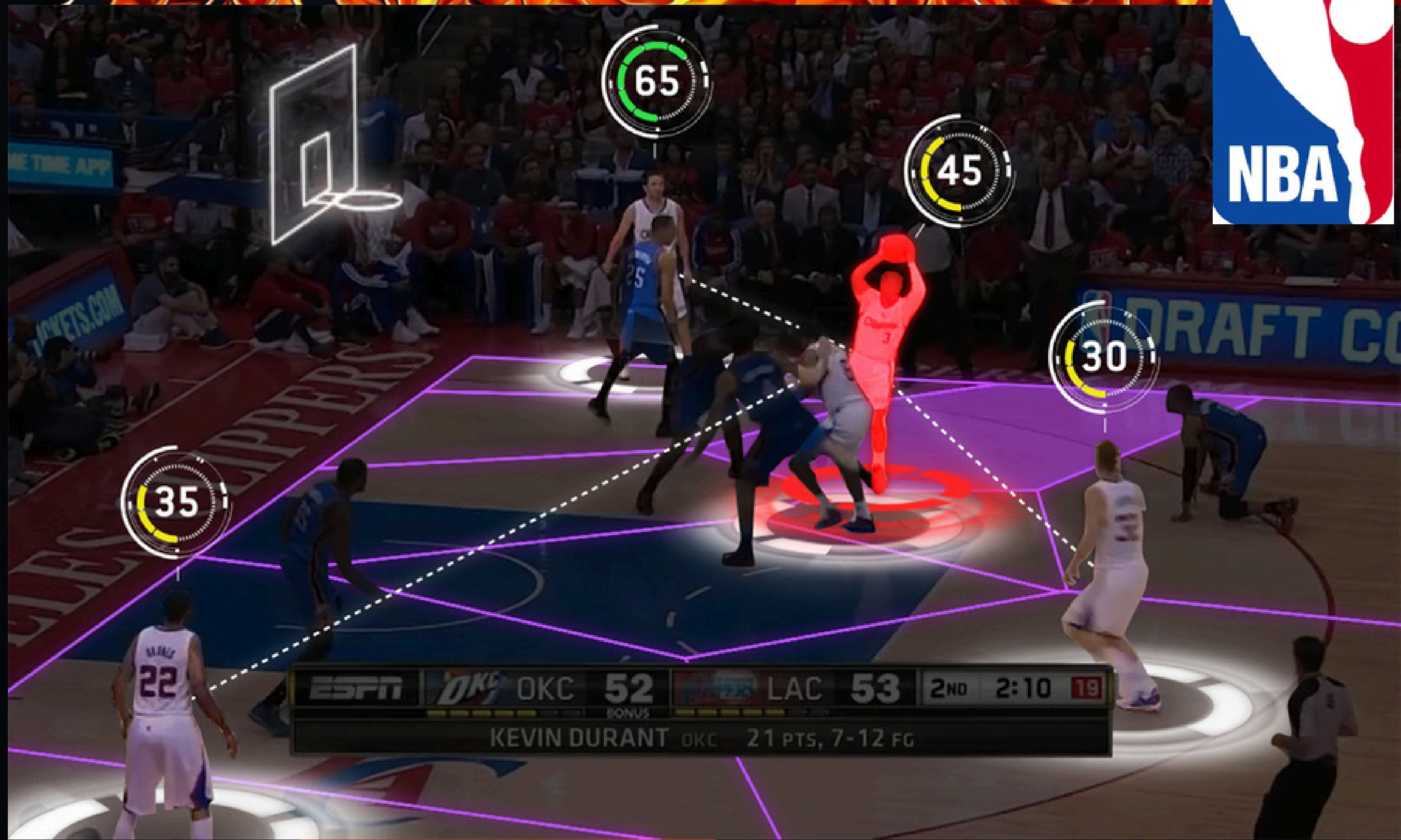
NBA ANALYSIS USING DEEP LEARNING



Vedanth Sirimalla
Kriti C Parikh

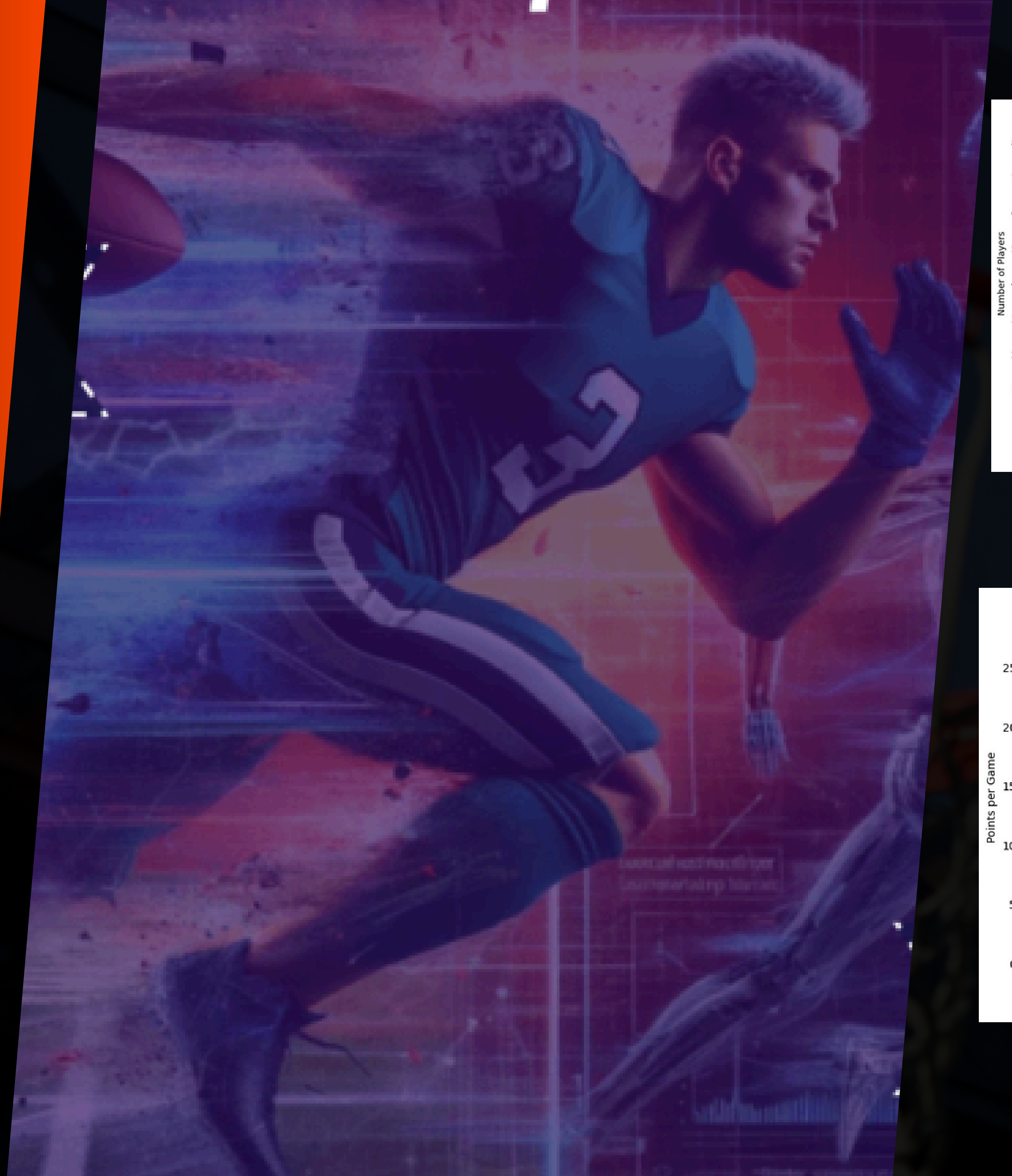
PROBLEM STATEMENT

• Develop deep learning models to predict NBA player Points Per Game (PPG)
• Compare effectiveness of 6 neural network architectures
• Evaluate model accuracy in forecasting player scoring output
• Identify most effective deep learning approach for NBA performance prediction

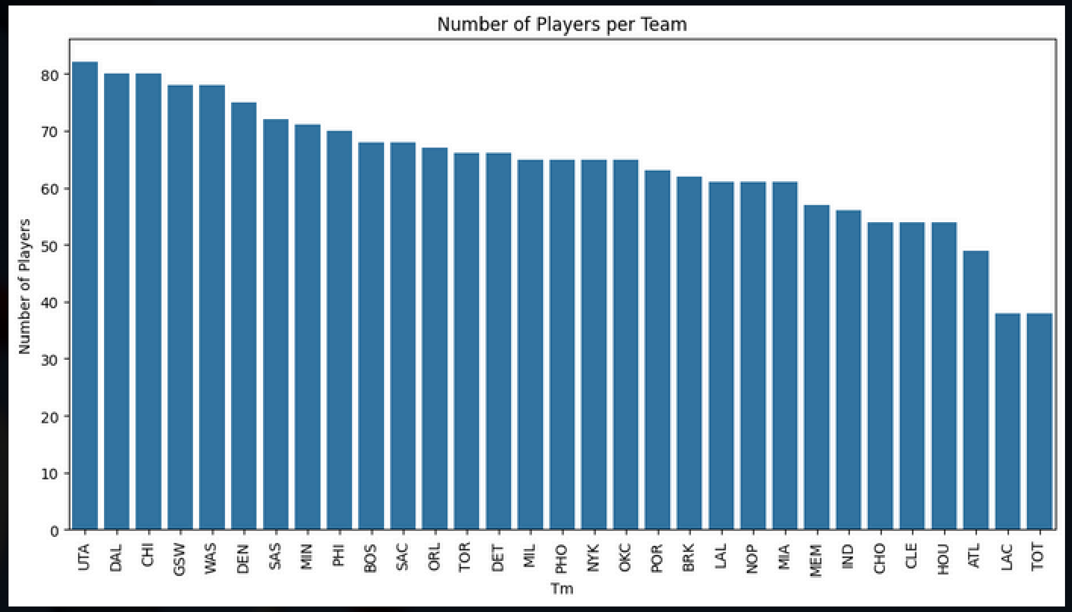
DATA PREPARATION

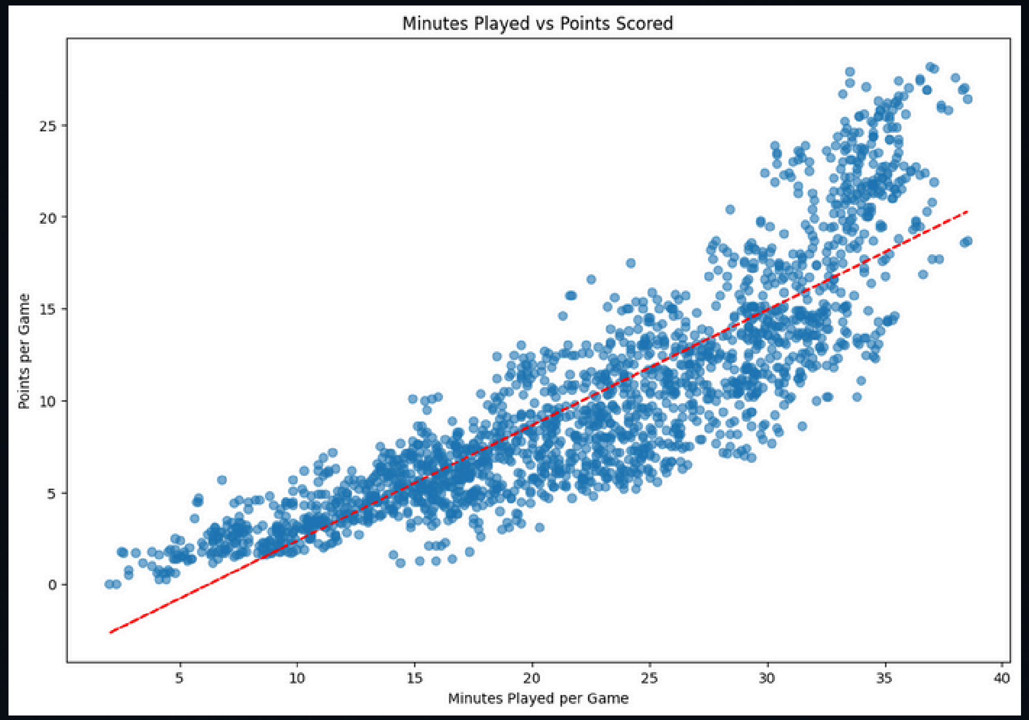
- Data Source: NBA player statistics for 2023-2024 season.
  - 5,522 initial rows, 29 features
- Data Cleaning: Removed 1,612 duplicate rows
- Handled missing values
- Final dataset: 3,910 rows
- Feature Engineering:
  - Created PointsPerMinute, TotalRebounds, ShootingEfficiency
- Data Normalization: Applied Standard Scaler to numerical features
- Outlier Removal: Used IQR method

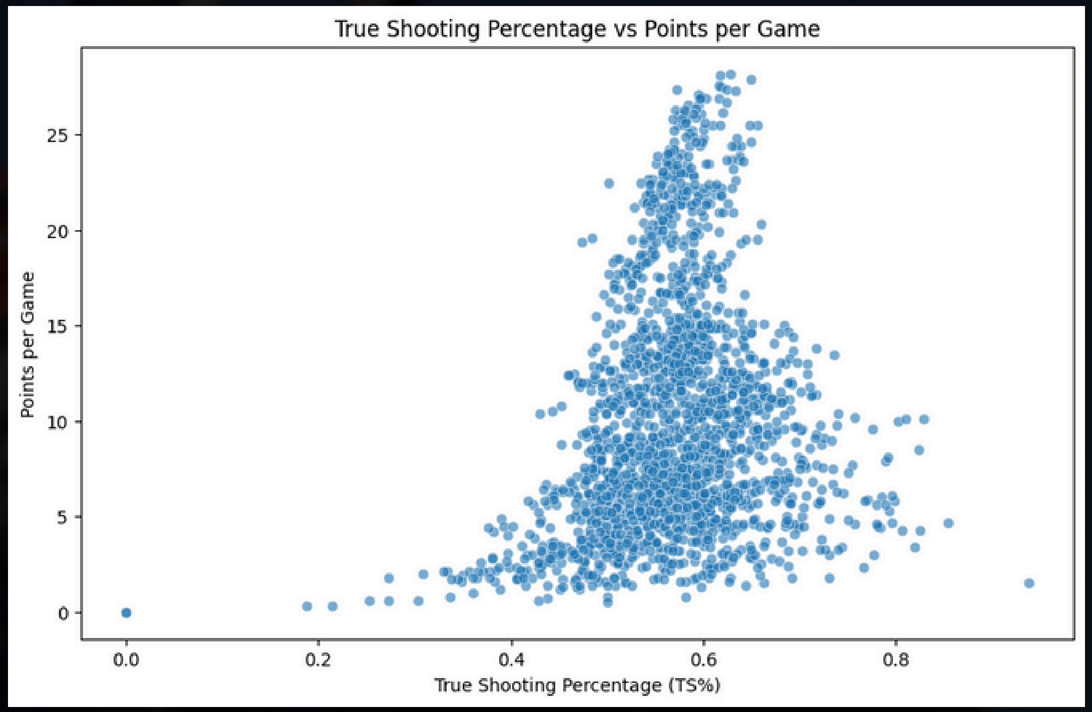
AI in Sports

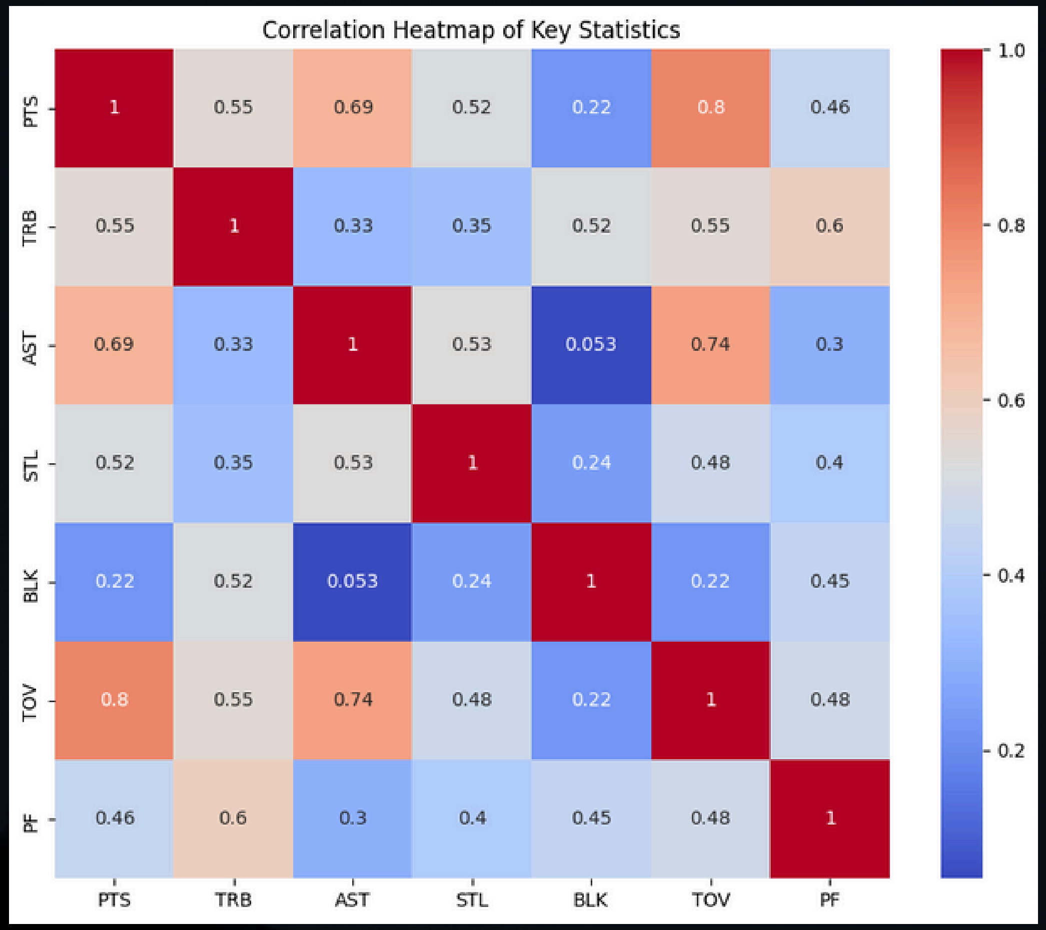


EXPLORATORY DATA ANALYSIS

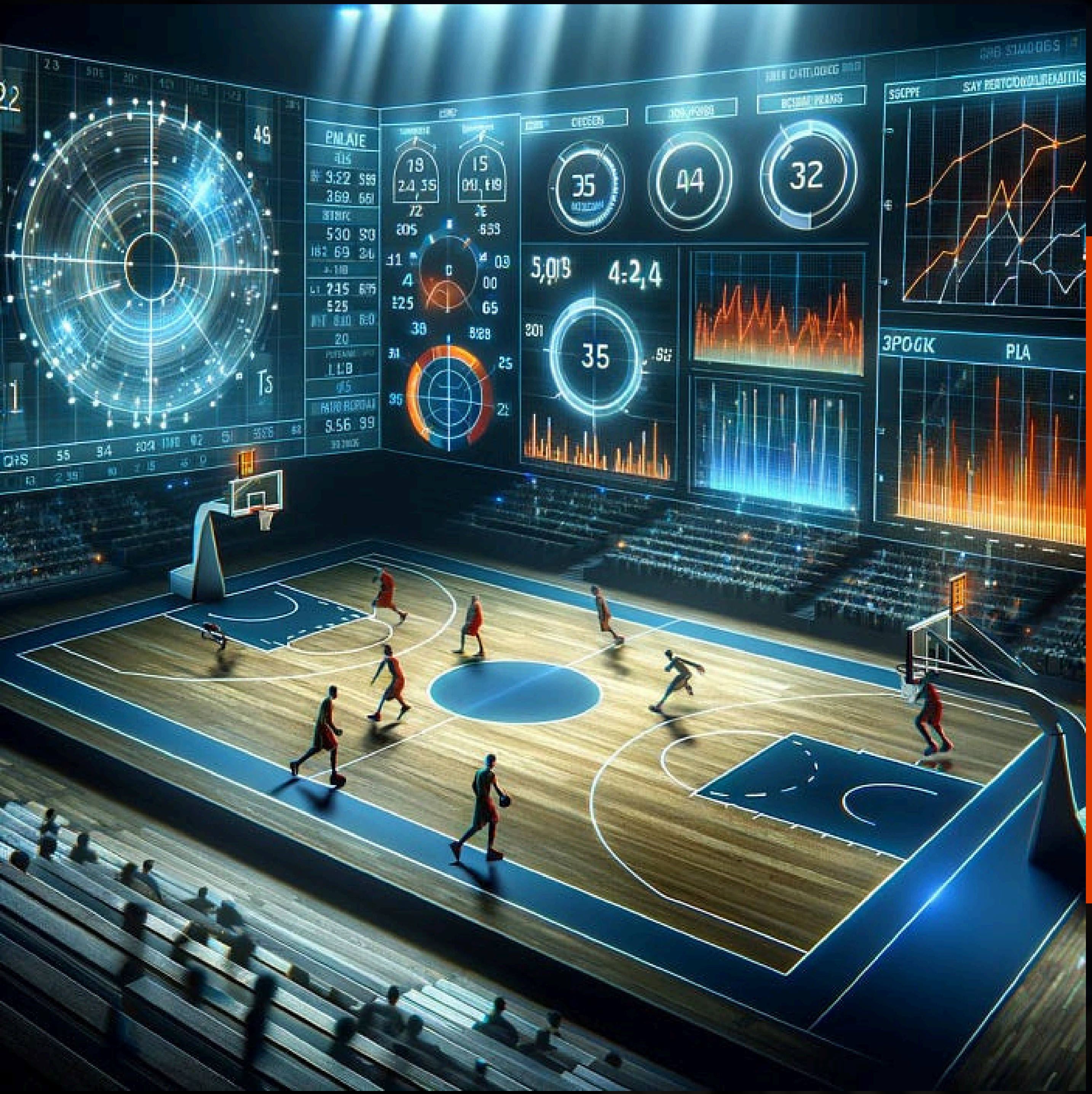








DEEP LEARNING MODELS



1. LSTM
Input shape: (1, num_features)
Architecture:
- LSTM(64, return_sequences=True)
- LSTM(32)
- Dense(16, activation='relu')
- Dense(1)
Additional: Early stopping with patience=10

2. MLP
Input shape: (num_features,)
Architecture:
- Dense(64, activation='relu')
- Dropout(0.2)
- Dense(32, activation='relu')
- Dropout(0.2)
- Dense(16, activation='relu')
- Dense(1)

3. CNN
Input shape: (num_features, 1)
Architecture:
- Conv1D(64, kernel_size=3, activation='relu')
- MaxPooling1D(pool_size=2)
- Flatten()
- Dense(50, activation='relu')
- Dense(1)

4. BiLSTM
Input shape: (1, num_features)
Architecture:
- Bidirectional(LSTM(64, return_sequences=True))
- Bidirectional(LSTM(32))
- Dense(16, activation='relu')
- Dense(1)

5. GRU
Input shape: (1, num_features)
Architecture:
- GRU(64, return_sequences=True)
- GRU(32)
- Dense(16, activation='relu')
- Dense(1)

6. Transformer
Input shape: (1, num_features)
Architecture:
- MultiHeadAttention(head_size=256, num_heads=4)
- Dropout(0.1)
- LayerNormalization()
- Dense(4, activation='relu')
- Dense(input_shape[-1])
- GlobalAveragePooling1D()
- Dense(1)

- Training
- Optimizer: Adam
- Loss function: MSE
- Training epochs: 100
- Batch size: 32
- Validation split: 20%

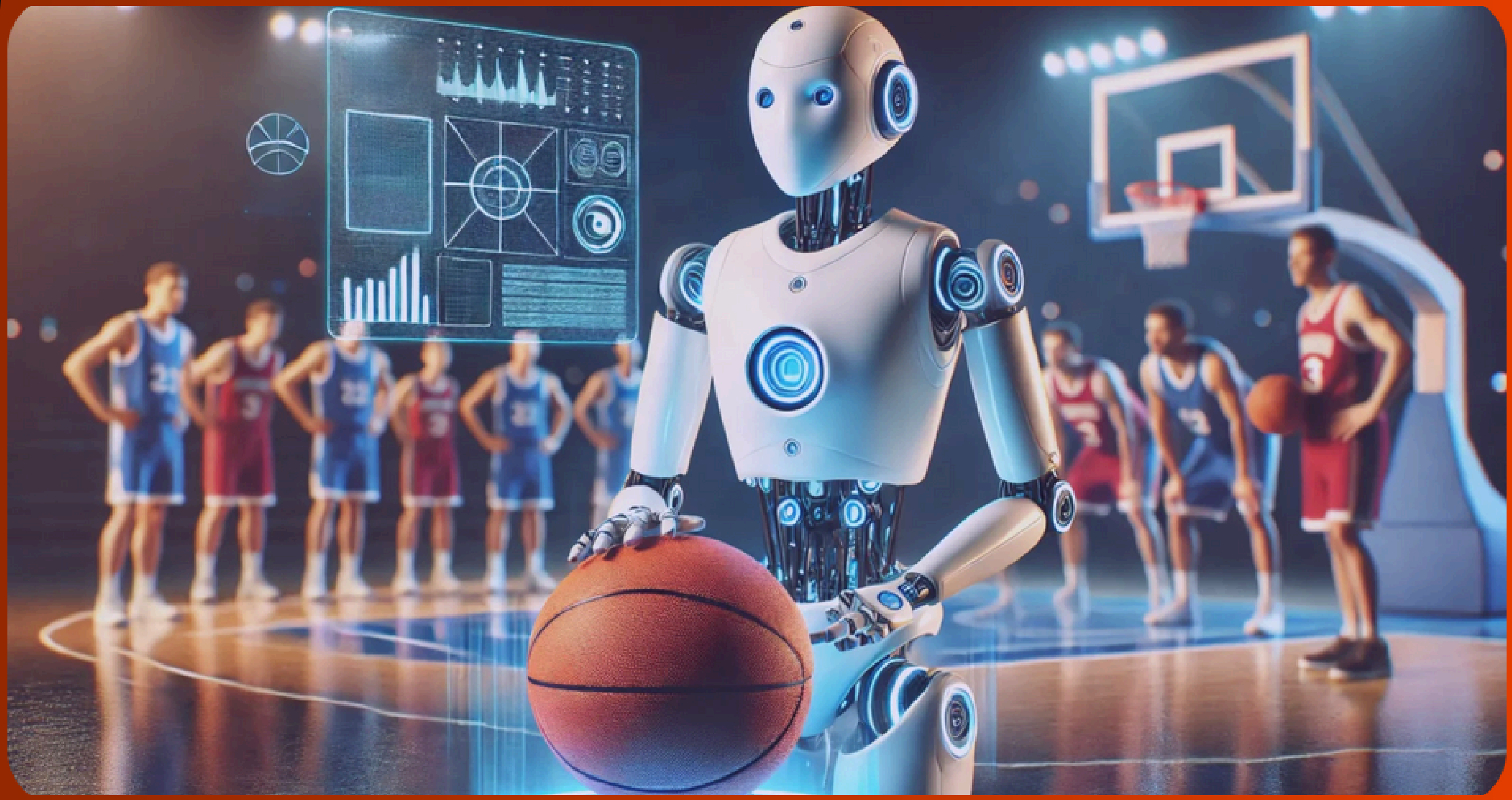
MODEL PERFORMANCE COMPARISON

	Model	MSE	RMSE	R2
0	LSTM	3.527810	1.878247	0.892709
1	MLP	3.845483	1.960990	0.883048
2	CNN	4.386096	2.094301	0.866607
3	BiLSTM	3.278220	1.810586	0.900300
4	GRU	3.320824	1.822313	0.899004
5	Transformer	5.137527	2.266611	0.843753

Case Study:

```
Selected player: Dante Exum
Actual PPG: 4.40
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 77ms/step
LSTM predicts 4.17 PPG
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 93ms/step
MLP predicts 3.99 PPG
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 170ms/step
CNN predicts 4.24 PPG
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 103ms/step
BiLSTM predicts 4.38 PPG
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 86ms/step
GRU predicts 4.10 PPG
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 121ms/step
Transformer predicts 3.47 PPG
```

CONCLUSION : Slam Dunking the Data



Fun Facts from Our Analysis:

- The Turnover Tango: Points and turnovers dance together with a 0.80 correlation - showing that even the best scorers fumble sometimes!
- The Efficiency Sweet Spot: Our scatter plot reveals that better shooters score more points (who knew?), but some players break the mold.
- Team Size Shuffle: From Utah's packed roster (~80 players) to smaller lineups (~40 players), teams play their own numbers game.

The MVP (Most Valuable Predictor):

- BiLSTM took home the trophy with an R^2 of 0.900.
- Predicted Dante Exum's 4.40 PPG with impressive accuracy (4.38).
- Other models brought their A-game too, most staying within half a point.

What We Learned:

Just like a well-executed fast break, our deep learning models showed that with the right combination of statistics and smart algorithms, we can predict player performance with impressive accuracy.

Whether you're a coach planning your next game strategy or a fantasy basketball enthusiast looking for that winning edge, these insights prove that in basketball, as in data science, it's all about making the right moves at the right time.

THANK YOU

Any questions?

REFERENCES

- https://www.kaggle.com/datasets/bryanchungweather/nba-player-stats-dataset-for-the-2023-2024
- https://www.simplilearn.com/tutorials/deep-learning-tutorial/deep-learning-algorithm
- https://viso.ai/deep-learning/deep-neural-network-three-popular-types/
- https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be
- https://www.sloansportsconference.com/research-papers/using-deep-learning-to-understand-patterns-of-player-movement-in-the-nba
- https://arxiv.org/abs/2111.09695
- https://www.tandfonline.com/doi/full/10.1080/24751839.2021.1977066#d1e1319
- https://github.com/luke-lite/NBA-Prediction-Modeling
- https://link.springer.com/article/10.1007/s10115-024-02092-9