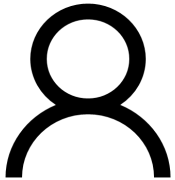
NBA match predictor

- Swaraj Kaondal
- Utsav Lal

Team 1

Winner?

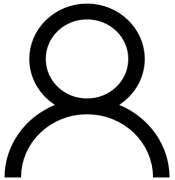
Team 2



Points: 65
Steals: 72
Blocks: 23

Points: 10
Steals: 54
Blocks: 32

Points: 27
Steals: 56
Blocks: 62

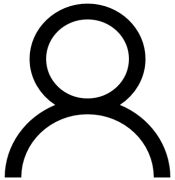


Points: 37
Steals: 22
Blocks: 47

Points: 45
Steals: 32
Blocks: 17

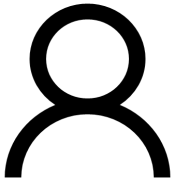
VS

Points: 66
Steals: 32
Blocks: 53

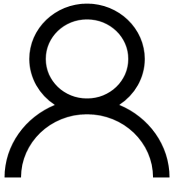


Points: 45
Steals: 12
Blocks: 9

Points: 34
Steals: 39
Blocks: 56



Points: 55
Steals: 42
Blocks: 23



Points: 57
Steals: 34
Blocks: 33

Previous Work

Forecasting NBA Champions

- Jun Yu Chen work involved exploring ANNs for forecasting NBA champion for a season
- Used a simple neural network with 3 hidden layers with 33 features
- Tackled class imbalance using case sensitive training, where a higher penalty is given for misclassifying the minority class
- Achieved test accuracy of 93% overall
- Correctly predict the winner of the 2022 - 2023 season
- Proposed methods like Synthetic Minority Over-sampling Technique to be used in future to tackle class imbalance

Win, Losses and Attributes' Sensitivities in the Soccer World Cup 2018

- Amr Hassan, Abdel-Rahman Akl, Ibrahim Hassan and Caroline Sunderland work involved predicting winners of 2018 soccer world cup
- Used Radial Basis functions for their artificial neural networks
- Achieved 72.7% win prediction accuracy.
- Concluded that distance covered by a team without the ball and average distance covered by a team as a whole were the 2 most important features while predicting
- Discussed about the possible drawbacks of their system and possible solutions

Data and Data Preparation

CMU Dataset - Player stats

Player Id	Year	Team	Points	Steals	Blocks	Shots attempted	…
1	1965	LAL	230	154	124	312	.
2	1984	OKH	194	160	197	491	.
3	1992	ARL	188	128	136	234	.
4	2001	BKN	153	203	158	255	.
.	.	.	.	.	.	.	.

Kaggle Dataset - Game data

Game Id	Date	Team 1	Team 2	Winner
1	10 Jun, 1987	LAL	MEM	Team 1
2	23 May, 1976	BOS	CHA	Team 1
3	05 Dec, 1997	OKC	PHI	Team 2
4	30 Aug, 1980	PHX	BOS	Team 2
.		.	.	.

Kaggle Dataset - Play by Play data

Game Id	Action	Player 1	Player 1 Team	Player 2	Player 2 Team
1	Steal	10001	LAL	20001	MEM
1	Block	10005	LAL	20010	MEM
1	Score	10005	LAL	-	-
2	…	…	…	…	…
.					

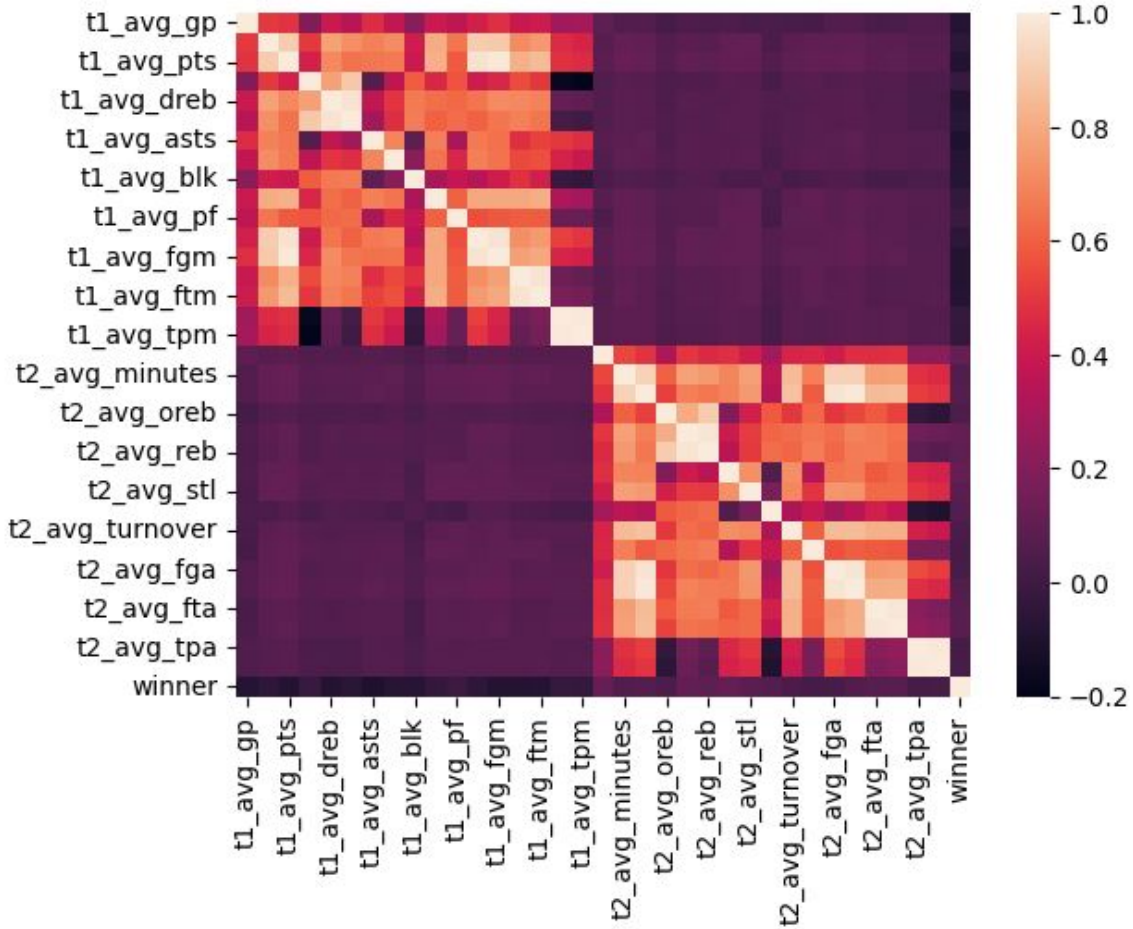
Game Id	Team 1	Team 2
1	[10001, 10005, 10006, …]	[20001, 20010, …]
2	…	…

Final Dataset

Game Id	T1_Points	T1_Steals	T1_Blocks	…	T2_Points	T2_Steals	T2_Blocks	…	Winner
1	18.51	34.23	24.63	…	18.23	38.32	12.45	…	T1
2	25.23	29.67	43.34	…	24.56	12.53	34.6	…	T1
3	22.12	17.89	21.26	…	28.55	51.45	34.21	…	T2
4	.29.34	.35.67	26.45	…	19.42	12.56	28.42	…	T2

34 Features (17 of each team) with a **binary target variable**.
8777 game data

Feature Correlation



Naive Bayes Model

Why Naive Bayes?

- Simple
- Easy to implement
- Good performance in categorical data
- Easily helps to identify features which might be influencing predictions

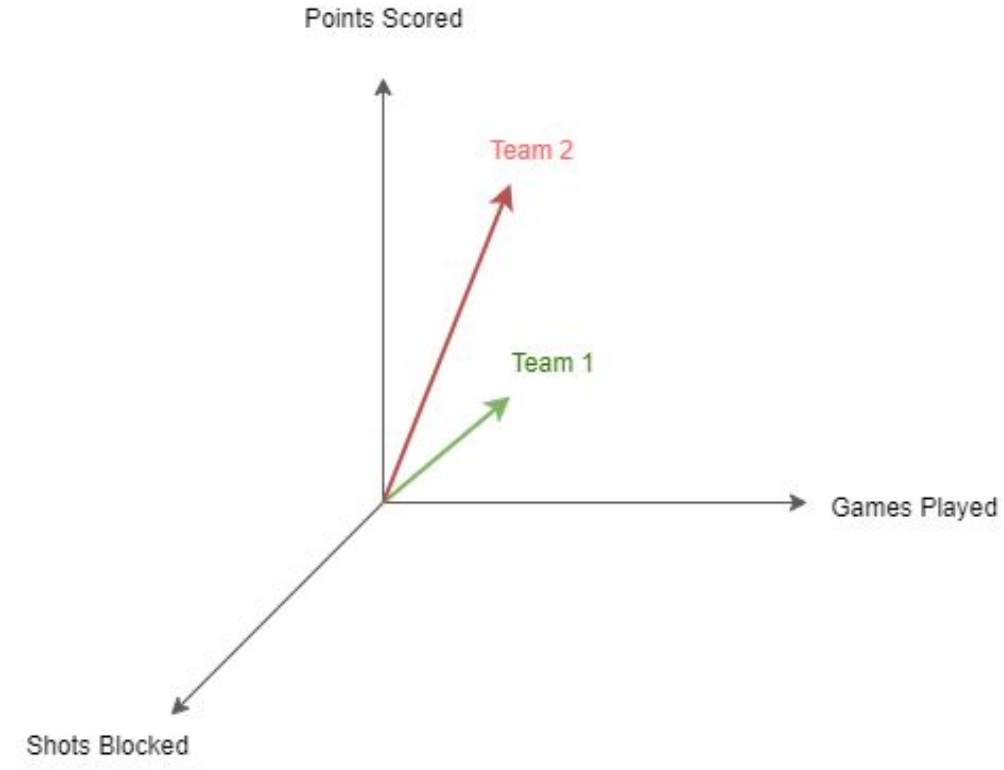
Result

	Predicted	
Actual	Team 1 Win	Team 2 Win
Team 1 Win	549	401
Team 2 Win	380	426

	Precision	Recall	F1-Score	Support
Team 1 Win	0.59	0.58	0.58	950
Team 2 Win	0.52	0.53	0.52	806
Accuracy			0.56	1756
Macro Avg	0.55	0.55	0.55	1756
Weighted Avg	0.56	0.56	0.56	1756

L2-Norm Comparator

L2-Norm Comparator



$$|\mathbf{x}| = \sqrt{\sum_{k=1}^{n} |x_k|^2},$$

Results

	Predicted	
Actual	**Team 1 Win**	**Team 2 Win**
Team 1 Win	2987	2307
Team 2 Win	1941	2719

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Team 1 win	0.61	0.56	0.58	5294
Team 2 win	0.54	0.58	0.56	4660
Accuracy			0.57	9954
Macro Avg	0.57	0.57	0.57	9954
Weighted Avg	0.58	0.57	0.57	9954

Artificial Neural Network

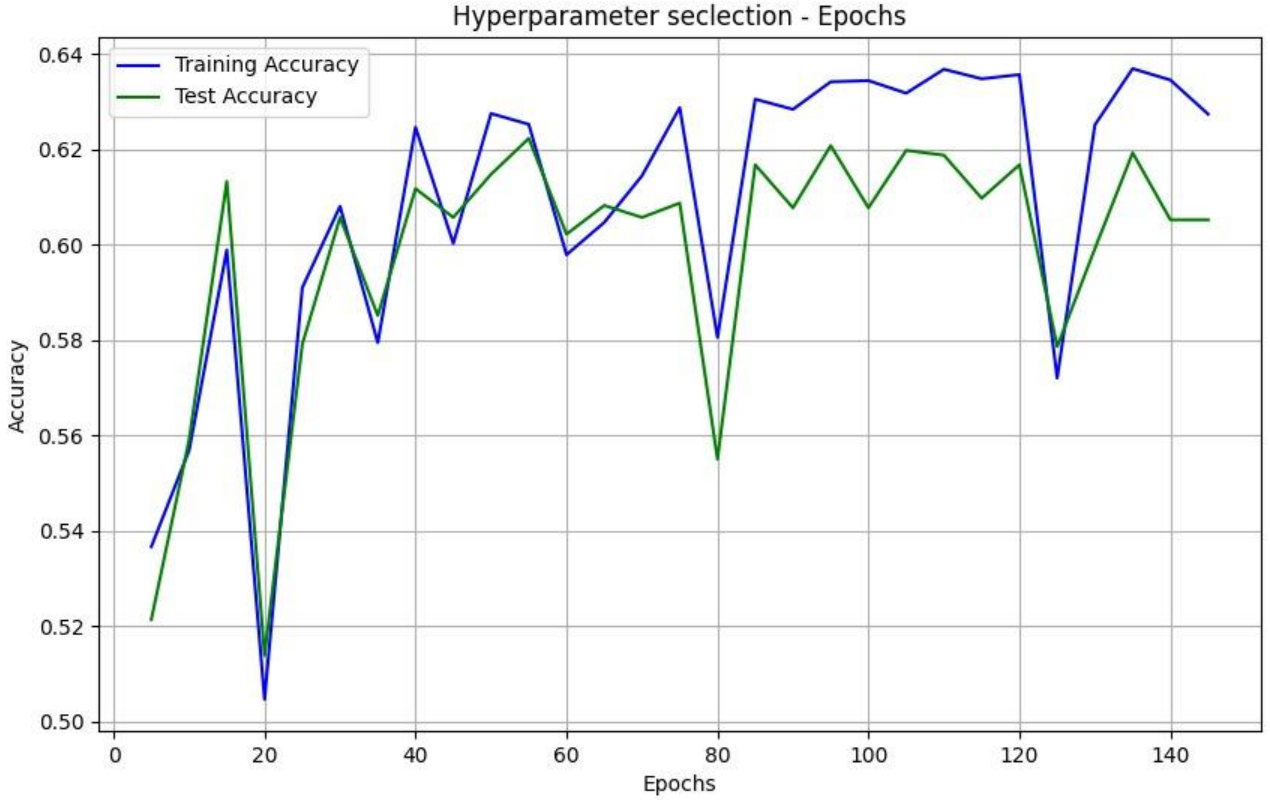
Grid Search

- Number of hidden layers - 1, 2 and 3
- Number of neurons - 34, 68 and 102
- Activation Layer - Relu and Tanh
- Optimizers - Adam, RMSProp and SGD

Number of total configurations = 648

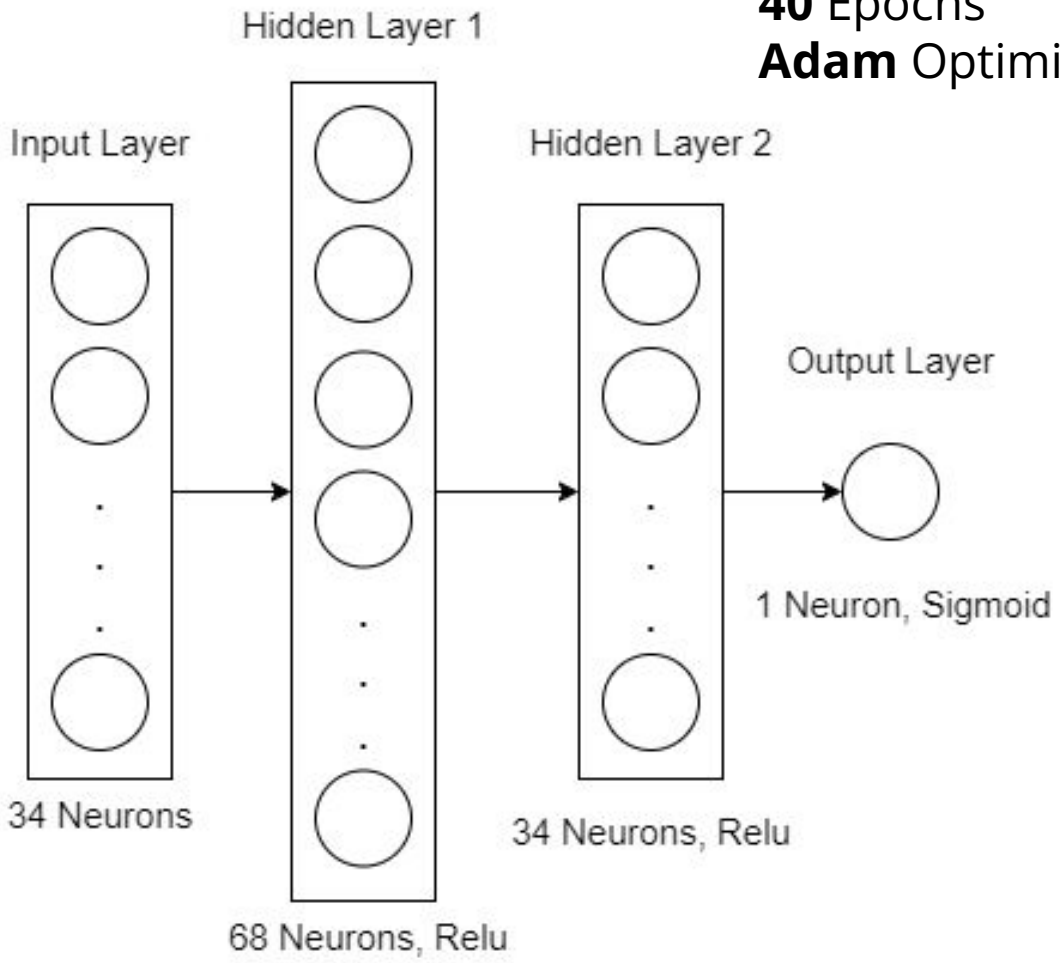
Epochs per combination = 10

Epochs - Tuning



Best Parameters

40 Epochs
Adam Optimizer



Results

	Predicted	
Actual	**Team 1 Win**	**Team 2 Win**
Team 1 Win	965	100
Team 2 Win	696	230

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Team 1 Win	0.58	0.91	0.71	1065
Team 2 Win	0.70	0.25	0.37	926
Accuracy			0.60	1991
Macro Avg	0.64	0.58	0.54	1991
Weighted Avg	0.63	0.60	0.55	1991

To conclude

Summing it up together

- We predicted neural network performance to be much better
- Possible reasons could be insufficient data which caused less generalization
- Another possible reason could be data mismatch between the two datasets
- In the end, NBA is a fierce competition there are always upsets and using pure statistics might not be the best way to predict a winner of a match

References

- Bruin Sports Analytics. [n. d.]. Decoding the Game: Forecasting NBA Champions with Neural Network Algorithms — bruinsportsanalytics.com. https://www.bruinsportsanalytics.com/post/nba-champs-neural-network. [Accessed 03-03-2024].
- School of Computer Science Carnegie Mellon University. [n. d.]. NBA statistics data. http://www.cs.cmu.edu/~awm/10701/project/databasebasketball2.0.zip
- Amr Hassan, Abdel-Rahman Akl, Ibrahim Hassan, and Caroline Sunderland. 2020. Predicting wins, losses and attributes' sensitivities in the soccer world cup 2018 using neural network analysis. Sensors 20, 11 (2020), 3213.
- Xiaohu Tang, Zhifeng Liu, Taizhao Li, Wenbin Wu, and Zhenhua Wei. 2018. The application of decision tree in the prediction of winning team. In 2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS). IEEE, 239–242.