Job Recommender System User Guide

Group 7 Member

1. Daniel Tan
2. Leonard Loh
3. Aaron Kueh Hee Kheng

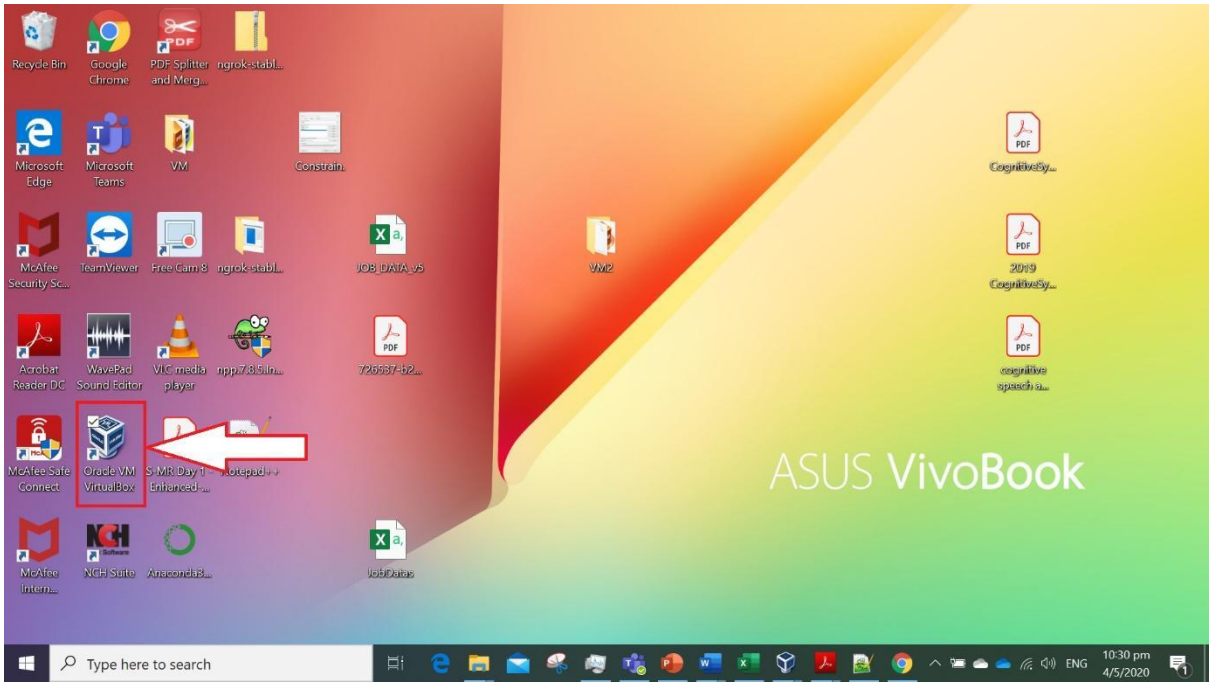
ISS-VM Installation

Requirements:

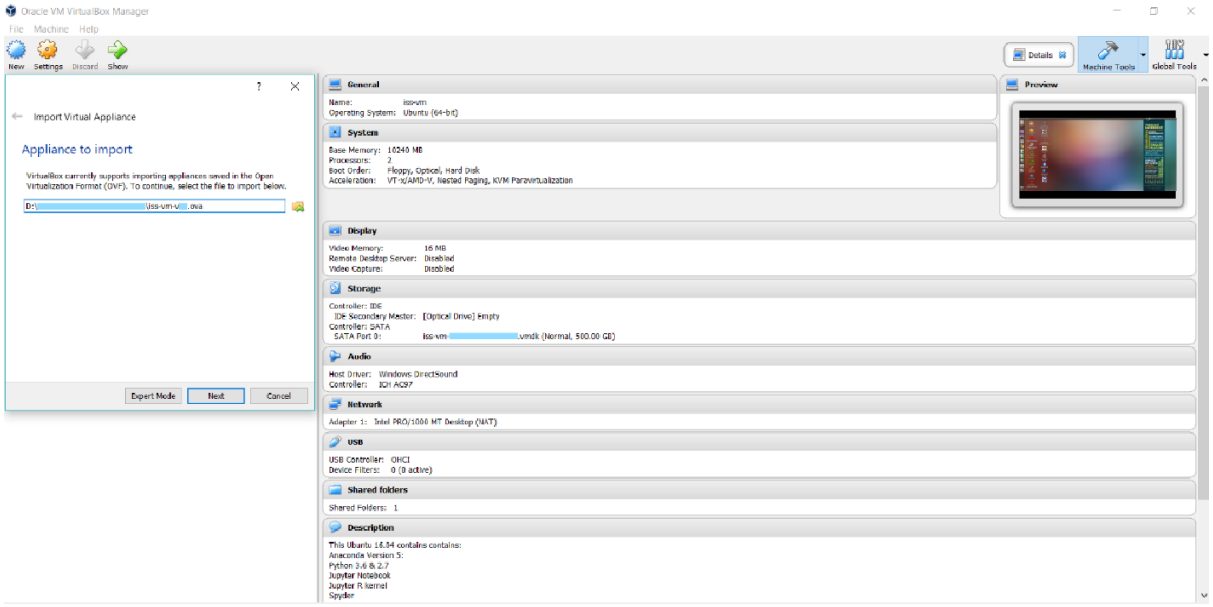
- ISS-VM Ubuntu 16.04
- Tool KIE 7.12
- Google Chrome

Procedure to install ISS-VM:

1. Download and install Virtualbox software (recommended version 5.2.20): https://www.virtualbox.org/wiki/Downloads
2. Download iss-vm virtual machine (an Appliance) from:
    1. http://bit.ly/iss-vm-v20a (part 1 about 11 GB in file size)
    2. http://bit.ly/iss-vm-v20b (part 2 about 11 GB in file size)
    3. http://bit.ly/iss-vm-v20c (part 3 about 10 GB in file size)
3. **[Note] Please check/ensure the 'virtualization' option is enabled in your computer's BIOS/hardware**
4. Put all three zip files in the same folder; select the first file iss-vm-vNN.zip.001. Use tools like 7-zip to unzip the folder. (https://www.7-zip.org/download.html)
5. Start Virtualbox software



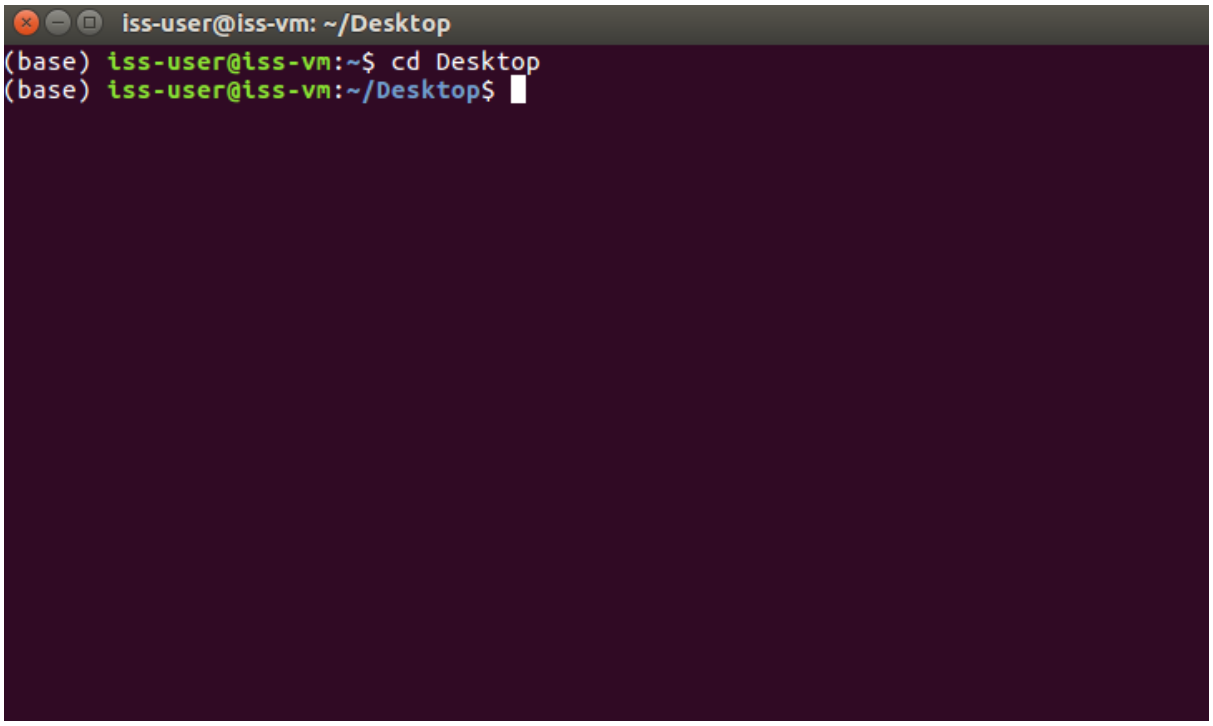
6. Click File ->Import Appliance



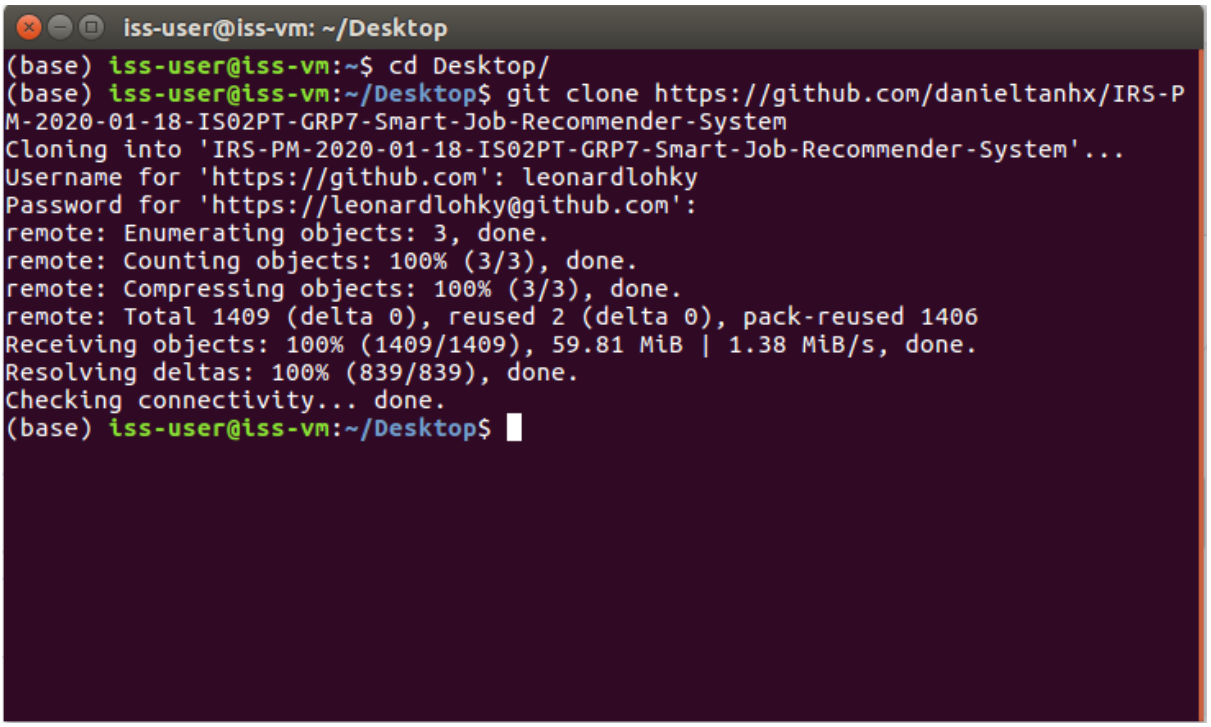
7. Click “Start” to launch iss-vm

Deploying in KIE jBPM 7.12

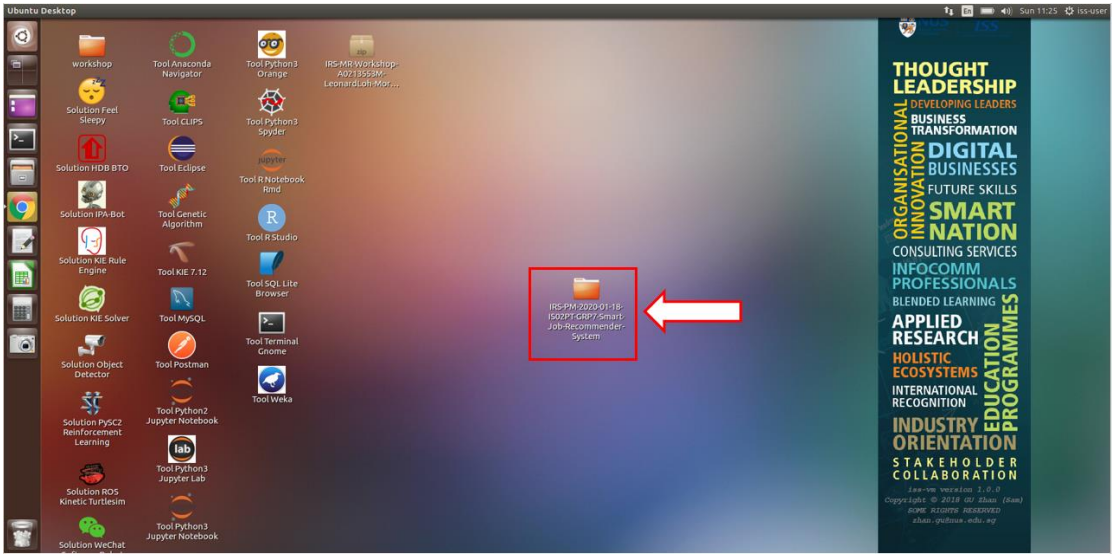
1. Open a new terminal. Navigate to the Desktop by typing “cd Desktop” in the terminal



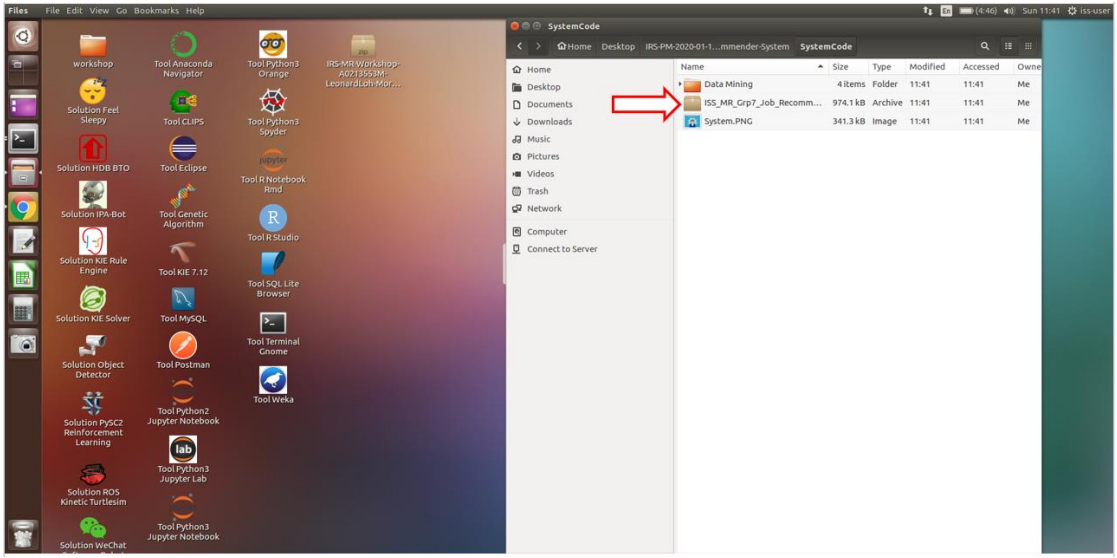
2. Clone the Git repository by typing “git clone https://github.com/danieltanhx/IRS-PM-2020-01-18-IS02PT-GRP7-Smart-Job-Recommender-System” in the terminal



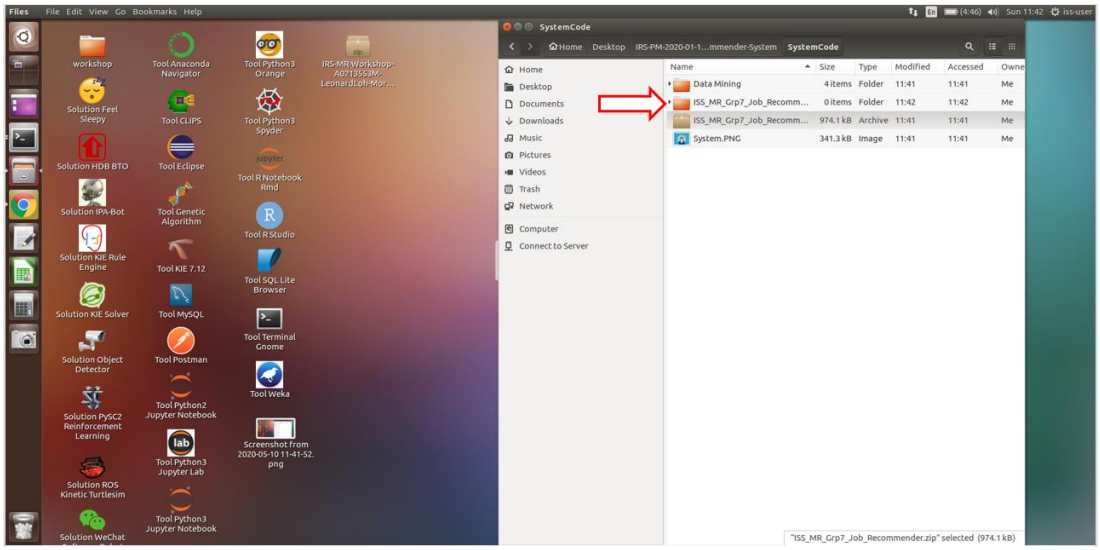
3. The cloned repository will appear on the Desktop screen as a folder named “IRS-PM-2020-01-18-IS02PT-GRP7-Smart-Job-Recommender-System”



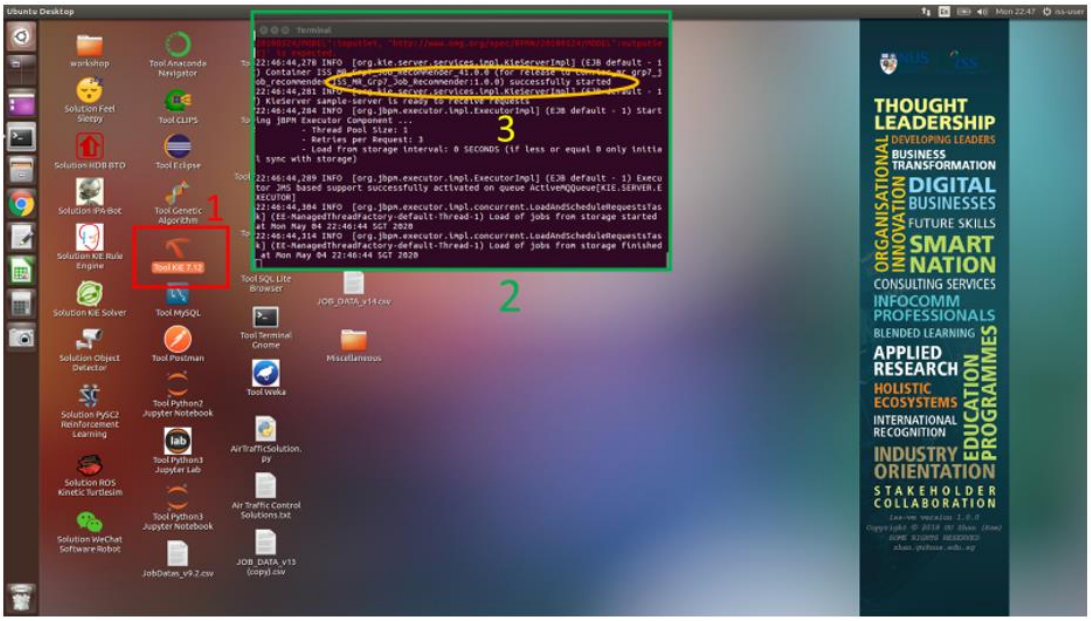
4. Enter the “IRS-PM-2020-01-18-IS02PT-GRP7-Smart-Job-Recommender-System” folder and navigate to the “SystemCode” subfolder. You will find a ZIP folder named “ISS_MR_Grp7_Job_Recommender.zip”



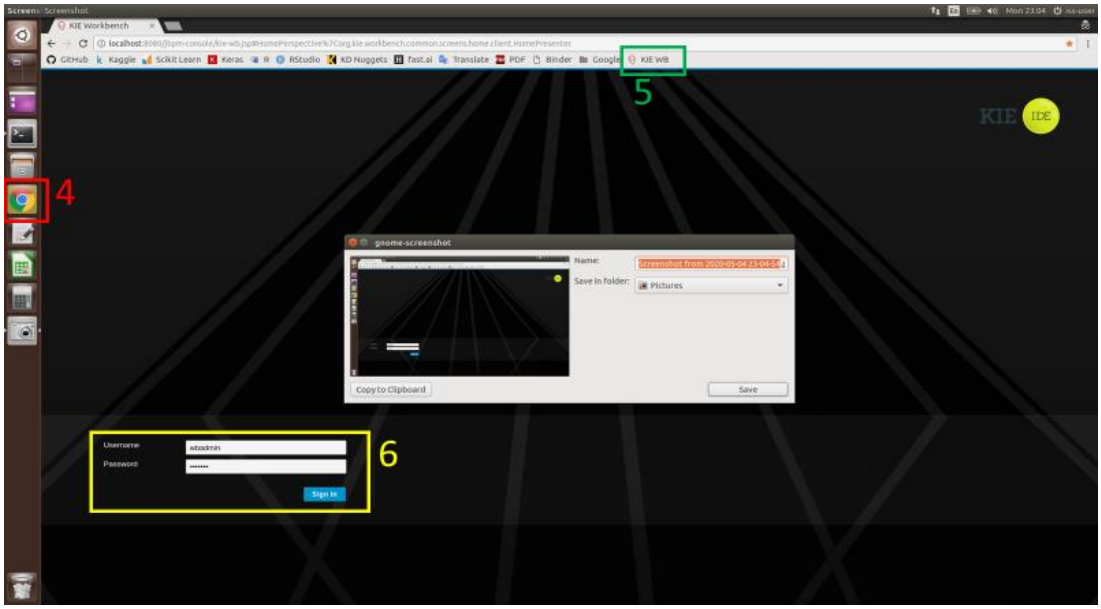
5. Extract the contents of the folder to the same location where the ZIP folder is located



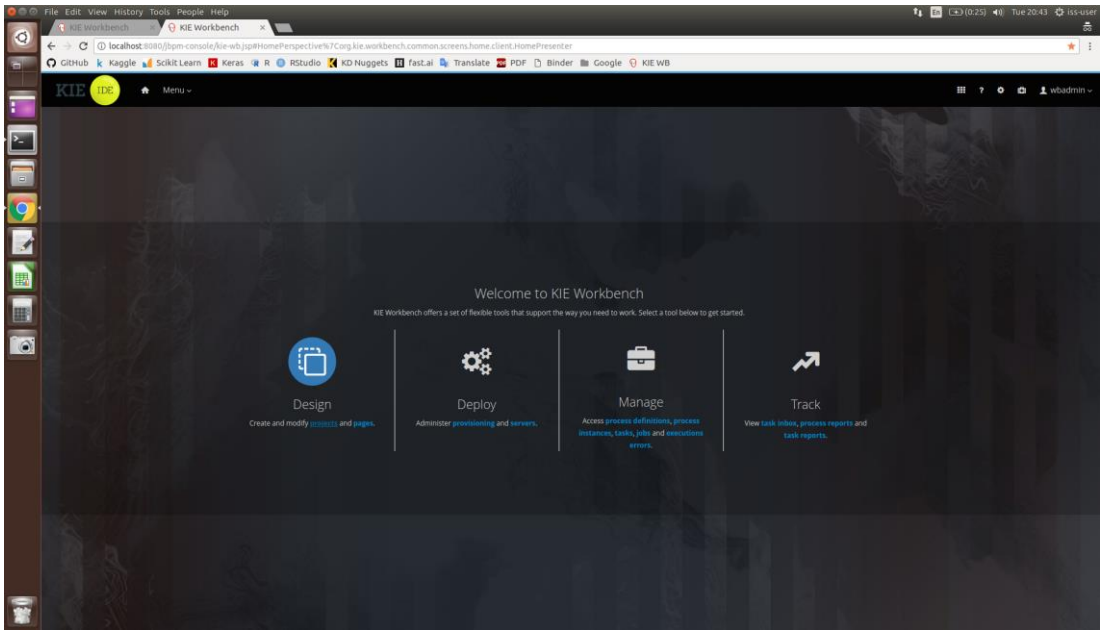
6. Click on Tool KIE 7.12, then wait for the terminal to pop out "successfully started" or "localhost:8080 successfully registered kind of words".



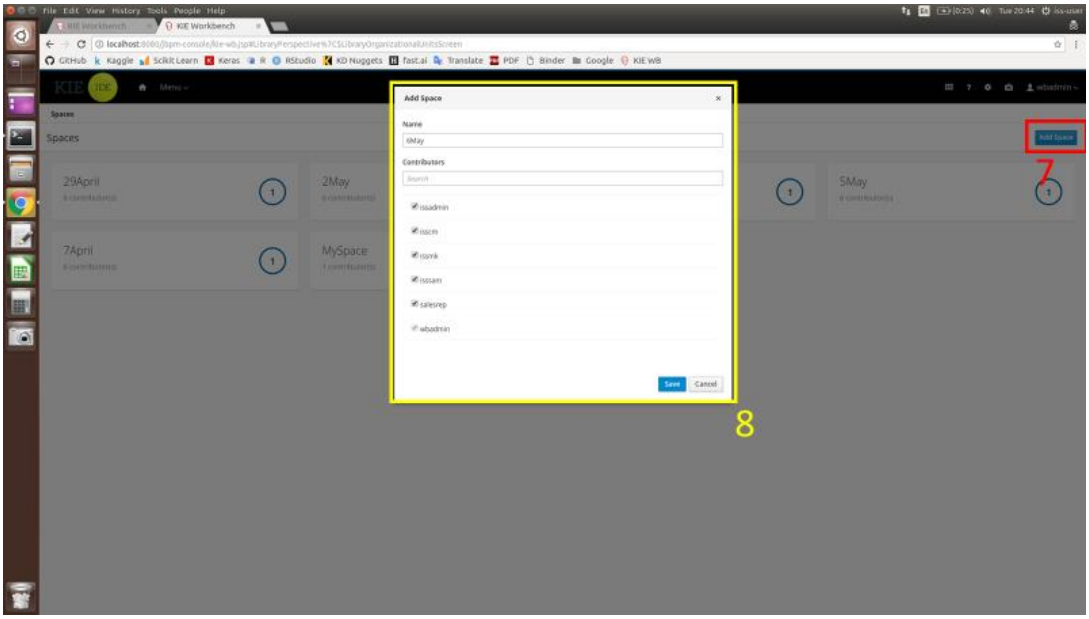
7. Launch Google Chrome and click on the KIE WB bookmark (labelled in Box 5). Once loaded, the below screen will appear. Sign in using username and password “wbadmin” and “wbadmin” respectively.



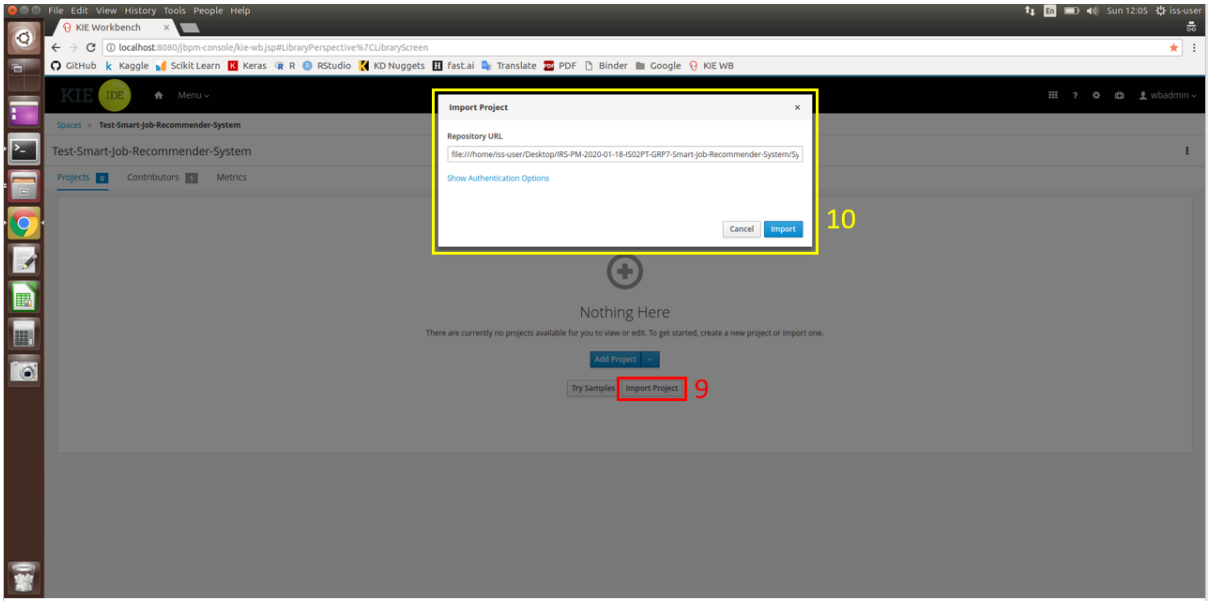
8. Click on Design



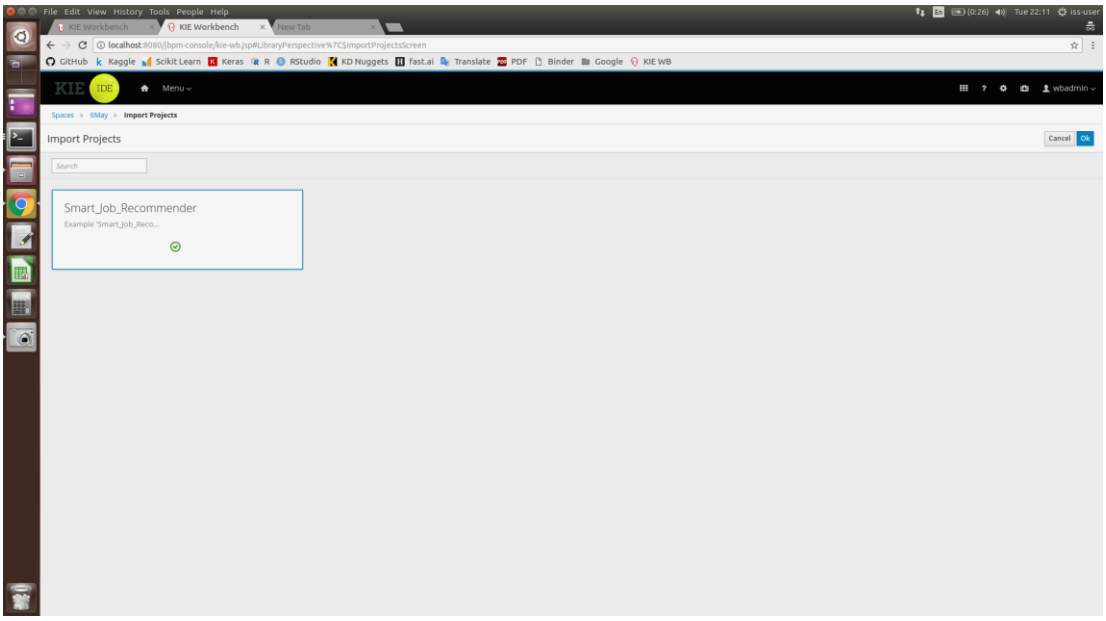
9. Press on the “Add Space” button (labelled in Box 7), then fill in the name of the space accordingly and click the necessary contributors. After all these are completed, press “Save”



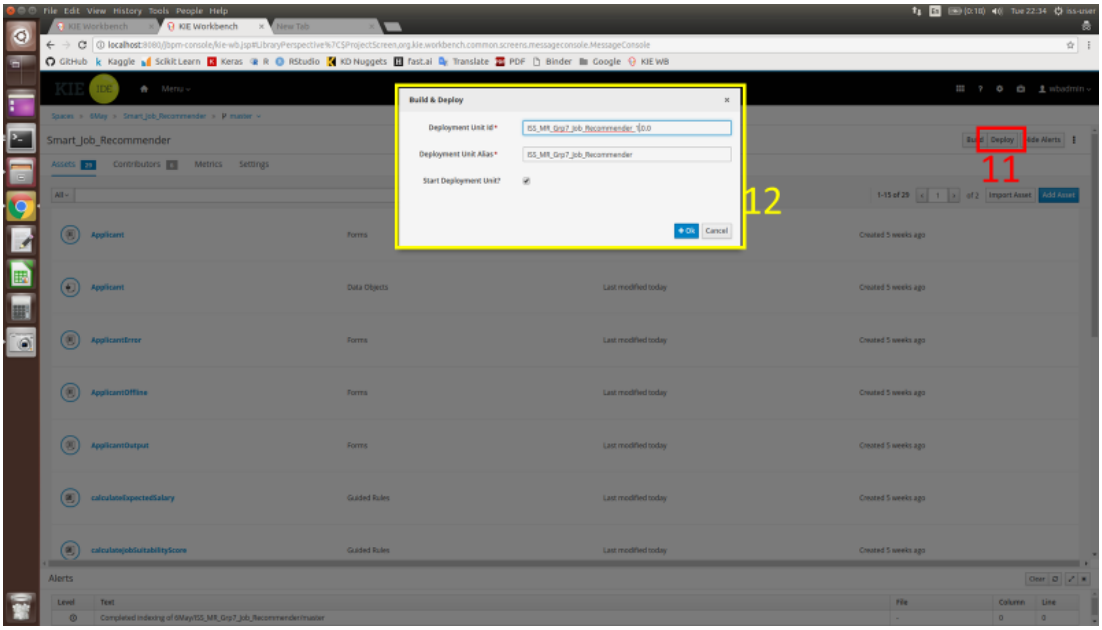
10. In the newly created space in KIE, click on “Import Project” (labelled in Box 9). Type “file:///home/iss-user/Desktop/IRS-PM-2020-01-18-IS02PT-GRP7-Smart-Job-Recommender-System/SystemCode/ISS_MR_Grp7_Job_Recommender” into the repository URL as shown in Box 10. Press “Import” after that



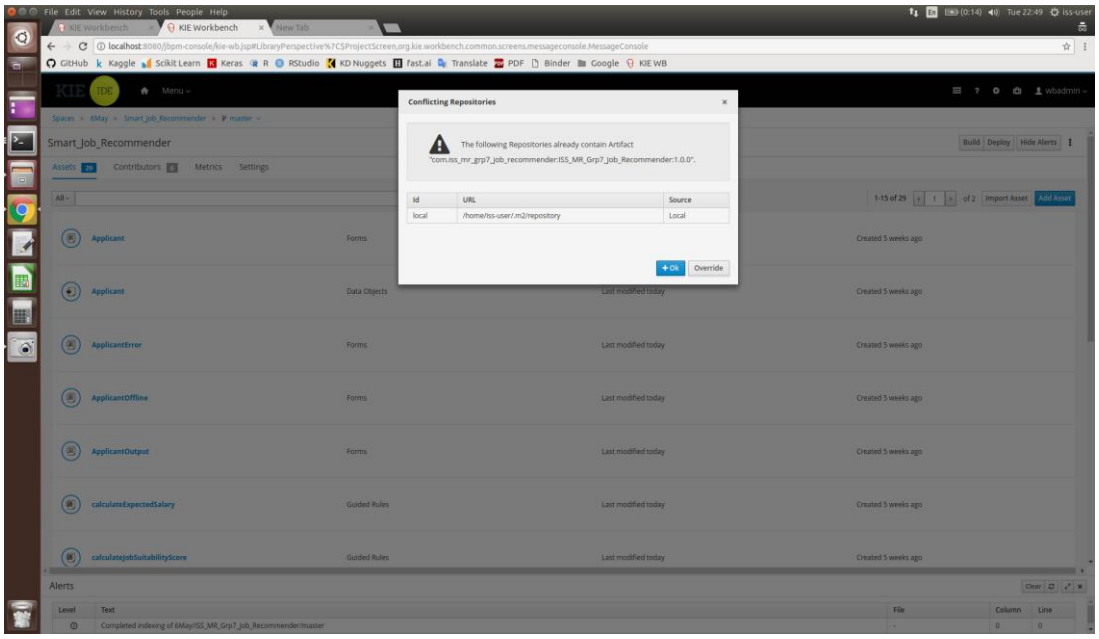
11. Once imported, click on the project “Smart_Job_Recommender” to select it. Then press “Ok” on the top right corner.



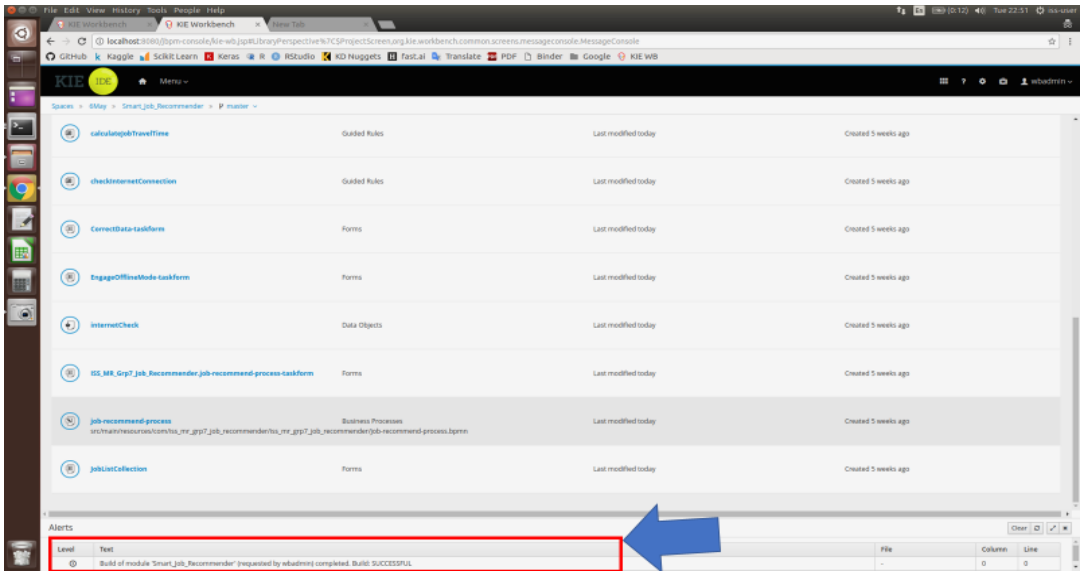
12. After importing successfully, click on the “Deploy” button (labelled in Box 11), which will bring up Box 12. Type in a Deployment Unit ID accordingly. After that, press “+Ok”.



13. After pressing “+Ok”, the below screen will pop out. To continue, press “Override”.



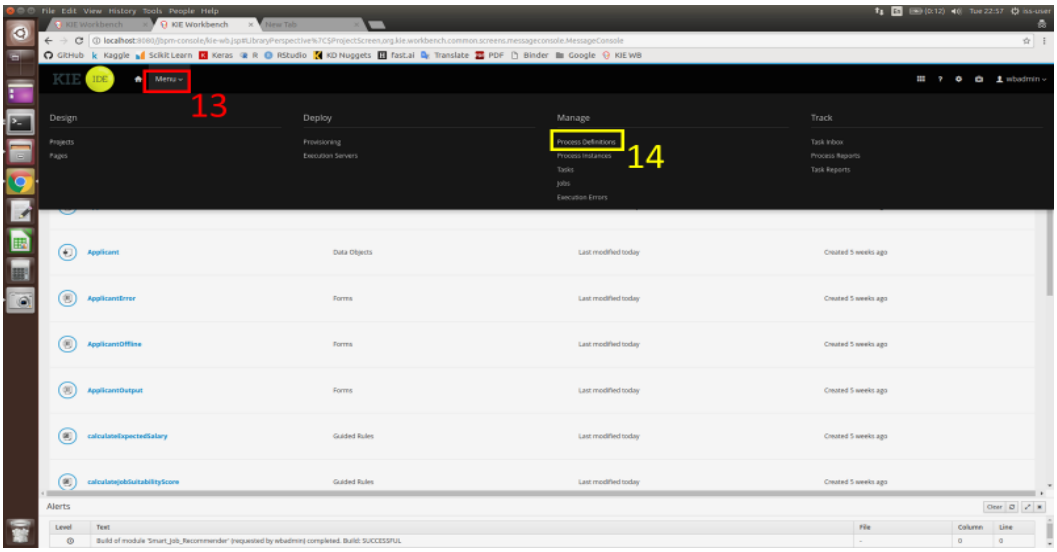
14. Once the deployment/build is completed, the text in level shown below will indicate Build: SUCCESSFUL.



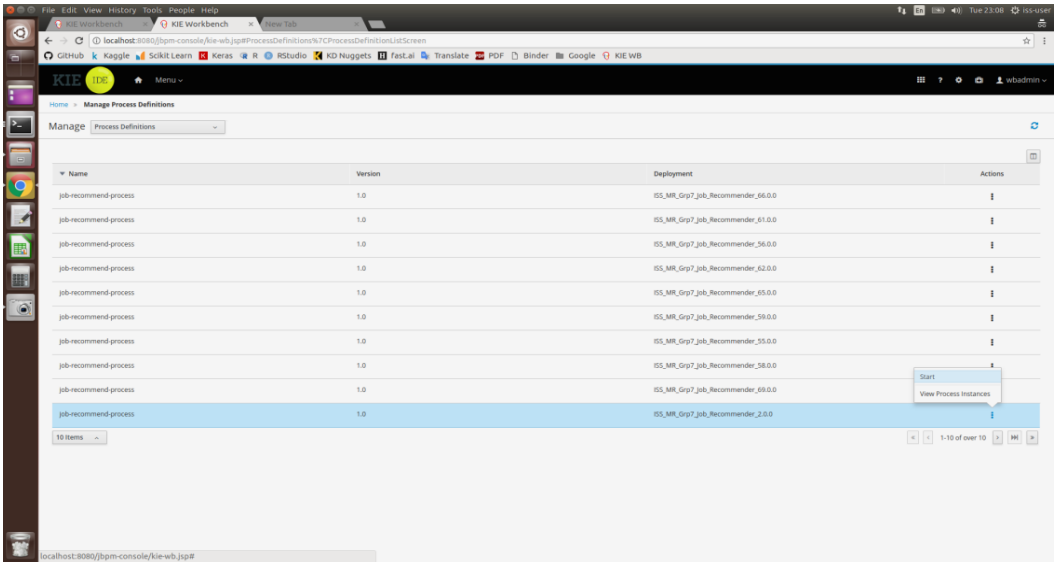
Running the System in KIE (Online Mode)

Online mode requires an internet connection. This is for the system to calculate an estimated travel time required for each job location depending on the user's ZIP code.

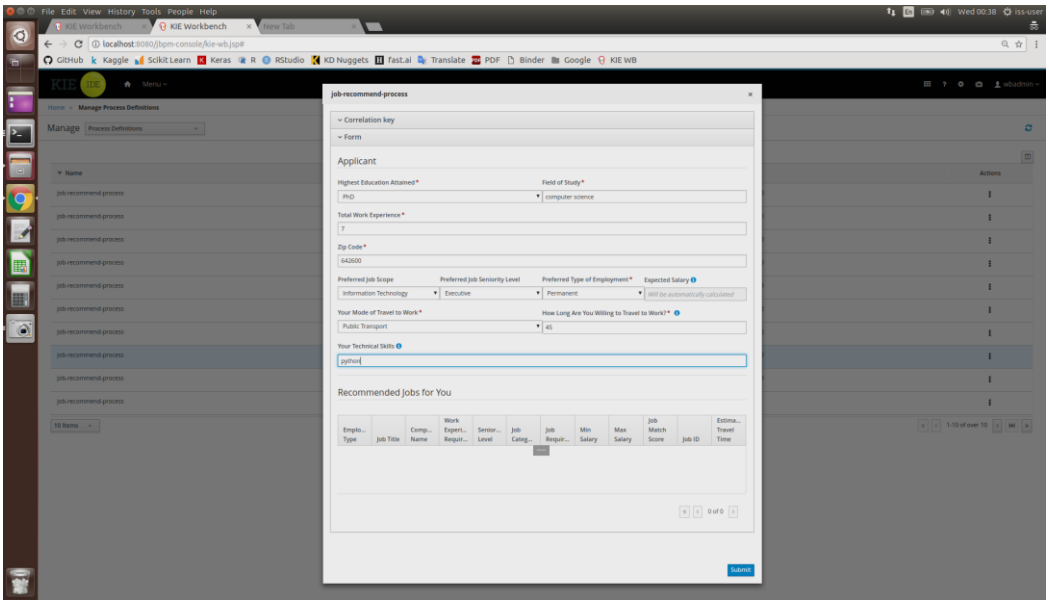
1. Press "Menu" (labelled in Box 13), and then click on "Process Definition" (labelled in Box 14)



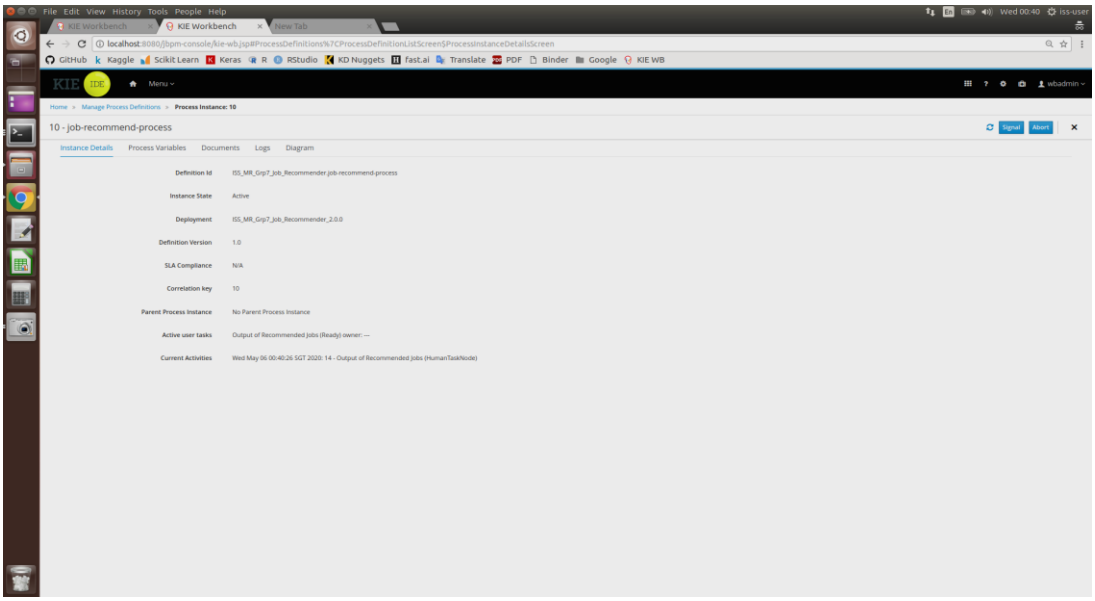
2. Select the relevant server that was deployed earlier to start the system



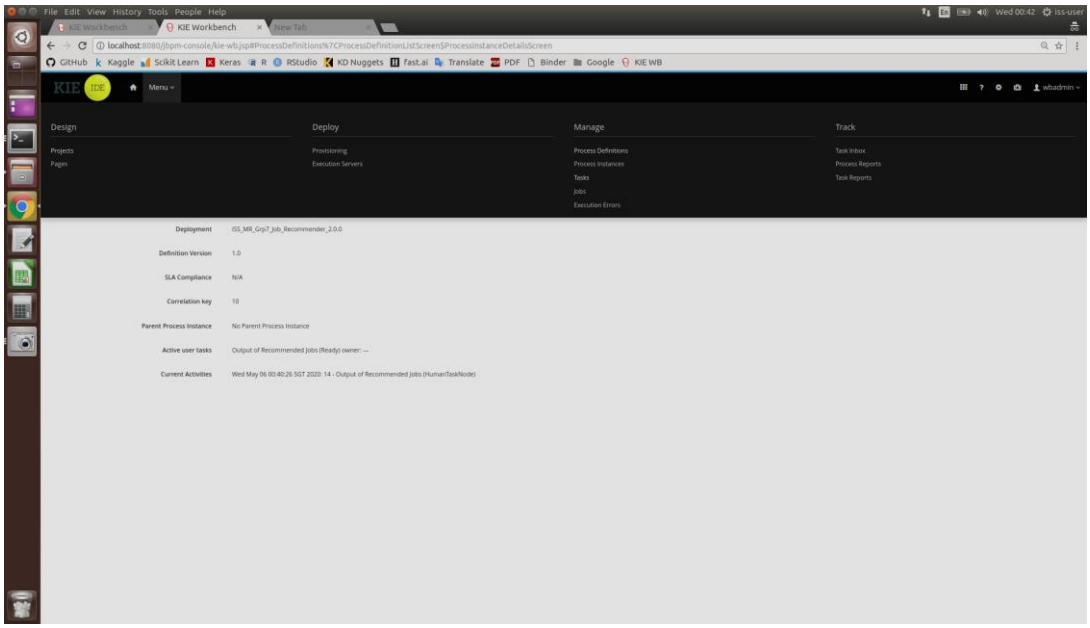
3. Once started, an applicant form for the user to fill in their details will appear. After filling up the form, press the “Submit” button ONCE. This is to avoid accidentally submitting multiple copies of the same task to the system. Please wait patiently as the system is processing the submitted information.



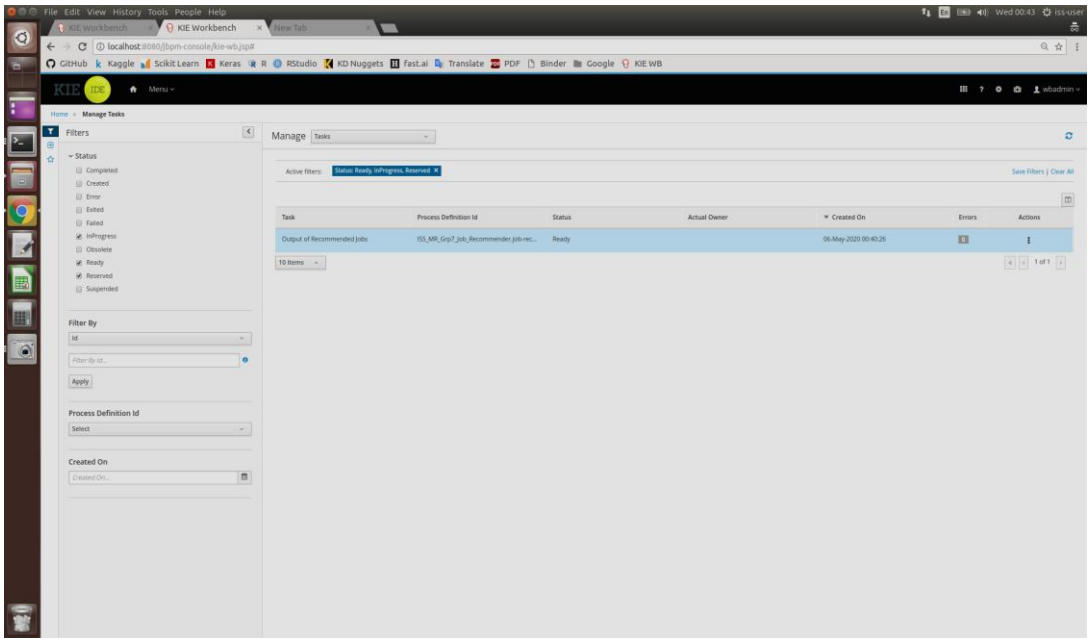
4. Once the user’s information has been processed, the below page will appear.



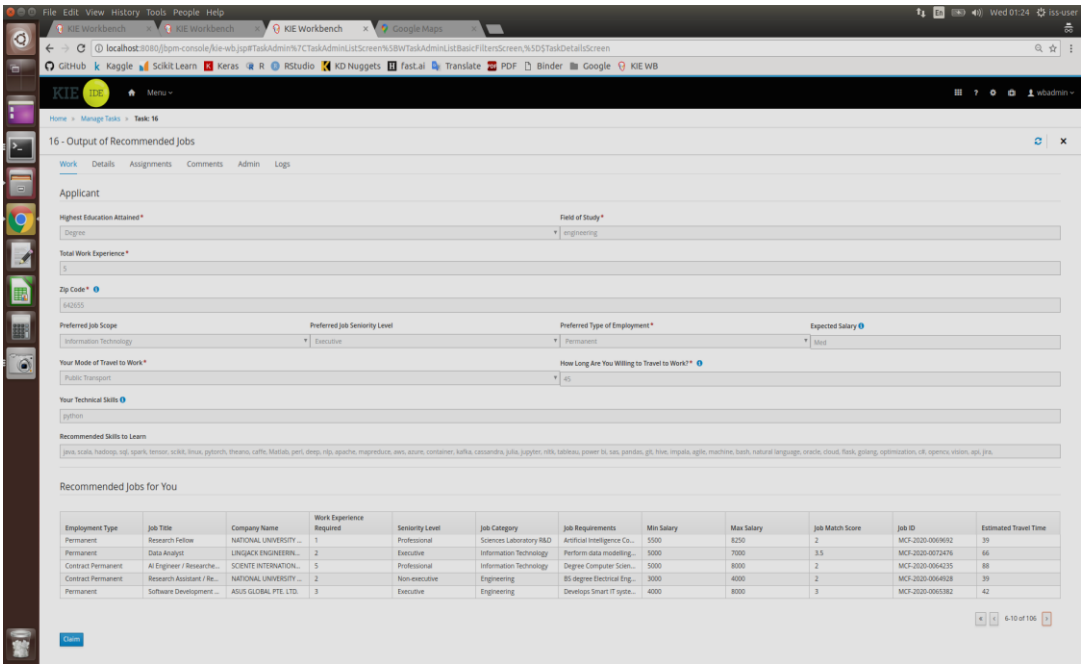
5. Click on “Menu”, then click on “Task”.



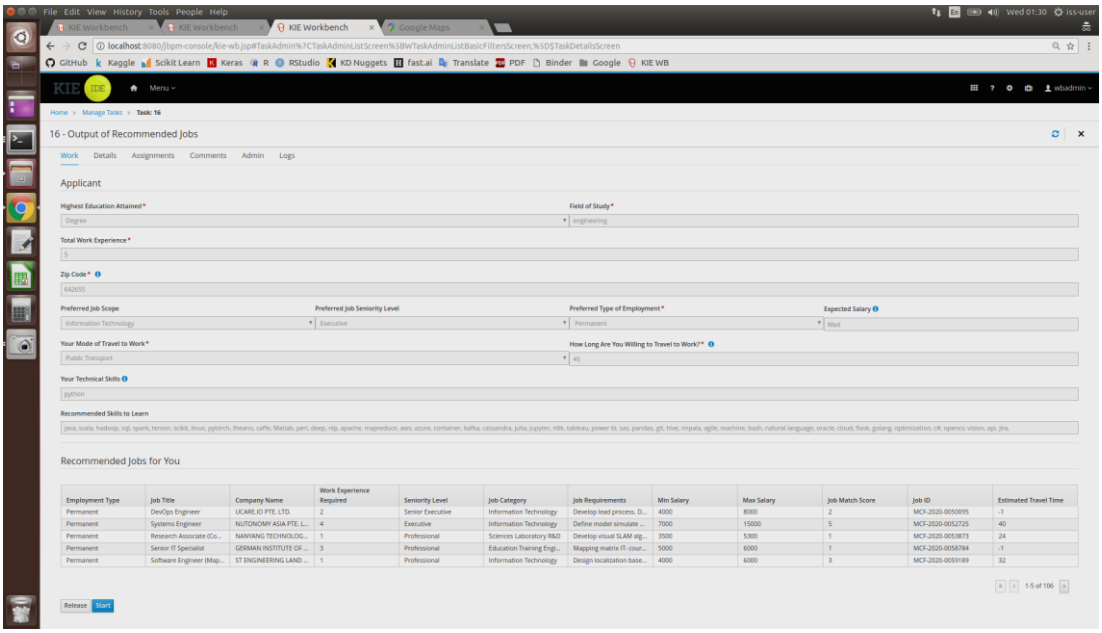
6. After completing Step 5, the below page should appear. To continue, click on on the task “Output of Recommended Jobs”



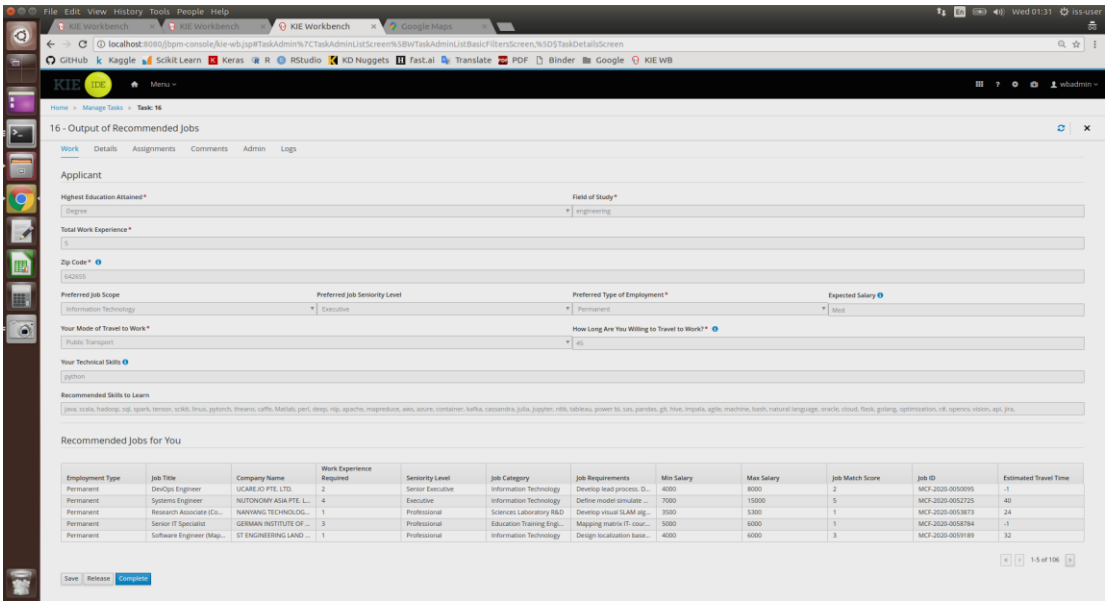
7. Press the “Claim” button at the below of the page



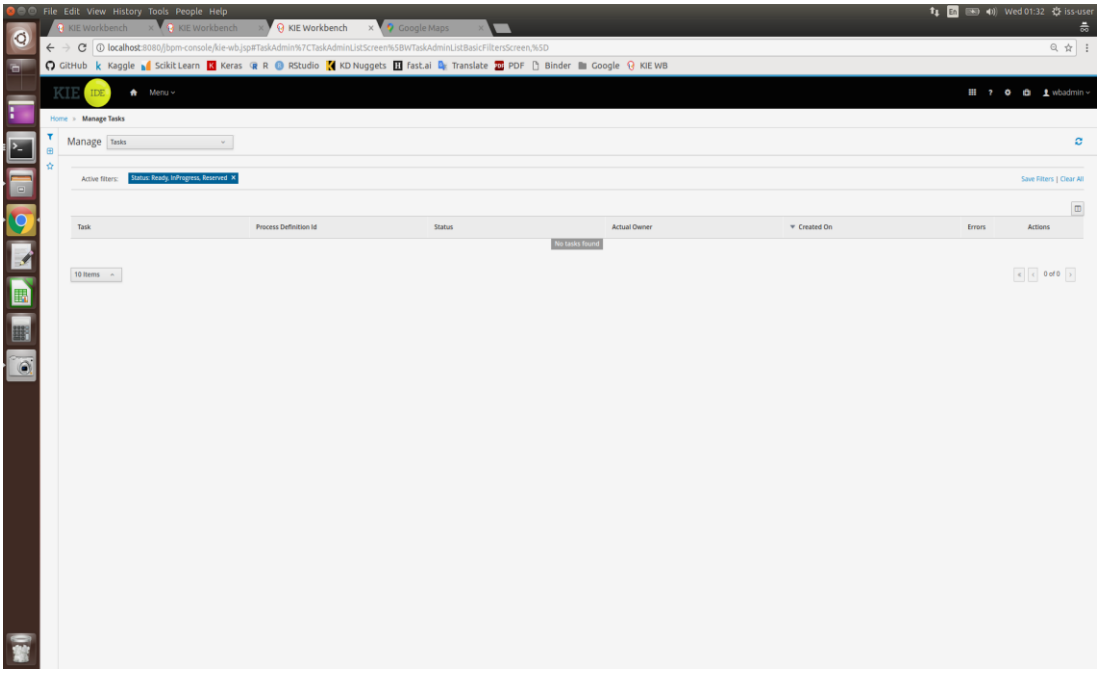
8. Press “Start”



9. After reviewing the output of recommended jobs, press “Complete” to end the task instance.



10. Process completed

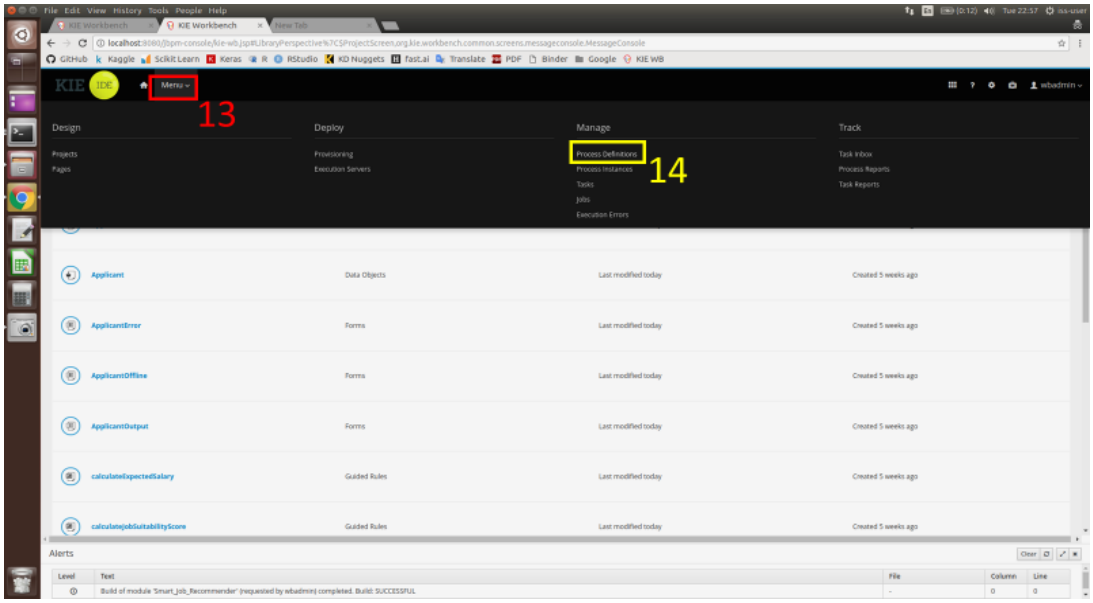


Running the System in KIE (Offline Mode)

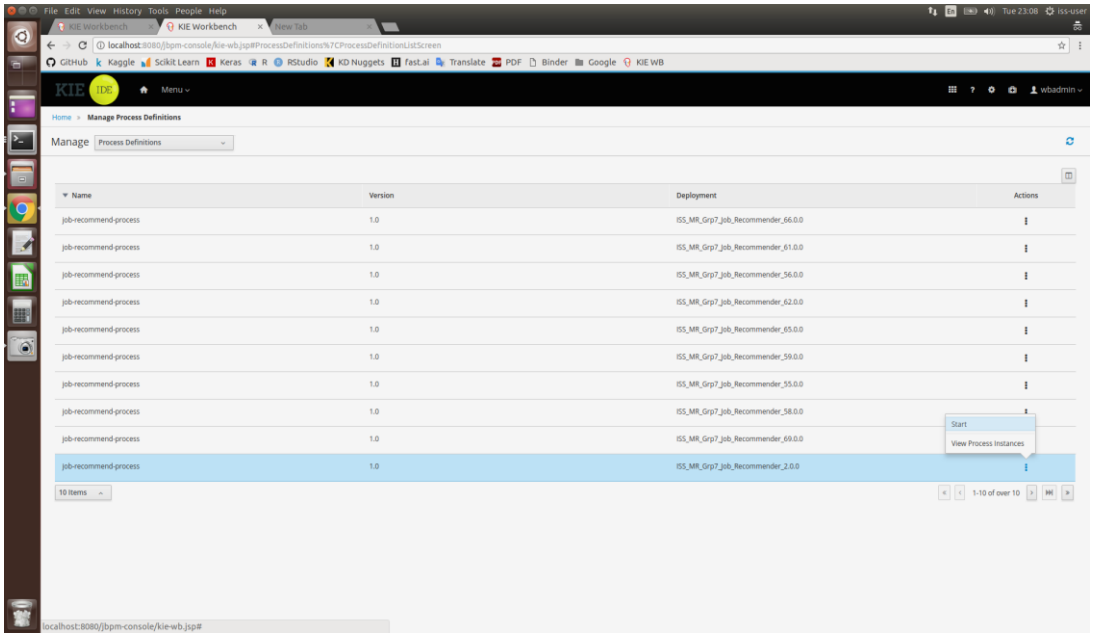
In the event an internet connection cannot be established, the system will run in Offline Mode. In this mode, the system will not be able to calculate an estimated travel time required for each job location according to the user's chosen ZIP code. Instead, users will be prompted to select one of three possible ZIP codes available in the system's database.

The procedure to operate the system in Offline Mode is as follows:

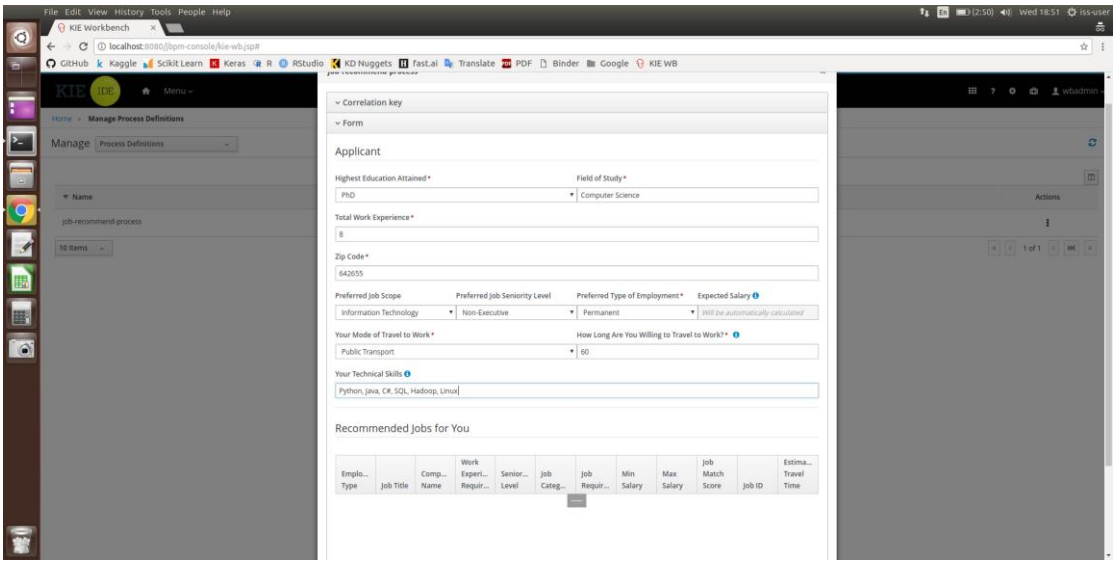
1. Press "Menu" (labelled in Box 13), and then click on "Process Definition" (labelled in Box 14)



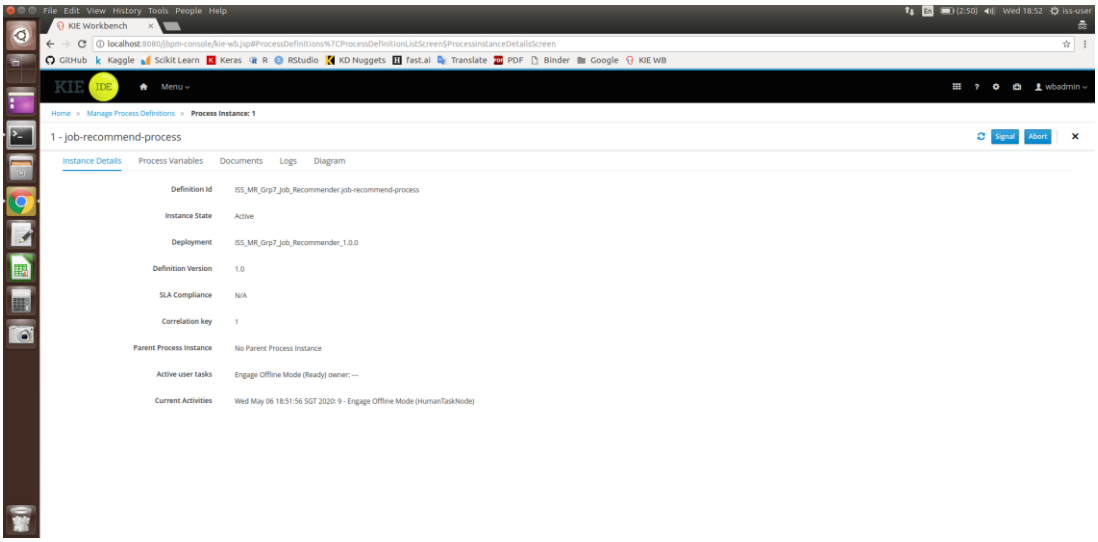
2. Select the relevant server that was deployed earlier to start the system



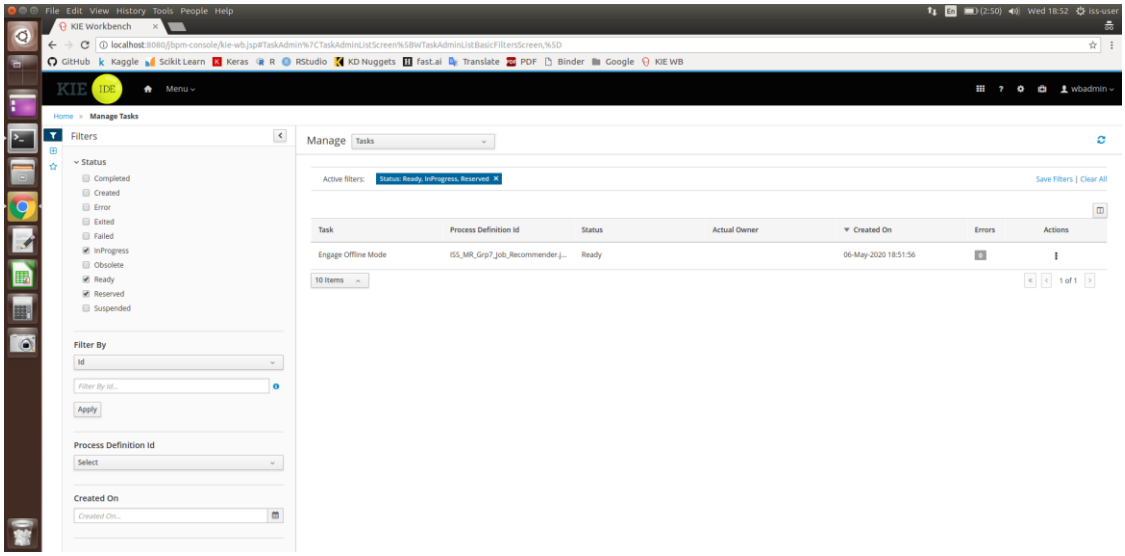
3. Once started, an applicant form for the user to fill in their details will appear. After filling up the form, press the “Submit” button ONCE. This is to avoid accidentally submitting multiple copies of the same task to the system. Please wait patiently as the system is processing the submitted information



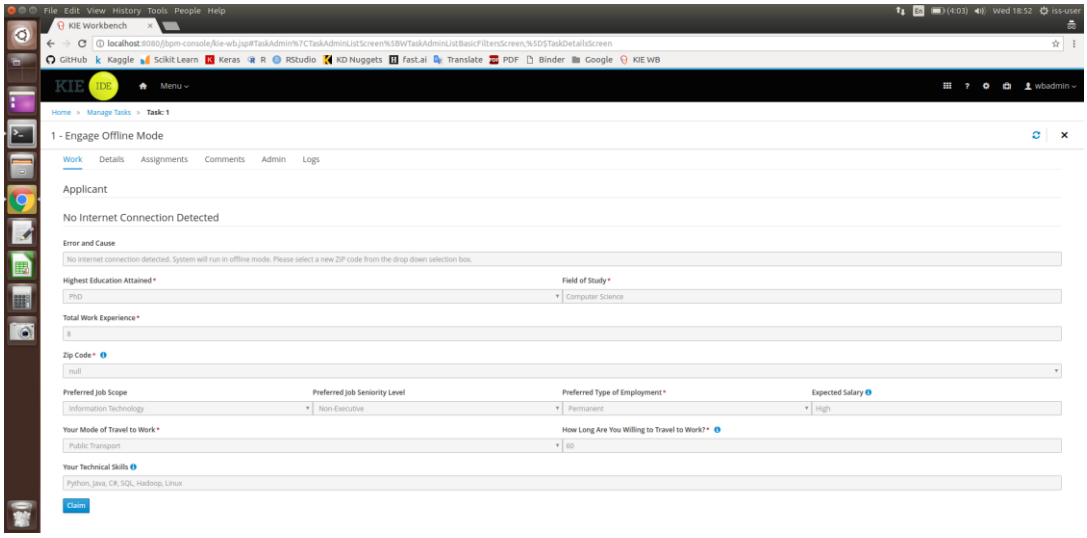
4. As no internet connection is established, the system will notify the user that it will be running in Offline Mode as shown in the page below



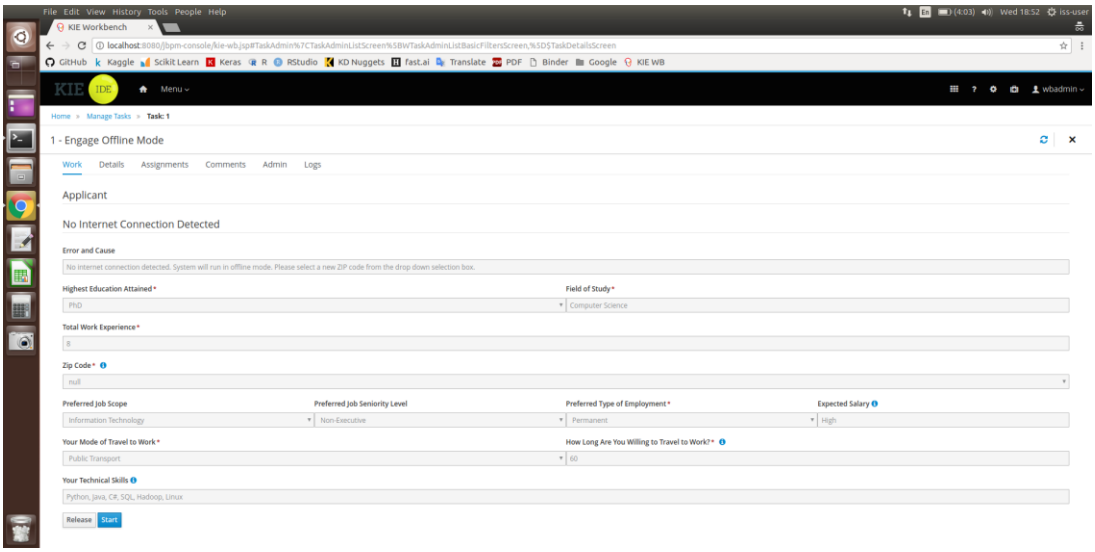
5. Click on “Menu”, then click on “Task”. This will cause the following page to appear. To continue, click on the task “Engage Offline Mode”



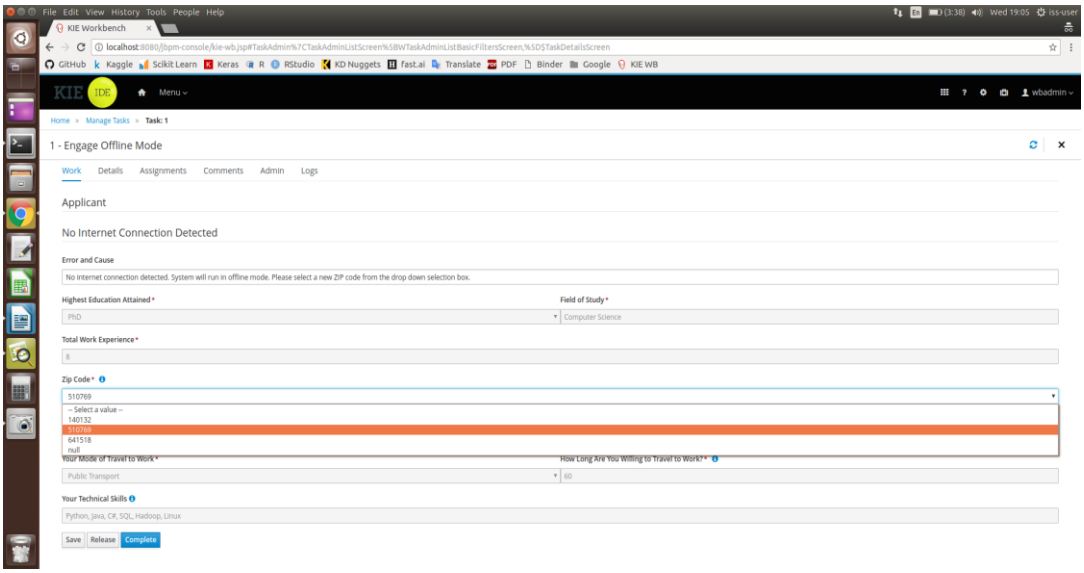
6. After Step 5, the following form will appear. Click on “Claim”



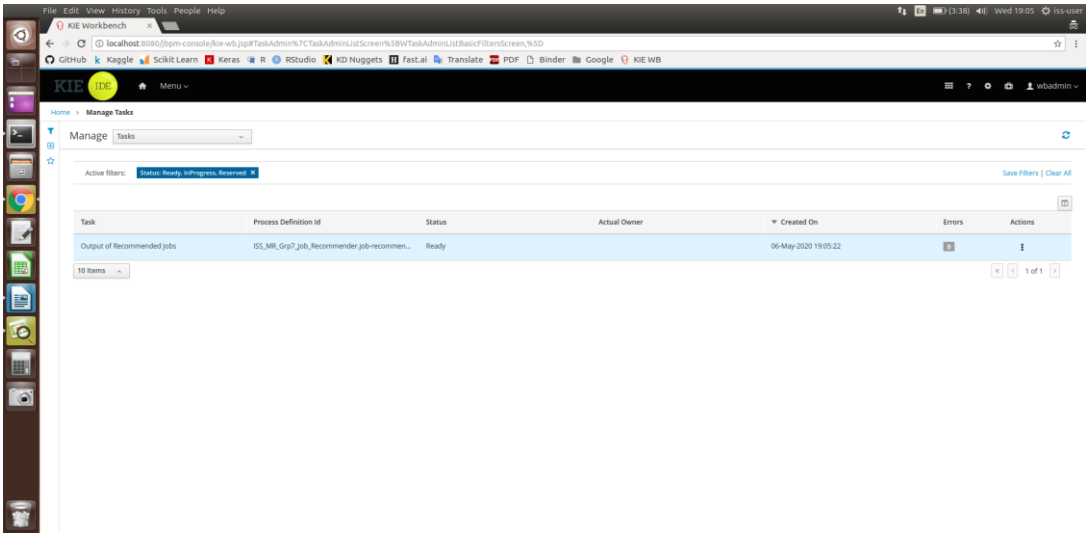
7. Click on “Start”



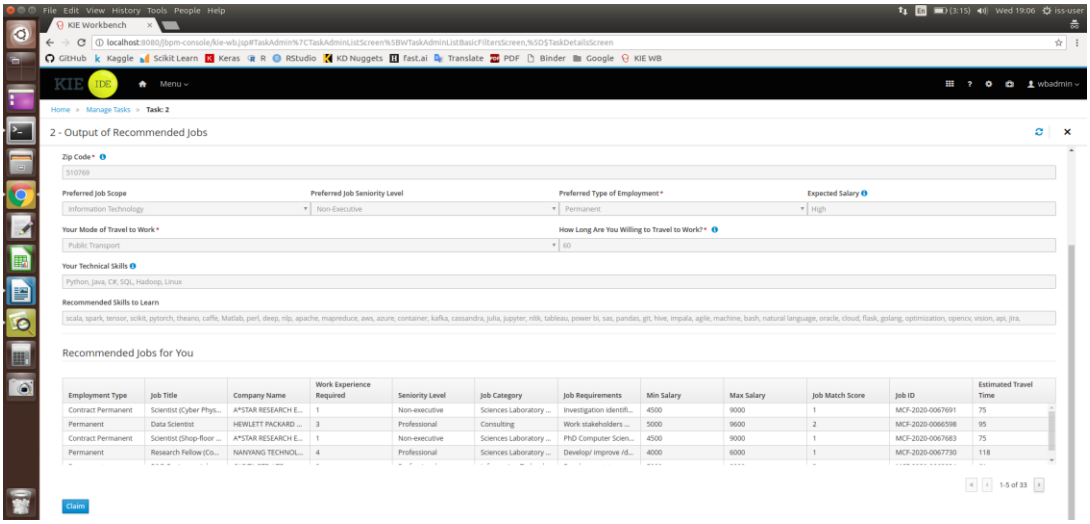
8. Click on the “Zip Code” box, which will cause a drop down menu to appear. Select one of three possible ZIP codes. Once done, click on “Complete”



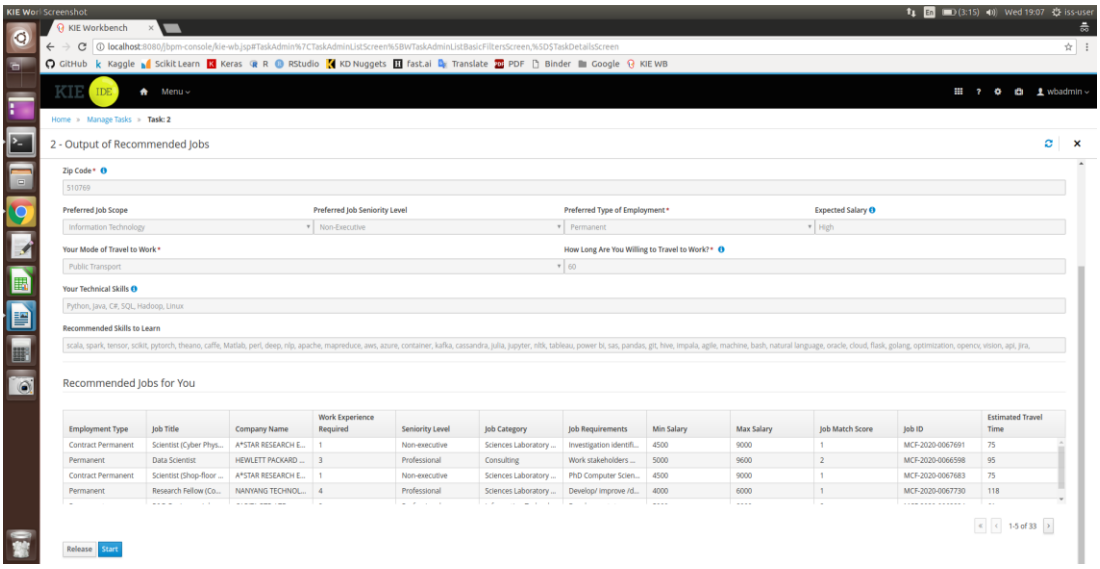
9. Click on “Menu”, then click on “Task”. This will cause the following page to appear. To continue, click on the task “Output of Recommended Jobs”



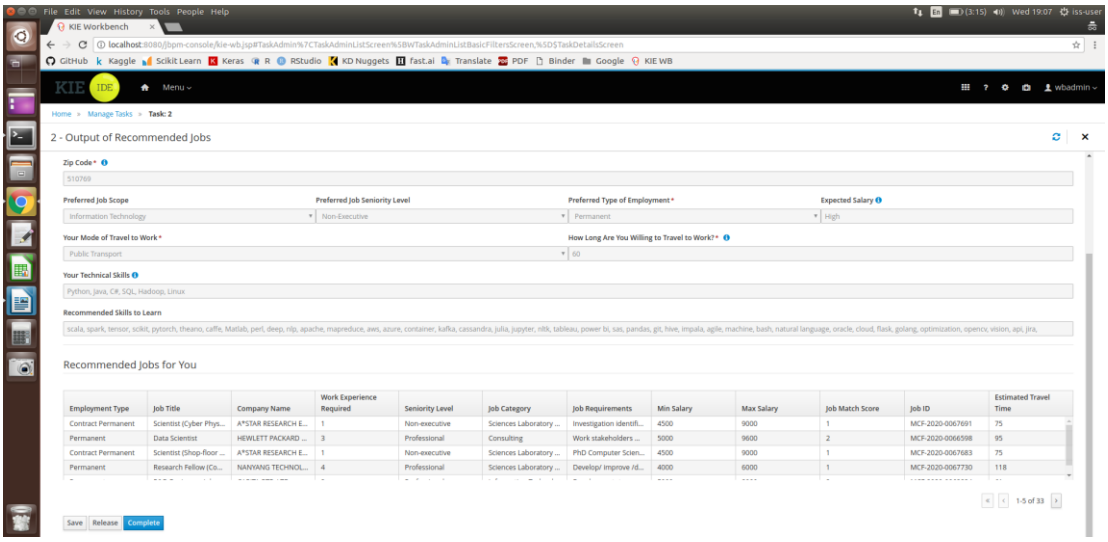
10. Press the “Claim” button at the below of the page



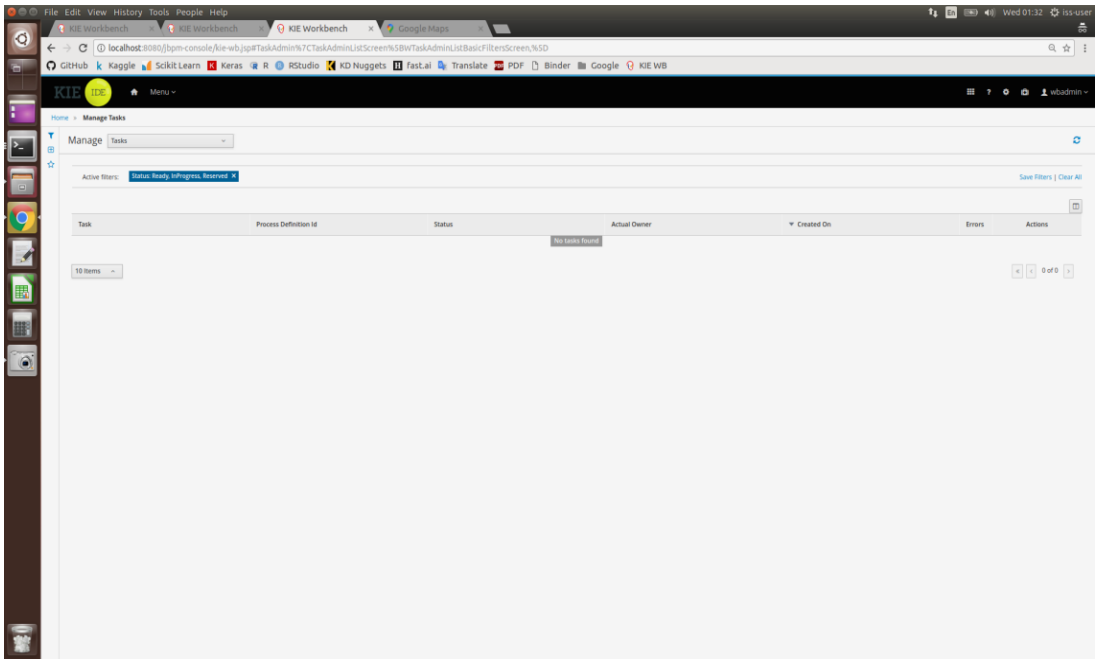
11. Press “Start”



12. After reviewing the output of recommended jobs, press “Complete” to end the task instance.



13. Process completed



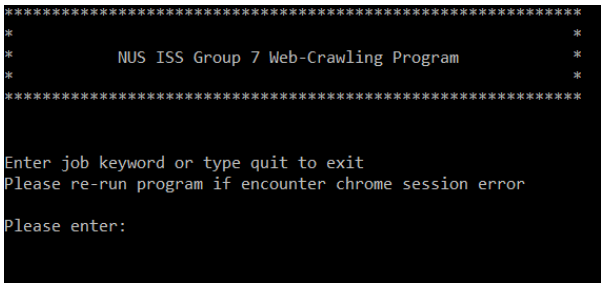
Running Data Mining algorithm in Python

Requirements:

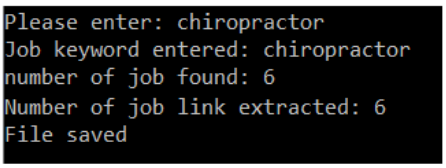
- Anaconda3
- Jupyter Lab
- Google Chrome browser version 81.0.4044
- Chromedriver version 81.0.4044
    - Saved to (\SystemCode\Data Mining\code\chromedriver)
- Anaconda Python3 environment and Library
    - Selenium
    - BeautifulSoup4
    - Scikit-learn

Web Crawling (online)

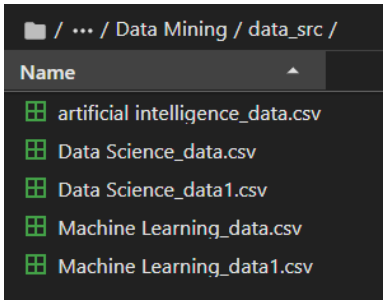
1. Open anaconda3 prompt
2. Navigate to project source code folder (..\SystemCode\Data Mining\code)
3. Type python “Web Crawling.py” to run program



4. Enter job search keyword or type quit to exit the program
5. Job crawled will be saved inside data_src folder



6. If encounter chrome session error, please re-run the program
7. Web-crawled data will be used for data mining and machine learning part.



8. Please take note popular job keyword might have few hundreds of records. Web-crawling will need few hours to crawl all records.

9. If Job not found, user will have to key in with another keyword.

```
Enter job keyword or type quit to exit
Please re-run program if encounter chrome session error

Please enter: Astronaut
Job keyword entered: Astronaut
Job not found, please try with other keyword
Job not found, please re-run
```

Data Mining (Offline)

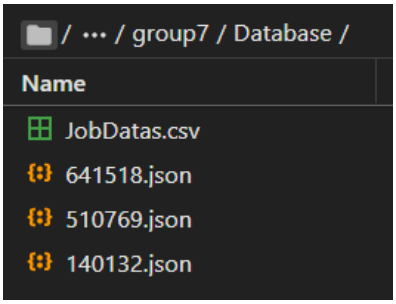
1. Open anaconda3 prompt
2. Navigate to project source code folder (..\SystemCode\Data Mining\code)
3. Type python “Data Mining.py” to run program
4. Program will ingest all crawled file inside the data_src folder to perform data mining process.

```
************************************************************
*                                                          *
*            NUS ISS Group 7 Data Mining Program           *
*                                                          *
************************************************************


Source file(s) loaded: ['../data_src\\artificial intelligence_data.csv', '../data_src\\Data
Science_data.csv', '../data_src\\Data Science_data1.csv', '../data_src\\Machine Learning_dat
a.csv', '../data_src\\Machine Learning_data1.csv']


Data ingestion and cleaning: Complete
Salary feature: Complete
Skills and qualification extraction: Complete
Number of words before filter: 68149
Number of words after filter: 45953
File saved
```

5. The program will generate an csv file named “JobDatas.csv” to the Database folder



6. The “JobDatas.csv” file will be consumed by the KIE program

Machine Learning (Offline)

1. Open anaconda3 prompt.
2. Navigate to project source code folder (..\SystemCode\Data Mining\code)
3. Type python “Machine Learning.py” to run program.
4. Program will consume the crawled data from data_src folder, process and predict the salary range.

```
************************************************************
*                                                          *
*            NUS ISS Group 7 Salary Prediction ML          *
*                                                          *
************************************************************


Data ingestion and cleaning: Complete
Feature engineering: Complete


Salary prediction: Complete
```

5. Model prediction report (classification report, confusion matrix and feature importance) will be generated.

```
classficication report:

              precision    recall  f1-score   support

         Low       0.81      0.97      0.88       231
         Med       0.72      0.62      0.67        37
        High       0.80      0.44      0.57        89

    accuracy                           0.80       357
   macro avg       0.78      0.68      0.71       357
weighted avg       0.80      0.80      0.78       357



confusion matrix:

            Pred Low  Pred Med  Pred High
Actual Low        224         3          4
Actual Med          8        23          6
Actual High        44         6         39
```

```
Feature Importances

                      coef  abs coef
Year_Experience   0.433057  0.433057
java scala        0.093971  0.093971
engineering       0.076283  0.076283
phd               0.067935  0.067935
design            0.050304  0.050304
statistical       0.047727  0.047727
architecture      0.047501  0.047501
Executive         0.043291  0.043291
master            0.043003  0.043003
perl              0.023512  0.023512
implementations   0.021084  0.021084
optimization      0.019473  0.019473
automation        0.017714  0.017714
python            0.015145  0.015145
```

6. Decision tree rule will be generated and saved as .txt file inside output folder (\Data Mining\output). Figure below shows the example of tree rules generated.

```
Decision tree rules:
|--- Year_Experience <= 3.50
|   |--- java scala <= 0.50
|   |   |--- phd <= 0.50
|   |   |   |--- Year_Experience <= 1.50
|   |   |   |   |--- Executive <= 0.50
|   |   |   |   |   |--- class: Low
|   |   |   |   |--- Executive >  0.50
|   |   |   |   |   |--- class: Med
|   |   |   |--- Year_Experience >  1.50
|   |   |   |   |--- architecture <= 0.50
|   |   |   |   |   |--- class: High
|   |   |   |   |--- architecture >  0.50
|   |   |   |   |   |--- class: Med
|   |   |--- phd >  0.50
|   |   |   |--- engineering <= 0.50
|   |   |   |   |--- statistical <= 0.50
|   |   |   |   |   |--- class: Med
|   |   |   |   |--- statistical >  0.50
|   |   |   |   |   |--- class: High
|   |   |   |--- engineering >  0.50
|   |   |   |   |--- Year_Experience <= 1.50
|   |   |   |   |   |--- class: High
|   |   |   |   |--- Year_Experience >  1.50
|   |   |   |   |   |--- class: High
|   |--- java scala >  0.50
|   |   |--- class: High
```