Syracuse University

IST 718
Final Project

Team 4

Brandon Bergstrom

Mark R Paradis

Charles Vanleuvan

George Smith

Project Overview

- Analyze historical data on PFL fights
- What is the PFL? (Professional Fighters League)
  - Mixed Martial Arts League:
    - Mixed martial arts (MMA) is a full-contact combat sport that allows a wide variety of fighting techniques and skills from a mixture of other combat sports to be used in competition. The rules allow usage of both striking and grappling techniques while standing and on the ground
- Analyze historical data on PFL fights
- Data attributes include Punching, Kicking, Grappling, etc.
- Data was collected by scraping Glorykickboxing.com using the Beautiful Soup package in Python

Professional Fight League Data

- SmartCage™
  - Proprietary technology for recording Realtime fight analytics
  - New levels of interactivity with fans and sportsbooks
- Cagenomics™
  - Delivery of SmartCage data for informative fight analysis
  - Started with Strike Speed
  - Kicks Speed to be delivered in 2022
  - On the horizon
    - “The Burn” – Total Calories burned by fighter in a match
    - Distance traveled – Movement and control of the ring by fighter
- Data and Technology Monetization
  - Selling IP to other MMA and fighting leagues and become premier fight data vendor

Business Questions

- Can we help fighters identify weaknesses in opponents?
- Are there long-term trends in the use of historical data?
- What are the key factors that leads to a fighter winning a fight ?
- Can we profit off fight predictions?

Goals

- Create a model that finds weaknesses in fighters and predicts a winner
  - Why this is important? Two reasons: provide analytics to fighters and earn money betting
- Focus on certain features, unique data, or unique models in order to produce a profit
  - Through feature engineering, identify the fighting stats that have the most influence to win probability.
- Make a profit using our data, features, and model
  - Develop a model that can repeatedly earn positive returns over the course of an entire betting cycle.
  - Use the bookmaker's odds against them to identify fights that have higher than average expected payout

Observe

54 Attributes

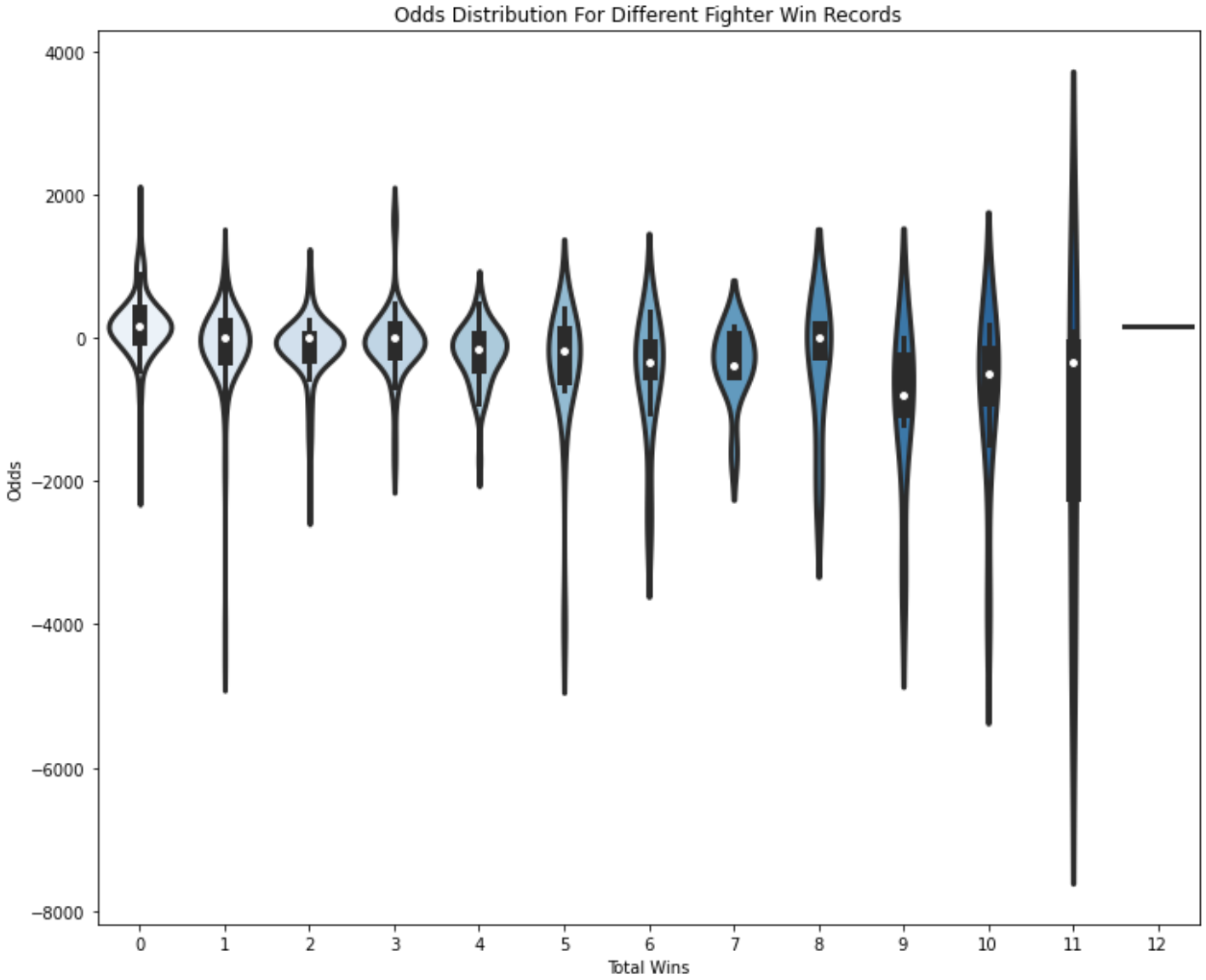
```
Index(['Sig. Str. Leg', 'Sig. Str. Leg Attempts', 'Opp Sig. Str. Leg',
       'Opp Sig. Str. Leg Attempts', 'Sig. Str. Ground Landed',
       'Sig. Str. Ground Attempts', 'Opp Sig. Str. Ground Landed',
       'Opp Sig. Str. Ground Attempts', 'Total Strikes Landed',
       'Total Strikes Attempts', 'Opp Total Strikes Landed',
       'Opp Total Strikes Attempts', 'Total Takedowns Landed',
       'Total Takedowns Attempts', 'Opp Total Takedowns Landed',
       'Opp Total Takedowns Attempts', 'Submission Total Attempts',
       'Opp Submission Total Attempts', 'Takedown %', 'Opp Takedown %',
       'Sig. Str. Leg %', 'Opp Sig. Str. Leg %', 'Sig. Str. Ground %',
       'Opp Sig. Str. Ground %', 'Total Strikes %', 'Opp Total Strikes %',
       'KnockDown Total', 'Opp KnockDown Total', 'KnockDown Total Difference',
       'Ground Time', 'Opp Ground Time', 'Standing Time', 'Opp Standing Time',
       'Dominant Positions', 'Opp Dominant Positions', 'Top Strike Speed',
       'Opp Top Strike Speed', 'Height', 'Weight', 'Reach', 'Leg Reach',
       'Points', 'Inner Zone', 'Opp Inner Zone', 'Outer Zone',
       'Opp Outer Zone', 'Sig. Str. Per KnockDown',
       'Opp Sig. Str. Per KnockDown', 'Sig. Str. Punches Diff',
       'Sig. Str. Diff', 'Sig. Str. Leg Diff', 'Sig. Str. Ground Diff', 'Odds',
       'Total Wins'],
      dtype='object')
There are 54 attributes.
```

Observe

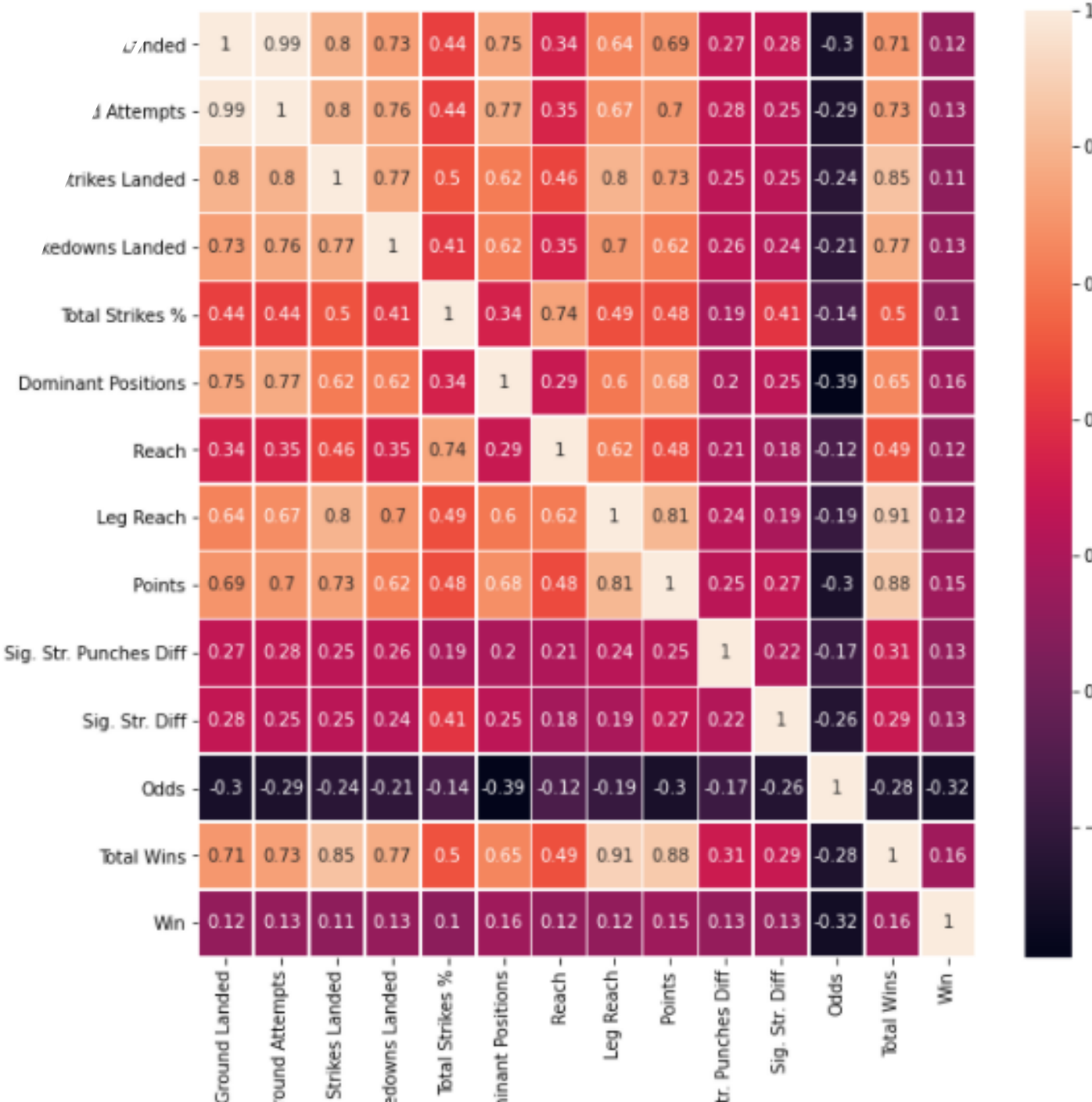
210 Fighters

```
['Francimar Barroso' 'Daniel Gallemore' 'Alexandre Almeida' 'Lee Coville'
 'Magomed Idrisov' 'Steven Siler' 'Caio Alencar' 'Kelvin Tiller'
 'Marcos Galvao' 'Nazareno Malegarie' 'Jared Rosholt' 'Valdrin Istrefi'
 'Alex Nicholson' 'Jake Heun' 'Timur Valiev' 'Max Coga' 'Lance Palmer'
 'Bekbulat Magomedov' 'Andre Harrison' 'Jumabieke Tuerxun' 'Josh Copeland'
 'Jack May' 'Dan Spohn' 'Bozigit Ataev' 'Chris Wade' 'Natan Schulte'
 'Thiago Tavares' 'Robert Watley' 'Maxim Grishin' 'Jason Butcher'
 'Rakim Cleveland' 'Rashid Yusupov' 'Vinny Magalhaes' 'Jamie Abdallah'
 'Islam Mamedov' 'Yuki Kawana' 'Smealinho Rama' 'Brandon Halsey'
 'Ronny Markes' "Sean O'Connell" 'Efrain Escudero' 'Jason High'
 'Kayla Harrison' 'Brittney Elkin' 'Brian Foster' 'Ramsey Nijem'
 'Luiz Firmino' 'Will Brooks' 'Bojan Velickovic' 'Jonatan Westin'
 'Danillo Villefort' 'Abus Magomedov' 'Rex Harris' 'Andre Lobato'
 'Louis Taylor' 'Anderson Goncalves' 'Bruno Santos' 'Sadibou Sy'
 'Joao Zeferino' 'Paul Bradley' 'Herman Terrado' 'Magomed Magomedkerimov'
 'John Howard' 'Gasan Umalatov' 'Shamil Gamzatov' 'Eddie Gordon'
 'Rick Story' 'Yuri Villefort' 'Abubakar Nurmagomedov' 'Pavlo Kusch'
 'Jake Shields' 'Ray Cooper III' 'Shawn Jordan' 'Philipe Lins'
 'Arthur Estrazulas' 'Artur Alibulatov' 'Rashid Magomedov' 'Carlton Minus'
 'Jozette Cotton' 'Muhammed Dereese' 'Leroy Johnson'
 'Ramazan Kuramagomedov' 'Robert Hale' 'Umar Nurmagomedov'
 'Saidyokub Kakharamonov' 'Emiliano Sordi' 'Caio Magalhaes' 'Mike Kyle'
 'Handesson Ferreira' 'Alexandre Bezerra' 'Johnny Case' 'Moriel Charneski'
 'Roberta Samad' 'Andre Fialho' 'Chris Curtis' 'Gamzat Khiramagomedov'
 'Glaico Franca' 'Genah Fabian' 'Bobbi Jo Dalziel' 'David Michaud'
 'Sarah Kaufman' 'Morgan Frier' 'Zane Kamaka' 'Larissa Pacheco'
 'Gadzhi Rabadanov' 'Damon Jackson' 'Movlid Khaybulaev' 'Jeremy Kennedy'
 'Luis Rafael Laurentino' 'Bao Yincang' 'Loik Radzhabov' 'Ylies Djiroun'
 'Akhmed Aliev' 'Carlao Silva' 'Peter Petties' 'Nate Andrews'
 'Alex Gilpin' 'Ante Delija' 'Carl Seumanutafa' 'Jordan Johnson'
 'Mikhail Mokhnatkin' 'Sigi Pesaleli' 'Viktor Nemkov' 'Ali Isaev'
 'Denis Goltsov' 'Satoshi Ishii' 'Zeke Tuinei-Wily' 'Jesse Ronson'
 'Freddy Assuncao' 'Ben Edwards' 'Sidemar Honorio' 'Brendan Loughnane'
 'Matt Wagy' 'Nikolai Aleksakhin' 'Daniel Pineda' 'David Alex Valente'
 'Anthony Pettis' 'Clay Collard' 'Marcin Held' 'Lazar Stojadinovic'
 'Bubba Jenkins' 'Sheymon Moraes' 'Joilton Lutterbach' 'Raush Manfio'
 'Mikhail Odintsov' 'Anthony Dizy' 'Jo Sungbin' 'Tyler Diamond'
 'Alexander Martinez' 'Rory MacDonald' 'Curtis Millender' 'Jason Ponet'
 'Gleison Tibau' 'Chris Camozzi' 'Antonio Carlos Jr.' 'Tom Lawlor'
 'Cezar Ferreira' 'Nick Roehrick' 'Marthin Hamlet' 'Fabricio Werdum'
 'Renan Ferreira' 'Mariana Morais' 'Mohammed Usman' 'Brandon Sayles'
 'Julija Pajic' 'Bruno Cappelozza' 'Cindy Dandois' 'Kaitlin Young'
 'Laura Sanchez' 'Olena Kolesnyk' 'Taylor Guardado' 'Magomed Umalatov'
 'Kyron Bowen' 'Claressa Shields' 'Bobby Moffett' 'Olivier Aubin-Mercier'
 'Jesse Stirn' 'Arman Ospanov' 'Cory Hendricks' 'Jordan Young'
 'Aleksei Kunchenko' 'Muhammed DeReese' 'Chandler Cole' 'Klidson Abreu'
 'Jamelle Jones' 'Micah Terrill' 'Darrell Horcher' 'Brett Cooper'
 'Tyler Hill' 'Leandro Silva' 'Michael Lombardo' 'Hopeton Stewart'
 'Elvin Espinoza' 'Stuart Austin' 'Marina Mokhnatkina' 'Claudia Zamora'
 'Miranda Barber' 'Amanda Leve' 'Christian Lohsen' 'Jonas Flok'
 'Jason Knight' 'Alejandro Flores' 'Carl Deaton' 'Jacob Kilburn'
 'Brandon Jenkins' 'Abigail Montes' 'Omari Akhmedov' 'Julia Budd'
 'Don Madge' 'Nate Williams']
There are 210 fighters.
```

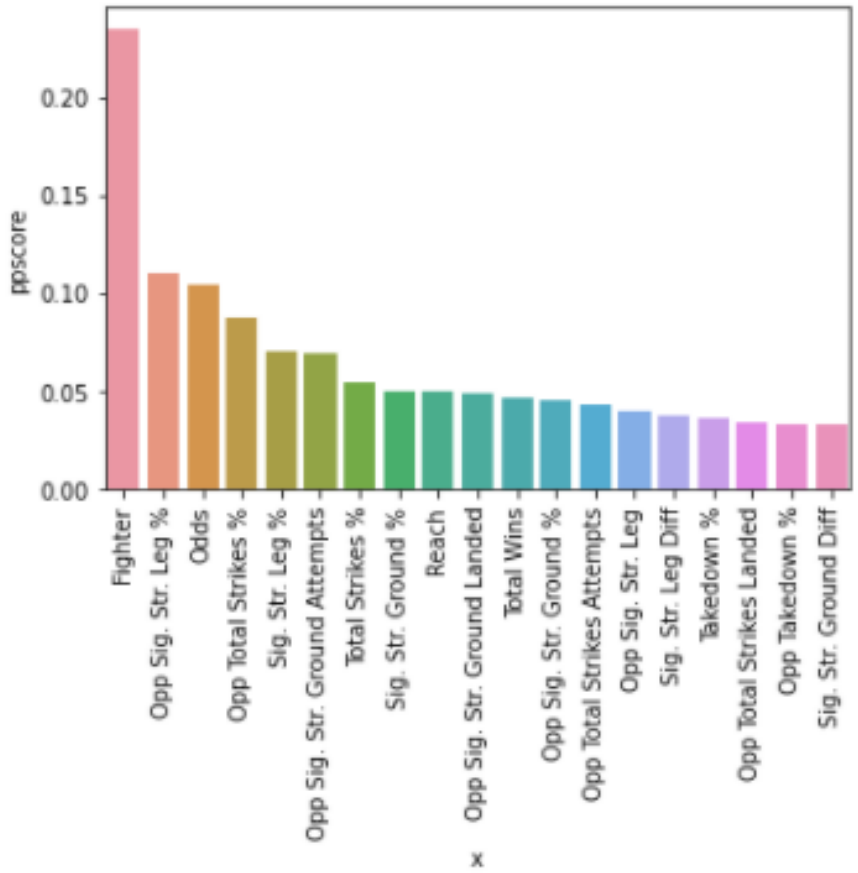
Analysis


Odds Distribution For Different Fighter Win Records
Odds
4000
2000
0
−2000
−4000
−6000
−8000
0
1
2
3
4
5
6
7
8
9
10
11
12
Total Wins

Analysis


nded 1 0.99 0.8 0.73 0.44 0.75 0.34 0.64 0.69 0.27 0.28 -0.3 0.71 0.12
Attempts 0.99 1 0.8 0.76 0.44 0.77 0.35 0.67 0.7 0.28 0.25 -0.29 0.73 0.13
trikes Landed 0.8 0.8 1 0.77 0.5 0.62 0.46 0.8 0.73 0.25 0.25 -0.24 0.85 0.11
edowns Landed 0.73 0.76 0.77 1 0.41 0.62 0.35 0.7 0.62 0.26 0.24 -0.21 0.77 0.13
Total Strikes % 0.44 0.44 0.5 0.41 1 0.34 0.74 0.49 0.48 0.19 0.41 -0.14 0.5 0.1
Dominant Positions 0.75 0.77 0.62 0.62 0.34 1 0.29 0.6 0.68 0.2 0.25 -0.39 0.65 0.16
Reach 0.34 0.35 0.46 0.35 0.74 0.29 1 0.62 0.48 0.21 0.18 -0.12 0.49 0.12
Leg Reach 0.64 0.67 0.8 0.7 0.49 0.6 0.62 1 0.81 0.24 0.19 -0.19 0.91 0.12
Points 0.69 0.7 0.73 0.62 0.48 0.68 0.48 0.81 1 0.25 0.27 -0.3 0.88 0.15
Sig. Str. Punches Diff 0.27 0.28 0.25 0.26 0.19 0.2 0.21 0.24 0.25 1 0.22 -0.17 0.31 0.13
Sig. Str. Diff 0.28 0.25 0.25 0.24 0.41 0.25 0.18 0.19 0.27 0.22 1 -0.26 0.29 0.13
Odds -0.3 -0.29 -0.24 -0.21 -0.14 -0.39 -0.12 -0.19 -0.3 -0.17 -0.26 1 -0.28 -0.32
Total Wins 0.71 0.73 0.85 0.77 0.5 0.65 0.49 0.91 0.88 0.31 0.29 -0.28 1 0.16
Win 0.12 0.13 0.11 0.13 0.1 0.16 0.12 0.12 0.15 0.13 0.13 -0.32 0.16 1
Ground Landed
Ground Attempts
Strikes Landed
edowns Landed
Total Strikes %
inant Positions
Reach
Leg Reach
Points
tr. Punches Diff
Sig. Str. Diff
Odds
Total Wins
Win

Analysis



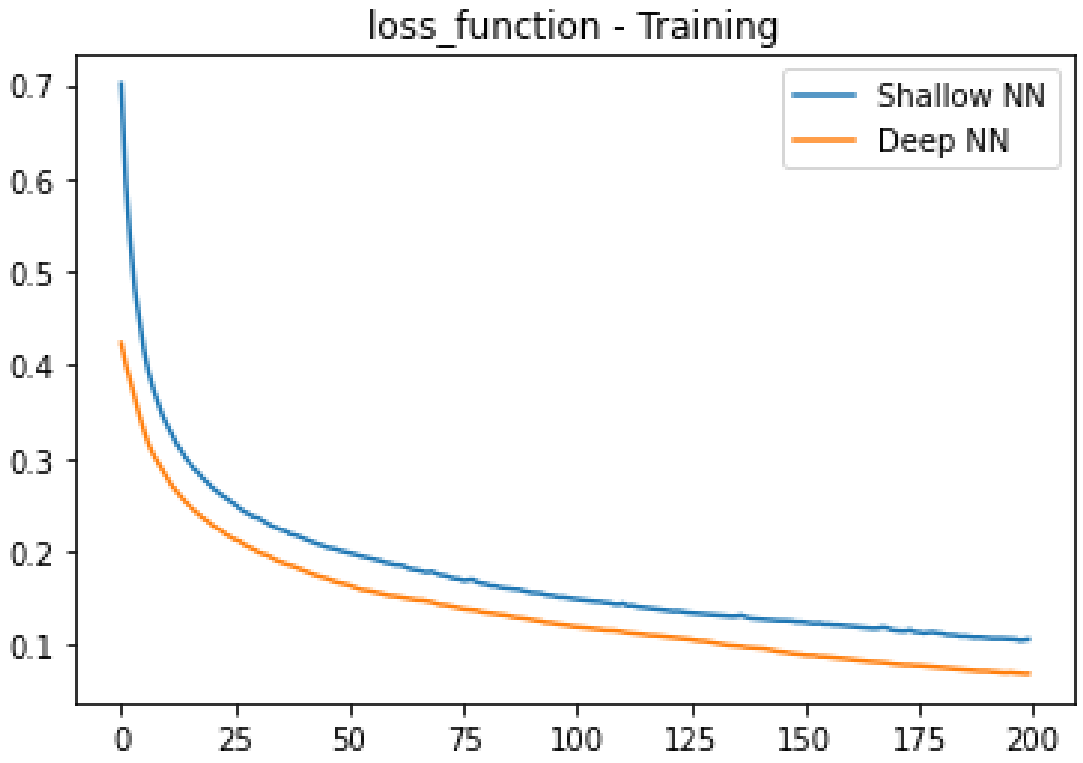
Statistical Significance

```
statistical_list = []


#Regression to find statistically significant columns for columns
from scipy.stats import pearsonr
def corrStats(column):
    return pearsonr(df['Win'],column)

predictors = df.drop(['Fighter','Height','Win'],axis=1)
for x in predictors.columns:
    var = corrStats(df[x])
    if var[1] <= 0.01:
        print(x,":",var[1])
        statistical_list.append(x)
```

```
Sig. Str. Ground Landed : 0.0014338432055135345
Sig. Str. Ground Attempts : 0.0012024254793710776
Total Strikes Landed : 0.005786383729122134
Total Takedowns Landed : 0.0009027497059453231
Total Strikes % : 0.008197739325794724
Dominant Positions : 3.247541601094928e-05
Reach : 0.0020989306301141164
Leg Reach : 0.0025488091405171252
Points : 7.952800376160851e-05
Sig. Str. Punches Diff : 0.001243942449304144
Sig. Str. Diff : 0.000687075893199883
Odds : 4.0737955266158824e-17
Total Wins : 3.0448312733329566e-05
```

TensorFlow

- TensorFlow is a type of deep learning neural network model
- When using these types of models, we can either create a:
  - Shallow Neural Net - 1 hidden layer
  - Deep Neural Net - 2 or more hidden layers
- The plot on the right depicts as the number of hidden layers increases the amount of loss decreases
- As a result, I used a deep neural network model to make my predictions



TensorFlow- Deep NN

- I received an accuracy score of .69 using 52 features in my model. I chose to drop certain features such as total wins
- I found that the predicted odds before each fight is the number 1 feature in the model. This means that whoever is creating these odds is doing a great job

Top 10 Features

```
[(0.1217231870634441, 'Odds'),
 (0.05530970100193494, 'Sig. Str. Diff'),
 (0.053025162185451925, 'Points'),
 (0.03372440101695282, 'Opp Total Strikes %'),
 (0.033152612822406426, 'Opp Total Strikes Landed'),
 (0.032613222979547475, 'Total Strikes %'),
 (0.03231284801299634, 'Total Strikes Landed'),
 (0.025131599718774305, 'Sig. Str. Leg Diff'),
 (0.024803420064386253, 'Total Strikes Attempts'),
 (0.024411239907425338, 'Opp Sig. Str. Ground %'),
```

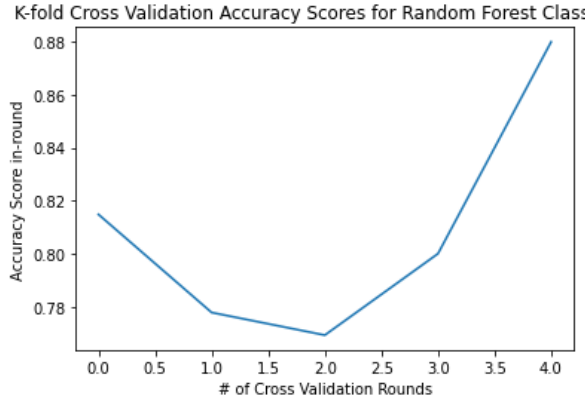
Accuracy

```
model_loss, model_accuracy = model.evaluate(X_test_scaled, y_test, verbose=2)
print(f"Normal Neural Network - Loss: {model_loss}, Accuracy: {model_accuracy}")
```

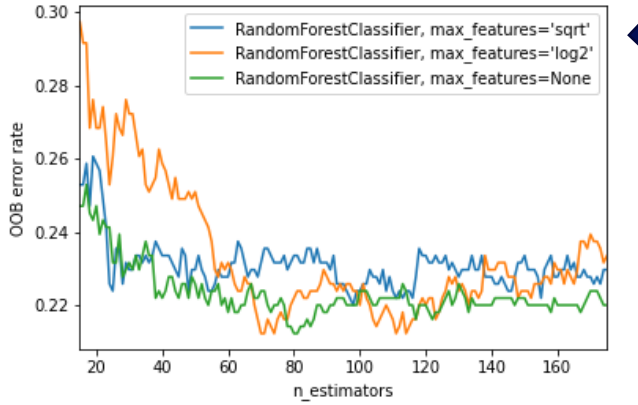
```
4/4 - 0s - loss: 0.7059 - accuracy: 0.6909 - 98ms/epoch - 25ms/step
Normal Neural Network - Loss: 0.7059475779533386, Accuracy: 0.6909090876579285
```

Random Forest

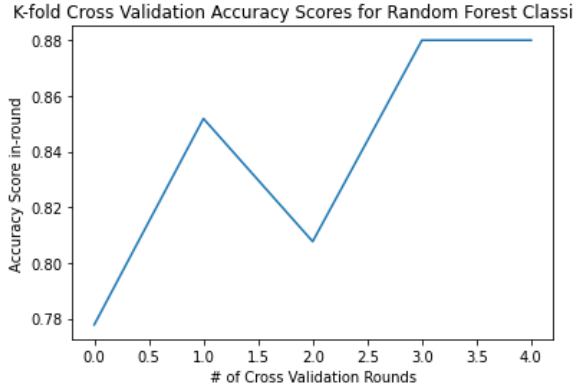
- The number of features in the dataset is large (54 features), which makes a good use case for decision tree modeling
- Random Forests utilize an ensemble method (i.e., majority vote) combined with bootstrapping the dataset to build trees that predict class
- 5-fold cross validation testing resulted in average accuracy score of 84% for correctly predicting whether a fighter would win a fight based on fighting features



81%



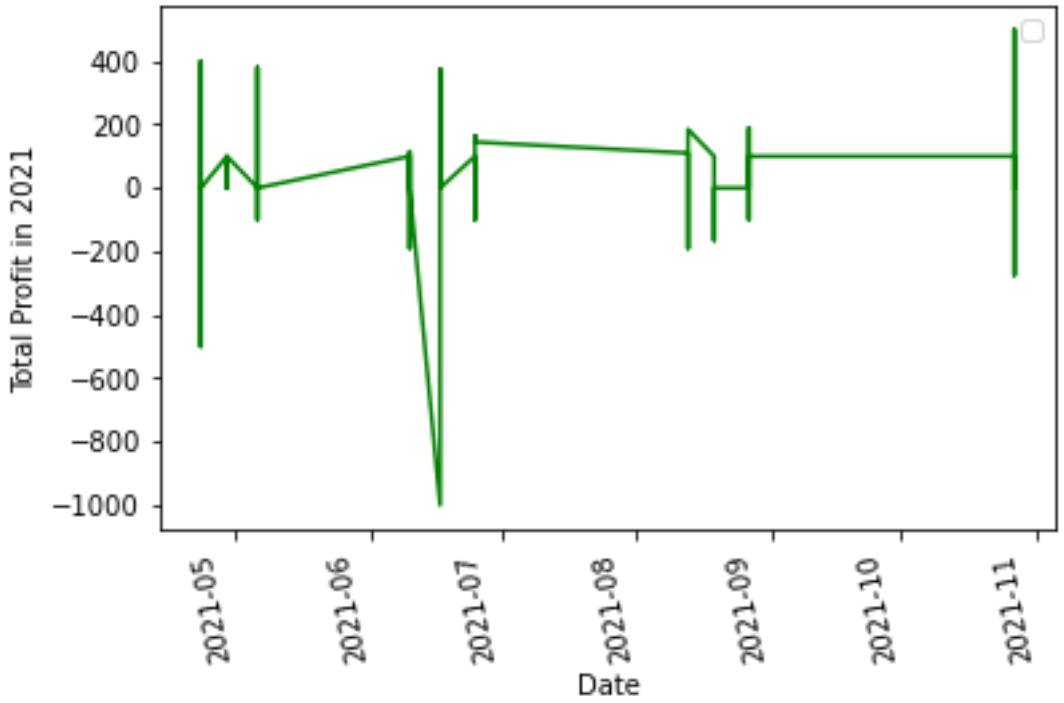
Model Tuning



84%

Keras - Convolutional Neural Network 1D

- Profit in 2021 testing at $5,412
- Accuracy is 80.61%




```
from keras.layers import Dropout
model = Sequential()
model.add(Conv1D(32, 2, activation="sigmoid", input_shape=(53, 1)))
model.add(Dropout(0.2))
model.add(Flatten())
model.add(Dense(16, activation="sigmoid"))
model.add(Dropout(0.2))
model.add(Dense(units = 2,activation = 'softmax'))


# Compile model
model.compile(loss='mean_squared_logarithmic_error', optimizer='adam', metrics=['binary_accuracy'])
```

Perspective

- 84% Fight prediction success rate is high for sports
  - Typically, 55-60% is considered best in class for predicting game outcomes using purely sport data
  - This means the addition of the moneyline odds provides an "x-factor" that vastly improves our model
- However, profitability is entirely dependent on the average odds from all bets wagered
- Model accuracy rate must be higher than the breakeven success rate to consider this a profitable, valid betting strategy
- **The average odds for all fighters in the data set is –190 (1.52 decimal odds)**
- **This gives a breakeven win rate of 65.75%**
  - **Our models outperform the sportsbooks by 18.25%**

American Odds	Decimal Odds	Break Even %
-110	1.91	52.4%
-120	1.83	54.5%
-130	1.77	56.5%
-140	1.71	58.3%
-150	1.67	60.0%
-160	1.63	61.5%
-170	1.59	63.0%
-180	1.56	64.3%
-190	1.53	65.5%
-200	1.50	66.7%
-210	1.48	67.7%
-220	1.45	68.8%
-230	1.43	69.7%
-240	1.42	70.6%
-250	1.40	71.4%
-260	1.38	72.2%
-270	1.37	73.0%
-280	1.36	73.7%
-290	1.34	74.4%
-300	1.33	75.0%