Baseball Hall of Fame Project

TTK28 Modeling with Neural Networks Exam

Olav Landmark Pedersen

26.11.2020

Goal

Determine if an eligible Major League Baseball (MLB) player will make it into the Baseball Hall of Fame (HoF) based on career statistics.

- BBWAA and Veterans Committee can nominate and elect individuals to HoF
- Subjective voting system
- MLB players are eligible after playing 10 years
- Player, managers, umpires and executives are eligible

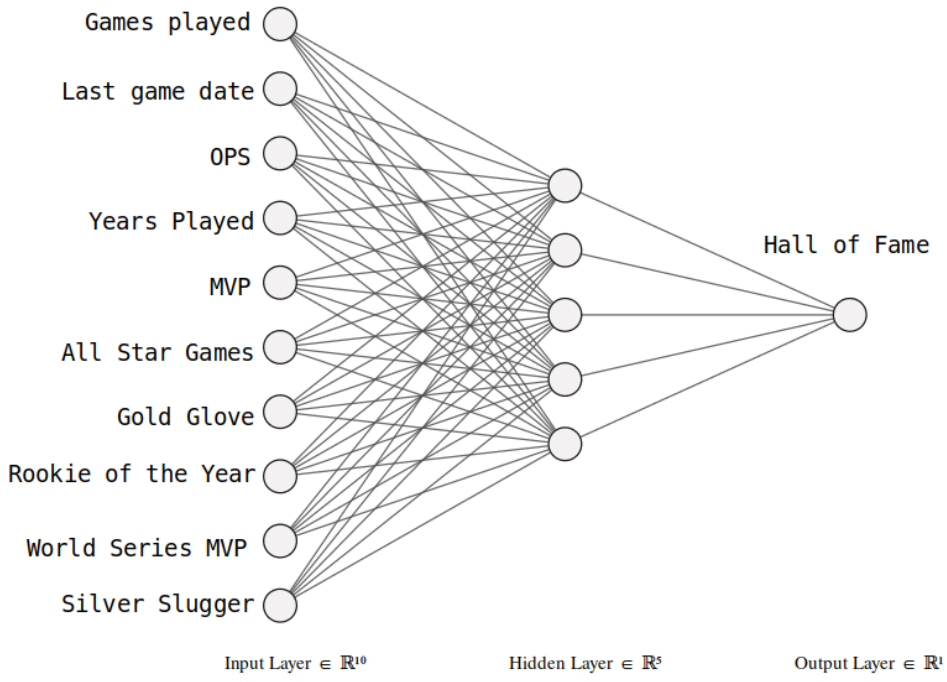
Dataset

- Sean Lahman's Baseball Archive[1]
- Imbalanced dataset:
  - 226 players in HoF
  - 3190 eligible players not in HoF
  - 14:1 ratio of non-HoF to HoF
- Hold-out split: Training 80%, Testing 20%
- Difference in quality features by positions
  - Good for batters (OPS, SLG, OBS, etc.)
  - Okay for pitchers (WHIP, K/BB, K/9, etc.)
  - Poor for other defense (FFRA, UZR, etc.)

[1] http://www.seanlahman.com/baseball-archive/

Model

- Depth: 3
- Width:
  - 10
  - 5
  - 1
- Activation:
  - ReLU
  - ReLU
  - Sigmoid
- Loss: Binary Cross-entropy
- Batch Size: 1
- Epochs: 50

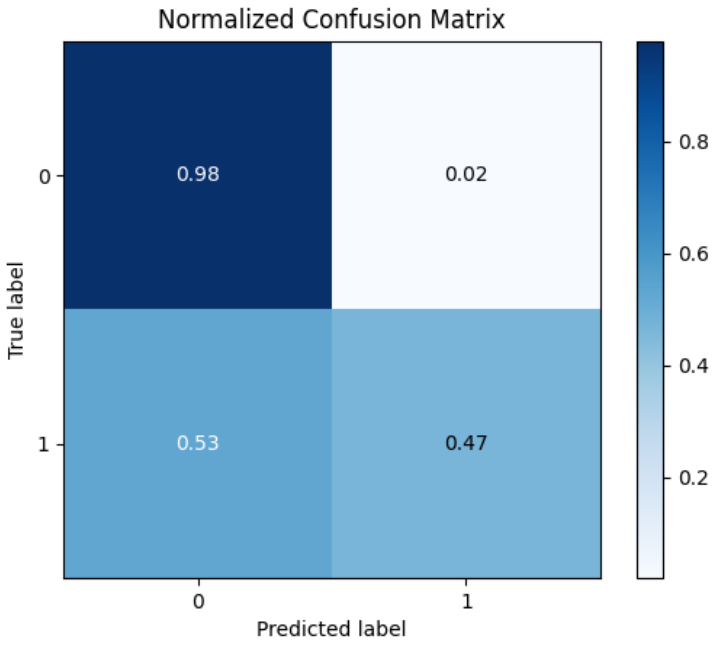


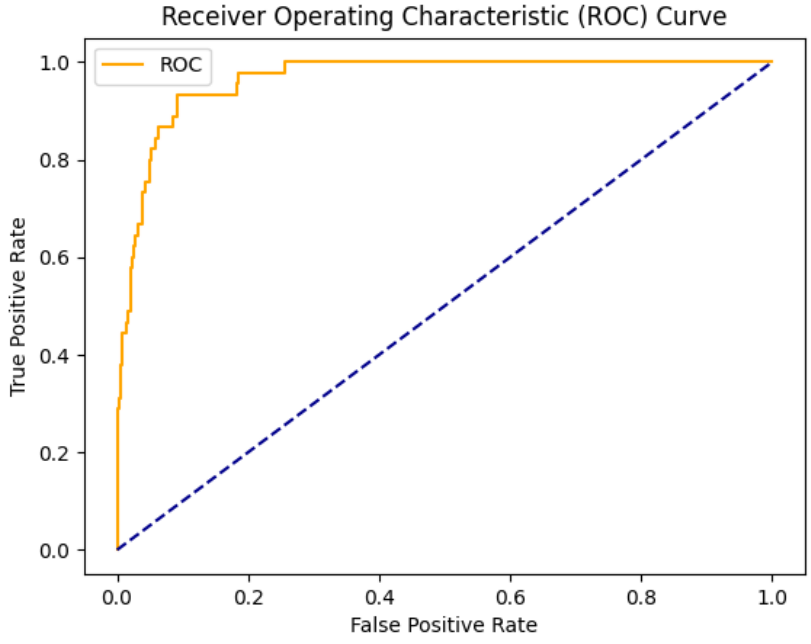
Implementation

- Fair amount of pre-processing
- Primarily used Keras and Scikit-learn tools
- Stratified 10-fold Cross-validation
- Final Test only done at the end.
- Handling and small imbalanced dataset:
  - Tried undersampling
  - Class Weights
  - Small Batch Size

Results

- Accuracy: 0.95
- AUROC: 0.73
- Precision: 0.66
- Recall: 0.47
- F1: 0.55





Thank you for your attention!