

# Prediction Model

This task is about creating a prediction model which helps to tell whether a superhero is human or not based on their super powers and their characteristics.

## Technologies

1. Python
2. Pandas, Numpy, sklearn
3. Jupyter Notebook

## Description

### 1. Data Set

The dataset is collected from the kaggle superhero data set which includes two sub data sets. One is for the different characteristics of the superhero and another power data describes the different powers of superheroes

### 2. Data Cleaning and Pre Processing

This data does not include any garbage values but it does have null values and string values for objects data. All the powers are categorical data and Hair color, Eye Color, Gender, Alignment, Skin Color, Publisher need to be converted into numerical data by making the different category of them as columns and then use them in prediction. We can use the dummies function from the pandas library to do so.

### 3. Machine Learning Technique

For making the model with greater accuracy , Random Forest Classifier is used for classifying the data to get better prediction.

## Results

When the dataset is used in decision tree technique the accuracy of the model is below 70% so as to make the model better and give it a higher accuracy , use random forest classifier so as to get the accuracy 83%