

UFC MATCH PREDICTOR

BY: AREF JADDA

THE FULL PROCESS



Data Cleaning

Machine Learning
with Scikit Learn

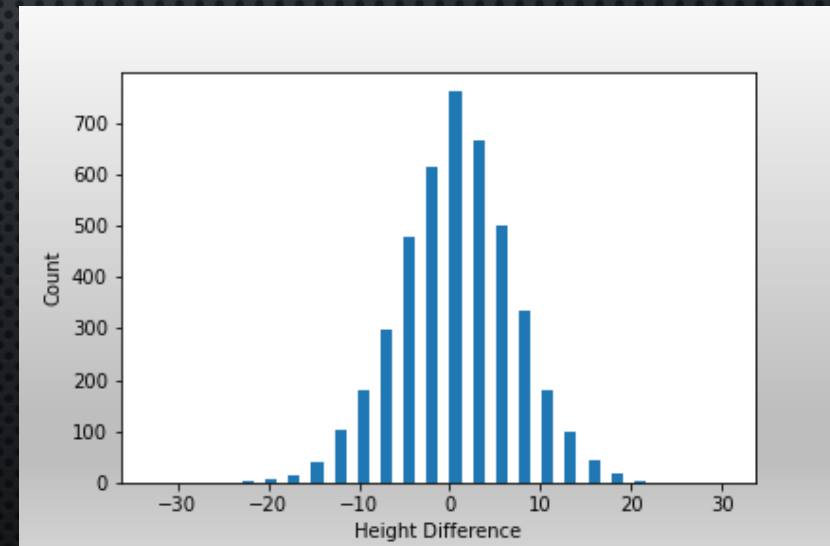
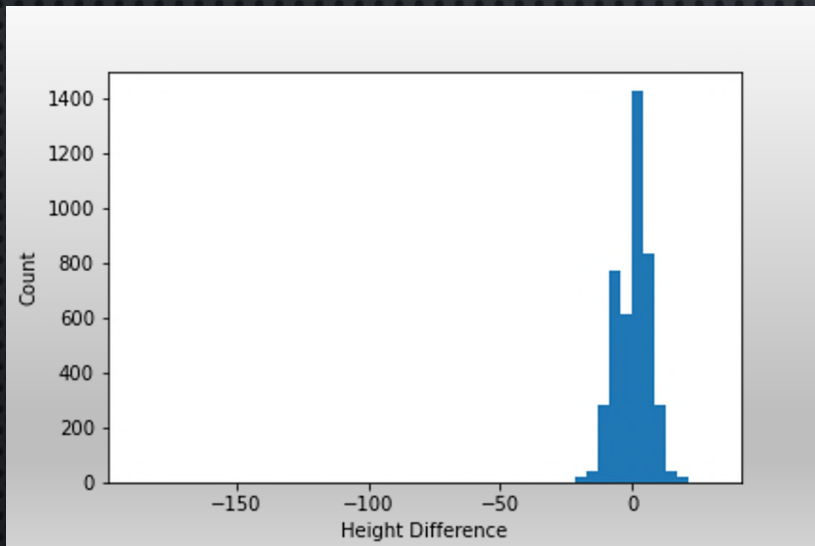
Deep Learning

Building An App



DATA CLEANING UPDATES

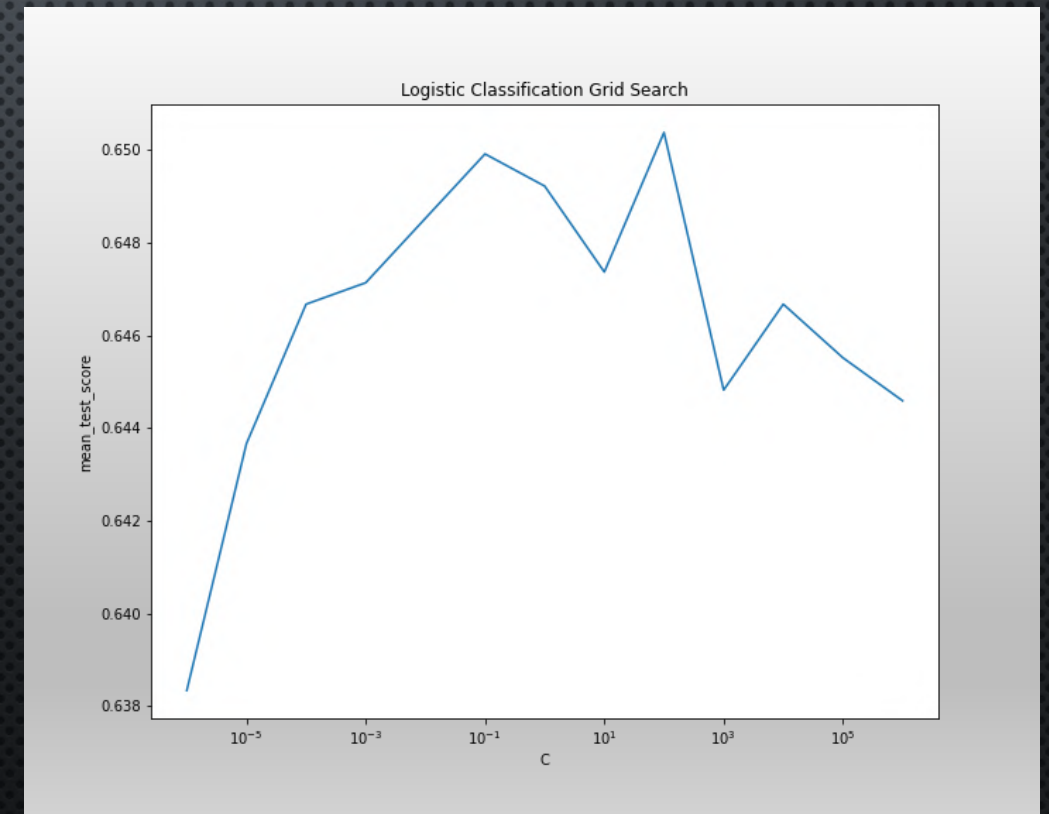
- Feature engineering
- Ex: 'reach_dif' = 'B_reach' - 'R_reach'
- Columns with collinearity higher than 70% were removed
- Some columns were normalized



LOGISTIC CLASSIFICATION MODEL

- Highest Accuracy Achieved: 69%

	Feature	Coefficient
65	win_by_Decision_Split_dif	0.157491
56	R_Southpaw	0.085700
19	Brazil	0.063082
49	Welterweight	0.061521
1	no_of_rounds	0.059643
66	win_by_TKO_Doctor_Stoppage_dif	0.056362
14	avg_td_dif	0.052930
38	USA	0.052726



XG BOOST

- Highest Accuracy Achieved: 71%

Test Set Scores:

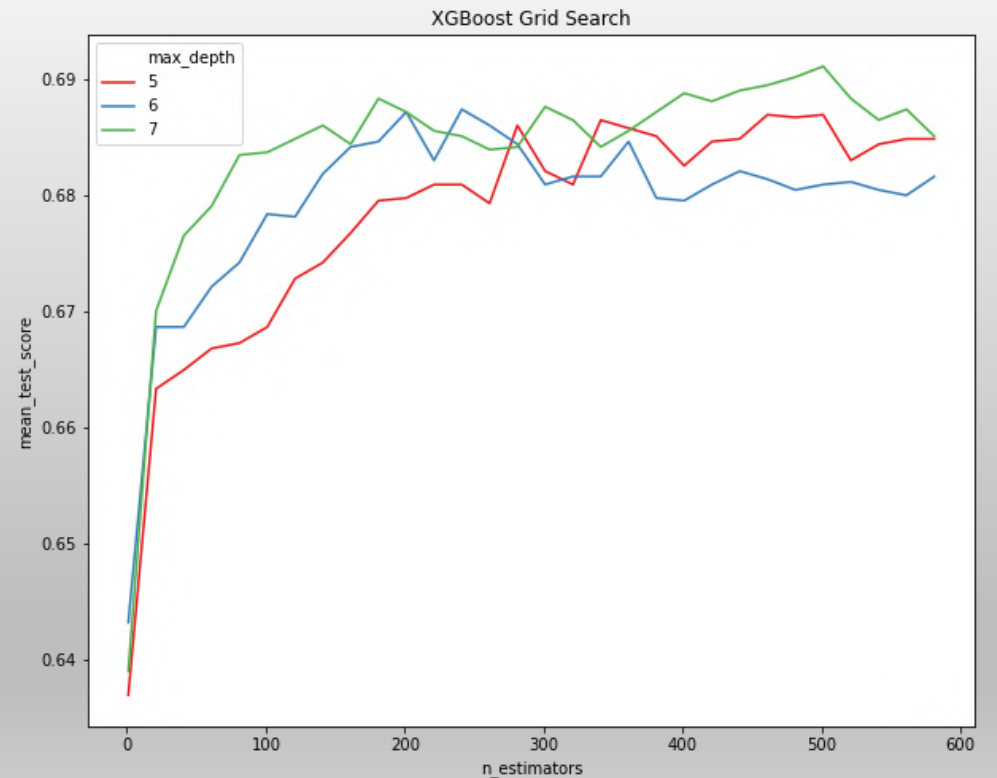
Random Forest score: 0.7577160493827161

Decision Tree score: 0.6697530864197531

AdaBoost score: 0.6697530864197531

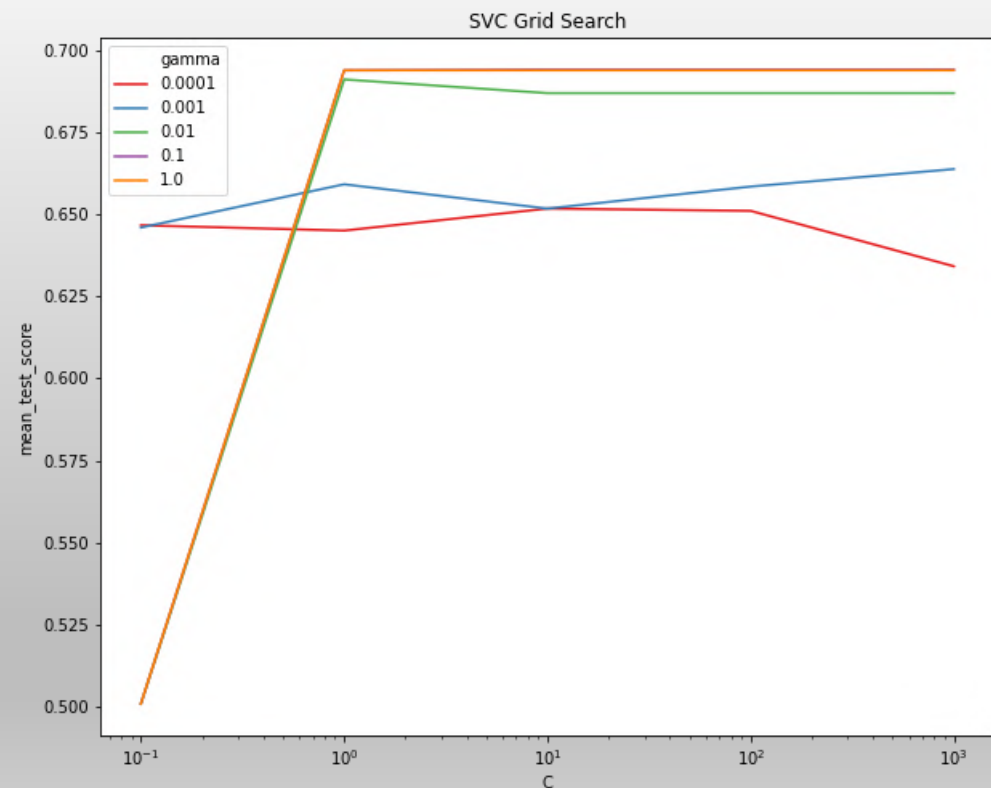
Gradient Boost score: 0.6851851851851852

XG Boost score: 0.6882716049382716



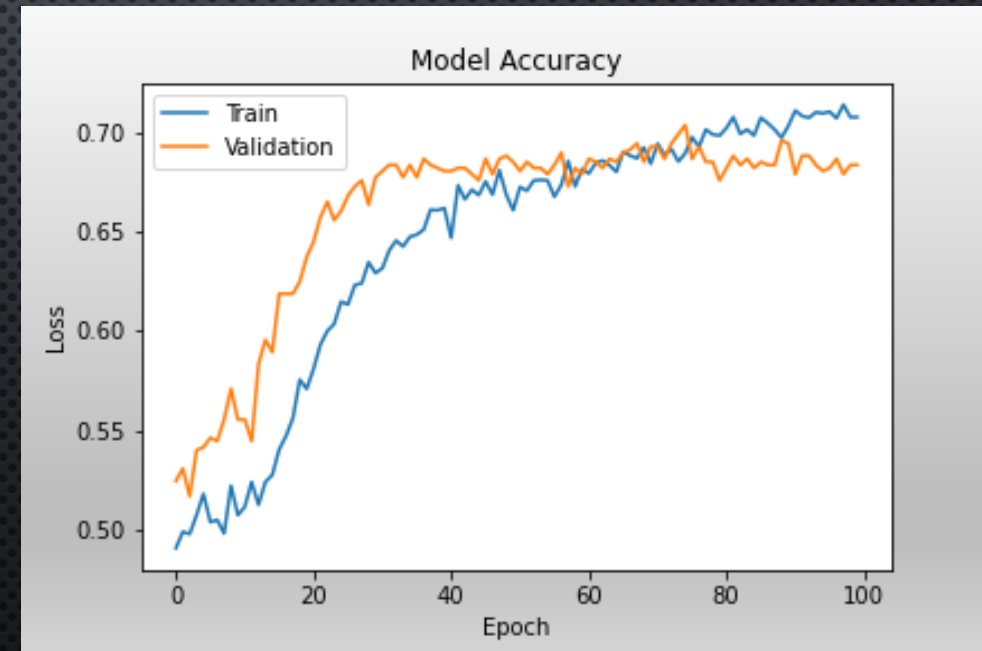
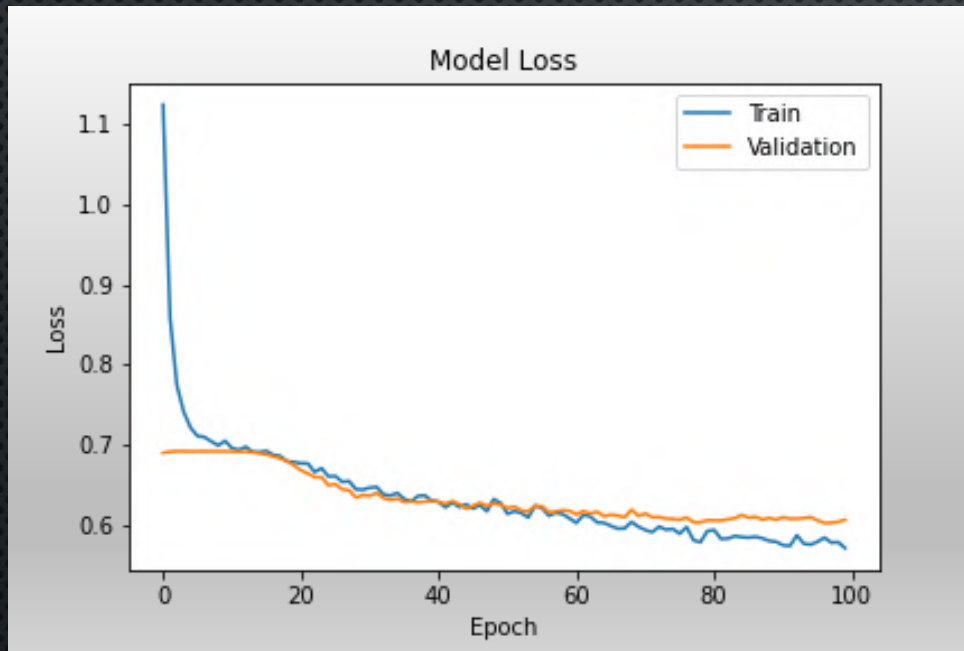
SUPPORT VECTOR CLASSIFIER

- Highest Accuracy Achieved: 74%
- Best model overall!

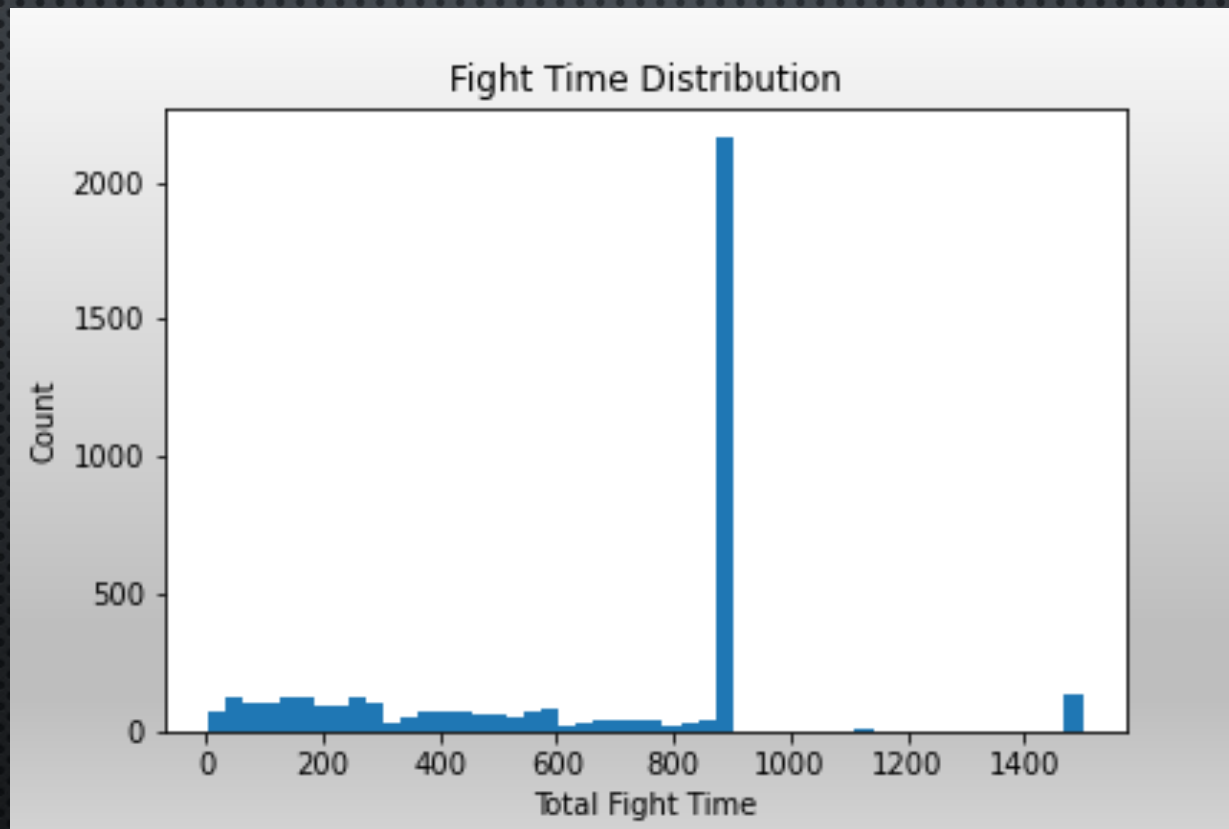


ARTIFICIAL NEURAL NETWORK

- Highest Accuracy Achieved: 70%



WHY I REMOVED ROUND PREDICTION



NEXT STEPS

During this week I will:

- Organize my findings and write the report
- Try to make a GUI with Streamlit

Some considerations for better future results:

- More data, as much as possible!
- Added columns such as:
 1. Main fighting style (grappler, striker, mixed)
 2. Original martial art (ex. They started with wrestling or karate)
 3. Do they have children?