

Implementasi Decision Tree untuk Gender Classification

Ide :

Penerapan decision tree learning untuk menentukan gender seseorang berdasarkan atribut-atribut yang sudah ditentukan. Mencari entropy dari atribut dari dataset yang telah diperoleh untuk dapat membuat decision tree yang nantinya dipakai dalam menentukan gender.

Proses pengerjaannya menggunakan pemrograman Python beserta library pandas dan sklearn untuk membantu proses klasifikasi dan tree. Setelah itu menggunakan library six ,IPython.display, dan pydotplus untuk melakukan visualisasi tree yang memiliki detail penjelasannya juga.

Setelah program memberikan hasil decision tree, maka user diminta untuk memasukkan beberapa input mengenai atribut tersebut dan program akan memberikan output berupa Male atau Female (berdasarkan hasil dari proses klasifikasi diawal).

Data Train :

Data diambil dari website kaggle.com yang berupa data mengenai gender classification. Link : <https://www.kaggle.com/hb20007/gender-classification>

Favorite Color	Favorite Music Genre	Favorite Beverage	Favorite Soft Drink	Gender
Favorite color (colors reported by respondents were mapped to either warm, cool or neutral)	Favorite broad music genre	Favorite alcoholic drink	Favorite fizzy drink	Binary gender as reported by the respondents
Cool 56% Warm 33% Other (7) 11%	Rock 29% Pop 26% Other (30) 45%	Doesn't drink 21% Beer 20% Other (39) 59%	Coca Cola/Pepsi 48% Fanta 21% Other (20) 30%	2 unique values
Cool	Rock	Vodka	7UP/Sprite	F
Neutral	Hip hop	Vodka	Coca Cola/Pepsi	F
Warm	Rock	Wine	Coca Cola/Pepsi	F
Warm	Folk/Traditional	Whiskey	Fanta	F
Cool	Rock	Vodka	Coca Cola/Pepsi	F
Warm	Jazz/Blues	Doesn't drink	Fanta	F
Cool	Pop	Beer	Coca Cola/Pepsi	F
Warm	Pop	Whiskey	Fanta	F
Warm	Rock	Other	7UP/Sprite	F