# CAN BOOKMAKERS PREDICT THE WINNER OF THE EUROVISION SONGCONTEST?

The annual Eurovision Songcontest ends with a final round of the 27 best contestants of the contest. Each year, bookmakers try to predict who is (not) going to win the contest. The lower the odds, the higher the chance of winning the competition. The higher the odds, the less likely a contestant will win and the more money you'll get. The outcomes of the odds can be seen as a prediction of the final positions of the competition. Different machine learning algorithms have been used to discover if the bookmakers are always right.

## DATA

The position is defined as the y-variable; the variable that will be predicted by the algorithm. The y-variable will rely on the X-variables. These are the defined as the odds.

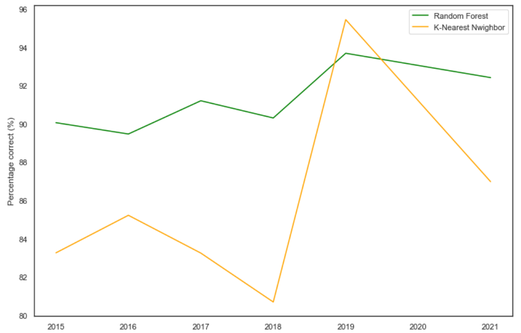
y X

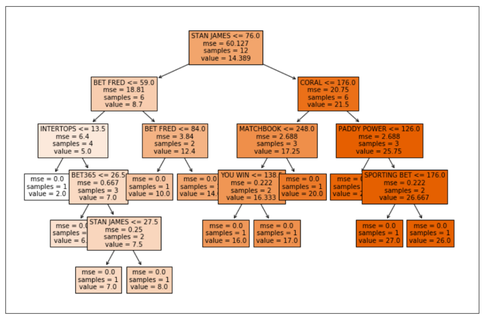
| | contestant | position | BET365 | UNIBET | YOU WIN | BOYLE SPORTS | CORAL | SKY BET | BET VICTOR | ... | BETFAIR SPORT | SPORTING BET | INTERTOPS | PADDY POWER | BET FRED | BETWAY | STAN JAMES | BWIN | MATCHBOOK | year |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Sweden Måns Zelmerlöw - Heroes | 1 | 2.1 | 2 | 2.38 | 2.25 | 2.1 | 2.1 | 2.1 | ... | 2.1 | 2 | 2.25 | 2 | 2 | 2.1 | 2.25 | 1 | 2.21 | 2015 |
| **1** | Russia Polina Gagarina - A Million Voices | 2 | 3.75 | 4.5 | 4 | 3.75 | 4.5 | 3.75 | 4 | ... | 4 | 4 | 4 | 4 | 4.5 | 4 | 3.5 | 4.75 | 5.16 | 2015 |
| **2** | Italy Il Volo - Grande Amore | 3 | 5 | 6 | 5 | 5 | 5.5 | 5.5 | 6 | ... | 6 | 5 | 4.25 | 5.5 | 6.5 | 4.5 | 4.5 | 5 | 6.74 | 2015 |
| **26** | Germany Ann Sophie - Black Smoke | 27 | 251 | 201 | 201 | 201 | 251 | 201 | 251 | ... | 151 | 151 | 151 | 201 | 101 | 151 | 126 | | 496 | 2015 |

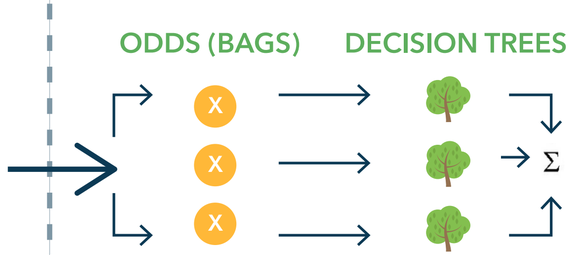
The data is retrieved with the Python webscraping library Selenium from the website https://eurovisionworld.com/odds/eurovision. The table shows data for the year 2015. To get the data from 2015-2019 and 2021, I created a loop to iterate through the years. Because the prediction should rely on the same X-variables, the dataset was cleaned by removing useless columns. I discovered each year included different odd categories, however, the values seemed inline with the other years, except for BET EXCHANGE. This column included very low and high values compared to the other odds. Therefore, this column was removed.

## PREDICTIONS

Two algorithms were running in the competition of the best algorithm for predicting the position based on odds: Random Forest and K-Nearest Neighbor. Because the data does not include categories, the regressor-version of both algorithms was used for the predictions. Random Forest (91.19 % average accuracy) performed better than K-Nearest neighbor (85.80% average accuracy).




The Random Forest Regressor algorithm was build with 100 trees. For each tree, the algorithm relies on the X-variables as an indicator ("bag") for finding if an odd-value is going to predict a low or high position of the contestant.

## DISCUSSION

The Random Forest model can predict 91.19% on average. This means that bookmakers can predict 91.19% right. The Random Forest model relied on pure X-variables. Therefore, no logarithmic transformation was taken into account. In order to improve the model, the log of the odds may improve the model. Furthermore, as we can conclude from the comparison, for each year the model performs different. In the future, another model (like K-Nearest Neighbor) might be a better predictor.