**KAGGLE-PUBG-placement**

A17-REPORT

**TEAM:**

Oskar Randmer

Gustav Nikopensius

**Project repository:** <https://github.com/AlcadoSs/IDSPUBGFinishPlacementPrediction>

**Kaggle link**: <https://www.kaggle.com/c/pubg-finish-placement-prediction/overview>

**Business understanding**

Our project does not have specific business aspect more like data science side to compare different ways in data preprocessing, different models and methods to reach lowest possible mean squared error (MSE) between predicted players placement and actual placement.

1. **Identifying your project goals**
   1. Background (Game description)

PlayerUnknown's Battlegrounds (PUBG) is an online multiplayer battle royale game developed and published by PUBG Corporation. Up to 100 players compete to be the last man or team standing on different maps. The game starts by players jumping out of a plane without any gear and parachuting down to the ground, where the players can search buildings, ghost towns and other sites to find weapons, vehicles, armor, and other equipment to compete against other players or/and teams. In every few minutes size of playable map reduces towards a random location to increase the chances of encounters between the players and reduce players count.

* 1. Project and data-mining goals
     1. Find most important factors, what defines player placement in final ranking.
     2. Data preprocessing- data cleaning, normalization, transformation and feature extraction and selection.
     3. Compare MSE of Simple linear regression, Ridge regression, Lasso regression, Random Forest and Lightgbm.
  2. Project and data-mining success criteria
     1. Reach below 0.05 MSE.

1. **Assessing your situation**
   1. Inventory of resources
      1. 2 students with laptops and PCs. Python and Jupyter Notebook.
   2. Requirements, assumptions, and constraints
      1. Finish project according to the project plan.
      2. Achieve set goals.
   3. Risks and contingencies
      1. Lack of knowledge in data science. Ask from instructors or try to find answer from Internet.
      2. Teamwork is not so good as expected. Communicate with teammate and it is not working out finish project by yourself.
      3. Problems completing set task. Communicate with teammate, maybe switch task for other with teammate.
   4. Terminology
      1. **Simple linear regression** - sklearn.linear\_model.LinearRegression
      2. **Ridge regression** - sklearn.linear\_model.Ridge
      3. **Lasso regression** - sklearn.linear\_model.Lasso
      4. **Random Forest** - sklearn.ensemble RandomForestRegressor
      5. **Lightgbm (Light Gradient Boosting Machine)** – LGBMClassifier
      6. **mean squared error (MSE)** - The mean squared error tells you how close a regression line is to a set of points. It does this by taking the distances from the points to the regression line (these distances are the “errors”) and squaring them. The squaring is necessary to remove any negative signs. It also gives more weight to larger differences. It is called the mean squared error as you are finding the average of a set of errors.

**Data understanding**

1. **Gathering data**
   1. Outline data requirements
   2. Verify data availability
   3. Define selection criteria
2. **Describing data**
3. **Exploring data**
4. **Verifying data quality**

**Project plan**

1. **Make a detailed plan** of your project with a list of tasks. There should be at least 5 tasks. Specify how many hours each team member is going to contribute to each task.
2. **List the methods and tools** that you plan to use. Add any comments about the tasks that you think are important to clarify.