National College of Ireland

Configuration Manual

MSc Research Project
Programme Name

Jnanashree Arkalgud Guruvachari
Student ID: 23174528

School of Computing
National College of Ireland

Supervisor: Yasantha Samarawickrama

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Jnanashree Arkalgud Guruvachari		
Student ID:	23174528		
Programme:	MSc in Cloud Computing	**Year:**	2024
Module:	MSc Research Project		
Lecturer:	Yasantha Samarawickrama		
Submission Due Date:	12-12-2024		
Project Title:	Resource Optimization in Cloud Data Centers using Machine Learning		
Word Count:	654	**Page Count:**	4

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Jnanashree A G

Date: 11-12-2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Jnanashree Arkalgud Guruvachari
Student ID: 23174528

1 Create a new notebook from Google Colab

1. Open Google Colab and click on “New Notebook” to create a new notebook at https://colab.google/

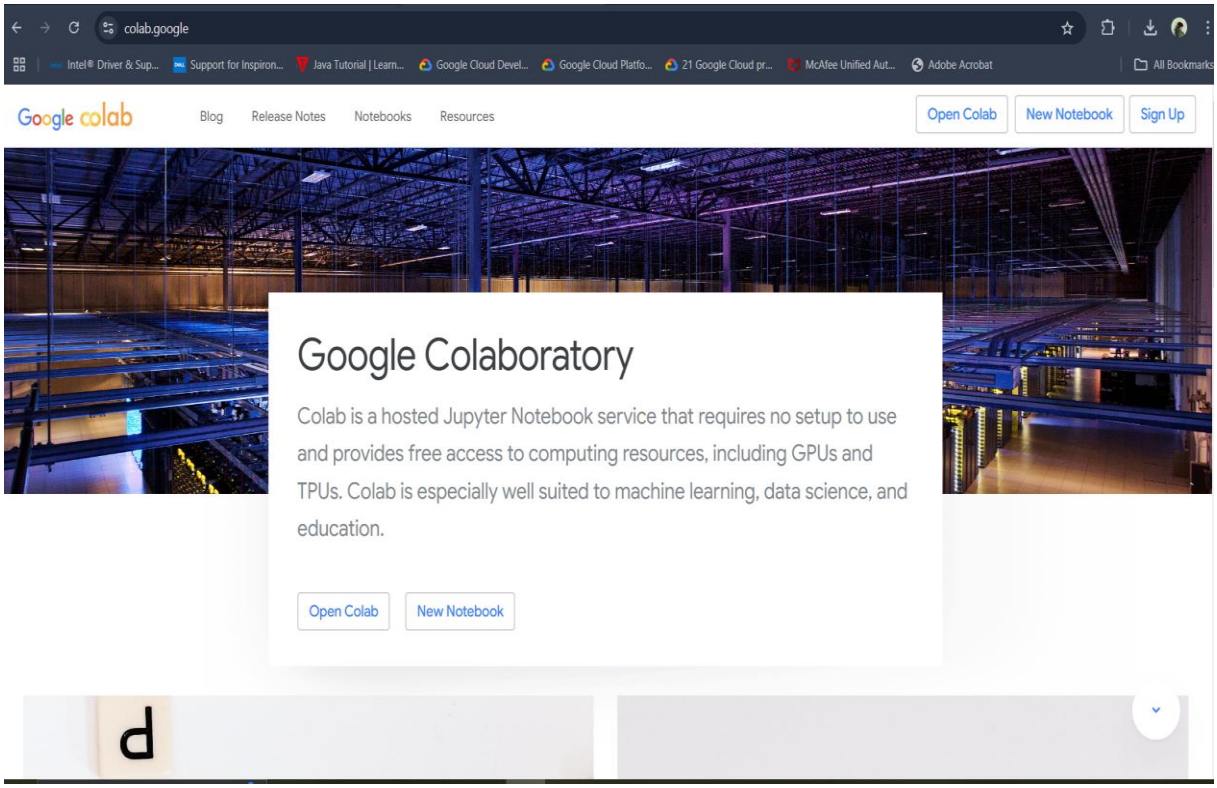

Figure 1: Google Colab Home page

2 Download the dataset from Kaggle

2. The dataset “Cloud Datacenter Workload” can be downloaded from Kaggle at https://www.kaggle.com/datasets/ashikhassan007/workloadtrace

   The dataset parameters must match the research requirements. The dataset downloaded has features such as Timestamp, CPU cores, CPU Capacity Provisioned, CPU usage, Memory Capacity Provisioned, Memory Usage, Disk read throughput, Disk write throughput, Network received throughput, and Network transmitted throughput.

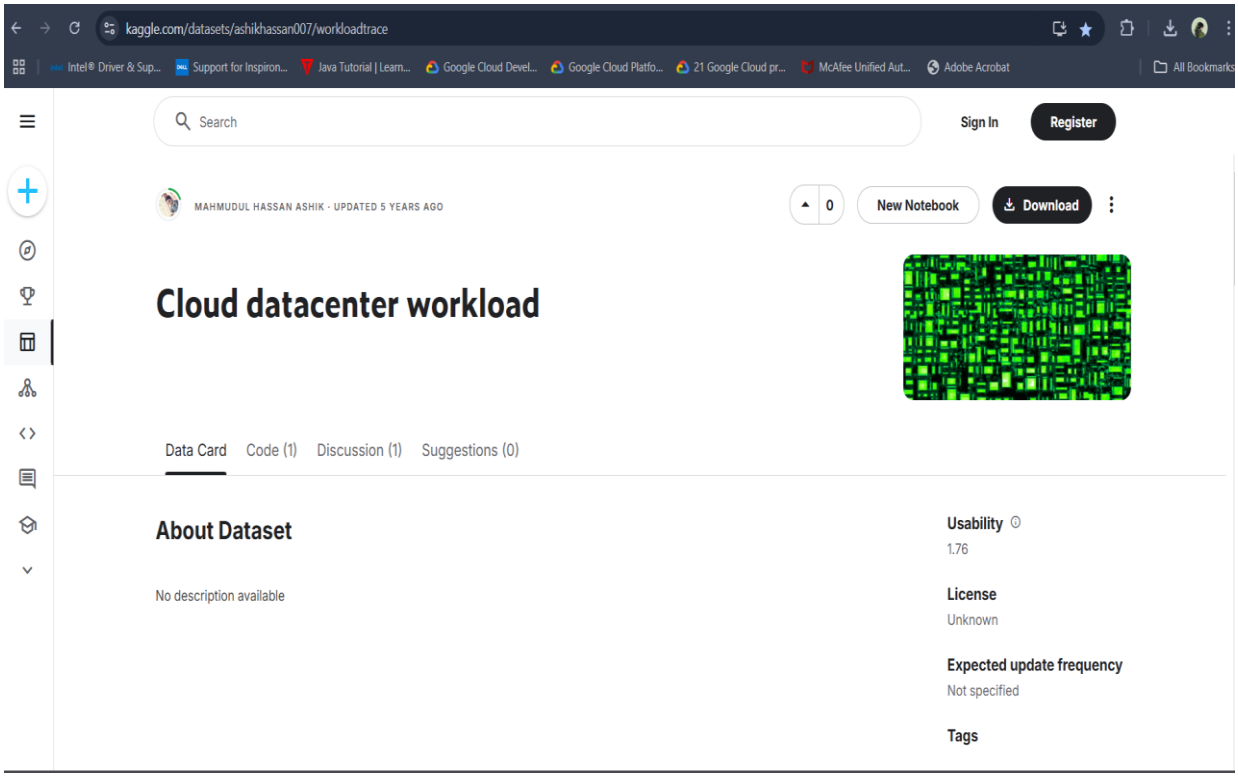

Figure 2: Dataset to download from Kaggle page

3 Importing Initial Utilities and Uploading the dataset

3. The initial import of pandas and numpy is done. Before adding the imports, the dataset must be uploaded to Google Colab from the Files section. Click on Upload file from the storage device.

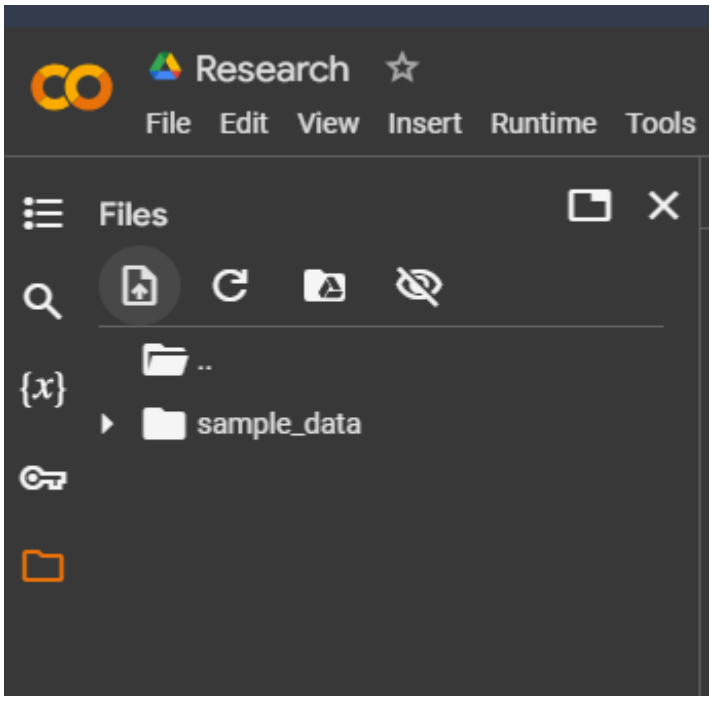

Figure 3: Uploading the dataset to Google Colab

```
[ ]  import numpy as np
```

```
[ ]  import pandas as pd
```

```
[ ]  import matplotlib.pyplot as plt
```

Figure 4: Initial utility import

4 Loading the dataset into a variable

4. The dataset must be loaded into a variable to use the data set.

```
[ ]  ds = pd.read_csv('3.csv')
```

Figure 4: Loading the dataset in Google Colab

5 Executing the command in Google Colab

5. To execute any command in Google Colab we can click the play button or shift + enter button.

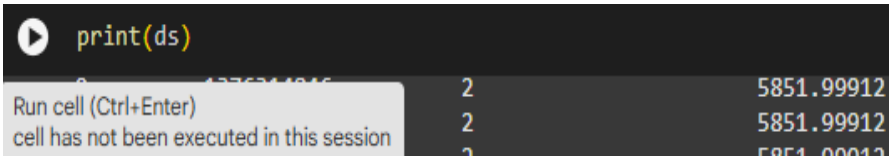

Figure 5: Execution of any command in Google Colab

6 Outlier detection using zscore method

6. Outliers are detected using zscore method and the utility “stats” is required to use the method must be imported.

```
[ ]  from scipy import stats
```

Figure 6: utility required for outlier detection

7 Linear Regression

7. The utilities required to implement the Linear Regression algorithm must be imported as shown in the below figure. The utilities required to visually represent the performance of the algorithm should also be imported.

```
[ ]  from sklearn.model_selection import train_test_split
     from sklearn.linear_model import LinearRegression
     from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
     import psutil
```

Figure 7: Utilities required to implement Linear Regression

8 Random Forest

8. The utilities required to implement the Random Forest algorithm and to represent the performance of the algorithm visually must be imported as shown in the below figure.

```
[ ]  import pandas as pd
     import numpy as np
     from sklearn.model_selection import train_test_split
     from sklearn.ensemble import RandomForestClassifier
     from sklearn.metrics import classification_report, confusion_matrix
     import matplotlib.pyplot as plt
     import seaborn as sns
```

Figure 8: Utilities required to implement Random Forest

9 Generative Adversarial Networks (GANs)

9. The utilities required to implement the Generative Adversarial Networks (GANs) and visually represent the performance of the algorithm must be imported as shown in the below figure.

```
[ ] import numpy as np
    import tensorflow as tf
    from tensorflow.keras import layers
    import pandas as pd
    from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay, precision_score, recall_score, f1_score, accuracy_score
```

Figure 9: Utilities required to implement GANs