**Write Up**

**3a) A summary on the objectives, features of your application (one paragraph)**

**3b) The dataset used, the source, the targets and the features.**

**3c) A summary of the results obtained (one-paragraph).**

**3d) Clear instructions on how to use your application (if appropriate, not required for Jupyter Notebook submissions).**

3a)

Since this is life or death matter. In my opinion it is okay to have slightly higher False-Positive Rates and make sure most or all of the True-Positives Rates Diabetic Patients are accounted for. It is okay to upset the Non-Diabetic Patients with False-Positive results as it does less harm as compared to False-Negative results which such judgements may cause a life. Thus, my hypothesis for this Dataset is that it will be in the High-Recall side.

The dataset taken from UCI had indicated that there were no missing indicators but I will do due diligence to understand the datasets clearer with numbers and visual-aids to make sense of it. To my understanding it should not make sense if there are ZEROs in any of the diagnosis attributes especially if it is a Blood Pressure attributes.

3b)

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes,

based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database.

In particular, all patients here are females at least 21 years old of Pima Indian heritage.

The web page indicated that there were no missing values in the dataset. As pointed out by a repository user, this cannot be true: **there are zeros in places** where they are biologically impossible, such as the blood pressure attribute. It seems very likely that zero values encode missing data. However, since the dataset donors made no such statement we encourage you to use your best judgement and state your assumptions.

Data Attributes

1. Number of times pregnant

2. Plasma glucose concentration a 2 hours in an oral glucose tolerance test

3. Diastolic blood pressure (mm Hg)

4. Triceps skin fold thickness (mm)

5. 2-Hour serum insulin (mu U/ml)

6. Body mass index (weight in kg/(height in m)^2)

7. Diabetes pedigree function

8. Age (years)

9. Outcome (0 or 1)

<https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database>

3c)

I had achieved a Best KNN Score of 0.7721840251252015 upon using Hyper Parameter Optimization in the 25th range, whereas prior to optimization I had achieved Best KNN score of 0.7532467532467533 in the 11th range when selected the range between 1 : 15.

\*The value of range is the total Outcome of Diabetic and Non-Diabetic patients that has been train\_test & train\_split.

My hypothesis has been met as the Recall score is the highest with the F1 score of 0.82.

3d)

\*\*\*Before Running the Jupyter Notebook

1. Make sure your are running Python 3.8.1.3

2. I have included the list of libraries to be installed in "requirements.txt"

3. Open Anaconda Prompt (Anaconda3)

4. Create a new environment with Python version stated in Step 1.

5. "Conda Activate" the created Environment in Step 4.

6. cd to the path where you have downloaded the zip file and extract the folder.

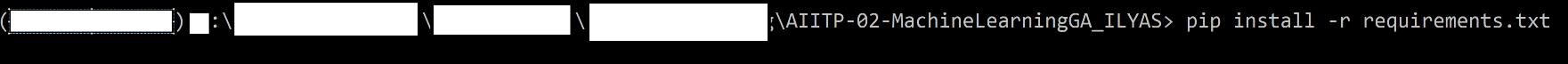
7. cd to "AIITP-02-MachineLearningGA\_ILYAS"

8. In the specified conda environment, type "pip install -r requirements.txt", (without quotations).

9. Open Anaconda, use VSCode to easily switch to the newly create enviroment OR kernel and you can RUN ALL the codes.

**<EXTRA Steps>**

10. I will copy this Write-Up to a separate document for screenshots. Fill in the blanks with you own Conda Environment and PATH.



11. “Application on” will be your new created conda environment. Run VS Code. Graphical user interface, application

Description automatically generated



Graphical user interface, text, application, email

Description automatically generated12. Select the Kernel which you had install the requirements.txt in and Run ALL…

