Kirinyaga University

SPC 2407 Machine Learning (Lesson Assignment)

DATE: 14/03/2023   **TIME: 1350hrs – 1600hrs**

Instructions: Attempt this assignment

Consider the given `diabetes2.csv` data-set. The diagnostic, binary-valued variable investigated is whether the patient shows signs of diabetes. Several constraints were placed on the selection of these instances from a larger database.

Number of Instances: 768
Number of Attributes: 8 plus class

For Each Attribute: (all numeric-valued)
- `preg`: Number of times pregnant
- `plas`: Plasma glucose concentration a 2 hours in an oral glucose tolerance test
- `pres`: Diastolic blood pressure (mm Hg)
- `skin`: Triceps skin fold thickness (mm)
- `insu`: 2-Hour serum insulin (mu U/ml)
- `mass`:Body mass index (weight in kg/(height in m)^2)
- `pedi`:Diabetes pedigree function
- `age`:Age (years)
- `class`: Class variable (0 or 1) – class value 1 is interpreted as "tested positive for diabetes"

Using Python's `Jupyter` notebook and Python's `scikit-learn` library, develop a simple regression model with this data.

Download the `.ipynb` file of your program and upload it here. **Note:** In Jupyter notebook, click on `File,` navigate to `Download as` menu choice. At this point, select `Notebook(.ipynb)` to download the file.

Note: You could use the discussion on this link https://towardsdatascience.com/how-to-build-your-first-machine-learning-model-in-python-e70fd1907cdd as a starting point.