Project 4 Group 5

The goal of the project is to use machine learning to correctly predict if a patient is going to show up to their appointment or not.

Source for the data is <https://www.kaggle.com/datasets/joniarroba/noshowappointments?select=KaggleV2-May-2016.csv> (Renamed no\_show\_data.csv and placed in the Resources file)

Goal One

As a group make a model from all the data except the PatientId, AppointmentID, ScheduledDay, AppointmentDay, and Neighbourhood that correctly predicts the outcome of whether a patient shows up to an appointment or not.

Goal Two

See if any of the individual issues is a good single-issue predictor of a no-show appointment and create visualizations to show the results. Create models without the single-issue to see how it changes the accuracy of the model from goal one.

Assign Predictor (Gender, Age, Scholarship, Hypertension, Diabetes, Alcoholism, Handicap, and SMS\_received). Look into if we can do it by day of the week?

Trevor- Gender and Age

Scott- Scholarship and Hypertension

Nick- Diabetes and Alcoholism

Cam- Handicap and SMS\_received (maybe day of the week).

**Goals we need to check off for full credit**

**Data Model Implementation (25 points)**

* A Python script initializes, trains, and evaluates a model (10 points)
* The data is cleaned, normalized, and standardized prior to modeling (5 points)
* The model utilizes data retrieved from SQL or Spark (5 points)
* The model demonstrates meaningful predictive power at least 75% classification accuracy or 0.80 R-squared. (5 points)

**Data Model Optimization (25 points)**

* The model optimization and evaluation process showing iterative changes made to the model and the resulting changes in model performance is documented in either a CSV/Excel table or in the Python script itself (15 points)
* Overall model performance is printed or displayed at the end of the script (10 points)

**GitHub Documentation (25 points)**

* GitHub repository is free of unnecessary files and folders and has an appropriate .gitignore in use (10 points)
* The README is customized as a polished presentation of the content of the project (15 points)

**Group Presentation (25 points)**

* All group members speak during the presentation. (5 points)
* Content, transitions, and conclusions flow smoothly within any time restrictions. (5 points)
* The content is relevant to the project. (10 points)
* The presentation maintains audience interest. (5 points)