

# Predicting NFL Scores

Nick Colvin

# Problem

- Motivation
  - \$7.6 billion wagered on 2022 Super Bowl
  - Applicable to other sports
- Challenges
  - Bookmaker models are sophisticated and accurate
  - What sort of data?
    - Odds, team/player statistics, weather
    - Granularity, more emphasis on more recent games
  - Risk tolerance – when to bet?
- Features
  - Dataset consists of 4204 NFL games from 2007-2022
  - **Original format:** date, VH, team, ML, open, close, final
  - **New format:** Home/away advantage, ID, ML, spread open, spread close, over/under open, over/under close
  - **Want to add:** record, rushing/passing/total yards gained/allowed, turnover margin, possession margin

# Methods

- Multilayer perceptron
  - Own implementation
  - Scikit-learn MLPRegressor
- Motivation
  - Input is somewhat high dimensional and real values
  - Output is a 2d vector of real values
  - Financial prediction
- Remove features, compare model behavior

# Results

- Own implementation performance
  - Hyperparameters: 1 hidden layer, 100 neurons, ReLU activation, Gaussian initialization, 400 iterations, 0.0002 learning rate
  - Average record of predicting the winner over 10 training sessions: 659-390-2
    - Standard deviation: 12.53
  - Average returns of ML bets: -\$49.68
    - Standard deviation: 21.64
- Scikit-learn performance
  - Hyperparameters: 1 hidden layer, 100 neurons, ReLU activation, 400 iterations, 0.0002 learning rate, SGD solver
  - Average record over 10 training sessions: 707-341-3
    - Standard deviation: 7.419
  - Average returns: -\$2.42
    - Standard deviation: 10.53
- Scikit prediction for TNF: Las Vegas Raiders 24 – Los Angeles Rams 20