

QBUS2820 Assignment 1

This version was compiled on September 29, 2020

Introduction

Data processing and exploratory data analysis

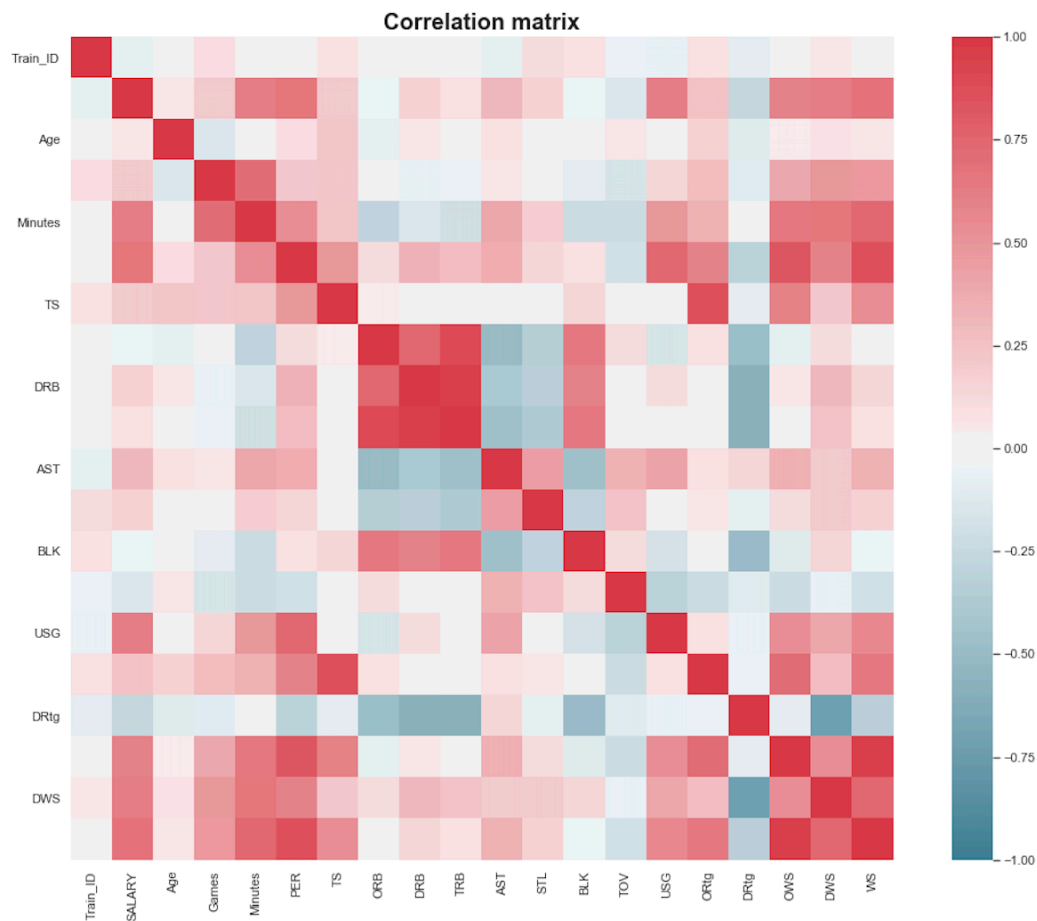


Fig. 1. Caption

Feature engineering

Methodology of the linear regression and kNN regression models

Variable	Number of neighbours	Validation error
DWS	20	4.1311
WS	8	4.1856
Minutes	9	4.2967
OWS	27	4.5046
PER	18	4.7217

(a) Top 5 K nearest neighbour models with the highest validation errors.

Polynomial degree	Validation error
[2, 3]	4.340583
[2, 3, 4]	4.350843
[2]	4.361685
[2, 3, 4, 5]	8.342030
[2, 3, 4, 5, 6]	8.407928

(b) Top 5 polynomial linear regression models with the highest validation errors.

Fig. 4. Validation errors of 10 models developed.

Methodology of the model that is not covered in this unit

Test set performance