UNIVERSITAS
GADJAH MADA
**Comparative Evaluation of Machine Learning and Deep Learning Models for Major League Baseball Game
Outcome Prediction**
Taichi Mori, Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR ISI

PRAKATA **iv**
DAFTAR ISI **v**
LIST OF TABLES **viii**
LIST OF FIGURES **ix**
LIST OF LISTINGS **x**
INTISARI **xi**
ABSTRACT **xiii**
CHAPTER I INTRODUCTION **1**
1.1 Research Background 1
1.2 Research Problem 3
1.3 Research Scope 3
1.4 Research Objectives 4
1.5 Research Advantages 5
CHAPTER II LITERATURE REVIEW **6**
2.1 Introduction 6
2.2 Machine Learning in Baseball Outcome Prediction 6
2.3 Feature Engineering and Data Representation 7
2.4 Comparative Evaluation of ML Models 8
2.5 Generalization to Other Sports and Research Gaps 10
2.6 Conclusion 10
CHAPTER III THEORETICAL BASES **11**
3.1 Introduction 11
3.2 Tabular Data 11
3.3 Supervised Learning 11
3.4 Machine Learning Algorithms in Sports Analytics 12
3.4.1 Logistic Regression 12
3.4.2 Decision Trees and Random Forests 12
3.4.3 Support Vector Machines (SVM) 12
3.4.4 Gradient Boosting Machines (XGBoost) 12
3.5 Deep Learning and Neural Networks 12
3.5.1 Neural Networks 12
3.5.2 1D Convolutional Neural Networks (CNN) 13
3.6 Concepts of Feature Engineering and Optimization 13
3.6.1 Feature Engineering 13
3.6.2 Hyperparameter Tuning 13
3.6.3 Sports Analytics and Data Characteristics 13
3.7 Performance Evaluation 13
3.8 Relevance to this Study 14

UNIVERSITAS
GADJAH MADA
Comparative Evaluation of Machine Learning and Deep Learning Models for Major League Baseball Game Outcome Prediction
Taichi Mori, Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

CHAPTER IV RESEARCH METHODOLOGY 15
4.1 Research Description . . . 15
4.2 Research Flowchart . . . 16
4.3 Tools and Materials . . . 18
4.3.1 Software Tools . . . 18
4.3.2 Libraries . . . 18
4.3.3 Materials . . . 19
4.4 Data Collection . . . 19
4.5 Data Preprocessing . . . 21
4.5.1 Data Cleaning . . . 22
4.5.2 Feature Engineering . . . 23
4.5.3 Encoding and Normalization . . . 23
4.5.4 Data Validation and Splitting . . . 23
4.5.5 Feature Selection . . . 24
4.6 Model Implementation . . . 25
4.6.1 Logistic Regression (Baseline Model) . . . 25
4.6.2 Decision Tree Classifier . . . 25
4.6.3 Random Forest Classifier . . . 25
4.6.4 Support Vector Machine (SVM) . . . 26
4.6.5 XGBoost Classifier . . . 26
4.6.6 Artificial Neural Network (ANN) . . . 26
4.6.7 One-Dimensional Convolutional Neural Network (1D-CNN) 26
4.7 Model Evaluation . . . 27
CHAPTER V IMPLEMENTATION 29
5.1 Overview . . . 29
5.2 Environment Setup . . . 29
5.3 Data Collection . . . 30
5.3.1 FanGraphs Sabermetrics via pybaseball . . . 30
5.3.2 Statcast Data via pybaseball . . . 33
5.3.3 Retrosheet Game Logs . . . 33
5.3.4 Lahman's Baseball Database . . . 34
5.3.5 Betting Odds Data . . . 34
5.4 Team Code Standardization . . . 35
5.5 Data Merging . . . 35
5.6 Missing Value Handling . . . 35
5.7 Feature Engineering – Rolling Metrics . . . 35
5.8 Final Dataset and Feature Summary . . . 35
5.9 Train-Test Split and Preprocessing Pipeline . . . 36
5.10 Model Implementation . . . 37
5.10.1 Logistic Regression . . . 37
5.10.2 Decision Tree . . . 38
5.10.3 Random Forest . . . 38
5.10.4 Support Vector Machine . . . 38
5.10.5 XGBoost . . . 38
5.10.6 Artificial Neural Network (ANN) . . . 39
5.10.7 One-Dimensional Convolutional Neural Network (1D-CNN) 39

UNIVERSITAS
GADJAH MADA
Comparative Evaluation of Machine Learning and Deep Learning Models for Major League Baseball Game Outcome Prediction
Taichi Mori, Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.id/

5.11 Model Evaluation Function 40
CHAPTER VI RESULTS AND DISCUSSION 41
6.1 Overview 41
6.2 Dataset Summary 41
6.3 Overall Model Performance 42
6.4 Analysis of Individual Model Performance 43
6.4.1 Logistic Regression 43
6.4.2 1D-CNN 44
6.4.3 Random Forest 44
6.4.4 Support Vector Machine 45
6.4.5 XGBoost 45
6.4.6 Decision Tree 45
6.4.7 Artificial Neural Network 45
6.5 Feature Importance Analysis 46
6.6 Comparison with Prior Studies 49
6.7 Discussion 50
CHAPTER VII CONCLUSION AND FUTURE WORK 52
7.1 Conclusion 52
7.2 Research Limitations 53
7.3 Future Work 54
REFERENCES 55
APPENDIX 57

UNIVERSITAS
GADJAH MADA
Comparative Evaluation of Machine Learning and Deep Learning Models for Major League Baseball Game Outcome Prediction
Taichi Mori, Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

LIST OF TABLES

2.1 Summary of Machine Learning Models Evaluated in Previous MLB Outcome Prediction Studies 9
2.2 Strengths and Weaknesses of Reviewed Machine Learning Approaches 9

4.1 Example of team-level sabermetrics used as model features (FanGraphs) 20

5.1 Development environment allocation by pipeline stage 30
5.2 Complete feature set used for model training (19 features, N = 26,822) 36
5.3 Confirmed dataset dimensions from implementation 37

6.1 Summary statistics of the unified MLB dataset 42
6.2 Comparative test-set performance of all seven models (sorted by AUC-ROC) 42
6.3 Top 10 feature importances from the Random Forest model 48
6.4 Accuracy comparison across prior studies and this research 49

UNIVERSITAS
GADJAH MADA
**Comparative Evaluation of Machine Learning and Deep Learning Models for Major League Baseball Game
Outcome Prediction**
Taichi Mori, Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

LIST OF FIGURES

2.1 Performance comparison of Machine Learning Models across prior studies 7

4.1 Research flowchart of the study 17
4.2 FanGraphs explanation and formula of Weighted Runs Created Plus (wRC^+), a park-adjusted measure of offensive performance. 20
4.3 Example of Statcast player profile (Baseball Savant) containing xwOBA, xSLG, exit velocity, sprint speed, and other advanced metrics used to construct player- and team-level features. 21

5.1 Output of FanGraphs team sabermetric data retrieved via pybaseball (first rows shown). Each row represents one team-season combination with wOBA, OPS, wRC+, and FIP values. 32
5.2 Confirmed Retrosheet dataset shape (26,822 rows, 8 columns) and first rows of the parsed game log. Each row represents one team's perspective for a single game, with the binary `target_win` label derived from the score comparison. 34
5.3 Confirmed dataset split dimensions. The training set contains 21,457 observations and the test set contains 5,365 observations, each with 19 features. 37
5.4 ANN (MLPClassifier) cross-validation scores and test-set evaluation output from VS Code. The model achieved a mean CV AUC of approximately 0.557 and a test-set AUC of 55.75%. 39
5.5 1D-CNN model architecture and training output. The network comprises one Conv1D layer (32 filters, kernel size 2), max-pooling, flattening, a dense layer (64 units, ReLU, dropout 0.3), and a sigmoid output. Training used early stopping on validation loss with patience 5, running for up to 30 epochs with a batch size of 32. . . . 40

6.1 ROC curves for all seven models on the hold-out test set (N = 5,365). Logistic Regression achieves the highest AUC (0.615), followed closely by the 1D-CNN (0.613). The ANN performs worst (AUC = 0.557), and all curves lie between 0.55 and 0.62, reflecting the inherent difficulty of predicting MLB game outcomes from team-level features. 43
6.2 Top 10 feature importances from the Random Forest model (mean decrease in impurity). Betting odds features (moneyLine and projectedRuns) dominate all other predictors by a substantial margin, followed by the three engineered rolling features. 47

UNIVERSITAS
GADJAH MADA
**Comparative Evaluation of Machine Learning and Deep Learning Models for Major League Baseball Game
Outcome Prediction**
Taichi Mori, Drs. Edi Winarko, M.Sc., Ph.D.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

LIST OF LISTINGS

A.1 Library imports used throughout the project (Phase 1: Colab; Phase 2: VS Code) . . . 57
A.2 FanGraphs data collection via pybaseball . . . 58
A.3 Statcast data collection and team-level aggregation . . . 58
A.4 Retrosheet game log parsing function . . . 59
A.5 Lahman database processing . . . 60
A.6 Betting odds data processing . . . 61
A.7 Team abbreviation standardization mapping . . . 61
A.8 Sequential left-join merging of all data sources . . . 62
A.9 Missing value imputation strategy . . . 63
A.10 Rolling performance feature engineering . . . 63
A.11 Train-test split first, followed by per-partition preprocessing . . . 64
A.12 Shared cross-validation and evaluation helper functions . . . 66
A.13 Logistic Regression implementation . . . 67
A.14 Decision Tree implementation . . . 67
A.15 Random Forest implementation . . . 67
A.16 SVM implementation . . . 68
A.17 XGBoost implementation . . . 68
A.18 ANN (MLPClassifier) implementation . . . 69
A.19 Data loading for 1D-CNN from locally saved CSV files . . . 69
A.20 Reshaping feature matrices into 3D tensors for Conv1D input . . . 70
A.21 1D-CNN architecture, compilation, and training . . . 70
A.22 Unified evaluation function used for all seven models . . . 71