CS 539 Midterm report

By Enbo Tian

**literature review**:

This project is considered the game of *Pokémon*, with developed by Game Freak and published by Nintendo, in 1996. Pokémon became more and more popular over time, with that Nintendo eventually produced animated TV shows, movies, trading card games, and various comics. The game *Pokémon Go* got a favorable reception. When Switch was put on sale in 2017, *Pokémon* on Switch and *Pokémon Go* got popular again, since the data on Switch and Phone can be connected.

**Data Source:**

• **Type\_1.** Primary type of the Pokémon. It is related the nature, with its lifestyle and with the movements it can learn for the fighting time. This categorical value can take 18 different values: Bug, Dark, Dragon, Electric, Fairy, Fighting, Fire, Flying, Ghost, Grass, Ground, Ice, Normal, Poison, Psychic ,Rock, Steel, and Water.

• **Type\_2.** Pokémon can have two types, but not all of them do. The possible values this secondary type can take are the same than the variable Type\_1.

•**Total.** The sum of all the base battle stats of a Pokémon. It should be a good indicator of the overall strength of a Pokémon. It is the sum of the next six variables. Each of them represents a base battle stat. All the

Battle stats are continuous yet integer variables, i.e. the number of values they can take is infinite in theory, or just very big in the practice.

• HP. Base health points of the Pokémon. The bigger it is, the longer the Pokémon will be able to stay in a fight before they faint and leave the combat.

• Attack. Base attack of the Pokémon. The bigger it is, the more damage its physical attacks will deal to the enemy Pokémon.

• Defense. Base defense of the Pokémon. The bigger it is, the less damage it will receive when being hit by a physical attack.

• Sp\_Atk. Base special attack of the Pokémon. The bigger it is, the more damage its special attacks will deal to the enemy Pokémon.

• Sp\_Def. Base special defense of the Pokémon. The bigger it is, the less damage it will receive when being hit by a special attack.

• Speed. Base speed of the Pokémon. The bigger it is, the more times the Pokémon will be able to attack to the enemy.

• Generation. The generation where the Pokémon was released. It is an integer between 1 and 6, so it is a

Numerical discrete variable. It could let us analyze the development or the growth of the game through the years.

• is\_Legendary. Boolean indicating whether the Pokémon is legendary or not. Legendary Pokémon tend to be stronger, to have unique abilities, to be hard to find, and to be even harder to catch.

• Color. Color of the Pokémon according to the Pokédex. The Pokédex distinguishes between ten colors: Black, Blue, Brown, Green, Grey, Pink, Purple, Red, White, and Yellow.

• hasGender. Boolean indicating the Pokémon can be classified as male or female.

• Pr\_Male. In case the Pokémon has Gender, the probability of its being male. The probability of being female is, of course, 1 minus this value. Like Generation, this variable is numerical and discrete, because although it is the probability of the Pokémon to appear as a female or male in the nature, it can only take 7 values: 0, 0.125, 0.25, 0.5, 0.75, 0.875, and 1.

• Egg\_Group\_1. Categorical value indicating the egg group of the Pokémon. It is related with the race of the Pokémon, and it is a determinant factor in the breeding of the Pokémon. Its 15 possible values are: morphous, Bug, Ditto, Dragon, Fairy, Field, Flying, Grass, Human-Like, Mineral, Monster, Undiscovered, water\_1, Water\_2, and Water\_3.

• Egg\_Group\_2. Similarly, to the case of the Pokémon types, Pokémon can belong to two egg groups.

• hasMegaEvolution. Boolean indicating whether a Pokémon can mega-evolve or not. Mega-evolving is property that some Pokémon have and allows them to change their appearance, types, and stats during a combat into a much stronger form.

• Height\_m. Height of the Pokémon according to the Pokédex, measured in meters. It is a numerical continuous variable.

• Weight\_kg. Weight of the Pokémon according to the Pokédex, measured kilograms. It is also a numerical continuous variable.

• Catch\_Rate. Numerical variable indicating how easy is to catch a Pokémon when trying to capture it to make it part of your team. It is bounded between 3 and 255. The number of different values it takes is not too high notwithstanding, we can consider it is a continuous variable.

文本, 应用程序

描述已自动生成• Body\_Style. Body style of the Pokémon according to the Pokédex. 14 categories of body style are specified: bipedal\_tailed, bipedal\_tailless, four\_wings, head\_arms, head\_base, head\_legs, head\_only, insectoid, multiple\_bodies, quadruped, serpentine\_body, several\_limbs, two\_wings, and with\_fins.

表格

描述已自动生成

图形用户界面

中度可信度描述已自动生成

**Data Explore:**

To figure out the number Legendary Pokémon:

图表, 条形图

描述已自动生成

Type1 and type 2 Pokémon:

文本

描述已自动生成

图表, 条形图, 直方图

描述已自动生成

图表, 直方图

描述已自动生成

Eggs in each type:

文本

描述已自动生成

图表, 条形图

描述已自动生成

图表, 条形图

描述已自动生成

Heatmap:

文本

中度可信度描述已自动生成

图表, 条形图

描述已自动生成

Reduce no related variable:

图表

中度可信度描述已自动生成

图形用户界面, 文本, 应用程序

描述已自动生成

**Methodology:**

Currently I used linear regression and KNN, but both of the accuracy are not good. I’m going to try some other method.

图表, 散点图

描述已自动生成

文本

描述已自动生成

Confusion matrix:

图形用户界面, 文本

描述已自动生成

表格

中度可信度描述已自动生成