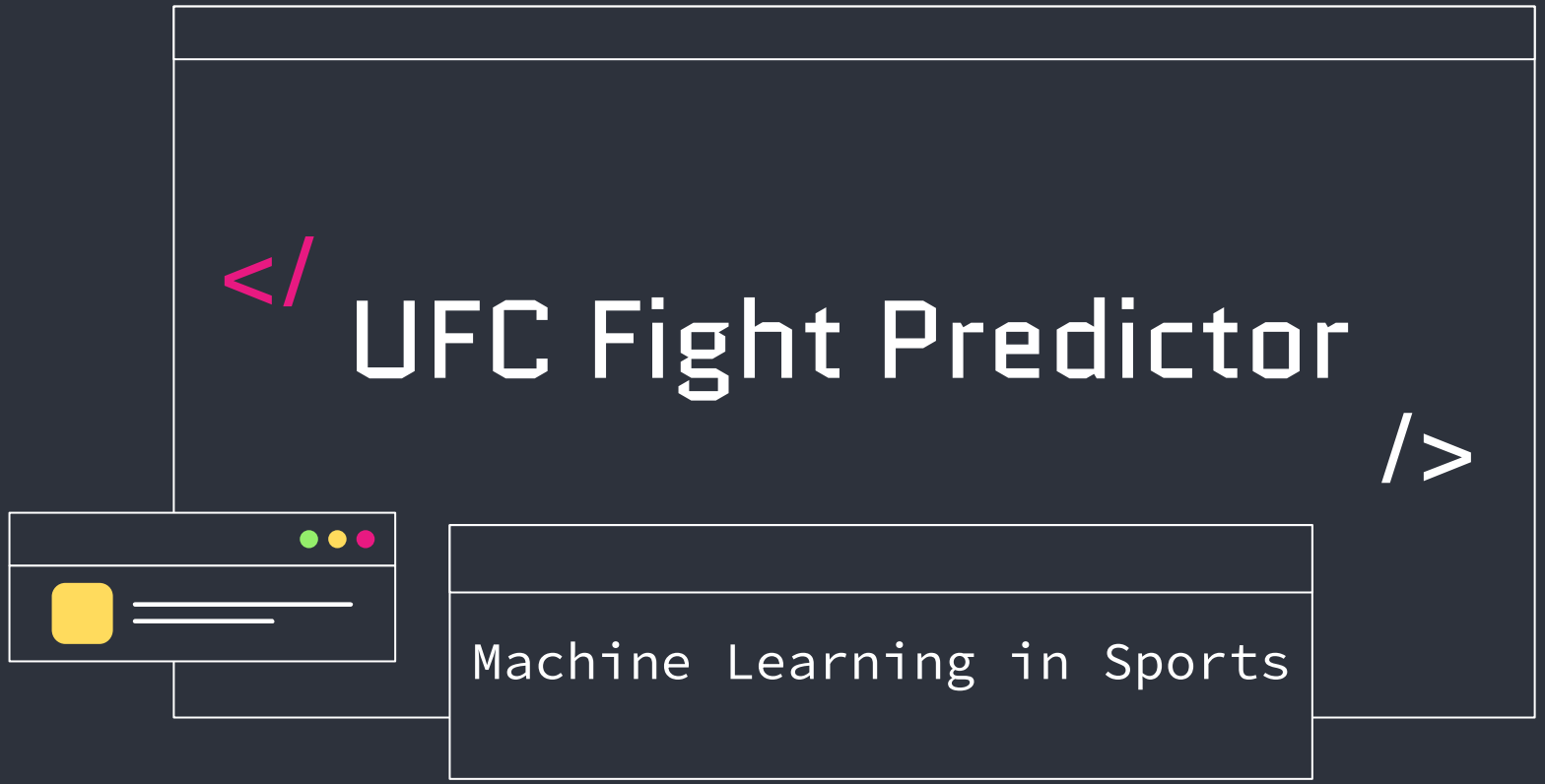


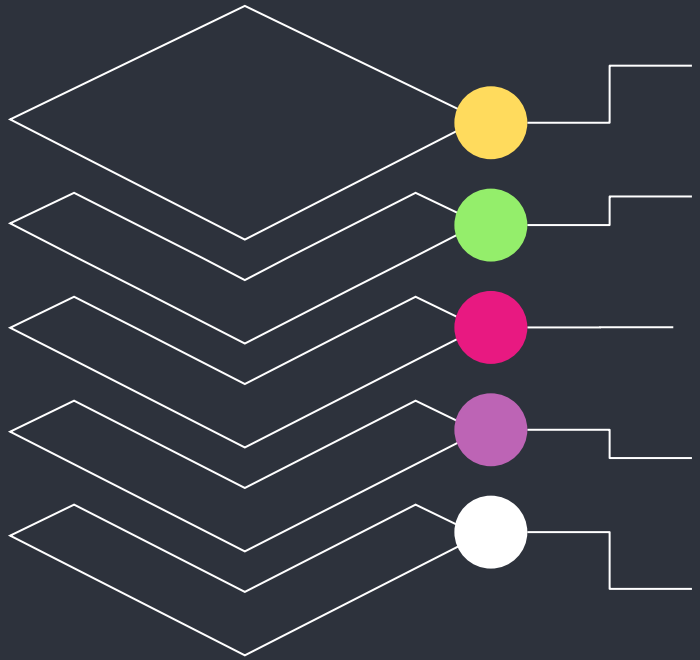
1 0 1 1 0 1 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 1 0 1 1 0 1

</Can something as unpredictable as MMA fights be predicted?

- MMA is a complex combat sport
- Nearly everything is allowed (punches, kicks, takedowns, grappling, …)
- Challenge: Create a machine learning model what predicts the Winner of a MMA fight as good as possible

1 0 1 1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1

</The data set



Source: www.ufcstats.com

Method: Web Scraping

7k+ Rows Every row represents a fight with corresponding fighter data

144 Features [Columns]

1994 to present

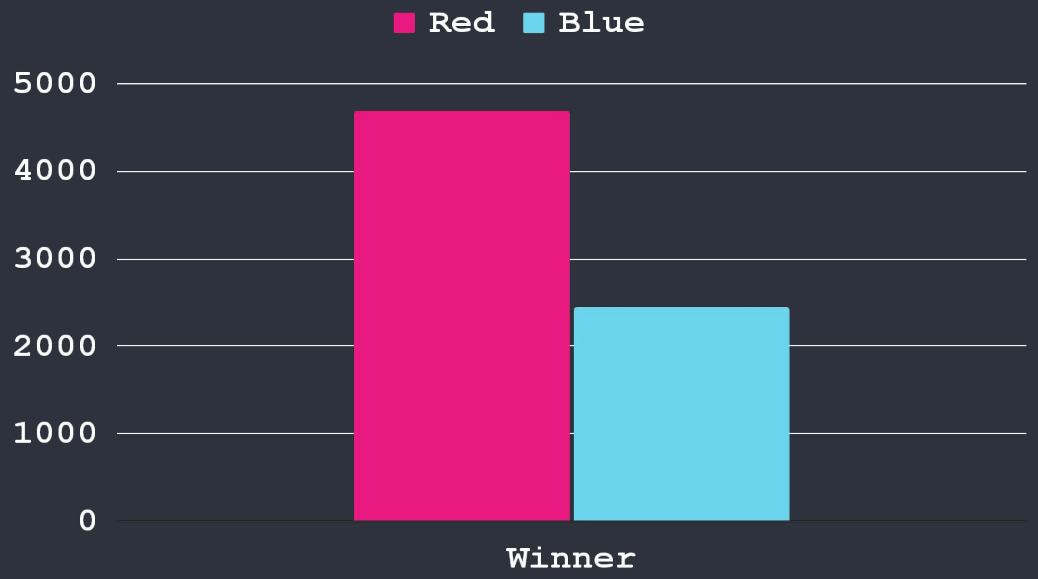
</Defining the machine learning problem

- Did the blue or the red fighter win?
  - Discrete target variable → classification
  - Labelled input and output data → supervised machine learning
- Draws are rare and ignored for the model

- What to consider for classification?
  - check data for imbalance
  - Normalize numerical features, one-hot-encode categorical features

1 0 1 1 0 1 1 0 1 1 0 1 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1 0 1 1 0 1 1 1 1 1 0 1

</Imbalance of the data



65,69%

34,31%

→ Slightly imbalanced data set

</Machine learning models used

Classification Models

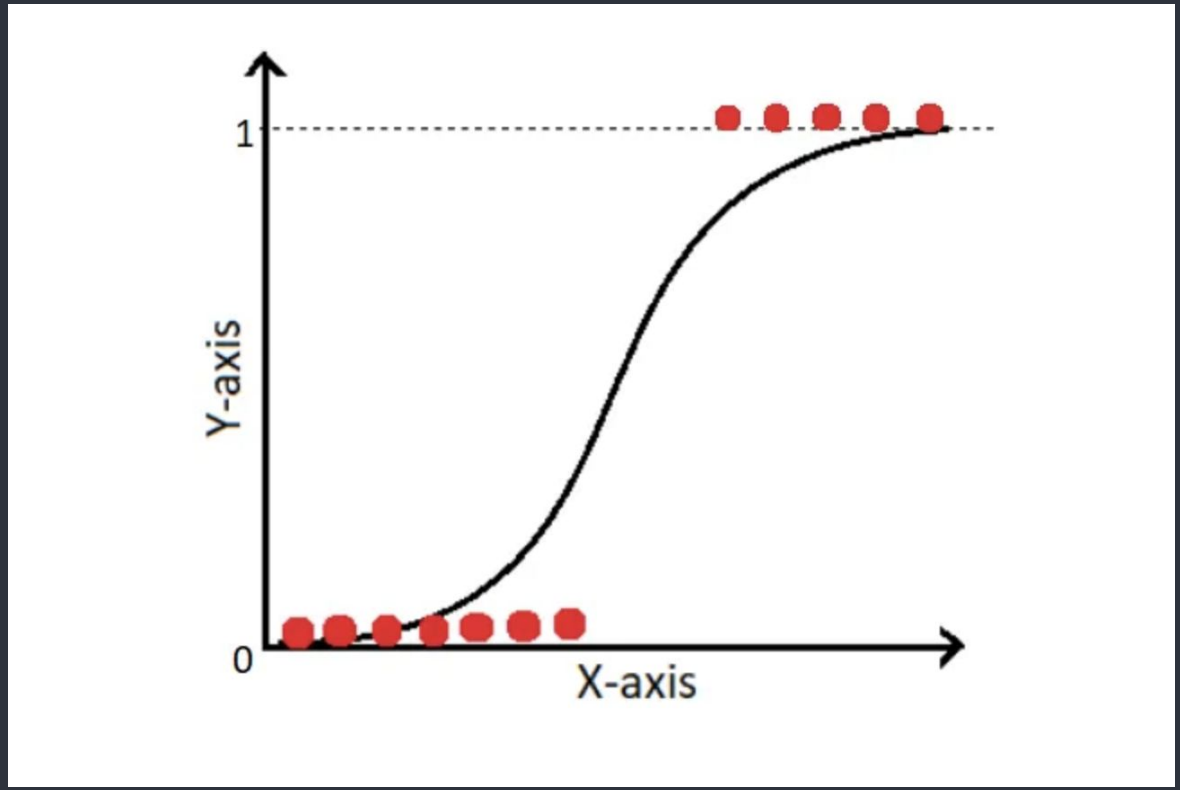
XGBoost

eXtreme Gradient Boosting

Logistic Regression

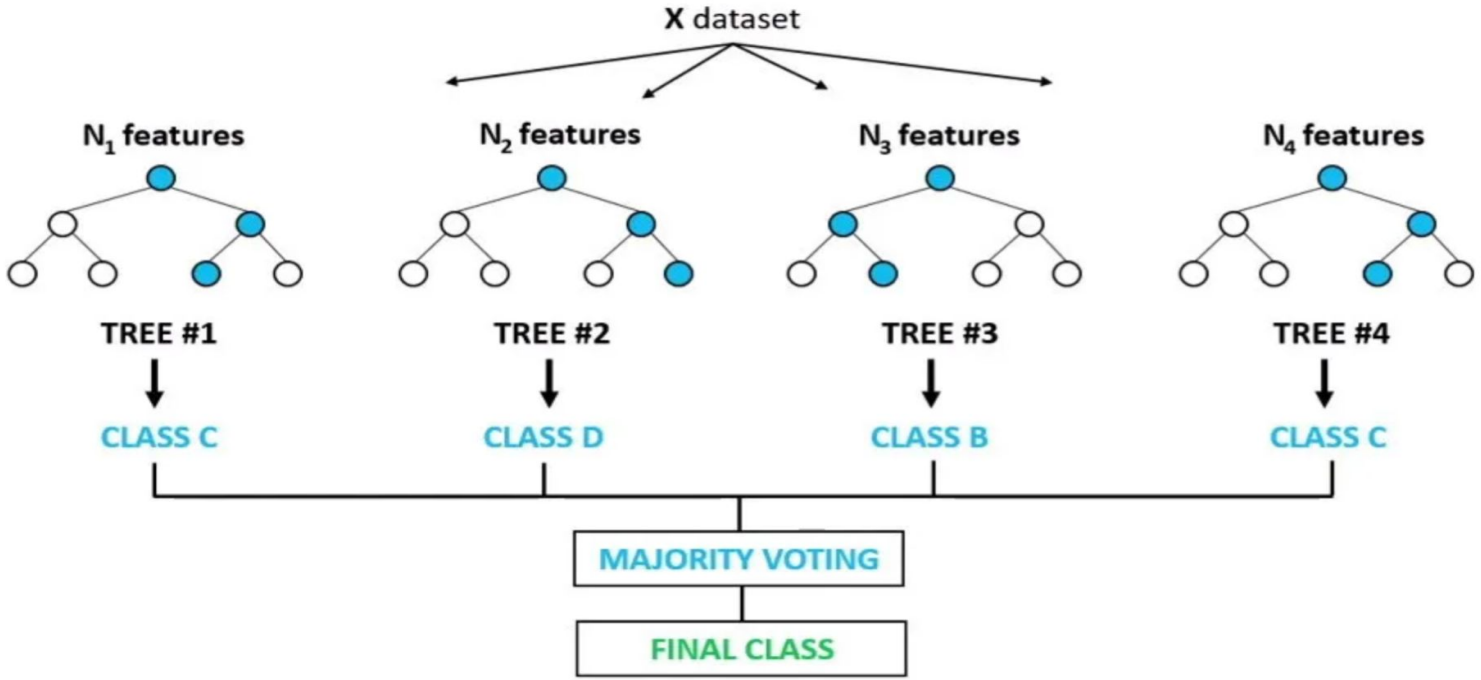
Random Forest Classifier

</Logistic Regression



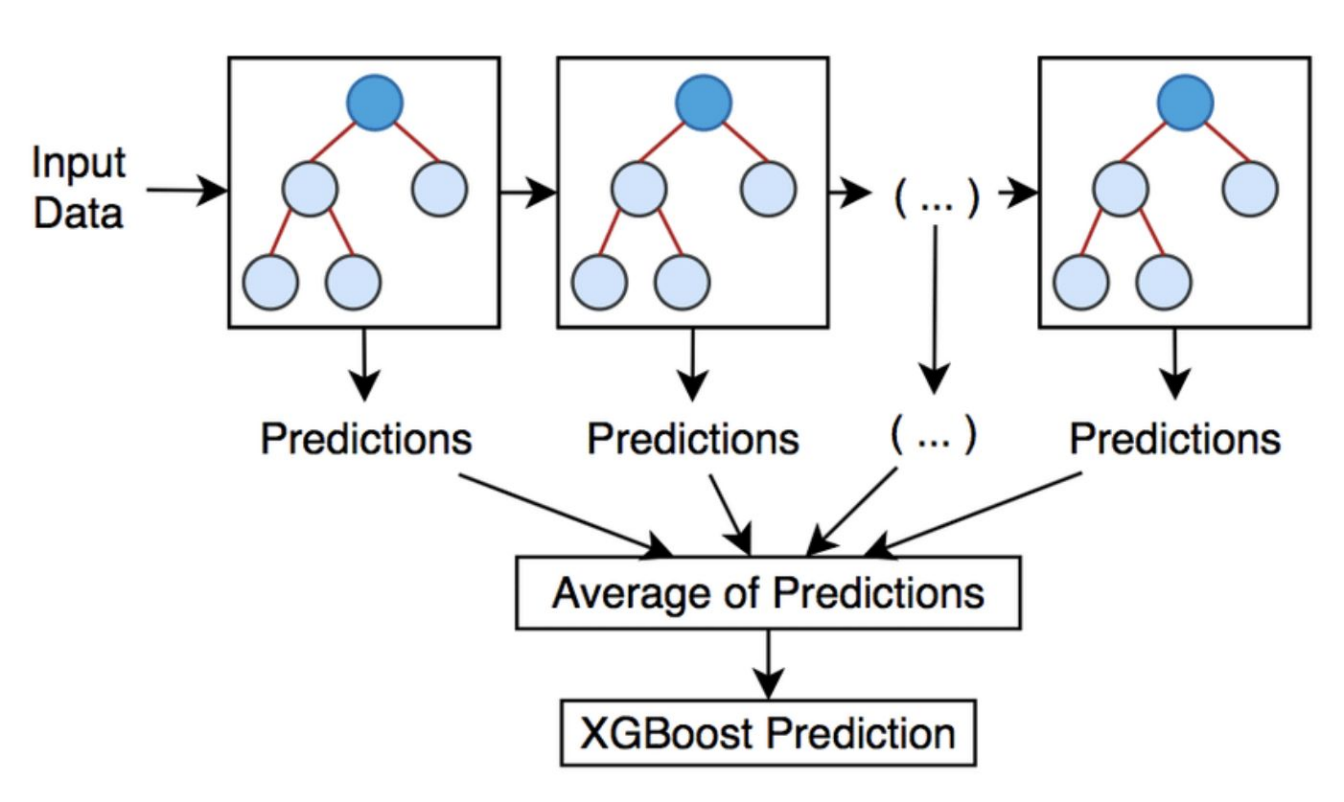
Source: https://medium.com/analytics-vidhya/the-math-behind-logistic-regression-c2f04ca27bca

Random Forest Classifier



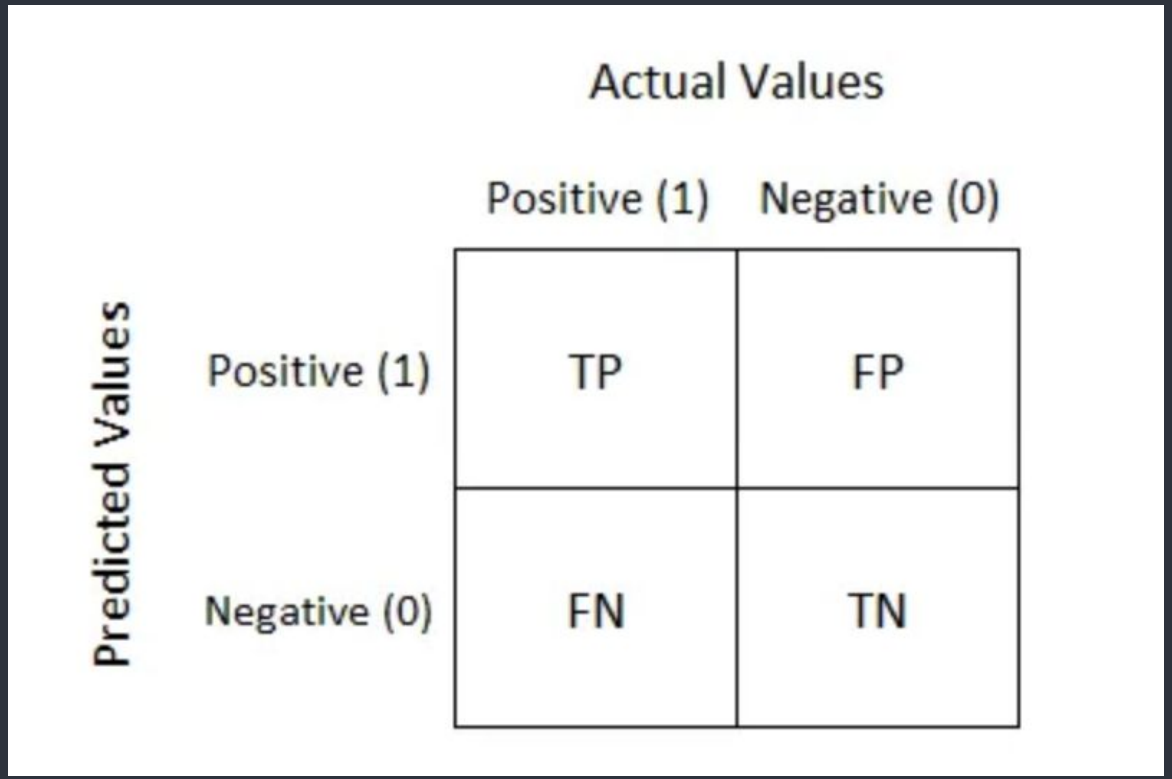
Source: https://medium.com/analytics-vidhya/random-forest-classifier-and-its-hyperparameters-8467bec755f6

</XGBoost



Source: https://www.researchgate.net/figure/XGBoost-model-Source-Self_fig2_350874464

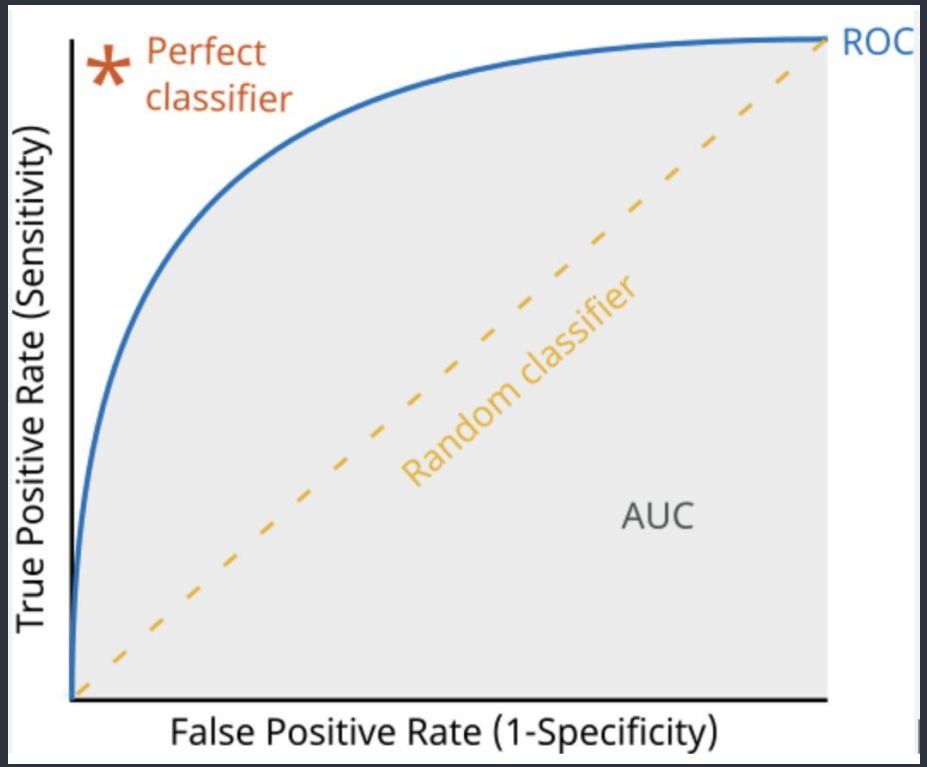
</Accuracy score



Accuracy Score =
(TP + TN)/
(TP + FP + FN + TN)

Source: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

</AUC-ROC Curve



The bigger the AUC (Area Under The Curve) the better the model

Source: https://www.mathworks.com/help/deeplearning/ug/compare-deep-learning-models-using-ROC-curves.html

</AUC + accuracy score of my models

{1/4} XGBoost V2

Accuracy: 0.6648
AUC Score: 0.7361

{2/4} Logistic Regression

Accuracy: 0.6596

{3/4} Random Forest

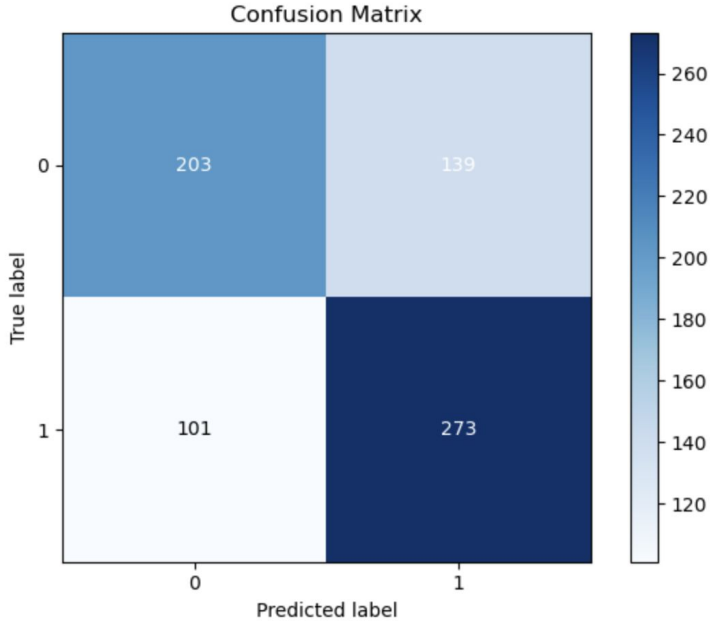
Accuracy: 0.6564

{4/4} XGBoost v1

Accuracy: 0.6453
AUC Score: 0.6515

</AUC-ROC curve

```
Model Report:
Accuracy (valid): 0.6648
AUC Score (valid): 0.7361
<Axes: title={'center': 'Confusion Matrix'}, xlabel='Predicted label', ylabel='True label'>
```



Source: Self

</Final evaluation

- Accuracy Score is around the probability of predicting the favored fighter (red fighter) will win
- Further feature engineering and model improvement is needed

</Feedback

- Things to in-/exclude for the future:
  - Explain the features, make it more tangible
  - Feature importance for the models
  - Shorten the part with the models. Maybe focus on one feature
  - Maybe exclude AUC