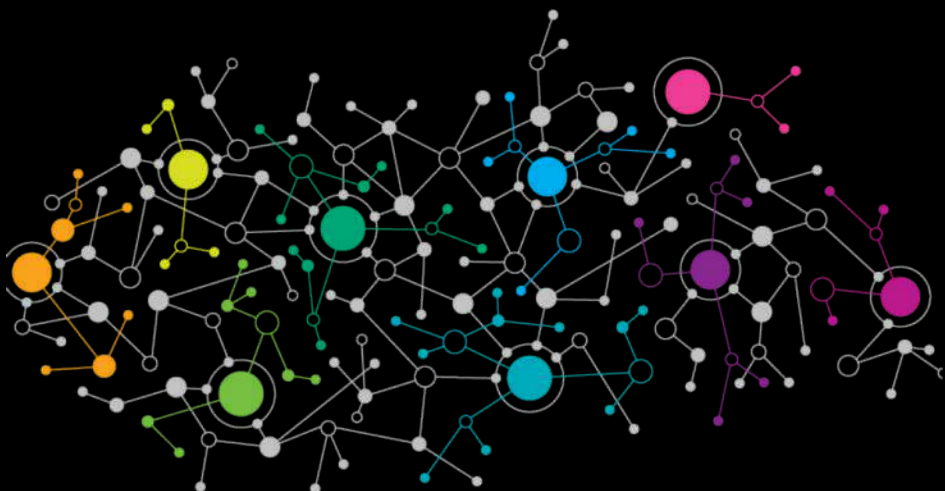


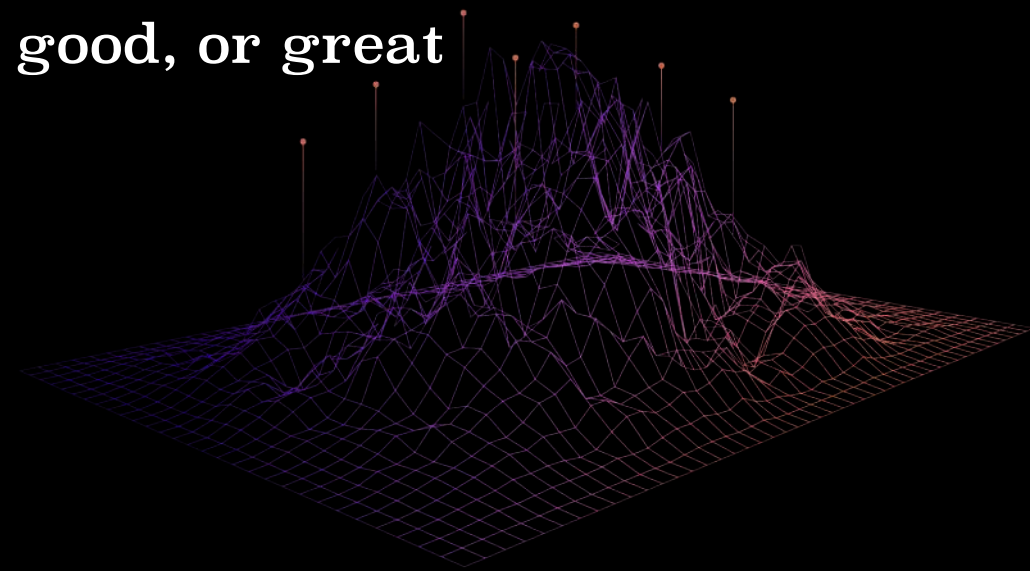
Supervised Learning: Predicting video games user review scores

Practical assignment developed for EIADC Class



- **Machine Learn Problem**

- In this practical work (assignment n° 2) we will analyze a dataset that contains various attributes like year of release, genre, number of follows, and the target attribute of prediction: rating of 6,000 games.
- The goal of this assignment is to predict if the average score users gave the video game a bad, mediocre, good, or great score.



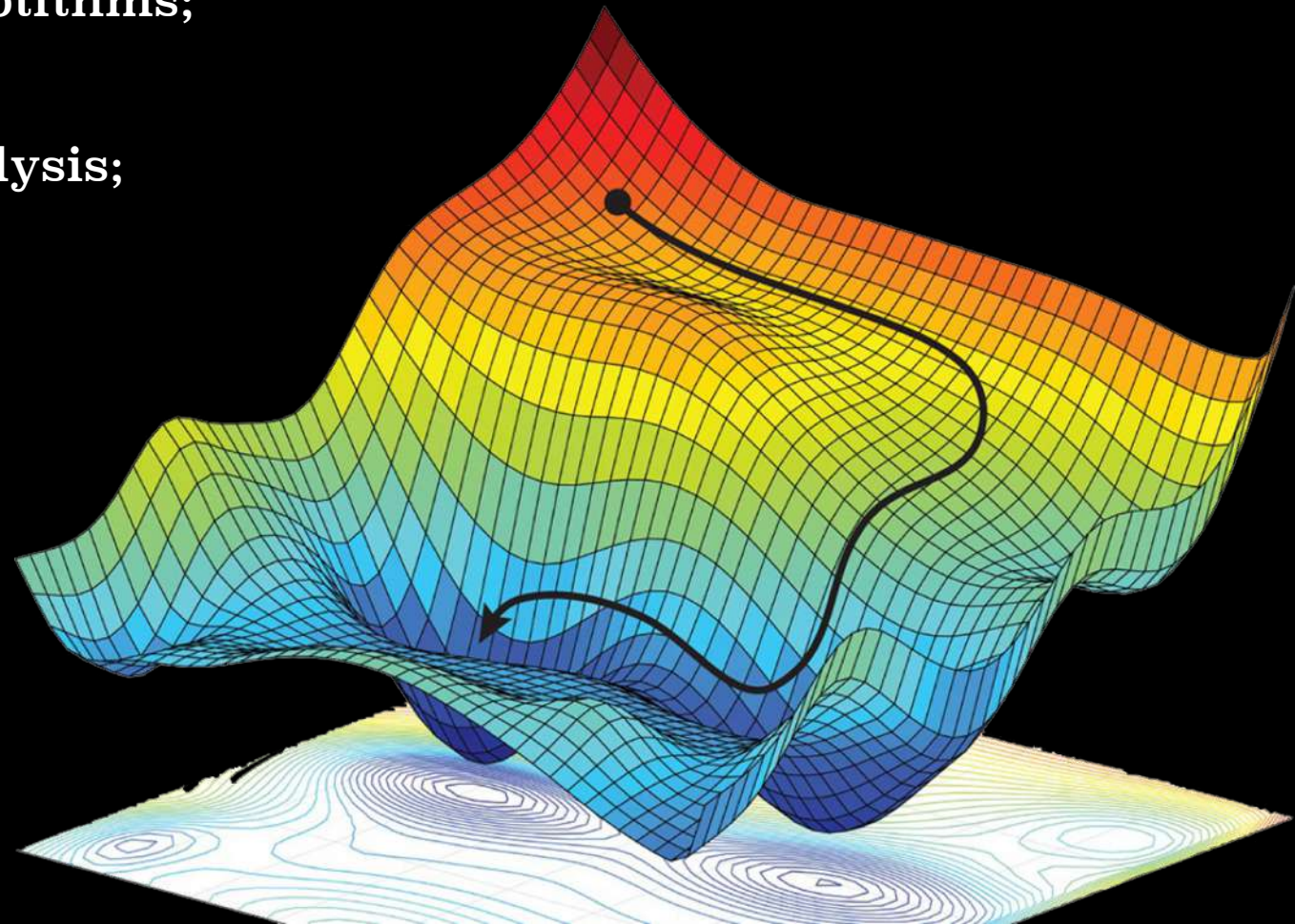
- Tools and Algorithms used

- Tools:

- Jupyter Notebook: integrated development environment (IDE) ;
- Scikitlearn: Machine Learning Algorithms;
- Seaborn: Plotting;
- Pandas: Manipulation and Data Analysis;
- Numpy: Mathematical Functions.

- Algorithms:

- Dummy classifier;
- Tree decision classifier;
- K-NN.



- Data preprocessing:

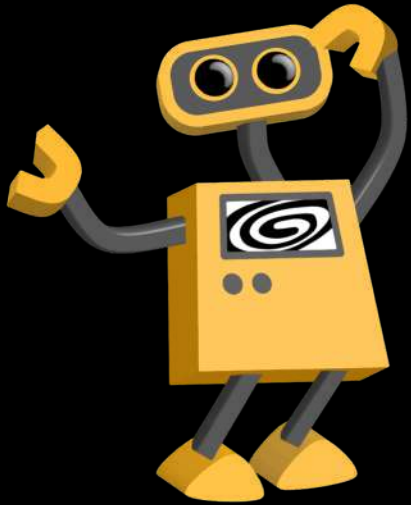
- Data Normalization;
- Removal of unwanted data with missing values;
- Visualization of data correlations.

Column	Non-Null Count		Dtype
-----	-----		-----
id	5753	non-null	int64
name	5753	non-null	object
category	5753	non-null	object
dics	5753	non-null	int64
expansions	5753	non-null	int64
year	5753	non-null	int64
follows	5753	non-null	int64
in_franchise	5753	non-null	bool

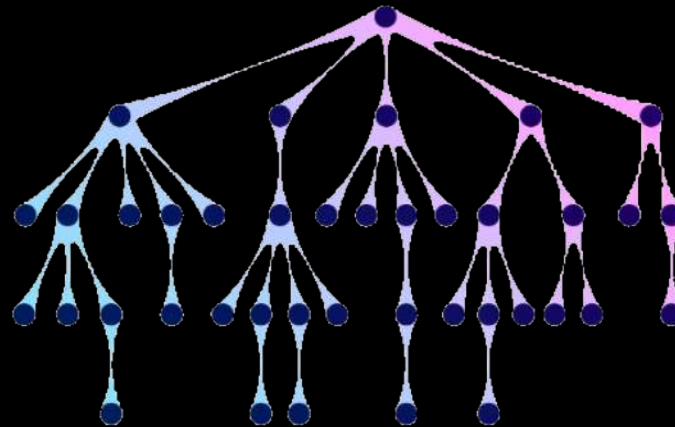


- Used Predicting Models:

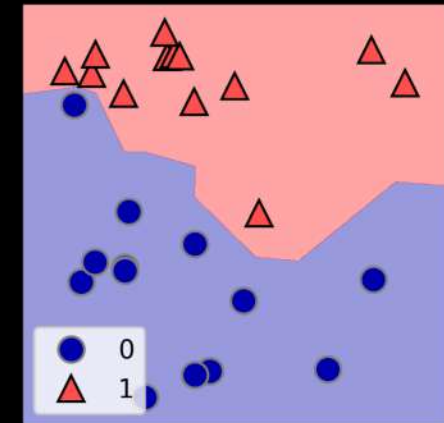
- Dummy Classifier: Used as a baseline to compare with the other models;



- Decision Tree Classifier: This model predicts the rating based on a set amount of decisions;

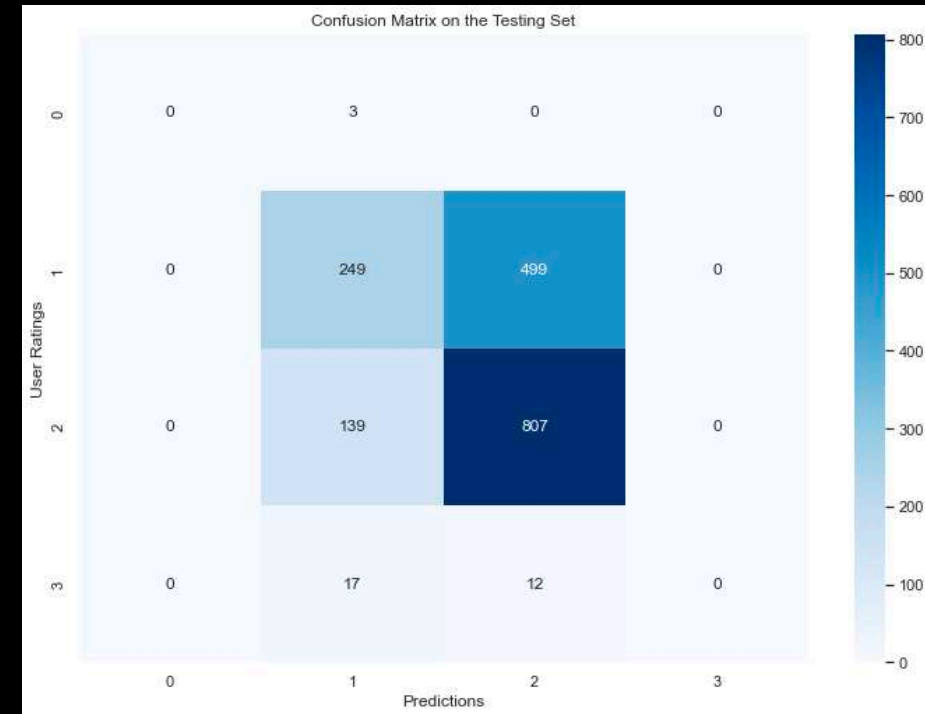
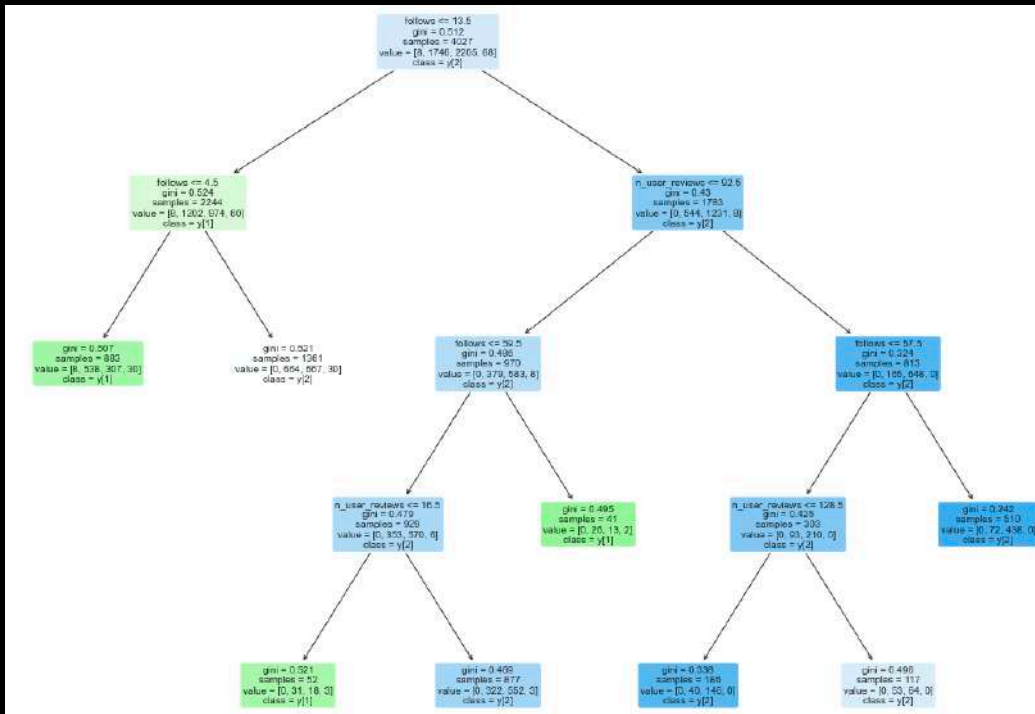


- K- Nearest Neighbors algorithm: The game is classified by a plurality vote of its neighbors, with the game's rating being assigned to the rating most common among its k nearest neighbors.



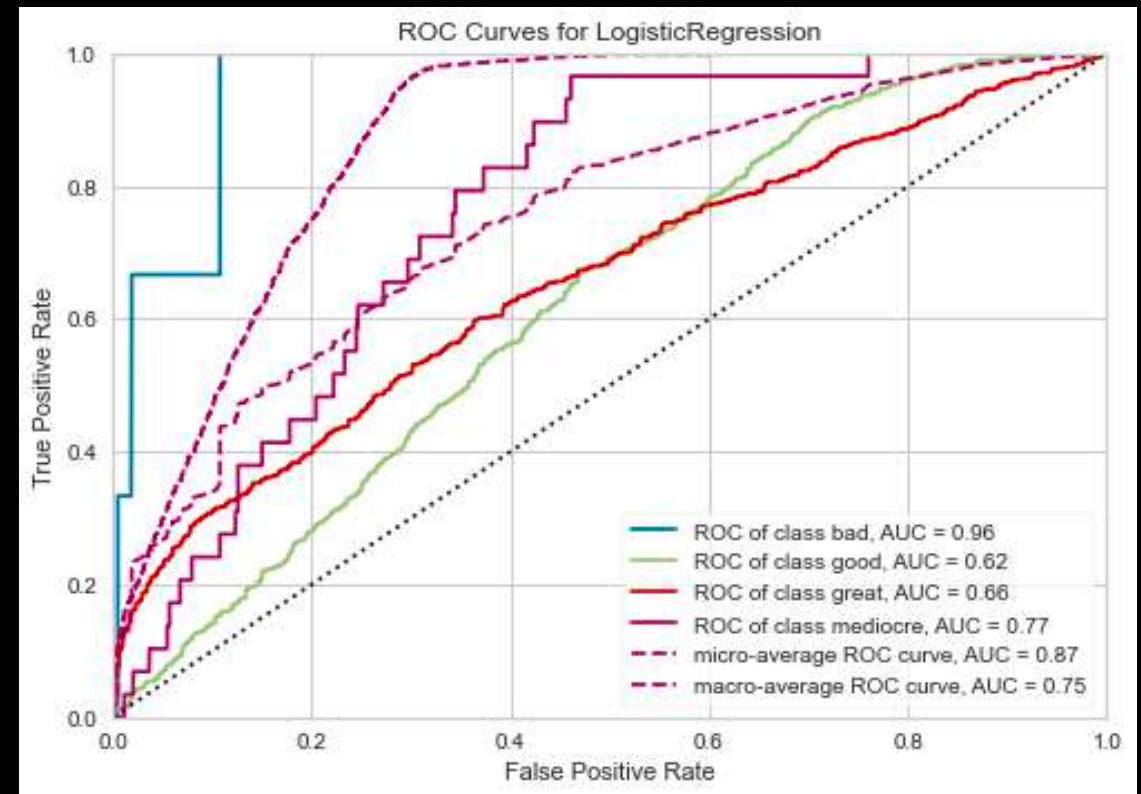
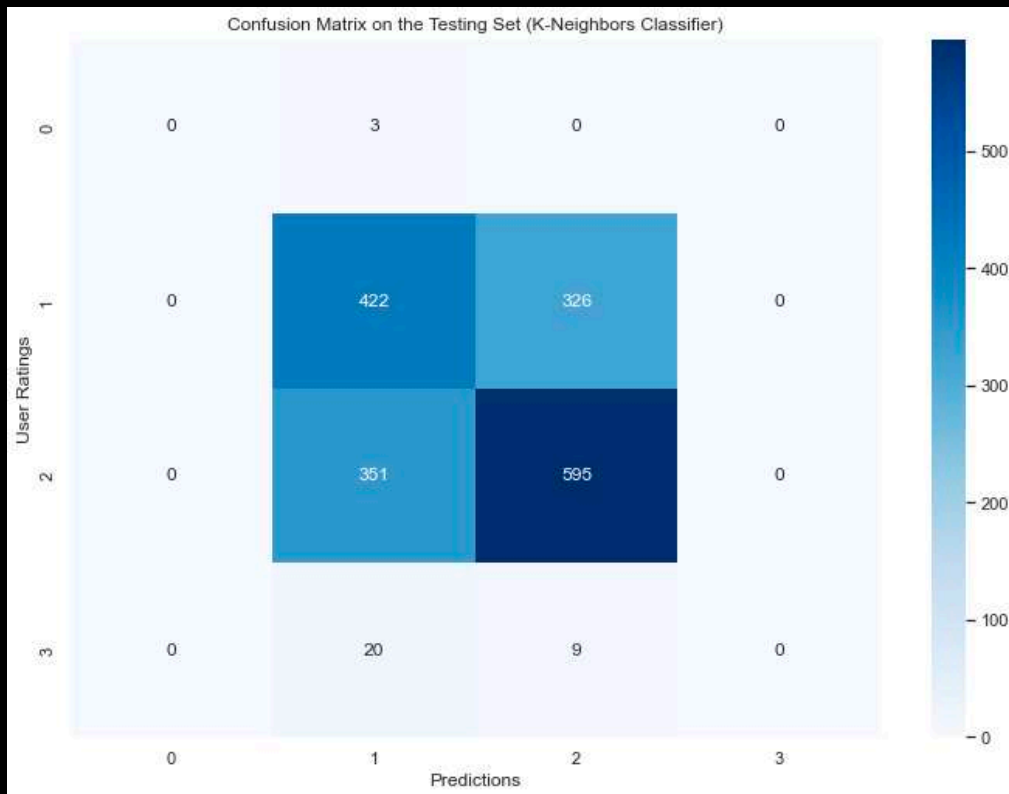
• Model vizualization and evaluation

- Dummy Classifier: The baseline accuracy was 55%.
- Tree Decision Classifier:
 - ✓ The average accuracy was 61%;
 - ✓ It failed the most predicting the worst games(bad and mediocre);
 - ✓ The average amount of leaves was 6.



- **K-NN:**

- ✓ The average accuracy was 62%;
- ✓ It would cluster more accurately between the good and great games;
- ✓ The average amount of neighbors was 85.



• Model Comparison

- Both Models compared with the dummy classifier did better with around 6-7% higher accuracy;
- The Models had difficulties predicting the rating of the low-end games, but generally predicted better the higher-end games, with K-NN being the best;



- In terms of code execution time, K-NN took far more time (35.2 s) than the Tree Decision Classifier (0.3 s) because the former has higher expensive real time execution;
- To conclude we can say that the K-NN had a better performance than the Tree Decision Classifier.

- # References

- Seaborn Documents: <https://seaborn.pydata.org/#> ;
- ScikitLearn Documents: <https://scikit-learn.org/stable/modules/classes.html> ;
- Decision Tree Classifier Tutorial:
<https://www.kaggle.com/code/prashant111/decision-tree-classifier-tutorial> ;
- Files available on classroom Moodle.