Term Project Proposal

Title: Machine Learning of Pokémon Datasets

Group member: Enbo Tian

Email: [etian@wpi.edu](mailto:etian@wpi.edu)

WPI ID: 250264656

Description:

This project is a new project for me. The dataset includes 21 variables per each of the 721 Pokémon of the first six generations, plus the Pokémon ID and its name, where each ID of Pokémon is unique.

To analysis the Dataset, I need to explore data first (1 week).

As I looked at the dataset, I think it is able to do some Logistic Regression with reduce the dimensionality. base on the variety of Pokémon, and some training and test and KNN (1 week).

One Pokémon have the data of ATK, defense, speed and so on, which can be clustering each Pokémon into 2 different types (is legendary or not), I think k-mean is good enough to do this (1 week).

It is also able to do some prediction with the evolution. I think this can be done with using Bayesian Gaussian (1 week).

It is also able to test the accuracy, comparing the accuracy with logistic regression (1 week).

Totally this project will take me 5 weeks to do, and it will take me one more week to write the report.

Reference：<https://www.kaggle.com/alopez247/pokemon>

Addition:

1. This is a new project for me all things were not done before.
2. I will cover some formula of Bayesian from our textbook Bishop, other literature review may cover the textbook from my other class, but I’m not sure. Most of the thing I can do without the books.
3. minimum achievable goal: Bayesian Regression will be done, and all steps before that will be done, such as explore data
4. challenge: I’ll trying to cover Bayesian Hierarchical Clustering, since it is a new topic for me, I need to search for how to do the programming.