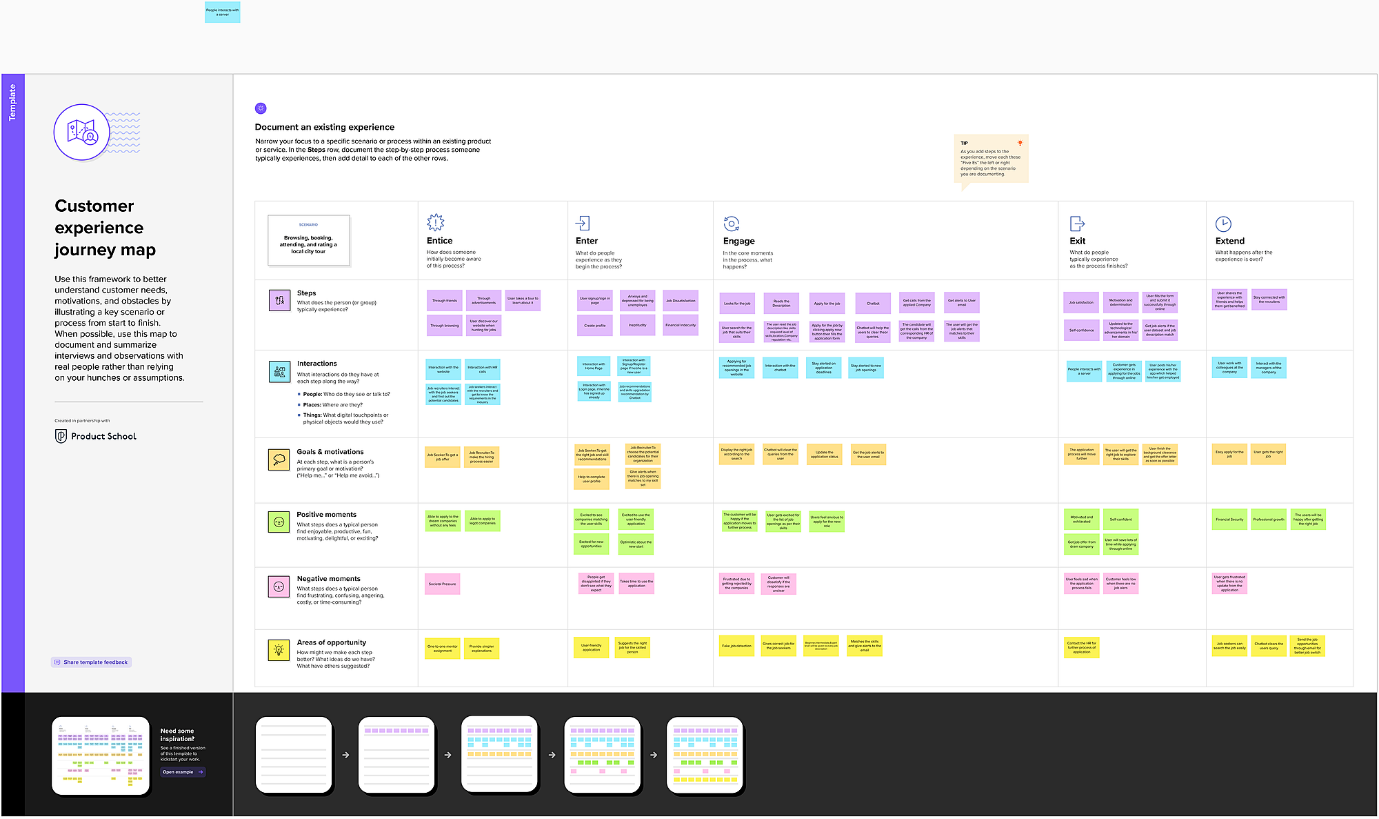
|  |  |  |
| --- | --- | --- |
| **MILESTONES** | **ACTIVITY** | **DESCRIPTION** |
| Ideation phase | Literature survey | Literature surveyon the selected project  and information gathering |
|  | Empathy map | Prepare empathy map to capture the user painand gains,  prepare a list of problem statement |
|  | Ideation | Organizing the brainstorming session and prioritize the top threeideas  based on feasibility |
| Project designphase 1 | Proposed solution | Prepare proposed solution document which includes novelty, feasibility of ideas,  business model,social impact,  scalability of  solution |
|  | Problem solution fit | Prepare problemsolution fit  documents |
|  | Solution architecture | Prepare solution architecture  document |
| Project designphase 2 | Customer journey map | Prepare customer journey map to understand the user interactions and experience  with theapplication |
|  | Functional requirements | Prepare functional and non-functional necessity  document |
|  | Data flow diagram | Prepare dataflow diagram  and userstories |
|  | Technology architecture | Draw technology architecture diagram |

|  |  |  |
| --- | --- | --- |
|  | Sprint delivery plan | Prepare springdelivery  plan |
| Setting-up app environment | Create IBM cloud  account | Sign up IBM cloudaccount |
|  | Create flask project | Getting started with the flask to create project |
|  | Install IBM cloud cli | Install IBMcommand line interface (CLI) |
|  | Docker CLI installation | Installing dockerCLI |

**Project Design Phase-II**

**Customer Journey Map**

|  |  |
| --- | --- |
| Date | 21 October 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

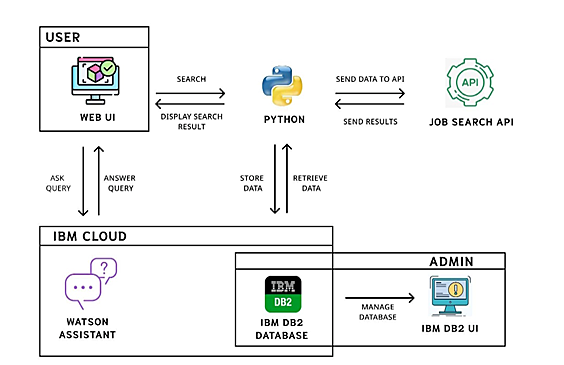
**Project Design Phase-II**

**Technology Stack (Architecture & Stack)**

|  |  |
| --- | --- |
| Date | 20 October 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Technology Architecture:**

The Deliverable shall include the architectural diagram as below and the information as per the table1 & table2



**Table-1: Components & Technologies:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Component** | **Description** | **Technology** |
| 1. | User Interface | How userinteracts with application e.g., Web UI, Mobile App,Chatbot etc. | HTML, CSS, JavaScript, Bootstrap |
| 2. | Application Logic-1 | Logic for a process in the application | Python |
| 3. | Application Logic-2 | Logic for a process in the application | IBM WatsonSTT service |
| 4. | Application Logic-3 | Logic for a process in the application | IBM WatsonAssistant |
| 5. | Database | Data Type,Configurations etc. | MySQL |
| 6. | Cloud Database | Database Serviceon Cloud | IBM DB2, IBM Cloudant etc. |
| 7. | File Storage | File storage requirements | IBM Block Storage or Other Storage Service or Local Filesystem |
| 8. | Infrastructure (Server / Cloud) | Application Deployment on Local System/ Cloud LocalServer Configuration:  Cloud ServerConfiguration: | Local, CloudFoundry, Kubernetes, etc. |

**Table-2: Application Characteristics:**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No** | **Characteristics** | **Description** | **Technology** |
| 1. | Open-Source Frameworks | List the open-source frameworks used | IBM cloud Kubernetes service |
| 2. | Security Implementations | List allthe security / access controls implemented, use of firewalls etc. | e.g., SHA-256, Encryptions, IAM Controls, OWASPetc. |
| 3. | Scalable Architecture | Justify the scalability of architecture (3 – tier, Micro-services) | Technology used |
| 4. | Availability | Justify the availability of application (e.g.,use of loadbalancers, distributed servers etc.) | Technology used |
| 5. | Performance | Design consideration for the performance of the application (numberof requests per sec, useof  Cache, use of CDN’s)etc. | Technology used |

**Project Design Phase-II**

**Data Flow Diagram & User Stories**

|  |  |
| --- | --- |
| Date | 12 October 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Data Flow Diagrams:**

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

|  |  |
| --- | --- |
|  |  |
|  |  |

**Example:**

**User Stories**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User Type** | **Functional**  **Requirement**  **(Epic)** | **User Story Number** | **User Story/ Task** | **Acceptance criteria** | **Priority** | **Release** |
| Customer (Mobileuser) | Registration | USN-1 | As a user,I can register for the application by entering my email, password, and confirming  my password. | I can access my account / dashboard | High | Sprint-1 |
|  |  | USN-2 | As a user, I will receive confirmation email onceI have registered for the application | I can receive confirmation email& click confirm | High | Sprint-1 |
|  |  | USN-3 | As a user, I can register for the application through LinkedIn | I canregister & accessthe  dashboard with Linkedin Login | Low | Sprint-2 |
|  |  | USN-4 | As a user, I can register for the application through Gmail | I can register and access  the dashboard through Gmailalso | Medium | Sprint-1 |
|  | Login | USN-5 | As a user, I can log into the application by entering email& password | I can log on to the  application through email idand password | High | Sprint-1 |
|  | Dashboard | USN-6 | As a user, I can login and chat with the chatbot | Once I logged on the  application I can chat with thechatbot | High | Sprint-3 |
| Customer (Web user) | Registration | USN-7 | As a user, I can log on and register the application for the services being provided | I can access my account / dashboard | High | Sprint-1 |
|  |  | USN-8 | As a user, I will receive confirmation email onceI have registered for the application | I can receive confirmation email& click confirm | High | Sprint-1 |
|  | Login | USN-9 | As a user, I can log into the application by entering email& password | I can log on to the  application through email idand password | High | Sprint-1 |
| Customer care executive | Should Regularize the Send grid service | USN-10 | As a executive and service operator of the service they should make sure that service provided are properly send and received by  the user. |  | High | Sprint-2 |
|  | Should monitor the chatbot regularly whether  working or not | USN-11 | As a executive to provide a quality based service chatbot is important for assisting if anyassistance is needed for the user |  | High | Sprint-2 |

**Project Design Phase-II**

**Solution Requirements (Functional & Non-functional)**

|  |  |
| --- | --- |
| Date | 13 October 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Skill / Job Recommender-Cloud Application Development |
| Maximum Marks | 4 Marks |

**Functional Requirements:**

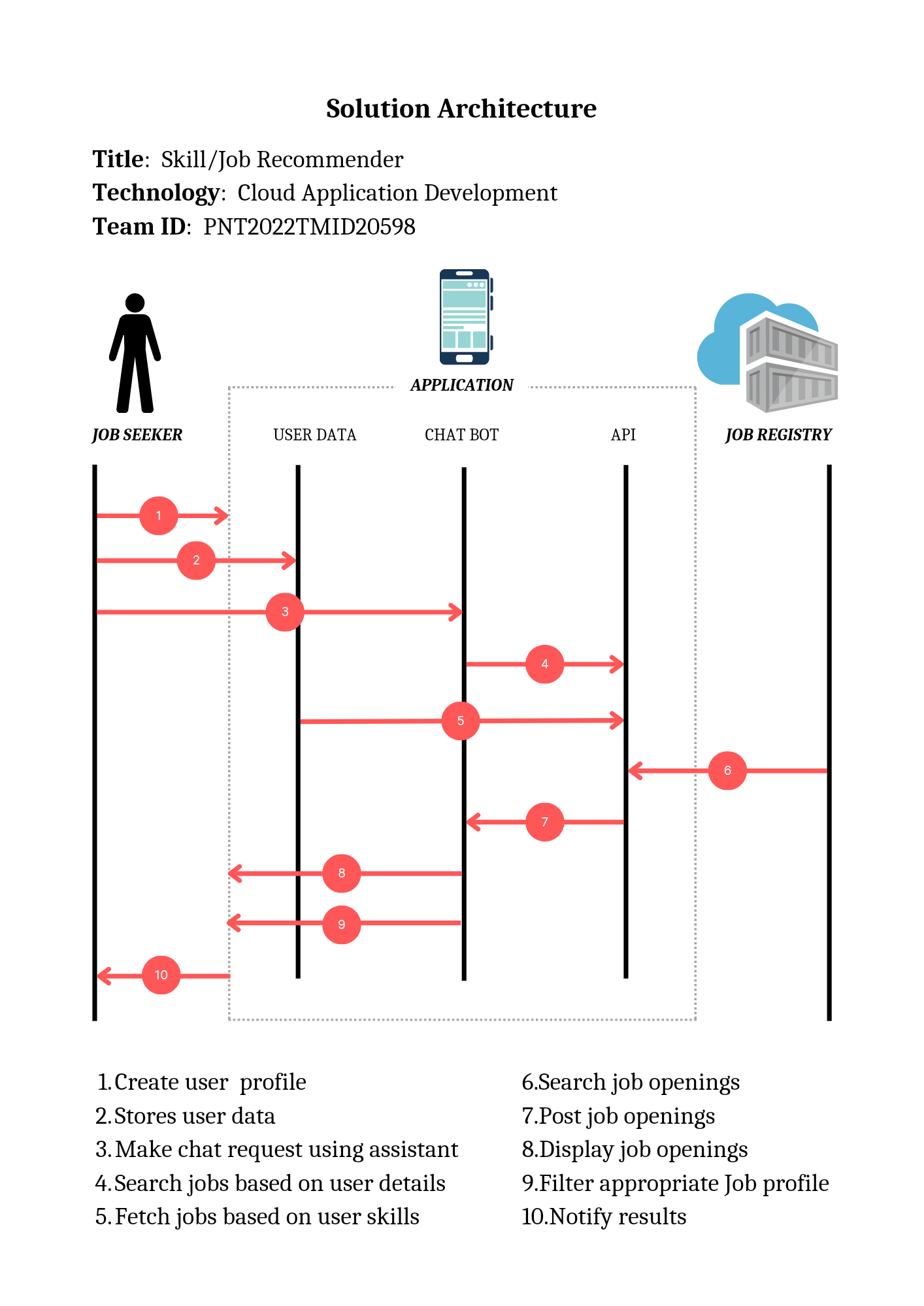
Following are the functional requirements of the proposed solution.

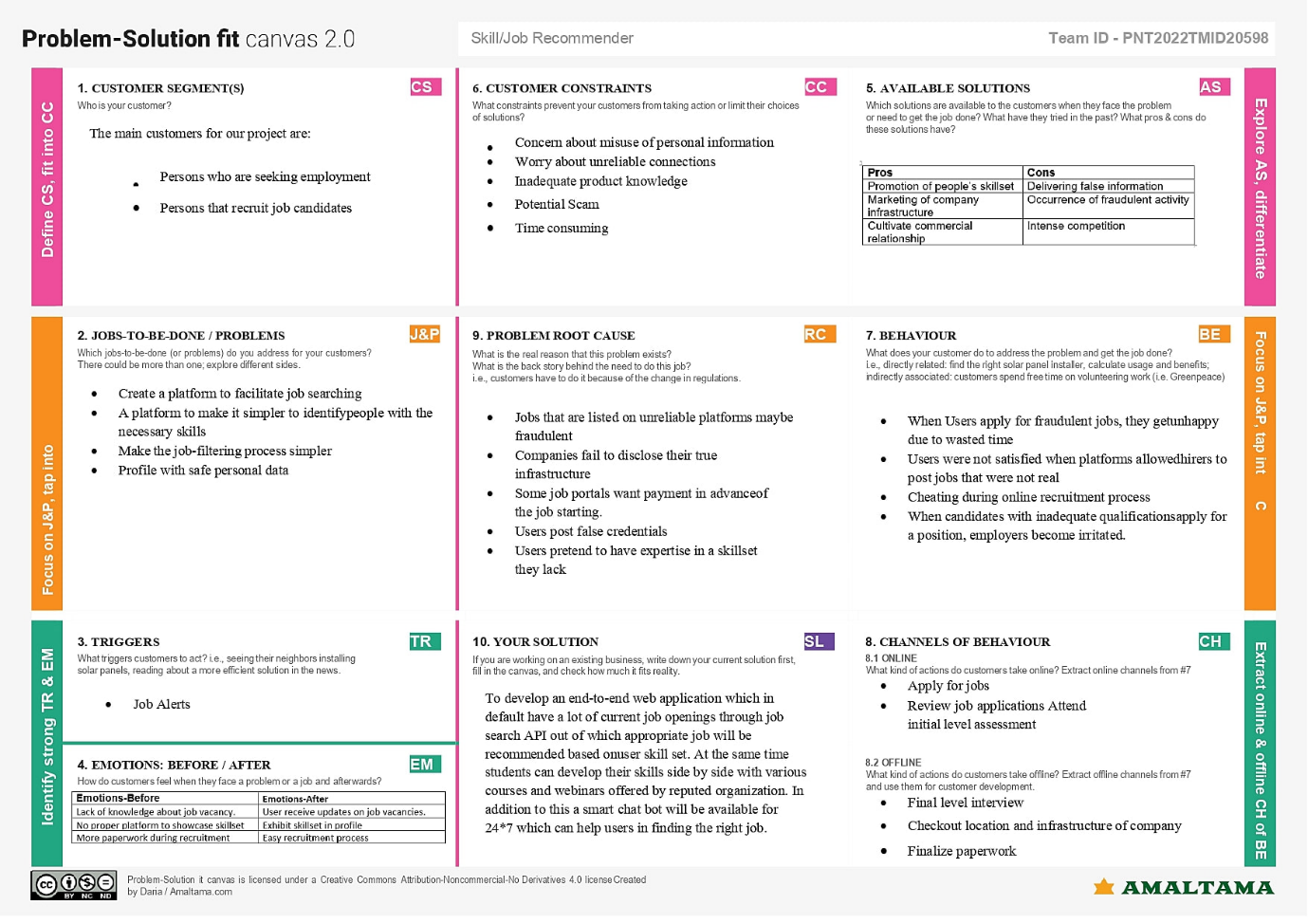
|  |  |  |
| --- | --- | --- |
| **FR No.** | **Functional Requirement (Epic)** | **Sub Requirement (Story / Sub-Task)** |
| FR-1 | User Registration | 1. Registration through Form 2. Registration through Gmail 3. Registration through LinkedIN |
| FR-2 | User Confirmation | 1. Confirmation via Email 2. Confirmation via OTP |
| FR-3 | Job profile display | Display job profiles based on availability, location, skills |
| FR-4 | Chatbot | A chat on the webpage to solve user queries and issue |
| FR-5 | Job registration | Copy of the company the user applied for with its registration/description details will be sent to the registered email id. |
| FR-6 | Logout |  |

**Non-functional Requirements:**

Following are the non-functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Non-Functional Requirement** | **Description** |
| NFR-1 | **Usability** | 1. The webpage will be designed in such a way that any non-technical user can easily navigate through it and complete the job registration work. (Easy and Simple design.) 2. Reduce information overload by generating personalized job suggestions. |
| NFR-2 | **Security** | 1. Using of SSL certificate (Python Flask to Cloud connect) will provide security to the project. 2. Database will be safely stored in DB2. |
| NFR-3 | **Reliability** | To make sure the webpage doesn’t go   down due to network traffic. |
| NFR-4 | **Performance** | 1. Focus on loading the webpage as quickly as possible irrespective of the number of user/integrator traffic. 2. Carry out an evaluation to quantify empirically the recommendation abilities of two state-of-the-art methods, considering different configurations, within the proposed framework |
| NFR-5 | **Availability** | 1. The scraper is set up to avoid duplicate job offers, thus all the job offers are unique. 2. To making the user reliable. This webpage will be available to all users (network connectivity is necessary) at any given point of time. 3. Made publicly available a new dataset formed by a set of job seekers profiles and a set of job vacancies collected from different job search engine sites. |
| NFR-6 | **Scalability** | 1. Increasing the storage space of database can increase the number of users. 2. Add some features in future to make the webpage unique and attractive |



**Project Design Phase-I**

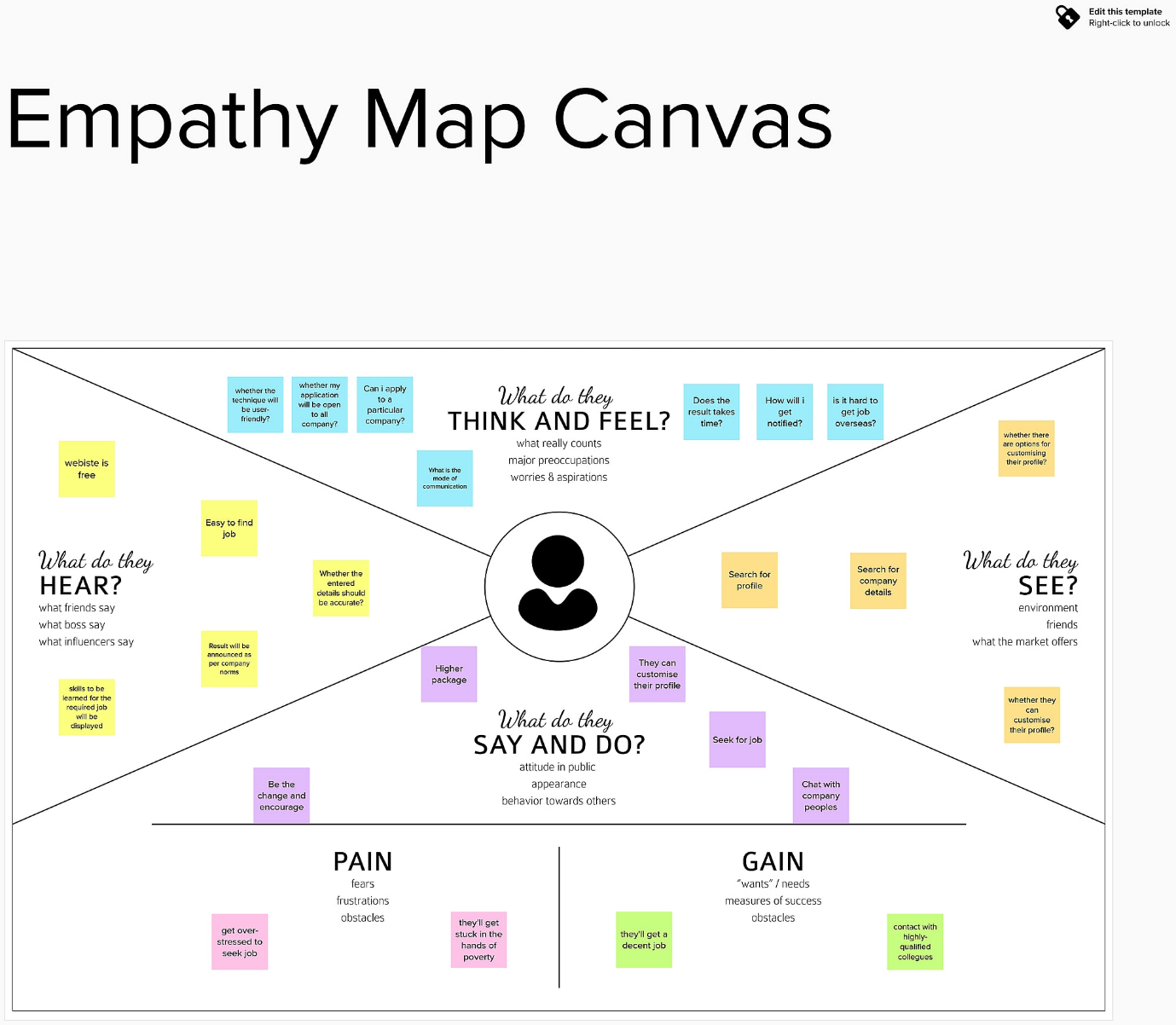
**Proposed Solution**

|  |  |
| --- | --- |
| Date | 24 September 2022 |
| Team ID | PNT2022TMID20598 |
| Project Name | Project – Skill and Job Recommender Application |
| Maximum Marks | 2 Marks |

**Proposed Solution Template:**

Project team shall fill the following information in proposed solution template.

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Parameter** | **Description** |
|  | Problem Statement  (Problems to be solve) | 1. Having better skills but wondering which job will best suits you? 2. We are givingopportunity to job Seekers. 3. User can access large no of data. 4. Having lots of skills but wondering which job will best suit you? Don’t need to worry! We have come up with a skill recommender solution through which the fresher or the skilled person can log in and find the jobs by using the search option or they can directly interact with the chatbot and get their dream. 5. To develop an end-to-end web application capable of displaying the current job openings based on the user skillset. The user and their information are stored in the Database. An alert is sent when there is an opening based on the user skillset. Users will interact with the chatbot and can get the recommendations based on their skills. We can use a job search API to get the current job openings in the market which will fetch the data directly from the webpage. |
|  | Idea/ Solution description | 1. To focuses on fit for feature. 2. To provide user what company expect. 3. Made publicly available a new dataset formed by a set of job seekers profiles and a set of job vacancies collected from different job search engine sites. 4. Put forward the proposal of a framework for job recommendation based on professional skills of job seekers. 5. Carried out an evaluation to quantify empirically the recommendation abilities of two state of the art methods, considering different configurations, within the proposed framework. 6. We thus present a general panorama of job recommendation task aiming to facilitate research and real-world application design regarding this important issue. |
|  | Novelty / Uniqueness | 1. We provide high Data Security. 2. We provide Mobile and computer both platforms. 3. The best position is suggested to any person according to her skills. While the position of known profiles is assumed to be correct, it should be noted that there are usually multiple advisable positions corresponding to a set of skills. A recommendation system should return a set of most likely positions and all of them can be equally valid. 4. The recommendation method we use is simply based on representing both positions and profiles as comparable vectors and seeking for each profile the positions with the most similar vectors. |
|  | Social Impact / Customer Satisfaction | 1. At last, we believe that two people with equal talent should have equal access to opportunity and we’re committed to making this vision reality through our project. 2. We are providing Friendly approach and employability. 3. Students will be benefited as they will get to know which job suits them based on their skills. |
|  | Business Model (Revenue Model) | 1. We are connecting you with other professionals also with companies and recruiters. Along with professionals, it also serves companies and even charges for providing certain premium services. 2. We can provide the application for job seekers in a subscription based and we can share the profiles with companies and generate the revenue by providing them best profiles |
|  | Scalability of the Solution | 1. Scalability is a custom training and organizational development firm dedicated to helping businesses scale. 2. Data can be scaled up and scaled down according to number of current job openings. |

Skill Job Recommender

Literature Survey

* 1. **Introduction**

The recommender system is becoming part of every business. The business tries to increase its revenue by raising the user’s interaction by recommending new items based onuser preferences. We have witnessed the rise of Netflix in the entertainment domain, using their strategies to implement a recommender systeminto their existingecosystem. But therehas been a minimal study in the hiring field from the perspective of a job seeker. To start any research, it is quintessential to review relevantwork in the domain and technology.

**1.2 Recommender Systems**

As discussed previously, RecSys are the system that analyses user preference history and catersthem with differentoptions of servicesrelated to the requirement. Recommender systems emerged as an independent research area in the mid-1990s([Ricci](" \l "_bookmark104) *[et al.](" \l "_bookmark104)*, [2011](#_bookmark104)). In recent years, the interest in recommender systems has dramatically increased. In the Rec- ommendation algorithm, it classifies into four types: Content-based filtering, Collaborative filtering, Rule-based, and Hybrid approaches ([Mobasher](#_bookmark100), [2007](#_bookmark100); [Al-Otaibi and Ykhlef](#_bookmark81), [2012](#_bookmark81)).

Collaborative Filtering (CF): Collaborative Filtering is a technique is based on the human ratings that are given to an item by a user and find similarity between different users who have given similar ratings to an items([Hu and Pu](" \l "_bookmark91), [2011](#_bookmark91)). The essential operation used here is the memory-based nearest neighbor approachto group users who have a similarinterest. As the volume of data grows gradually, there will be high latency in generating recommendations [Mobasher](" \l "_bookmark100) ([2007](#_bookmark100)); [Herlocker *et al.*](#_bookmark89) ([1999](#_bookmark89)). Collaborative filtering has an advantage over content-based filtering techniques, but due to the nature of the hiring process, a job cannot berated by the user and will not be possibleto create a similarity matrix.

Content-based filtering (CBF): These are the most subjective and descriptive based filtering. Content-based filtering can also be called as attribute-based recommender as it uses the explicitly defined property of an item. It is an approach to an information retrieval or machine learning problem. The assumption made in content-based filtering is that user prefersitem with similar properties. Content-based filtering recommends items to the user whose properties are similar to the item which the user has previously shown interest. [Mobasher](#_bookmark100) ([2007](#_bookmark100)) express that drawback of this filtering technique is their tendency to over-specialize in suggesting the item to a user profile as user profiles are relayed on an attribute of the previous item opted by the user. Nevertheless, in the job domain, the job listed in the job board be availableonly for few days; due to the nature of the domain,the tendency to over-specialize in recommending the same item would not be any problem in the job domain recommender system. In domains like entertainment, user preference are tends to change depending on various factors, but In Job domain, the user tends to look for the job where he can use his previous skills. New recommendation of jobs can be made when there is a change in user preference, i.e. if a user thinks to change his/her job domain by updating his new skills and the job domain if he/she wishes. Another scenario of new recommendation is when new jobs are listed in the database;system would identifythe properties of the job listed, such as

job domain and skills requiredfor the job and matcheswith the users with a high similarity score.

Rule-based Filtering (RBF): These filtering techniques depend upon decision rules such as an automaticor manual decisionrule that are manipulated to obtain a recommendation for the user profile. Currently,the E-commerce industryuses a rule-based filtering technique to recommend an item based on the demographic regionof a user, purchase history,and other attributes that can be used to profile an user. A drawback in rule-based filtering is user feeds the information to the system. These inputs are utilized as a description of a user profile or can be considered as a preference of a user, defined by the user. Thus the data acquired is prone to bias. With the age of the user’s profile, recommendation tends to hit the saturation and become static[Mobasher](#_bookmark100)([2007](#_bookmark100)).

Hybrid filtering (HF): As the title describe, its incorporation of multiple techniques to improve the performance of recommendation. The previously discussed recommendation technique has its weakness and strengths. In order to get a better recommendation and overcome the challenges posed by earlier techniques, this technique is sought after. All of the learning/model-based techniques suffer from cold-start in one or other form. It is a problemrelated to handling a new user or new item. These and other shortcomings of the CF,CBF, and RBF could be resolved by using hybrid filtering techniques [Burke](#_bookmark87) ([2007](#_bookmark87)); [Jain and Kakkar](#_bookmark92) ([2019](#_bookmark92)); [Dhameliya and Desai](" \l "_bookmark88)([2019](#_bookmark88)).

The surveys conducted by [Burke](#_bookmark86) ([2002](#_bookmark86)) and [Dhameliya and Desai](#_bookmark88) ([2019](#_bookmark88)) have identified different types of hybridfiltering techniques that could be used by integrating CF, CBF, and RBF.

* + 1. Weighted: The similarity score obtained from different recommendation components are couplednumerically to get one betterrecommendation.
    2. Mixed: Recommendations obtained from different recommending techniques are put together and presented as one recommendation.

* + 1. Switching: choosing one among the recommendation components based on the scenar- ioswhere it suitsbest.
    2. Feature Combination: Attributes derived from diverse knowledge origins are fused and supplied to a recommendation algorithm.
    3. Feature Augmentation: One recommendation technique is used to compute a set of attributes of user or item, which is then part of the input to the next recommenda- tion technique. Two or more recommendation techniques are serialised to get on recommendation.
    4. Cascade: Recommending systems are given strict priority, with the lower priority ones breaking ties in the scoring of the higher ones. Here one Recsys technique refines recommendation of another.

There had been attempts to develop a recommendation system by several researchers. One such implementation was done by [Rafter *et al.*](#_bookmark102) ([2000](#_bookmark102)). They had devised a hybrid Recsys CASPER for Job finding search engine. They had implemented an automated collaborative filtering module and personalized case retrieval modulein their job recommendation system.ACF module utilized user behavior information such as read time and activity on the page during his time on the system to profile the user. Similarity measure such as the Jaccard index and other clustering algorithms was used for similar groupinguser against targetuser. Their other module PCR finds the similarity betweenthe user’s query and jobs in the system. The module computes similarity with a target user’s query and jobs from the job case base using different similarity measures. This systemhas faced sparsityand scalability problems.

* 1. **Natural languageprocessing**

These are the times that can be considered as an era of data. Every keystroke hit on twitter,online news, or in a research paperis recorded somewhere on the internet. All these generated data are availablefor the analysis through many means. In this abundanceof data, Text data holds the majority of the share. Most of these text data are in an unstructured form. To put the abundance of text data into a perspective, a trillion-plus query per year is being handled by Google, and Whatsapp handles 30+ billion messages per day. That being said, how do we extract information from the unstructured text data or how can we make machine understand what the text is about? To answer all the questions, Text analysis is a most sought after technique to extract useful information from the text data. Text analysis can be performed by utilizing techniques such as Natural language processing. Natural languageprocessing is a process of information retrieval from unstructured data. It refersto the utilization of computers to process natural language([Brants](" \l "_bookmark84), [2003](#_bookmark84)). The advancement in the personal assistant, text summarizing, and methods to caption a subject is due to the successful research in the field of NLP. Search engines like google and other industry leaders utilize NLP to its full extent. The gap between industry and academia in the field of NLP is very minimal as there is an advancement in the NLP; the business has tried implementing and has broughtcloser to everyone’s life.

In Recsys for the hiring domain, the data we handle here is nothing other than text data. A user profile describes the details about user experience and skills he/she familiar with. On the other hand, the job listed has information as job title, skills required to fulfill the role. All these information is filled with text data. In this scenario, we utilize the Natural Language Processing to measure the similarity between Jobs by checking the similarity between the job title and job description of the listed job. Determining the text-similarity is an essential task in several industrial application such as query search, text summarizing and video tagging([Kenter and De Rijke](" \l "_bookmark95), [2015](#_bookmark95)). In earlierstudies, researchers have used

different approaches to identify similarity between the text by using edit distancealgorithm which is discussed by [Mihalcea *et al.*](#_bookmark98) ([2006](#_bookmark98)), lexical overlapping technique ([Jijkoun](" \l "_bookmark93) *[et al.](" \l "_bookmark93)*, [2005](#_bookmark93)) as this might work in most cases but can’t rely on these technique because of its frail nature([Kenter and De Rijke](" \l "_bookmark95), [2015](#_bookmark95)). In such cases, we rely on technique called word embedding. This is huge development in the field of distributional semantics. As this requires only a large amountof unlabelled word data. These words are represented in semantic spaceas a vector. That is, words that are semantically similar will stay close in the semantic space.In order to retrieve terms that based on the similarity between two terms, we can utilize most well know method called word2vec a vector space model then we can use cosine similarity to measure the similarity between them ([Shrestha](#_bookmark108), [2011](#_bookmark108); [Barrón-Cedeno *et al.*](#_bookmark82), [2009](#_bookmark82)). Thismodel can also be used to determine similarity between the sentences([Barzilay and Elhadad](" \l "_bookmark83), [2003](#_bookmark83)). It’s a group related model which is used to produce word embedding and these are set of languagemodelling and featurelearning techniques of NLP where words are mapped to real values in the vector. Typically word2vec takes large set of words which is called corpus as a input and producesvector space with dimensions being in hundreds([Mikolov](#_bookmark99) [*et al.*](#_bookmark99), [2013](#_bookmark99)). Once vector space model is generated we can use similarity measuringmethod to determine the distance or how similar is the word with which we are comparing. To find similarity in vector spacewe can use similarity measureslike Cosine similarity and Jaccard similarity.

* + 1. **Jaccard Coefficient**: Jaccard Coefficient is a methodto compare elementsof two sets to identify which elements are shared between two sets and which are distinct. It’s similarity measure for two sets of data with result ranging from 0% to 100%. Two sets can be said similar, when result is close to 100% . Formula for Jaccard Index is as shown below([Sternitzke and Bergmann](" \l "_bookmark110), [2009](#_bookmark110)),

* + 1. **Cosine similarity**: Cosine similarity is also a measure to find similarity between two sets of non zero vector. It is a weighted vector space model utilized in the process of information retrieval. The similarity is measured by using euclidean cosine rule,i.e., by taking inner product space of two nonzero vector that measures the cosine of the anglebetween the two vectors. If the angle between two vectors is 0 deg , then the cosine of 0 is 1; Meaning that the two non zero vectors are similar to each other.In order to weight the words we have used the well-known word2vec vector space model([Rong](#_bookmark106), [2014](#_bookmark106); [Herremans and Chuan](#_bookmark90),[2017](#_bookmark90)).

**2.2        Inferences**

Based on all the researchmethodologies and techniques reviewed in this chapter, the CF technique cannot be considered as it does not satisfythe aims of the research. As the datasetof the user does not hold the information of rating againsta particular job, we will not be able to create a rating matrix that requires for CF technique. Instead, I have chosen to implement content-based filtering. I used multiple attributes in the user data to create a user profileand recommend the job to those profileswhich have a high similarity score received from cosine similarity. Also, i have given higherweights to job skills when compared to the job domain of the user while computing similarity scores between user profile and job.