

Supervised Learning

Top Hits Spotify

Inteligência Artificial

Grupo 12_1B:

- André Flores - up201907001
- Diogo Faria - up20907014
- Tiago Rodrigues - up201906807

Especificação

A partir de um conjunto de dados sobre músicas do Spotify vamos utilizar algoritmos de aprendizagem supervisionada, especificamente de classificação, para poder prever a sua popularidade.

No conjunto de dados que nos foi dado, existe informação sobre 2000 músicas, dividida entre 18 diferentes parâmetros.

Algoritmos e Ferramentas

Algoritmos: Decision Trees, K Nearest Neighbor e Support-Vector Machine.

Ferramentas: Python 3, Jupyter Notebook, pandas, Seaborn, Matplotlib e Scikit-learn.

Trabalho Implementado

Foi realizado um pré-processamento sobre os dados, sendo que se fez a verificação de dados nulos, os quais não se encontraram nenhum, e também a eliminação de dados de erro no campo de 'genre', onde se encontravam valores como: 'set()'.

Sendo que vamos utilizar algoritmos de classificação e o atributo que pretendemos que os modelos possam prever está dado com valores entre 0 e 100, foi necessário uma transformação sobre eles, sendo que se agregaram em 4 diferentes valores: 'Low', 'Average', 'High' e 'Very High', de forma a que cada valor apresentasse um conjunto de dados semelhante.

Trabalho implementado (Cont.)

De forma a reduzir o número de atributos a utilizar, criamos uma matriz de correlação que nos permitisse agregar atributos.

No entanto, sendo que a maior correlação encontrada foi de apenas 0.65/1, decidimos que nenhum dos atributos poderiam ser agregados.

