**DEEP LEARNING – 046211**

**Winter 2024-2025**

**Final Project Report**

Project Name:

**NBA Stats Transformer: Game Insights with Deep Learning**

Shadi Safadi - 322253089

Idan Bason - 211643630

Date: 4.2.2025

**Introduction**

**Project Goal**

The primary goal of this project is to enhance predictive modeling and performance analytics in NBA games through comprehensive data collection, preprocessing, and model development. By integrating detailed game statistics and opponent data, we aim to develop accurate predictive models that can forecast game outcomes, aiding data-driven decision-making in sports analytics.

**Motivation**

The motivation behind this project stems from the increasing importance of data analytics in sports. Accurate predictions of game outcomes can provide valuable insights for teams, coaches, analysts, and bettors. Furthermore, understanding the factors that influence game results can help teams make informed strategic decisions, improve player performance, and enhance fan engagement. The ability to predict game outcomes accurately can also offer a competitive edge in a highly dynamic and competitive sports environment.

**Previous Work**

Previous work in this domain has primarily focused on utilizing traditional statistical methods and basic machine learning techniques to predict game outcomes. Researchers have explored various factors such as team performance, player statistics, historical game results, and situational factors. While these approaches have shown promise, they often lack the ability to capture the complex interactions between different factors and the temporal dynamics of thegames. Recent advancements in machine learning and deep learning have opened up new possibilities for developing more sophisticated predictive models that can handle large and complex datasets.

**Method**

**Explanation of the Algorithm**

The predictive models developed in this project leverage the pytorch-tabular library, which provides a framework for building tabular data models using PyTorch. The primary model architecture used is the Category Embedding Model, which is well-suited for handling categorical and numerical features. The model is trained using a supervised learning approach, with the target variable being the game outcome (win/loss).

**Architecture**

The Category Embedding Model architecture consists of several key components:

* Embedding Layers: These layers convert categorical features into dense vector representations, capturing the underlying relationships between categories.
* Fully Connected Layers: These layers process the combined embeddings and numerical features through multiple layers of neurons, applying non-linear transformations to learn complex patterns in the data.
* Dropout Layers: Dropout is used to prevent overfitting by randomly setting a fraction of the input units to zero during training.
* Output Layer: The final layer produces the predicted probabilities for the target classes (win/loss).

**Loss Function**

The loss function used for training the model is the Binary Cross-Entropy Loss, which is appropriate for binary classification tasks. This loss function measures the difference between the predicted probabilities and the actual labels, penalizing incorrect predictions more heavily.

**Hyperparameter Optimization**

Hyperparameter optimization is performed using Optuna, a powerful optimization framework that systematically searches for the best hyperparameter configurations. This involves defining a search space for hyperparameters such as learning rate, dropout rate, hidden layers, hidden dimensions, and batch size. Optuna evaluates different configurations through multiple trials and identifies the best-performing set of hyperparameters.

**Experiments and Results**

**Dataset**

The dataset for this project is collected from the NBA API, focusing on the most recent 100 games for each team. This ensures that the data is up-to-date and relevant. The dataset includes a wide range of features, such as team performance metrics, game context, and opponent statistics. Opponent stats are integrated into the dataset by merging the main game data with opponent team statistics, allowing for a comprehensive view of each game.

**Experiments**

Several experiments are conducted to validate the predictive models and their performance:

1. Baseline Model: A baseline model is developed using basic statistical features and evaluated for its predictive accuracy.
2. Feature Engineering: Additional features such as opponent stats are integrated into the dataset, and their impact on model performance is evaluated.
3. Hyperparameter Tuning: The impact of different hyperparameter configurations is assessed using Optuna to identify the best-performing model.
4. Model Comparison: The performance of the Category Embedding Model is compared to other machine learning models such as Logistic Regression and Random Forest.

**Results**

The results of the experiments demonstrate significant improvements in predictive accuracy with the integration of opponent stats and hyperparameter optimization. The best-performing model achieves a high accuracy score on the test set, outperforming the baseline model and other traditional machine learning models. The confusion matrix and loss curves further validate the model's effectiveness in predicting game outcomes. The visualizations generated using Optuna's tools provide insights into the importance of different hyperparameters and the optimization process.

**Conclusion/Future Work**

**Conclusion**

This project successfully demonstrates the potential of advanced machine learning techniques and comprehensive data integration in predicting NBA game outcomes. The use of pytorch-tabular for model development, combined with Optuna for hyperparameter optimization, results in accurate and robust predictive models. The integration of opponent stats further enhances the model's ability to capture the dynamics of each game, leading to improved performance.

**Future Work**

Future work can focus on several areas to further enhance the predictive models:

* Feature Expansion: Incorporating additional features such as player-level statistics, injury reports, and real-time game data can provide more granular insights and improve predictive accuracy.
* Temporal Dynamics: Developing models that capture the temporal dynamics and trends over time can provide a deeper understanding of team performance and game outcomes.
* Explainability: Enhancing the explainability of the models by using techniques such as SHAP (SHapley Additive exPlanations) can help stakeholders understand the factors driving the predictions.
* Real-Time Predictions: Implementing real-time prediction capabilities to provide live insights during games can offer valuable information for coaches, analysts, and fans.

In conclusion, this project lays a strong foundation for utilizing data analytics and machine learning in sports analytics, providing valuable insights and accurate predictions for NBA games. With continued advancements and refinements, these techniques can revolutionize the way sports data is analyzed and utilized for decision-making.

**Ethics Statement:**

**Student Names:** Shadi Safadi, Idan Bason

**Project Title:** NBA Stats Transformer: Game Insights with Deep Learning

**Describe your project:** This project is focused on analyzing historical NBA game data to enhance predictive modeling and performance analytics. The goal is to collect, preprocess, and augment game statistics with additional information, enabling comprehensive analysis and accurate predictions.

**2a.**

1. **NBA Teams and Coaches: Teams and coaching staff may use the predictions to inform game strategies, player rotations, and training programs.**
2. **Sports Bettors and Gambling Industry: Bettors and gambling platforms may rely on the predictions to make informed bets or set odds for NBA games.**
3. **NBA Fans and Viewers: Fans may use the predictions to enhance their engagement with the sport, discuss outcomes, or participate in fantasy leagues.**

**2b.**

1. **NBA Teams and Coaches:**
   * **"This project uses historical NBA game data to predict future game outcomes. The predictions are based on statistical patterns and trends, which can help you make data-driven decisions about game strategies, player performance, and training adjustments. However, the predictions are probabilistic and should be used as a supplementary tool rather than a definitive guide."**
2. **Sports Bettors and Gambling Industry:**
   * **"The project provides predictions for NBA game outcomes based on historical data and statistical analysis. These predictions can help you make more informed betting decisions or set more accurate odds. However, the predictions are not guaranteed and should be used alongside other factors, such as player injuries or team dynamics, to make well-rounded decisions."**
3. **NBA Fans and Viewers:**
   * **"This project predicts NBA game results using historical data and advanced analytics. As a fan, you can use these predictions to enhance your understanding of the game, engage in discussions, or participate in fantasy leagues. Keep in mind that the predictions are based on probabilities and may not always reflect the actual outcome due to the unpredictable nature of sports."**

**2c.**

1. **NBA Teams and Coaches:**
   * **Responsible Party: The project team, including data scientists and sports analysts, should explain the model to teams and coaches. They can provide detailed insights into how the predictions are generated and how they can be integrated into decision-making processes.**
2. **Sports Bettors and Gambling Industry:**
   * **Responsible Party: The project team, in collaboration with gambling platforms or sports analysts, should communicate the predictions and their limitations to bettors. This ensures transparency and helps bettors understand the risks and uncertainties involved.**
3. **NBA Fans and Viewers:**
   * **Responsible Party: The project team, along with media outlets or sports commentators, should explain the predictions to fans. This can be done through articles, social media posts, or broadcasts, ensuring that fans understand the probabilistic nature of the predictions and how to interpret them.**

**3a.**

**We think that the most important thing to add is custom disclaimers that are as relevant as possible to the generated material, and that resonates with the user. In addition, there should be more clear laws and restrictions on the model developers and users, and clearly stating them in the disclaimer.**