

Redwine quality predictor

Easv 2021 - Machine learning

Introduktion

Ved hjælp af machine learning modellen "Random Forest Regressor" vil jeg forudsige kvaliteten af rødvin.

Datasæt

<http://archive.ics.uci.edu/ml/datasets/Wine+Quality>

Datasættet dækker over portugisiske vine af typen "Vino Verde". Der er 11 målbare parametre på vinene (eks: alkohol) og 1 parameter "kvalitet" som giver en score på vinen fra 1 til 5.

Repository for opgaven

https://github.com/KimRamgaard/Exam_Project_2020

Noter at Wine_Quality_2020_Old.ipynb, kun er med af historiske grunde.

Beskrivelse af opgaven

Jupyter notebooken er delt op i 4 hovedgrupper:

Download and analyze

Her downloader jeg datasættet og analysere det, så jeg har en generel ide om hvad der skal forberedes inden at machine learning algoritmen køre igennem. Det giver også belæg for at kunne forudsige det endelige output.

Data Preparation

Forbered datasættet, her bliver predicts og label-sættet som algoritmen kan bruge til at "gætte" værdierne defineret.

Predict-Værdierne bliver også skaleret så algoritmen har nemmere ved at bruge dem.

Random Forest Regressor

Her starter machine learning modellen "Random Forest Regressor" med predefined hyperparameter med at træne, hvorefter output bliver analyseret

Tweaking Hyper Parameters

Her forsøger jeg at forbedre resultatet ved at justere på hyperparametrene.

Konklusion

Random forest regressor ser ud til at virke fint på datasættet. Det er muligt at jeg har problemer med datasættets størrelse, men ud fra testresultaterne ser der ikke ud til at være under/overfitting.

I forhold til resultatet fra opgaven, er min konklusion at alkohol er den største faktor når der bliver dømt på kvalitet. Det er nok også tilfældet herhjemme.