Basketball Offense Optimization

DHRUV SRIVASTAVA AND YASH SANGTANI

Let's first get a rough idea of what Basketball really is....

WHAT IS BASKETBALL?

WHAT IS BASKETBALL?

Problem Statement

01

During an offensive possession, given an opponent's defensive scheme, what is the decision has the highest probability of success?

Enhance team performance and strategy formulation

02

At each point in the game, can we weigh or quantify value of each player's future actions based on in-game scenario.

Metric Formulation and Computation

03

Develop an extensive toolkit for coaches, players, and administration to analyze basketball games and offensive schemes to make better plans/decisions.

Research and Analysis

Reviewing prior research

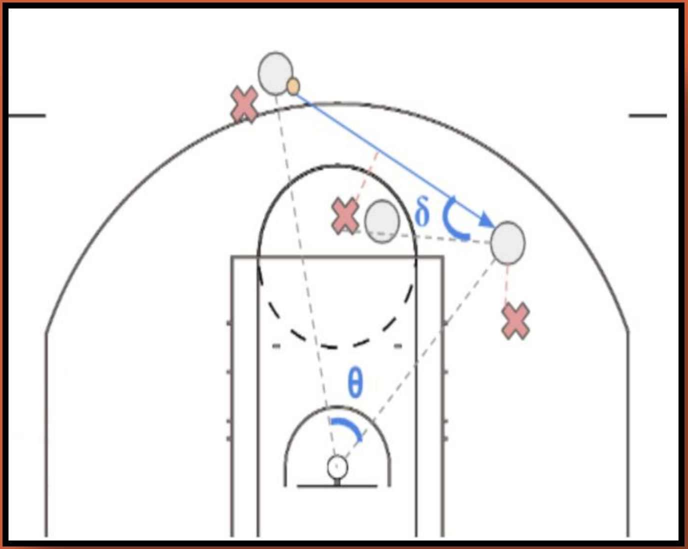
Paper	Author(s)
A Method for Using Player Tracking Data in Basketball to Learn Player Skills and Predict Team Performance.	Brian Skinner, Stephen J. Guy
Optimizing the best play in basketball using deep learning	Leili Javadpour, Jessica Blakeslee.
Applying Deep Learning to Basketball Trajectories	Rajiv C. Shah, Rob Romijnders
Classifying NBA Offensive Plays Using Neural Networks	Kuan-Chieh Wang, Richard Zemel
Generating Defensive Plays in Basketball Games	Chieh-Yu Chen, Wenze Lai
Expected Possession Value: An Evaluation Framework for Decision-Making, Strategy, and Execution in Basketball	Ivan C. Jutamulia
"How to Get an Open Shot": Analyzing Team Movement in Basketball using Tracking Data	Patrick Lucey, Alina Bialkowski,
PREDICTING SHOT MAKING IN BASKETBALL LEARNT FROM ADVERSARIAL MULTIAGENT TRAJECTORIES	Mark Harmon, Patrick Lucey

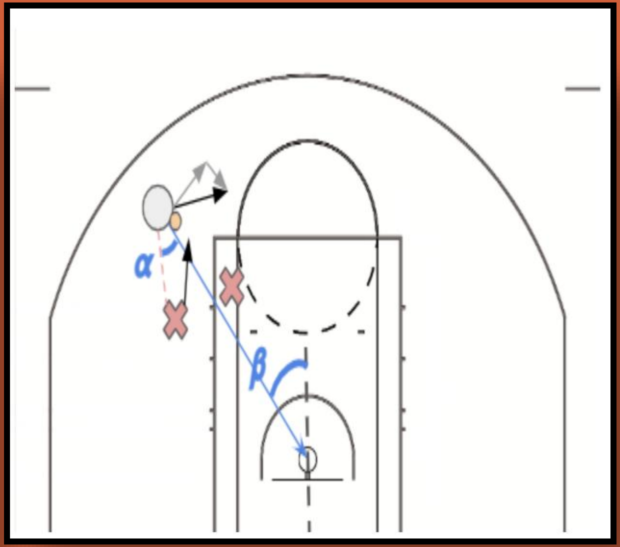
Reviewing prior research

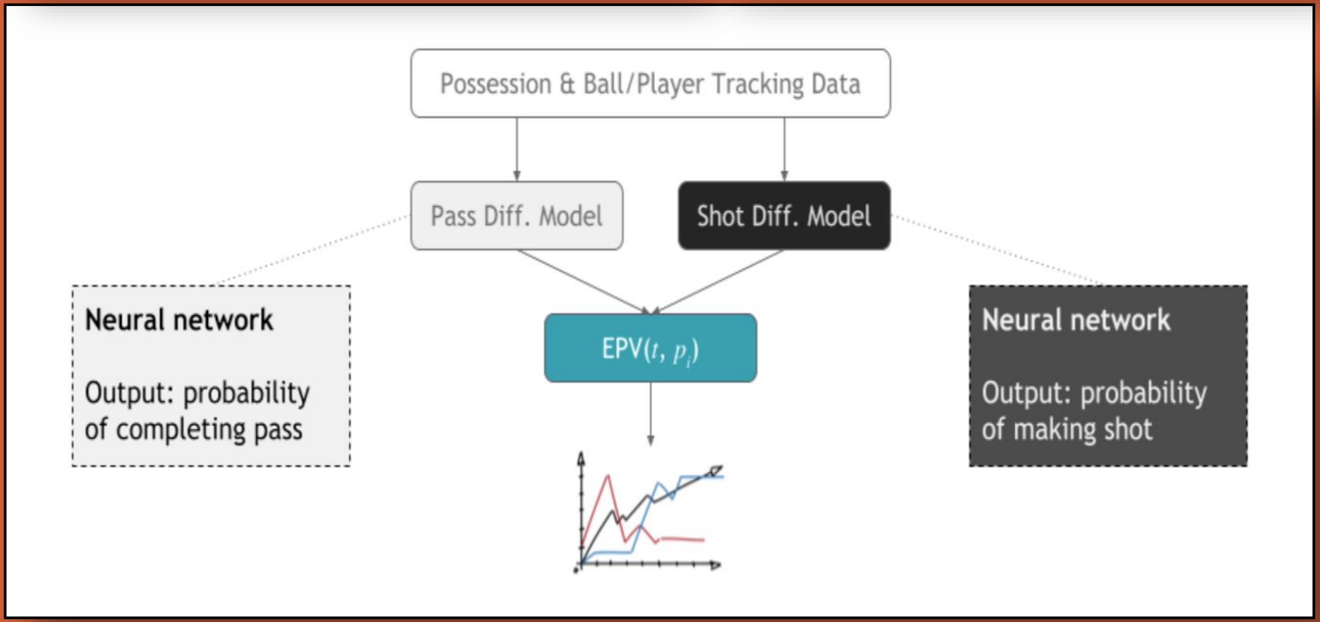
Paper	Author(s)
Identifying Basketball Plays from Sensor Data; towards a Low-Cost Automatic Extraction of Advanced Statistics	Adria Arbues Sanguesa, Thomas B. Moeslund
Optimizing Offensive Gameplan in the National Basketball Association with Machine Learning	Eamon Mukhopadhyay,
Possession Sketches: Mapping NBA Strategies	Andrew C. Miller, Luke Bornn
Perceived Hotness Affects Behavior of Basketball Players and Coaches	Yigal Attali
Crunch time in the NBA – The effectiveness of different play types in the endgame of close matches in professional basketball	Jan Christmann, Max Akamphuber
Use of Machine Learning to Automate the Identification of Basketball Strategies Using Whole Team Player Tracking Data	Changjia Tian, Varuna De Silva
Decision making for basketball clutch shots: A data driven approach	Yuval Eppel, Mor Kaspi
An Analysis of Sports News in the Era of Big Data- Visual Data News with NBA Coverage as an Example	Kai Gao, Li Tang, and Jialin Lu

Expected Possession Value: An Evaluation Framework for Decision-Making, Strategy, and Execution in Basketball

Ivan C. Jutamulia's thesis introduces Expected Possession Value (EPV), a metric using NBA tracking data to assess basketball decision-making. EPV evaluates player and team performance by quantifying and analyzing scoring opportunities, distinguishing effective strategies from execution. Jutamulia's work highlights the potential of advanced analytics to enhance decision-making and team performance in professional basketball, offering valuable insights for optimizing strategies and improving execution.





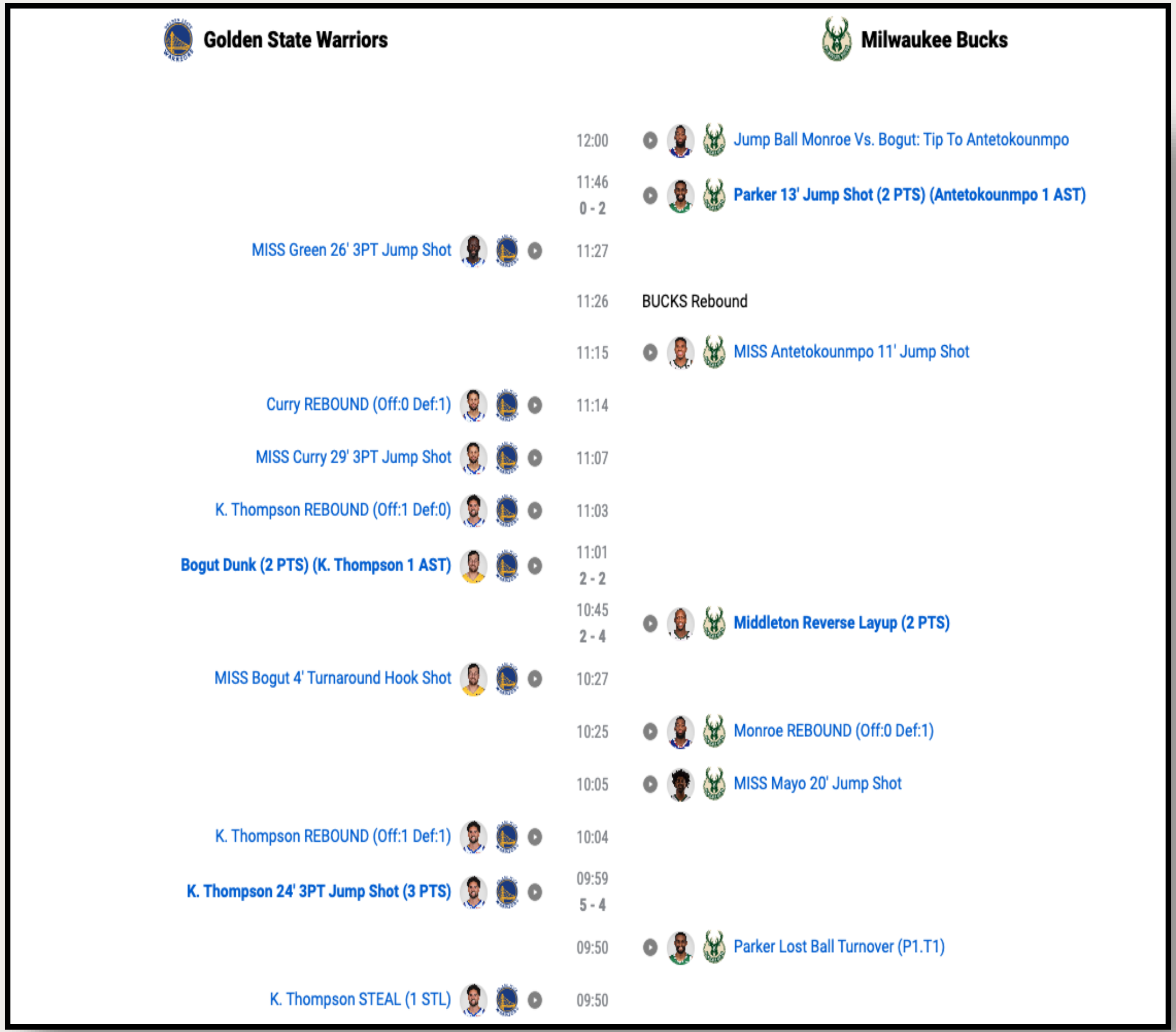


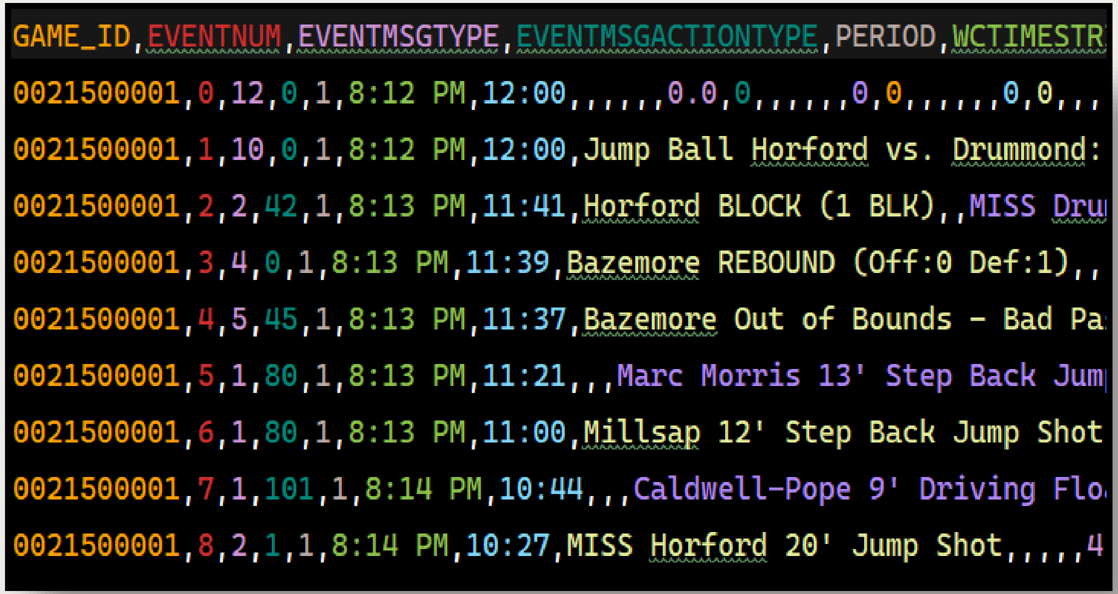
Primary Dataset

```
"data_sets": {
  "BasketballGame": {
    "gameId", "gameDate",
     "events": [ {
                              "eventId", "visitor": {"name", "teamId", "abbreviation",
                              "players":
[{ "lastName", "firstName", "playerId", "jersey", "position"}]
                    },
                              "home": {"name", "teamId", "abbreviation",
                              "players":
[{ "lastName", "firstName", "playerId", "jersey", "position"}]
                    },
          "moments": [
            [
              "quater",
              "Timestamp",
              "quarter_time",
              "shot_clock",
              [
                [
                  "Player ID",
                  "Team ID",
                  "X-Coordinate",
                  "Y-Coordinate",
                  "Z-Coordinate"
                ] ... for 11 such lists (1 for the ball & 10 players)
              ]          ]          ]   }}
```

The tracking data provided by Sports VU involves comprehensive snapshots of the basketball court captured by six cameras at intervals of 0.04 seconds. Each moment within the dataset comprises coordinates detailing the positions of players and the ball on the court.

Secondary Dataset





The dataset contains play-by-play event data from an NBA basketball game. It includes fields such as GAME_ID, EVENTNUM, EVENTMSGTYPE, and player-specific information like PLAYER1_NAME and PLAYER1_TEAM_CITY. Each record details game events like shots, rebounds, and turnovers, providing a comprehensive play-by-play breakdown with timestamps, player actions, and the game score at each event.

```
PLAYER_NAME,OFFENSIVE_RATING,DEFENSIVE_RATING
Aaron Brooks,99.0,109.0
Aaron Gordon,114.0,105.0
Aaron Harrison,77.0,103.0
Adreian Payne,81.0,108.0
Al Horford,113.0,101.0
Al Jefferson,105.0,102.0
Al-Farouq Aminu,105.0,107.0
Alan Anderson,107.0,108.0
Alec Burks,104.0,107.0
Alex Len,96.0,107.0
Alex Stepheson,105.0,105.33333333333333
Alexis Ajinça,100.0,107.0
Allen Crabbe,114.0,110.0
Alonzo Gee,108.0,110.0
Amir Johnson,117.0,102.0
Andre Drummond,103.0,98.0
Andre Iguodala,115.0,105.0
```

Ratings (NBA_API)

ShotsLog (NBA_API)

```
,GAME_EVENT_ID,PLAYER_ID,PLAYER_NAME,TEAM_ID,TEAM_NAME,PERIOD,MINUTES_REMAINING,SECONDS_REMAINING,EVENT_TYPE,ACTION_TYPE,SHOT_TYPE,SHOT_ZONE_BASIC,SHO
,21500014,20,202083,Wesley Matthews,1610612742,Dallas Mavericks,1,9,4,Missed Shot,Jump Shot,3PT Field Goal,Above the Break 3,Right Side Center(RC),24+ ft
,21500014,168,202083,Wesley Matthews,1610612742,Dallas Mavericks,2,11,2,Missed Shot,Jump Shot,3PT Field Goal,Above the Break 3,Left Side Center(LC),24+ f
,21500014,217,202083,Wesley Matthews,1610612742,Dallas Mavericks,2,8,10,Made Shot,Jump Shot,3PT Field Goal,Above the Break 3,Right Side Center(RC),24+ ft
,21500014,340,202083,Wesley Matthews,1610612742,Dallas Mavericks,3,10,57,Made Shot,Jump Shot,3PT Field Goal,Above the Break 3,Right Side Center(RC),24+ f
,21500014,358,202083,Wesley Matthews,1610612742,Dallas Mavericks,3,9,33,Missed Shot,Jump Shot,2PT Field Goal,Mid-Range,Right Side(R),16-24 ft.,21,182,110,
```

Data and Feature Preprocessing

SportsVu Game Logs

Extracting chunks from large, unreadable JSON data

Parsing them, interpreting, and storing parts as game objects to recall in future.

Adjusting for any idiosyncrasies in time stamps, value count, etc.

Play by Play Data

Integrating them with the game logs to extract meaningful events like shot made or missed.

Using it in tandem with the game logs to extract features for our models.

Others

Create a pipeline to output player-specific stats to further bolster the feature vector and thus model performance.

Create a workaround unlogged players, events, etc.

Ensuring the features are easy and fast to store and access.

Pass Features

dist	Distance of the pass
passer_basket_dist	Distance from passer to the basket
ball_end_basket_dist	Distance from the target to the basket
basket_angle	Angle of passer --> basket --> target
closest_def_dist_passer	Closest defender distance to the passer
closest _def_dist_ball_end	Closest defender distance to the target
closest_def_trajectory	Closest perpendicular defender distance to pass trajectory
closest_def_angle_passer	Closest defender angle to passer w.r.t. pass trajectory
closest_def_angle_ball_end	Closest defender angle to target w.r.t. target trajectory
closest_def_trajectory_angle	Angle of passer --> target --> most obstructive defender
backcourt	Whether the pass was made entirely on backcourt
shot_clock	Time remaining on the shot clock

Pass Features

1. Processing the Tracking Data
2. Filtering out Possessions
3. Extracting Passes
4. Extracting unsuccessful passes using Play-by-Play Data
5. Calculating Euclidean distance to and the ball man and the distance features.
6. Using arccos function to calculate the angle features in radians.
7. Checking the nearest basket at start of possession for backcourt pass.

backcourt	dist	passer_basket_dist	ball_end_basket_dist	basket_angle	closest_def_dist_passer	closest_def_dist_ball_end	closest_def_trajectory	closest_def_angle_passer
0	7.695607672854687	6.305958620424015	1.402065672142358	0.14721286077449508	4.367718892110615	3.6794257450042367	1.1715256112004648	2.8700439883930966
0	12.304162510402728	25.393425771291678	22.590962207132755	2.6360105052497755	7.169871981904558	5.996277571335397	2.332130897876174	2.81029738367207
1	14.78264133766696	26.91068023338875	14.56414809669965	2.728010822188732	5.662852755537621	9.550784742208357	1.734502807686103	2.83029426244166
1	12.10129222688222	15.821083165602158	25.034741935640163	2.744788230893842	2.7753837496101337	13.647808498378048	1.3899590895123424	1.072238796524626
1	13.750779857873518	17.712186363498436	18.817913755382133	2.3719871765012823	1.8423859411100592	13.311157357529058	0.6052768698376301	1.877369687301735
1	26.51308877775843	19.06361195347041	26.565862699276302	1.940741800145448	6.305620095922368	34.95438033103863	0.3892328405680542	2.881270383497844
1	22.731655470981426	25.60602802756609	25.60704156033063	2.221779612873466	8.319089123996688	19.697434111447105	0.10654583563400864	2.121600446766518
1	34.41328681280677	31.346671963902324	25.31908308673519	1.851272055750782	8.327191177972319	8.31806238101759	7.436357716423196	1.3157272320015112
1	30.557867730286745	32.34605147119197	6.570685202724296	1.9457095513310578	13.796425318527263	5.9139334477993595	0.8368326310102844	2.7318561165757385
0	19.615886840344487	10.682108455637396	22.770178859236044	2.10528172043821	12.203113295913464	6.163865226503578	0.21557946708290143	2.4378469490031844
1	12.573542682673017	28.80316304369366	20.470596367771016	2.7513420683183796	6.172267455684339	4.851498532546413	1.7839046900607856	2.5634577649871226
1	11.941845564317104	12.06427568571773	17.511197758103243	2.393080128409117	2.7999783262196947	10.551084547879421	1.599889626130193	2.1956945591914545
1	7.653327155590565	10.608436378854337	10.598498613393323	2.403153571238051	3.8664106910802905	2.0394766393611854	0.6635342120178849	1.549314743194876
1	16.00480048885646	10.745234774931628	23.62553648928422	2.5361372775921724	2.191856775658475	17.679452759429516	1.0101168736502715	0.7443466007772025
1	24.023007926621098	44.75480137944643	27.25561469516327	2.6658783292544417	7.953648797759431	4.389115688609723	0.4850516450559603	1.6468727213022176

Pass Difficulty Model

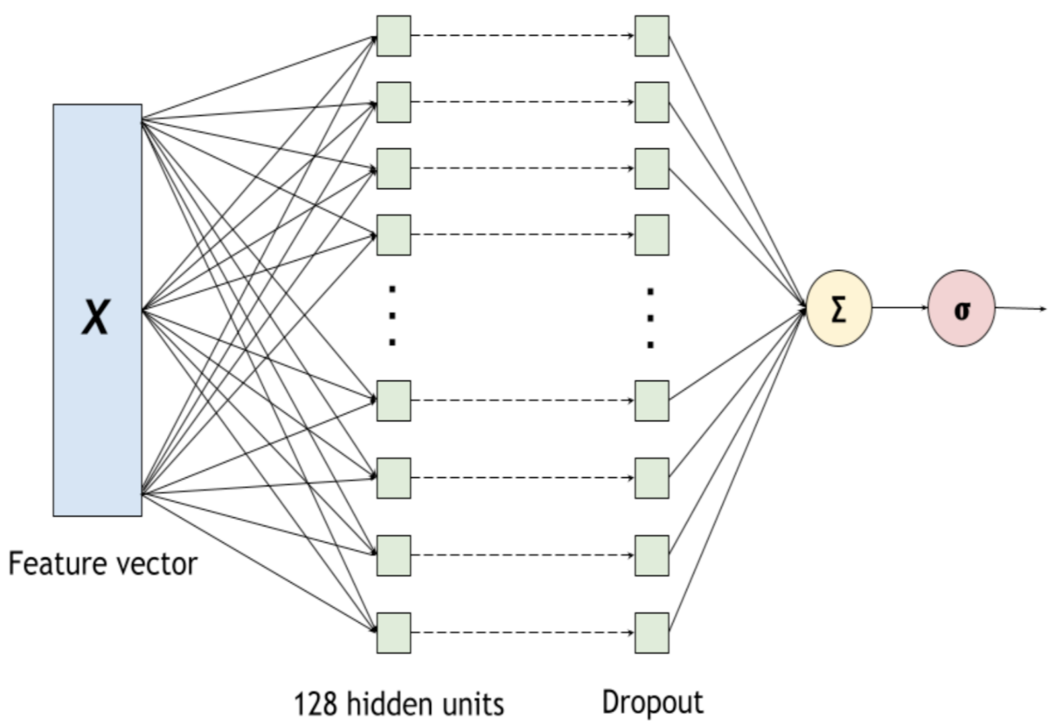


Figure 5-5: Pass difficulty neural network architecture.

Specifiations About the Model:

Using One fully-connected hidden layer with 128 units, each activated with a ReLU function.
One Output layer that is fully connected, and uses Sigmoid as a final output activation function.
Epochs = 5

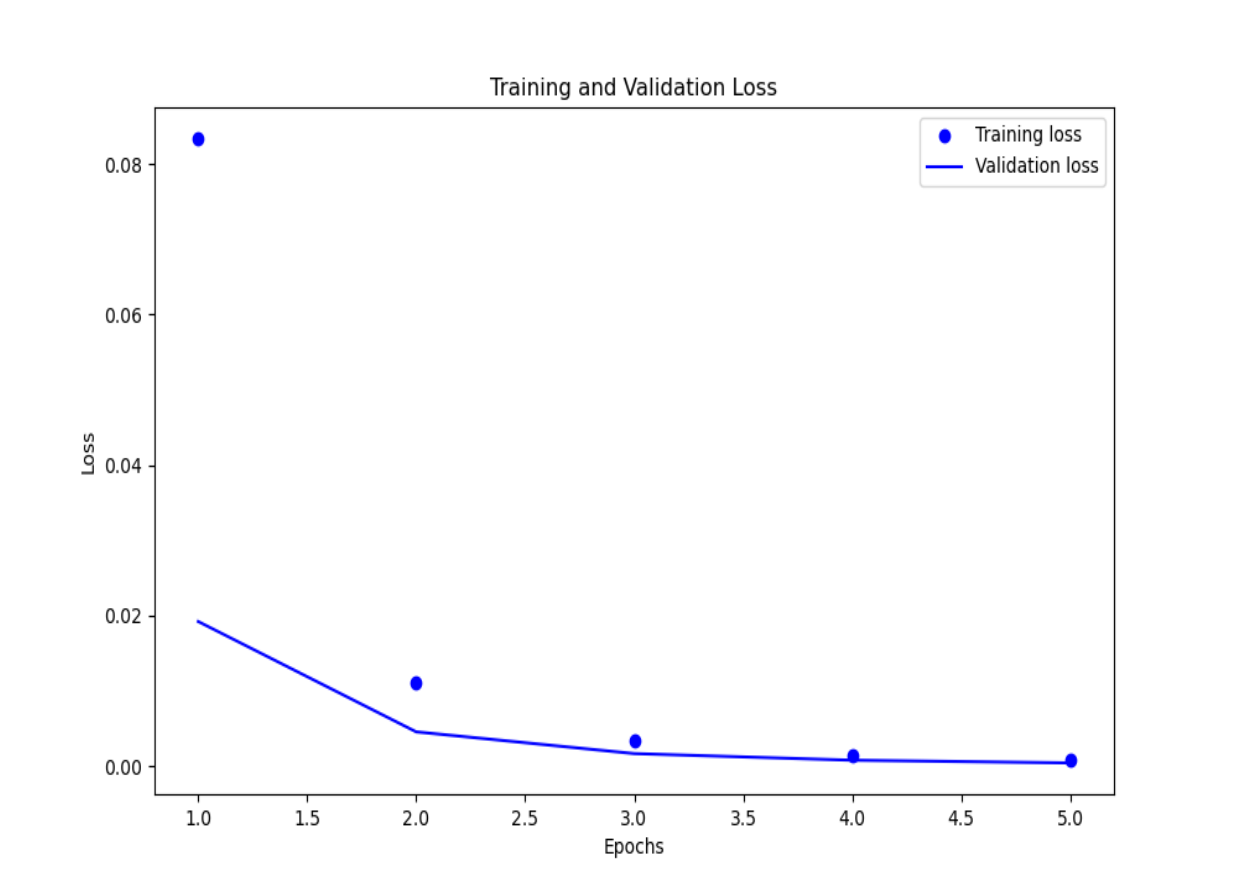
For regularization,

Train Test Split = 40%
Dropout = 0.2
Validation split = 40%
batch_size = 32

Trained the model with binary cross-entropy function.

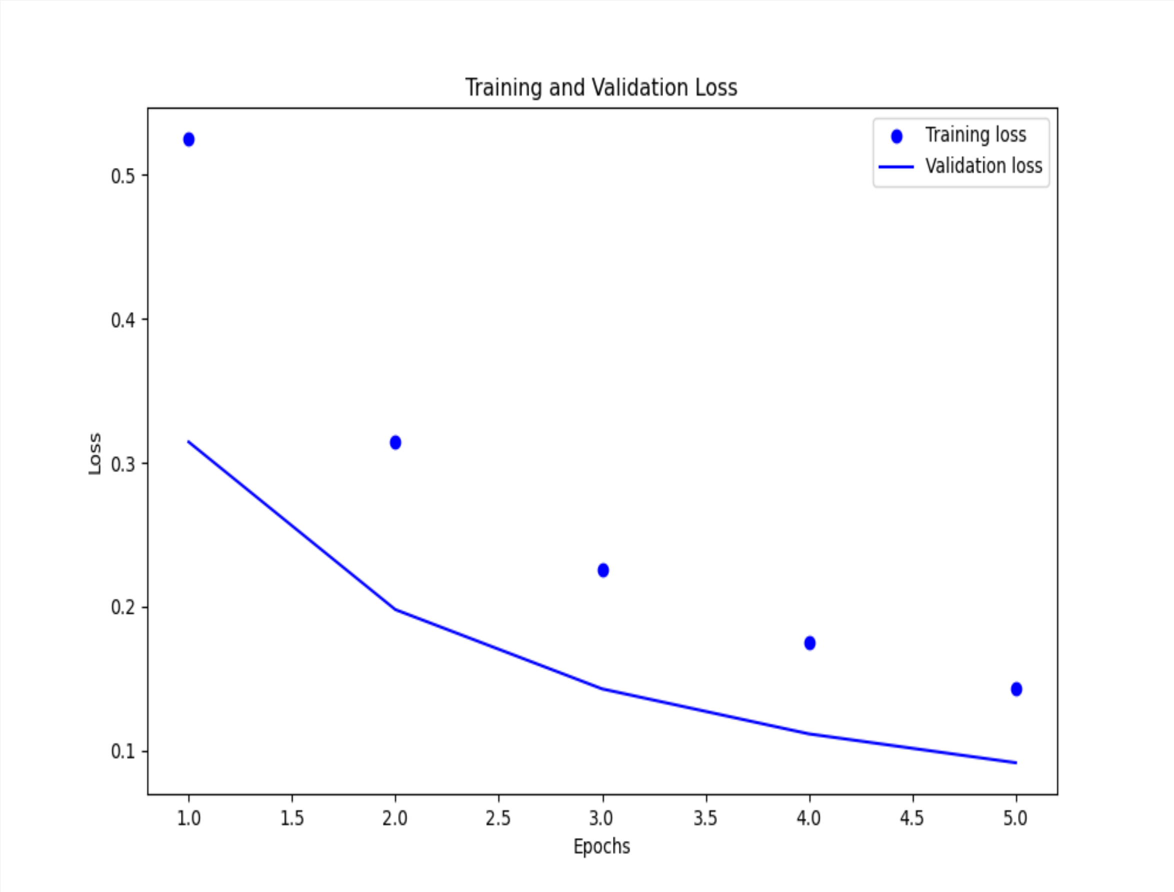
Pass Difficulty Model – Evaluation

Dropout - 0.2; units=128; Batch Size = 32



F1 Score: 1.0
Accuracy: 1.0
ROC AUC Score: 1.0

Dropout - 0.7; units=56; Batch Size = 560



F1 Score: 0.9852895148669797
Accuracy: 0.9769072717527246
ROC AUC Score: 0.9962791644729961

Proposed Shot Features (in research papers and other attempts.)

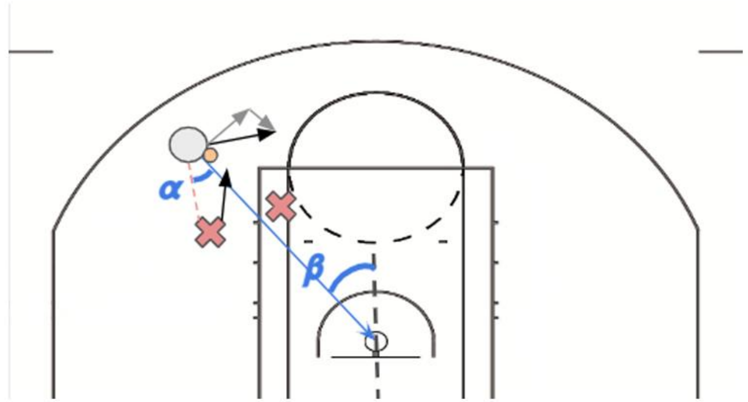


Figure 5-8: Diagram of geometric shot features. Blue arrow indicates trajectory of hypothetical shot. Black arrows represent movement vectors for players, decomposed into gray parallel and perpendicular components. In this example, the shooter is fading to their left while shooting, while the defender is moving towards them to contest. α is the angle of the closest defender with respect to the shot trajectory, and β is the angle of the shot with respect to center court.

Feature name	Description
dist	Distance of the shot (from shooter to basket)
x	x coordinate of shooter
y	y coordinate of shooter
shot_angle	Angle of the shot w.r.t. court center
closest_def_dist	Closest defender distance to shooter
closest_def_angle	Closest defender angle to shooter w.r.t. shot trajectory
num_close_defs	Number of defenders within 4 feet of shooter
shot_clock	Time remaining on the shot clock
shooter_par_vel	Parallel velocity of shooter w.r.t. shot trajectory

Added Shot Features

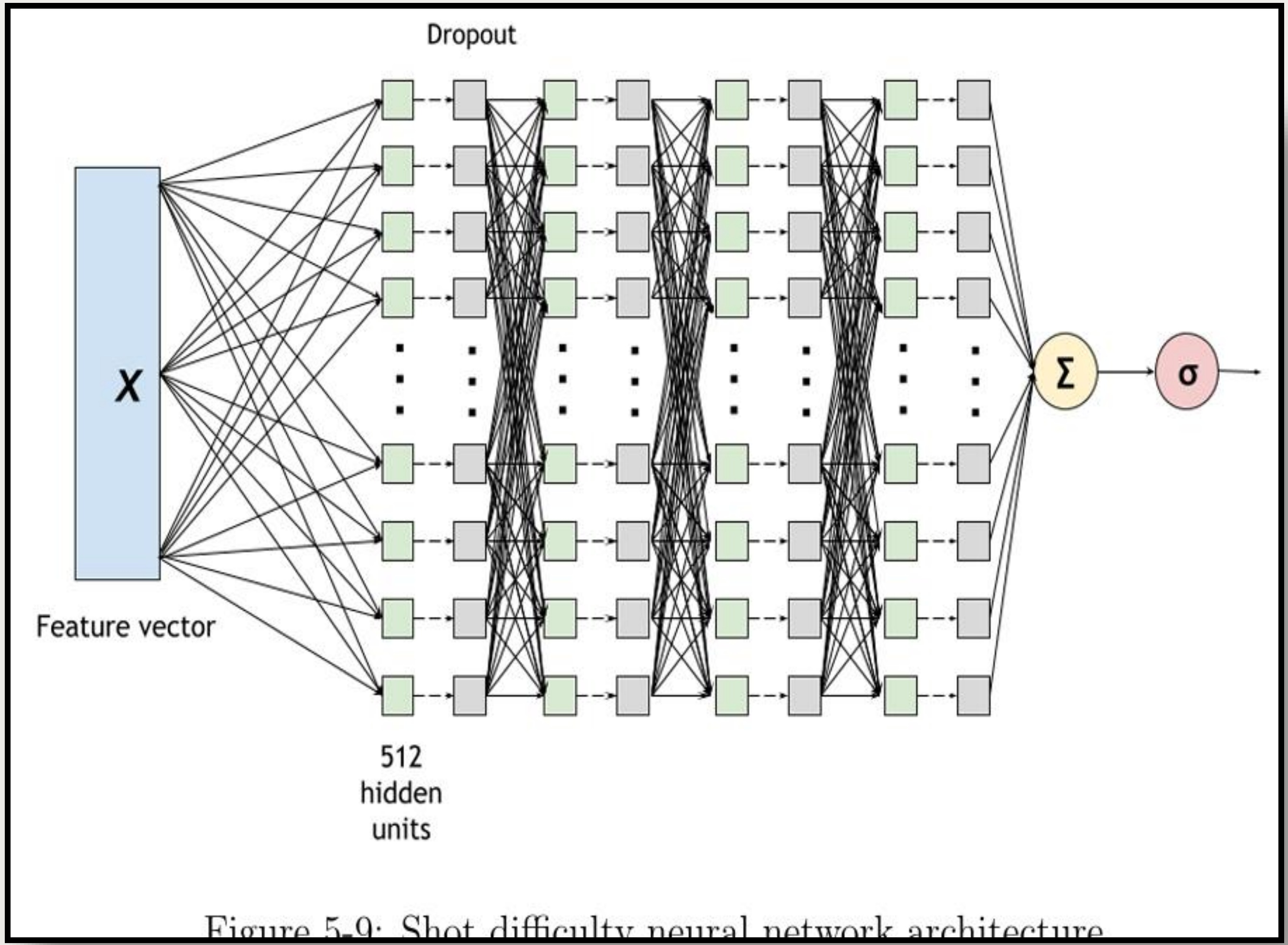
shooter_offensive_rating	Offensive rating of the shooter -> float
X	x coordinate of the shooter -> float
y	y coordinate of the shot -> float
fg_percentage_zone (utilizes zone_model)	field goal percentage of the shooter from the specific zone he is shooting from -> float
closest_defender_defensive_rating	defensive rating of the closest defender -> float
closest_defender_velocity	velocity of the closest defender in ft/msec -> float
score_margin	score margin of the game -> int
quarter	Quarter of the game -> int
minutes	Minutes passed into the quarter -> int
seconds	Seconds passed into the minute -> int
points_attempted (utlizes zone_model)	Points attempted from the zone

Shot Features

1. Processing the Tracking Data
2. Filtering out Shot Attempts
3. Extracting Basic Features
4. Calculating distance features using Euclidean function
5. Using arccos function to calculate the angle features in radians.
6. Using Zone Model to predict Zone.
7. Use it to extract player specific fg% from that zone
8. Finding and adding player ratings from NBA API

shooter_offensive_rating	dist	x	y	shot_angle	shooter_velocity	fg_percentage_zone	closest_defender	closest_defender_name	closest_defender_defensive_rating	closest_defender_dist	closest_de
115.0	48.86918493337596	5.28373	-23.86914	1.5721523964817816	0.012191721602177451	71.07438016528926	203999	Nikola Jokic	106.69356796116504	2.179366321410882	2.7494088
120.0	74.01490920479333	54.45817	-40.64393	1.0905939387072423	0.01023113112331667	40.74074074074074	2738	Andre Iguodala	105.0	9.800651653813633	0.7022177
115.0	53.78298428597097	29.96391	-22.82013	1.0954432966267422	0.0012233432507279844	39.53488372093023	2749	Jameer Nelson	112.0	2.363992950412498	1.5745170
120.0	51.000426783983876	87.04514	-25.97517	1.5393235514281887	0.002406838239786591	50.43103448275862	203110	Draymond Green	100.0	0.7481223669962028	2.8488854
115.0	50.35382009259278	33.68014	-16.628239999999998	0.9732429932443033	0.008430666621256246	38.46153846153847	202702	Kenneth Faried	107.0	1.9386639379995723	2.9774284
110.0	64.68542911669212	61.18241	-33.56395	1.1322394621606966	0.008045284787781658	43.75	202691	Klay Thompson	106.69356796116504	13.471969605907667	0.4160816
104.37388483373884	48.806098067819356	39.9155	-9.45666	0.7838212798315396	0.007396217466550127	39.8876404494382	203914	Gary Harris	111.0	10.734213405937114	1.1277439

Shot Difficulty Model



Specifiations:

Using 4 fully-connected hidden layer with 512 units, each activated with a Leaky ReLU function.
One Output layer that is fully connected, and uses Sigmoid as a final output activation function to convert it into a probability.
Epochs = 10

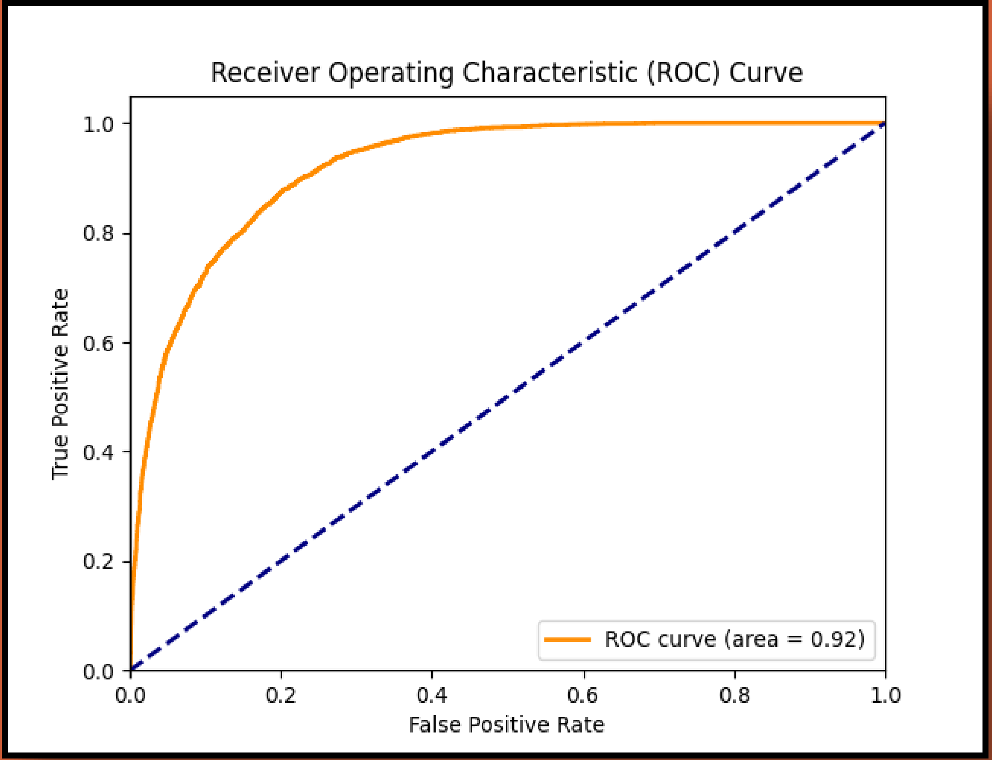
For regularization,

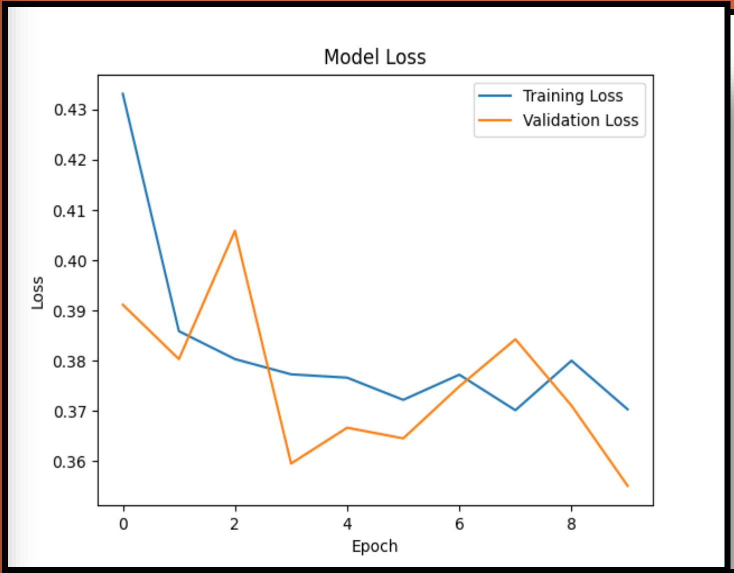
Train: Test Split = 80:20
Dropout = 0.2
Validation split = 80:20
batch_size = 32

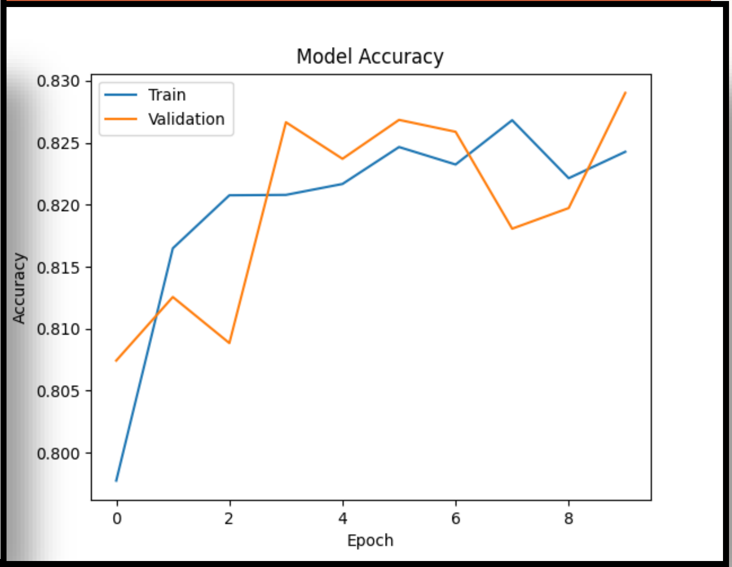
Trained the model with binary cross-entropy function.

Evaluation – Shot Model

```
1092/1092 ━━━━━━━━━━━━━━━━━━━━ 5s 5ms/step - accuracy: 0.8279 - loss: 0.3651 - val_accuracy: 0.8237 - val_loss: 0.3621
342/342 ━━━━━━━━━━━━━━━━━━━━ 0s 1ms/step - accuracy: 0.8211 - loss: 0.3719
Model Loss: 0.36629071831703186
Model Accuracy: 0.8246771097183228
```



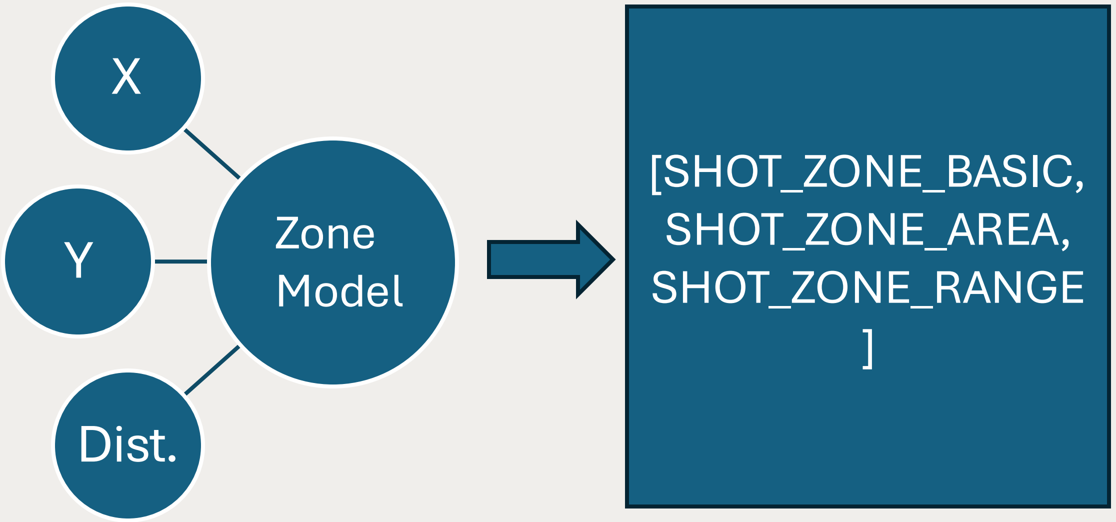




Average ROC – AUC : 0.63

```
F1 Score: 0.8202745286545959
```

Zone Model (Still Developmental)



About the Model:

Input layer connected to a dense hidden layer consisting of 128 units, each activated by ReLU.

A second dense layer with 64 units follows, again with ReLU activation.

There are three output layers: 'shot_zone', 'shot_area', and 'shot_range', each using softmax activation.

Epochs = 10

For regularization,

Train Test Split = 80:20
Dropout = 0.5
Validation split = 80:20
batch_size = 32

Trained the model with binary cross-entropy function.

```
601/601 ━━━━━━━━━━━━━━━━━━━━ 1s 1ms/step - loss: 2.7485 - shot_area_accuracy: 0.7400 - shot_range_accuracy: 0.6253 - shot_zone_accuracy: 0.5288
[2.7453575134277344, 0.7396689653396606, 0.6230353116989136, 0.5324763059616089]
```

Expected Possession Value (EPV)

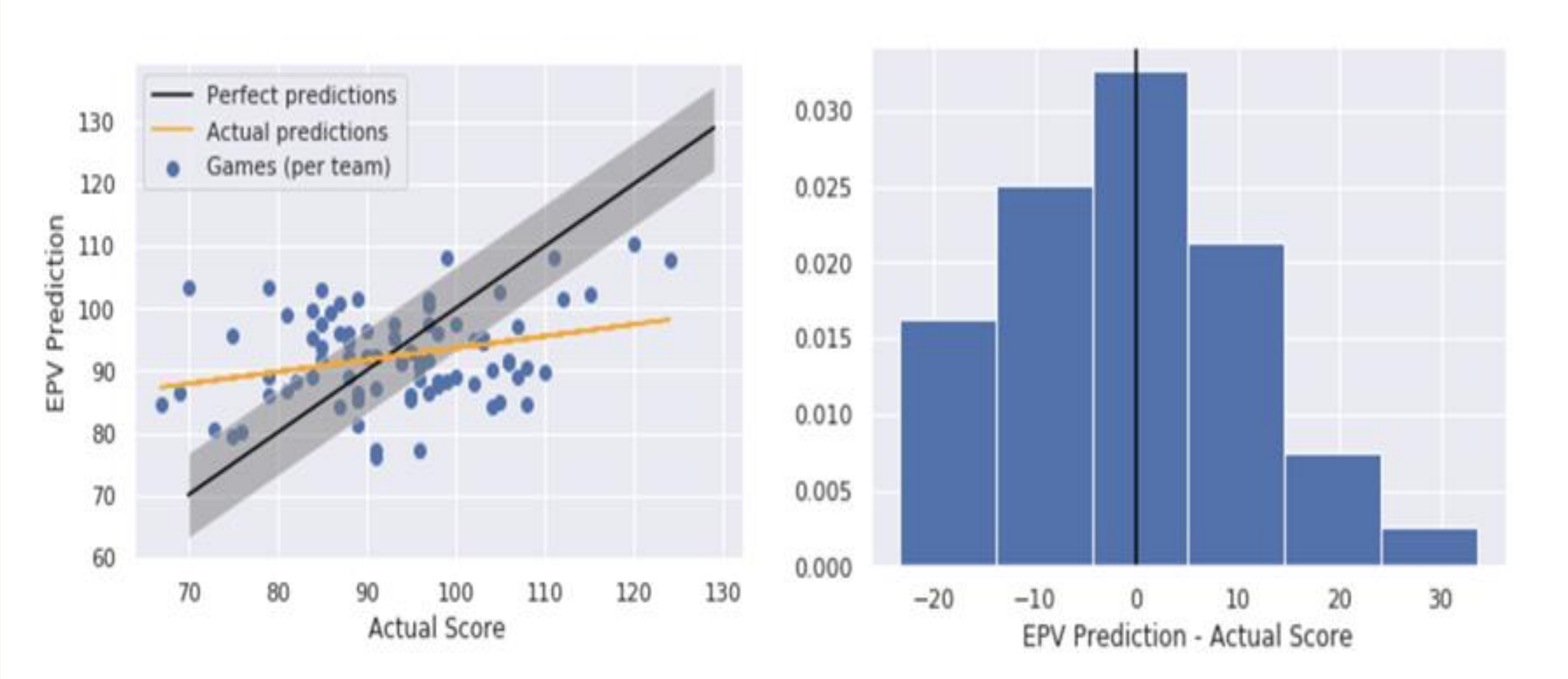
Formally, denoting the ballhandler at time t as $bh^{(t)}$, player i at time t as $p_i^{(t)}$, a completed pass as the arrow $\rightarrow$, and the point value of the shot as a variable v which can be either 2 or 3, EPV is calculated according to the equation:

$$EPV(t, p_i) = \mathbb{P}(bh^{(t)} \rightarrow p_i^{(t)}) \cdot \mathbb{P}(p_i^{(t)} \text{ scores}) \cdot v \tag{6.1}$$

where $\mathbb{P}(bh^{(t)} \rightarrow p_i^{(t)})$ is an output of the pass difficulty model, and $\mathbb{P}(p_i^{(t)} \text{ scores})$ is an output of the shot difficulty model.

EPV Evaluation

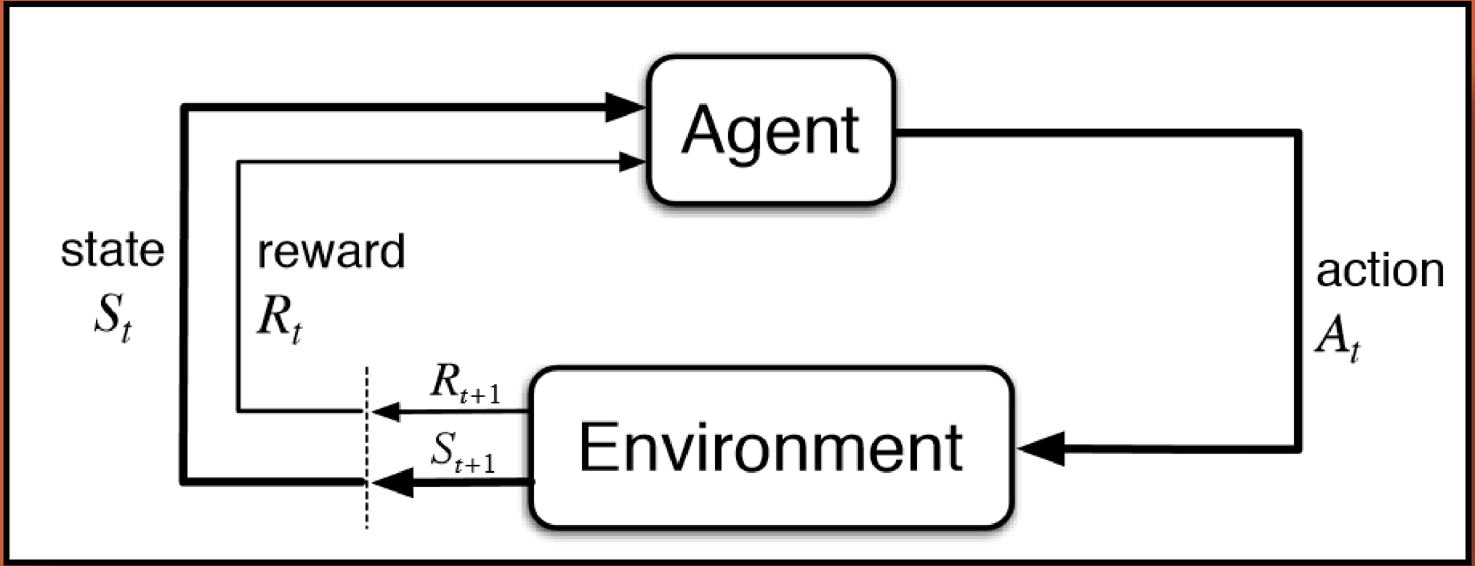
While we have already shown evaluation metrics of the pass and shot difficulty model as independent machine learning models, it is not so straightforward to do so with EPV as a whole.



Since EPV can be interpreted as points expected to be obtained from a possession, one strategy we can use is to aggregate computed EPV values from a large sample size of shots, comparing the expected total points and the actual total points scored from that collection of shots.

Deployability and Future Prospects

Agent Plays basketball: Till now we have computed the probability of making a successful pass, and the probability of shooting a successful shot. These probabilities are very useful. They can be used as the states and rewards for a Reinforcement Learning Problem. The states can be the probability with which you can transition to state B from A (Here A and B both are players) and then you goal of the agent could be to maximise the shot probability which can be interpreted as its reward. The agent cannot just pass the ball forever. If the shot clock runs out, the model is faced with a penalty.



Deployability and Future Prospects

This model that we have created can be used by the coaches for in depth analysis. This kind of analysis can be extended to the team scale, where they can evaluate team strategy in terms of the expected point value that their tactics can generate, regardless of how many points they are actually scoring. By leveraging EPV, they can identify opportunities that are being missed to potentially get more points out of a possesion, highlighting instances where game-plan and strategy has opened up good opportunities that are not being taken advantage of.


Follow the NBA Finals

Optimizing the best play in basketball using deep learning

The research focused on Division I women's basketball games, applying deep learning to predict plays that lead to the highest probability of a successful shot. The study emphasizes the potential of deep learning in sports analytics, specifically in enhancing game strategies by identifying optimal player positioning and play execution against various defensive setups.

It emphasizes the potential of machine learning in extracting actionable insights from complex sports data, offering a roadmap for applying similar methodologies to optimize strategies in other sports or game situations.

EPV of players

Player	EPV
1	1.65
2	1.45
3	1.38
4	1.21
5	1.05

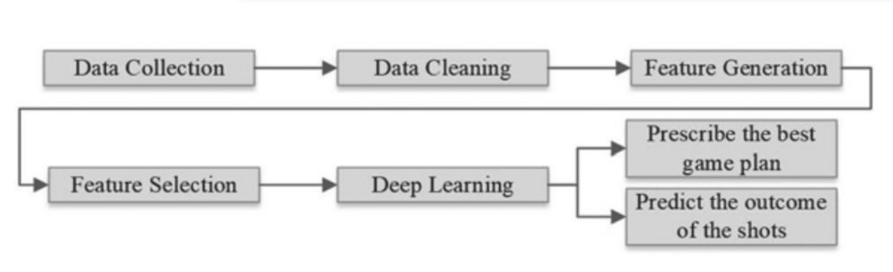


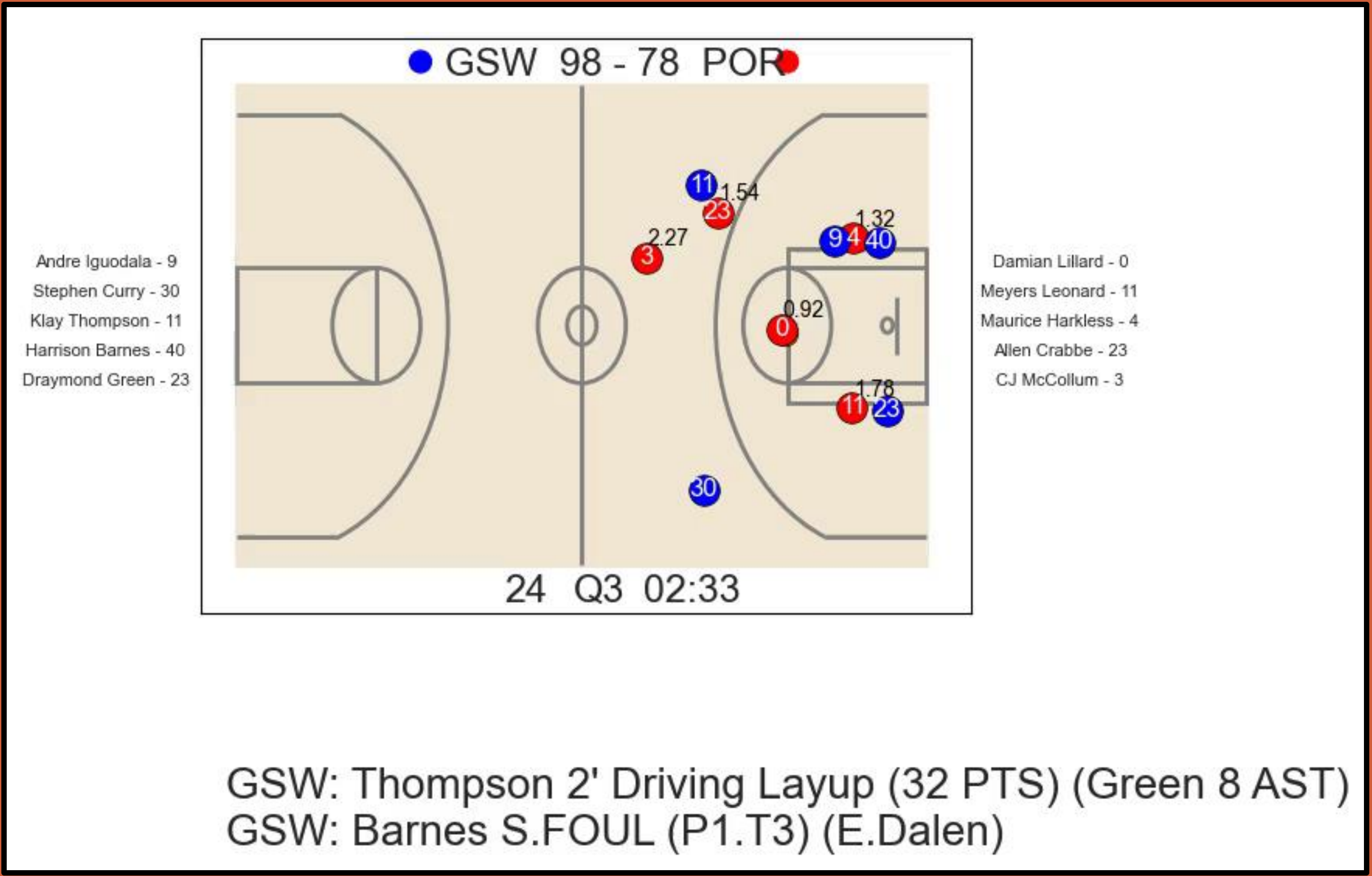
Table 1
Feature description

Attributes	Description
Player	Player who takes the shot
Play	The play that was run to get a player a shot
Defense Type	Whether the opponent's defense is in a zone or man-to-man
Defender Position	The location of defender
Screen	If a screen was used to get the player an open shot
Quarter	The quarter the shot was taken in
Seconds on Shot Clock	Number of seconds left on the shot clock
Number of Defenders	The number of defenders in the half court at the time of the shot
Location	The location shot was taken (out of 11 sports)
Hand	right or left
Shot type	Labeled as lay-up, dribble jumper, spot up, turn-around jumper (TAJ), floater, step back, or spin shot
Passes in half court	Number of passes prior to the shot
Minutes left in quarter	Minutes remaining in the quarter
2PA	2 points filed goals attempted
2PM	2 points filed goals made
3PA	3 points filed goals attempted
3PM	3 points filed goals made
FTA	free throws attempted
FTM	free throws made
Point difference	The point different of the game
Result	Make or miss

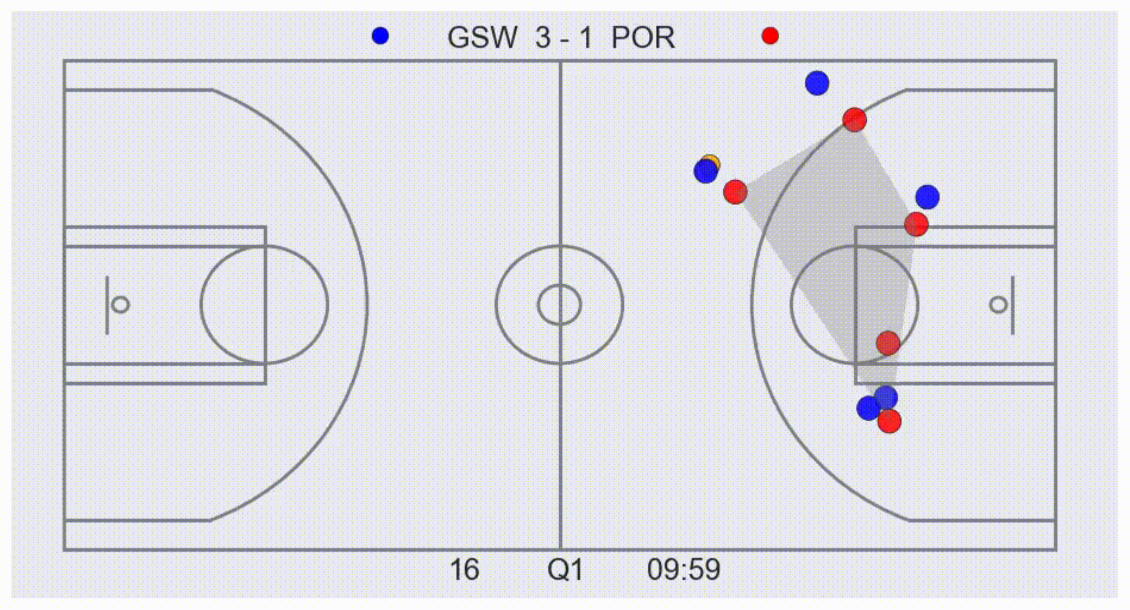
Epoch – 2 : Game Visualizer (with EPV)

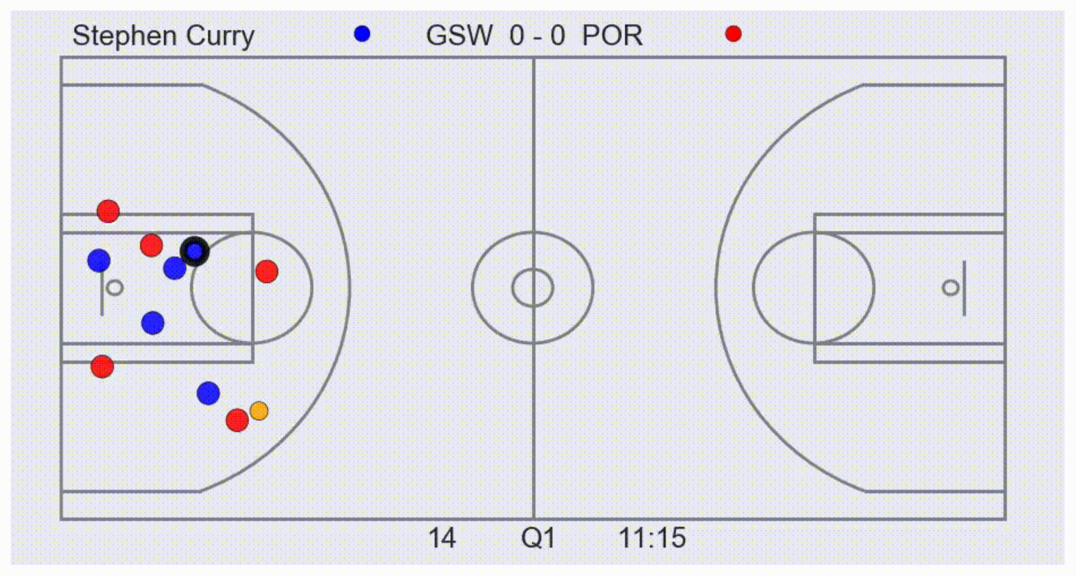
```
from simulate import Game

game = Game( date: '01.08.2016', team1: 'POR', team2: 'GSW')
game.watch_play( game_time: 1500, length: 5, show_epv=True)
```



This is the second iteration of the visualizer we developed. This iteration integrates the EPV and the visualizer. We can see the EPV of each player as the play progresses. This allows us and the coaches to analyze "What could have been the best move?" or "Was that the best play to run?"





A few more tools we have developed till now