**#Technologies Used**

##Data Cleaning and Analysis

Pandas will be used to clean the data and perform an exploratory analysis. Further analysis will be completed using Python.

## Database Storage

pgAdmin is the database we intend to use, and we will integrate PostgreSQL to display

## Machine Learning

To create our classifier we will use SciKitLearn a ML library. A classifier (algorithm) automatically sorts or categorizes data into one or more classes. Our training model and testing model will contain (data pertaining to year range from 1999 to 2012 equals training dataset) and (data pertaining to year range 2013 to 2017 equals test dataset).

Machine Learning is an application of artificial intelligence where a machine and or computer learns from past experiences (training data) and makes future predictions.

For this SciLearn we will be using Random Forest Classifier. This classifier combines multiple decision trees from a randomly selected subset of the training set. Votes are then collected from different decision trees to determine the final prediction.

Diagram: Random Forest Machine Learning

Diagram

Description automatically generated

## Dashboard

Will use Tableau for a fully functioning and interactive dashboard.

Alternative option us Dash a free Python library that we can develop a web-based and customizable and interactive dashboard within Python.