ACTA INFOLOGICA 2023;7(1):139-151
dergipark.org.tr/acin

İSTANBUL UNIVERSITY PRESS

DOI: 10.26650/acin.1180583

RESEARCH ARTICLE

Machine Learning Models on MOBA Gaming: League of Legends Winner Prediction

MOBA Oyunlarında Makine Öğrenimi Teknikleri: League of Legends Kazanan Tahmini

Kaan Arık[1]

ABSTRACT

The entertainment industry includes companies engaged in telecommunications services, television, music streaming, video games, and live events. Gaming has gained momentum in revenue growth in the entertainment industry over the past decade. This momentum has made the gaming industry one of the most popular areas of the entertainment industry. Official leagues have been teamed up with professional players, and the concept of e-sports has become widespread. MOBA (Multiplayer Online Battle Arena), which is a derivative of MMO (massively multiplayer online) games, is the name given to the games played on the Internet in which players destroy the opponent's base by dominating specific objectives on a map, usually with two teams of five players each. LoL (League of Legends) is one of the most popular MOBA games. Predicting winners in online games has become an essential application for machine learning models. This research aims to predict classification with machine learning methods of match winner with LoL player metrics. Key performance metrics and their impact on each game model were analyzed. The results show that winner prediction is possible in League of Legends, also, LightGBM (0.97), Logistic Regression (0.96), SVM and GBC (Gradient Boosting Classifier) (0.95) are outperformed with a high accuracy ratio. This paper will contribute to the classification research on topic of gaming with machine learning.

Keywords: Gaming, classification, machine learning, league of legends, MOBA

ÖZ

Eğlence endüstrisi, telekomünikasyon hizmetleri, televizyon, müzik, video oyunları ve canlı konserler gibi işlerle uğraşan alışılmadık derecede geniş bir şirket yelpazesini içerir. Oyun, son on yılda eğlence sektöründe gelir artışı ivmesi elde etmiştir. Bu ivme oyun sektörünü eğlence endüstrisinin en popular alanlarından biri haline getirmiştir. Profesyonel oyuncularla resmi ligler kurulmuş ve e-spor kavramı yaygın hale gelmeye başlamıştır. Çevrimiçi oyun türlerinden olan MMO (Massive Multiplayer Online) oyunların bir türevi olarak karşımıza çıkan MOBA (Multiplayer Online Battle Arena) internet üzerinde genellikle 5 kişilik 2 takımla bir harita üzerinde belirli yapıları domine ederek rakibin üssünü yok etme hedefiyle oynanan oyunlara verilen isimdir. LoL (League of Legends) bir MOBA oyunudur. Çevrimiçi video oyunlarında kazananların tahmini, makine öğrenmesi tabanlı tahmin modelleri için önemli bir uygulama haline gelmiştir. Araştırmanın hedefi LoL oyuncu metrikleriyle maç kazanma tahminin makine öğrenmesi yöntemleriyle sınıflandırma tahminidir. Önemli performans ölçütleri ve bunların her bir oyun modeli üzerindeki etkisi de analiz edildi. Sonuçlar, League of Legends oyununda kazanan tahmininin mümkün olduğunu göstermektedir. LightGBM (0.97), Lojistik regresyon (0,96), SVM ve GBC (0.95) başarım oranı ile öne çıkan algoritmalardır. Çalışmanın oyun alanında makine öğrenmesiyle sınıflandırma çalışmalarına katkı sağlayacağı düşünülmektedir.

Anahtar Kelimeler: Makine öğrenmesi, sınıflandırma, oyun, league of legends, MOBA

[1](Lect.) Sakarya Applied Sciences University, Department of Multidimensional Modeling and Animation, Sakarya, Turkiye

ORCID: K.A. 0000-0002-0930-8955

Corresponding author:
Kaan ARIK
Sakarya Applied Sciences University, Department of Multidimensional Modeling and Animation, Sakarya, Turkiye
E-mail address: kaanarik@subu.edu.tr

Submitted: 26.09.2022
Revision Requested: 23.03.2023
Last Revision Received: 02.05.2023
Accepted: 09.05.2023
Published Online: 15.06.2023

Citation: Arik, K. (2023). Machine learning models on moba gaming: league of legends winner prediction. *Acta Infologica, 7*(1), 139-151. https://doi.org/10.26650/acin.1180583

1. INTRODUCTION

In recent years, video games have been one of the popular fields of the entertainment industry for adults and children. Games started to be played in the 1960s and prime arcade games became widespread in the mid-1970s by introducing Pac-Man and Space Invaders by Atari (Donovan, 2010). During the 1980s and 1990s, Nintendo released handheld game consoles and gained popularity among a colossal gamer audience (Kent, 2001). With the development of specialized hardware, in the 2000s, Xbox and Sony launched their first-generation consoles that are still widely played today (Schreier, 2017). Since the introduction of smartphones and tablets between 2000 and 2010, games are becoming a product that can be reached instantaneously. And now, thanks to hardware performance and computer abilities, people can play games with realistic graphics anytime. People often use smart devices while online shopping, having fun, communicating with others, and spending more time with those devices than older generations. In 2001, Marc Prensky introduced the "Digital Natives" term and defined them as "ones at a higher level in terms of technical competence compared to their predecessors". Undoubtedly, most of the time these devices are used is for games and their community.

The entertainment industry includes companies engaged in telecommunications services, television, music streaming, video games, and live events (Nathan Reiff, 2022). Gaming has gained momentum in revenue growth in the entertainment industry over the past decade. Games are interactive and have succeeded thanks to ensuring more satisfaction than visual and written content(Robson & Meskin, 2016). For instance, no matter how much people like a movie in their life, they never watch that movie all the time, because movies open up an opportunity for linear expression. However, if one likes the game, they want to play it multiple times. Games are non-linear interactive structures that allow players to make in-game decisions, thus keeping their interest. Eventually, it's worth stating that games take the experience and storytelling in movies(Beverly Peders, 2018).

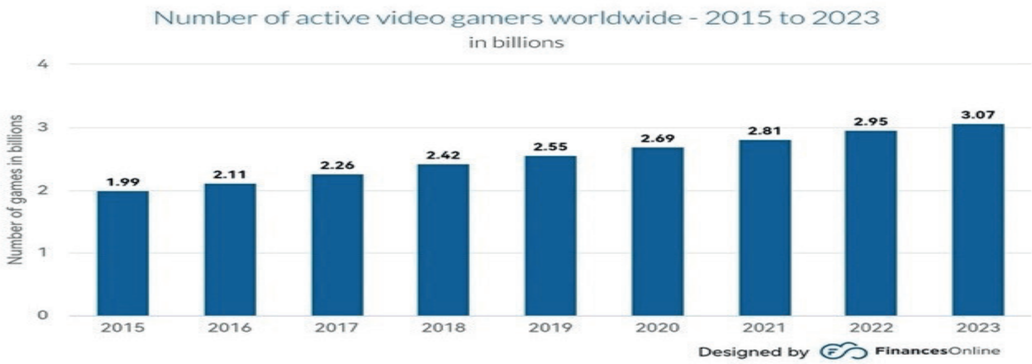


Figure 1. Number Of Active Players Between 2015-2023 in Billions

Recently, the interest in online games has increased with the ease of Internet access worldwide. At the beginning of 2022, there were 4.95 billion internet users worldwide, making up 62.5 percent of the world's population. According to an estimate for yearly growth of 5.6%, the number of video game players would rise from 2.69 billion in 2020 to 3.07 billion in 2023. In 2020, the global gaming industry brought in $159.3 billion, with the Asia Pacific region accounting for nearly half of that total. The number of active players and the global increase between 2015-2023 are given in Figure 1(Simon Kemp, 2022).

In addition, games are popular and yield high revenues; it's also paved the way for forming its own culture(Nestor Gilbert, 2022). Magazines, websites and internet content for digital games have been launched, and this situation has gone to the next level, becoming a structure where a virtual world exists over the Internet. Moreover, League of Legends, DOTA2, Counter-Strike Global Offense, Valorant and others are included in the MOBA games that are referred to as online games today.

PC games are one of the platforms in which video games are best experience. This platform stands out with the advantage of supporting high-quality resolution and offering more gameplay. Mostly, PC games outperform when compared to consoles, because PC games provide more customization through various accessories and options that can enhance the experience. For example, while you can use your keyboard and mouse to play games on a PC, you can also use it by integrating a game controller. It also offers a broader range of games, thanks to game-selling media such as Steam and Epic Store.

2. MOBAs AND MACHINE LEARNING

This section presents the definition and features of MOBAs and research on machine learning applications in the game industry.

2.1. MOBA (Multiplayer Online Battle Arena)

MMO (Massive Multiplayer Online) is a game many people can play simultaneously (Donovan, 2014). The reason for including the definition of MMO games is MOBA (Multiplayer Online Battle Arena) games partially use MMORPG mechanics (Galehantomo, 2015). An online multiplayer battle arena, or MOBA, is a real-time strategy video game genre. League of Legends, DOTA II, Heroes of the Storm, etc. are included in this category. Games are classified differently based on gameplay and shooting angles. The MOBA (Multiplayer Online Battle Arena) genre is included in the list of online games. MOBAs, an exciting subgenre of MMO games, have become an essential part of video games and a pop culture phenomenon in recent years. MOBAs typically comprise five players on each team, each controlling a single character. In some other MOBAs, player number changes regarding game mechanics and map. Unlike MMORPG (Massively Multiplayer Online Role-Playing Game) games, a MOBA game does not have a unified structure or many simultaneous players on a map; much of the strategy is developed around cooperative team play and individual characters.

MOBAs tend to have a real-time strategy element and revolve around a simple goal: to compete against opponents in a team and defeat them in combat. Players make instant progress and must have a constant Internet connection; otherwise, they can be accepted as AFK (Away from Keyboard) players. Game sessions usually are between 30 minutes and 1 hour, although they can be longer or shorter based on player progress. In MOBAs, champions with a specific skill kit are usually offered to players by game designers. Also, players focus on winning the game in different lanes by choosing these champions. LoL (League of Legends) is played on the 'Summoner's Rift' map. In addition, there are different forms (drakes, Baron Nashor, jungle camps, towers etc.) that can kill players, which is important for customizing their champion and purchasing different items. Often the task of these forms is to provide teams with extra skill points and buffs. Every player has league points in their profiles, and they were categorized into different leagues based on win/lose rates. Players are often located in different lanes, where they fight against rivals. It mainly consists of four corridors: top-lane, mid-lane, bot-lane, and jungle. Only bot-lane players can play as a duo. The primary source of money for players while playing is to have in-game resources by hitting the last hit to minions that are produced from the bases. Matches are divided into two groups, ranked and normal. While players get points only in a ranked match, they only gain experience in a normal match. LoL had 117 million monthly players as of 2022. Also, it has played a crucial role in the popularity of the new sports branch, called e-sports, in recent years (League of Legends Live Player Count and Statistics, 2022). League of Legends has official leagues with professional teams in countries such as America, Europe, China, South Korea, and Türkiye.

2.2. Machine Learning and Gaming

Customer loyalty is the persistent emotional bond between a firm and a customer. This relationship is evident when a consumer consistently purchases from you rather than one of your rivals (Oracle, 2020). When a customer has a good experience with you, loyalty develops naturally and aids in developing trust. In recent years, the gaming industry has placed more emphasis on the idea of loyalty, because businesses want users to play their games frequently and to pay them money. Companies examine the participants and apply the right policies in this procedure using statistical and artificial intelligence techniques (Alpaydin, 2020). The video game industry benefits from artificial intelligence in numerous ways, including player league classification, customer churn analysis, winner prediction, and the ability for players of the same level to compete in a match.

Machine learning is a branch of artificial intelligence that uses statistical models and algorithms to manage mechanical motion (Nilsson, 2010). Contrast this with traditional AI methods like search trees and expert systems. In recent years, machine learning has been one of the most popular areas in academia and practice. It has evolved into one of the frameworks researchers use to assess and categorize participants, much like games. Video games have utilized various artificial intelligence techniques, from PCG (procedural content generation) to NPC (Non-player Character) control. It's also actively used to make enemy characters playable against rivals. Using deep learning agents to compete against skilled human players in sophisticated strategy games is the most well-known example of how machine learning is used in video games. Machine learning has found significant use in games like Atari/ALE, Doom, StarCraft, and Minecraft. Machine learning also impacts games like Chess and Go that weren't designed to be played on screens (Justesen et al., 2019).

Machine learning algorithms perform quite well in game AI problems. Ensemble learning and tree-based models have been successful on various datasets. These are primarily solutions to classification problems. Issues such as player defection, winner/loser prediction, and CLV estimation are well mentioned in the literature. This section provides solutions to problems such as business analysis and classification. One of the most common problems in game AI is data access. High-budget game studios buy low-budget game studios and consolidate these companies into their ecosystems into a single hub. For this reason, we are very cautious about sharing data with researchers. However, in-game metrics data via API may be used for research within the guidelines set by the company.

3. DATASET AND METHOD

This section mentions the data used during research and the theoretical structures of implemented algorithms. There is information about the scraping dates of the data, data information, and how data was scraped. In addition, training and testing division and the pre-processing steps are discussed in this section.

3.1. Dataset

The data set consists of approximately 138,000 players and between 10.01.2022 – 12.09.2022 with 96 columns. The data was scraped via Riot API, keeping with Riot Games data privacy and framework policy(Tim Sevenhuysen, n.d.). The IDs of the players have been anonymized, and no-account information has been included in the research. Each game has ten players that target column winner (1)/loser (0). Attributes are divided into two groups player and in-game metrics. The attributes and explanations in the data are given in Appendix B.

3.2. Method

The research aims to predict the winner/loser classification using machine learning techniques based on player and in-game metrics. Data mining steps were chosen throughout the research based on the CRISP-DM model. CRISP-DM (Cross-Industry Standard Process for Data Mining) is a standardized methodology used to describe how business problems are solved with data-based solutions and to increase the efficiency of business applications. Due to CRISP-DM being a high-level methodology, the steps outlined in the model can be implemented in many different ways, sequences, and technologies to meet business needs. It consists of 6 steps: business understanding, data understanding, data preparation, modelling, evaluation, and deployment (IBM, 2021).

Table 1
Data partition for train and testing

Split Ratio	Train Set	Test Set	Total
%70 - %30	92979	39849	132828

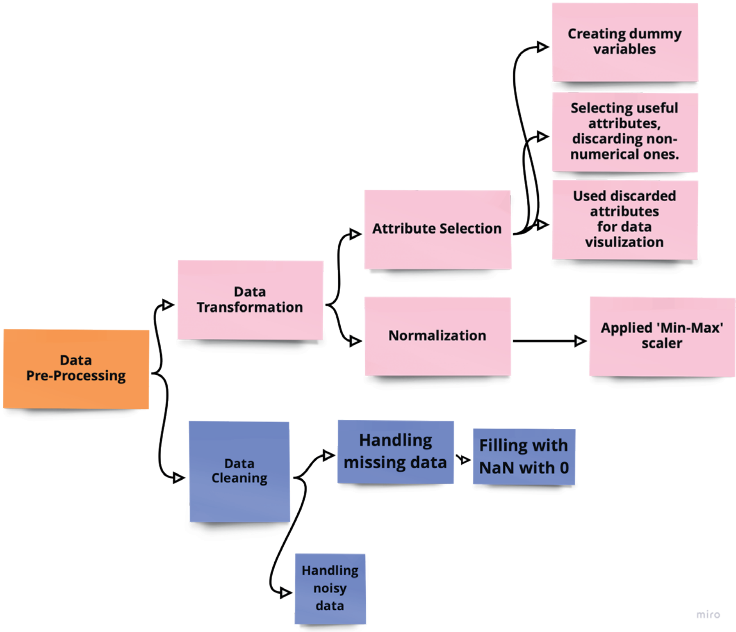


Figure 2. Diagram Of Data Preprocessing

Here are the data pre-processing steps:

· Non-numeric (date, group, id, status etc.) attributes were eliminated from the data as feature engineering.

· Attributes with correlations above 0.90 were excluded because they cause over-fitting in tree-based algorithms.

· In total, data has been reduced from 123 to 96 features.

· NaN values are marked as 0, not because the player does not have the opportunity to perform the tasks but because they didn't choose to do it, described as sparse data.

· When an AFK (Away from Keyboard) player is in-game, other players can start a new match by voting in the first 10 minutes without losing league points. In this respect, rows with a game session of less than 600 seconds were removed from the data.

· Data was normalized between a value of 0 and 1.

3.3. Machine Learning Algorithms

The ability to utilize the best method to turn a dataset into a model is known as machine learning. Machine learning is a collection of techniques for building models out of data. The most effective method (supervised, unsupervised, reinforcement learning, etc.) will vary depending on the data, the resources at hand, the nature of the data, and the predicted outcome of all processes (Han & Kamber, 2012). Ordinary programming algorithms simply tell the computer what to do. Machine learning algorithms contain simpler methods than nonlinear regression, giving more accurate results and outputs compatible with mathematical functions. It is often possible to talk about two methods: regression and classification. While making inferences on quantitative data (income level, plant height, etc.) with regression, classification also includes non-numerical variables (credit approval status, gender, number of rooms, etc.). Among algorithms implemented throughout the research, the best-performing ones were mentioned. LightGBM, Logistic Regression, Support Vector Machines, and Gradient Boosted Classifiers are algorithms with the best accuracy ratio for prediction. Performance metrics were explained in the findings section and discussed in the conclusion. Also, the confusion matrix and ROC curves for implemented algorithms have been added to Appendix A.

3.3.1. LightGBM (Light Gradient Boosting Machine)

As a component of the Microsoft DMTK (Distributed Machine Learning Toolkit) project, the boosting method LightGBM was created in 2017. Its benefits over other boosting methods include fast processing, handling large amounts of data, using fewer resources (RAM), a high prediction rate, parallel learning, and GPU learning support. A histogram-based method is used in LightGBM(Microsoft, 2022). Making the variables with continuous values lowers the computing cost. The computation and, thus, the number of branching directly relate to the training duration of the decision trees. This approach results in a decrease in both resource utilization and training time. Learning decision trees can be done in one of three ways: level-wise, depth-wise, or leaf-wise. Two strategies can be used when learning decision trees: level-by-level, depth-by-depth, or leaf-by-leaf. A level-oriented strategy keeps the tree in balance as it grows.

A hand-focused strategy, however, will continue to split hands and reduce losses. This feature sets LightGBM apart from other boosting algorithms. A leaf-oriented strategy results in a lower model error rate and faster learning. However, the leaf-oriented growth strategy makes the model prone to overfitting when the number of data is small. Therefore, this algorithm is more suitable for use with big data. Additionally, parameters such as tree depth and number of leaves can be optimized to prevent overfitting(Ke et al., n.d.).

3.3.2. Logistic Regression

Logistic regression is like a regression problem in which the dependent variable is categorical, and it is frequently applied to linear classification issues. A classification occurs here even though it is labelled regression(Wright, 1995).

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}} \qquad \text{Formula (1)}$$

An analysis of a dataset with one or more independent variables predicting an outcome is done statistically using logistic regression. The predicted class is measured with a binary variable (only two possible outcomes)(Kleinbaum & Klein, 2010). Although logistic regression has the word regression in its name, it is a classification algorithm. A few iterations of artificial neural networks and a logit function are applied, as seen in Formula 1(Alpaydin, 2020).

3.3.3. Support Vector Machines

Support Vector Machine is a supervised learning method generally used in classification problems. It draws a line to separate points placed on a plane, aiming to have this line at the maximum distance for the points of both classes. It suits complex but small to medium datasets (William Noble, 2006).

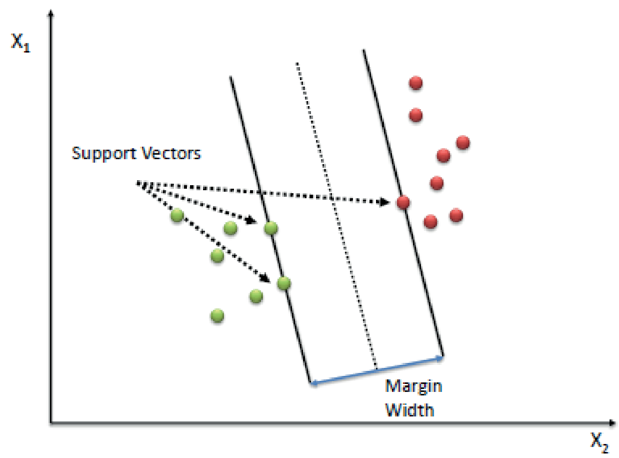


Figure 3. SVM Algorithm with Margins

The main purpose of classification problems is to decide in which class the future data will take place. A line separating the two classes is drawn to make this classification, and the region between ±1 of this line is called the margin. The wider the margin, the better the separation of two or more classes. At the same time, apart from the linear model, it performs data separation in different kernels(Géron, 2019).

3.3.4. Gradient Boosting Classifier

Gradient boosting classifiers update the classifiers and weighted inputs using the AdaBoosting algorithm in conjunction with weighted minimization. Reducing the loss, or the discrepancy between the actual class value of the training example and the predicted class value is the goal of Gradient Boosting classifiers. Although comprehension of the method for decreasing the classifier's loss is unnecessary, it works similarly to gradient descent in neural networks(Lemarechal, 2012).

In the case of Gradient Boosting Machines, the weights of the preceding learners are frozen or cemented in place, remaining intact when the new layers are added, and this is done each time a new weak learner is added to the model. The methods employed in AdaBoosting, where the values are modified when additional learners are added, differ from this. Gradient boosting algorithms' strength stems from the fact that they can be applied to situations involving more than just binary classification; they can also be used to solve regression and multi-class classification problems(Mason et al., 1999).

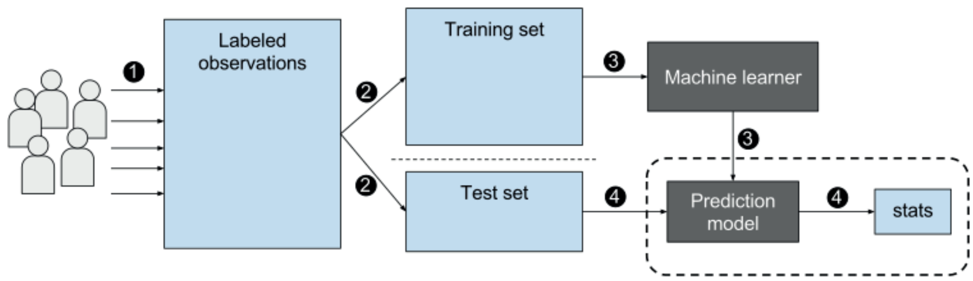


Figure 4. GBC (Gradient Boosting Classifier) Algorithm Diagram

A loss function is necessary for the Gradient Boosting Classifier to operate. Gradient boosting classifiers can handle a variety of standardized loss functions in addition to bespoke loss functions. However, the loss function must be differentiable. Regression techniques employ squared errors, whereas classification algorithms often use logarithmic loss. Any differentiable loss function may be used in gradient boosting systems instead of having to be derived specifically for each additional boosting procedure. A gradient-boosting model's additive component results from adding new trees over time without changing the values of the model's already-existing trees. The error between the specified parameters is minimized using a method akin to gradient descent.

3.4. Performance Metrics and Libraries

Accuracy is the percentage of samples classified as correct. Recall is a metric that shows how many of the transactions of positively predict. Precision shows how many of the values predicted as positive are positive. The F1 score measures a test's accuracy—the harmonic mean of precision and sensitivity (Géron, 2019)-(Stehman, 1997).

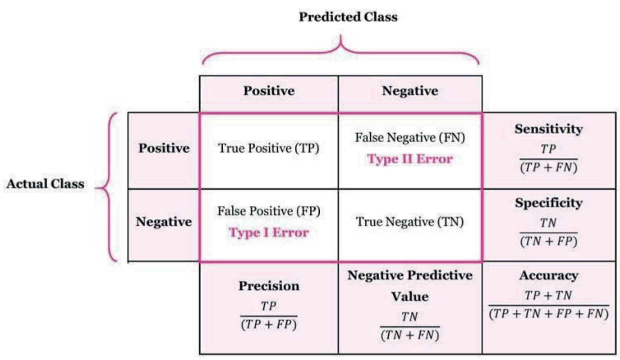


Figure 5. Performance Evaluation Metrics in the Classification Problem (Shajihan, 2020)

The research used ‘Pandas and NumPy’ for data processing and the ‘Scikit-Learn’ library in Python to applying machine learning algorithms. The applications were carried out on Google Collab, and ‘pyCaret’ library was used to evaluate the performance of the algorithms.

4. FINDINGS

In this section, classification algorithms are implemented on the dataset. Also, a parameter tuning was made for all the methods and parameters that could obtain the best results. Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k, which refers to the number of groups a given data sample will be split into. For ML algorithms, ‘k-fold cross-validation’ method has been deployed with k=10. LightGBM, Logistic Regression, Support Vector Machines, Gradient Boosting Classifier, Random Forest, Linear Discriminant Analysis, Ridge Classifier, Extra Tree, AdaBoost, and Decision Tree performed over 90%, as seen in Table 2. The best classifier in terms of accuracy is LightGBM. Therefore, when compared to runtime, it is evident that LightGBM outperforms results based on performance. It performs not only in terms of accuracy ratio, but also recall, precision and F1 score. LightGBM algorithm is a histogram-based algorithm that selects the best split in the sorted histogram. It speeds up training and reduces memory usage focusing on the accuracy of results. It provides faster and higher efficiency. It performs better with less memory usage and can handle large data sets. As a result, ensemble methods perform with high accuracy rates in classification in gaming (Arık et al., 2022).

Table 2

Performance metrics of algorithms in classification

Model	Accuracy	AUC	Recall	Prec.	F1	TT (Sec)
LGBM	0.968	0.996	0.971	0.966	0.969	4.8870
LR	0.955	0.988	0.956	0.954	0.955	8.6470
SVM	0.950	0.000	0.954	0.946	0.950	0.6480
GBC	0.949	0.990	0.953	0.946	0.949	89.9920
RF	0.9481	0.990	0.955	0.941	0.948	32.7530
LDA	0.9449	0.982	0.948	0.941	0.945	3.1410
RC	0.9416	0.000	0.946	0.937	0.941	0.3960
ETC	0.9393	0.987	0.949	0.930	0.939	19.2340
ADAB	0.9353	0.983	0.936	0.934	0.935	17.3250
DT	0.9263	0.926	0.926	0.925	0.926	4.6720
KNN	0.8879	0.948	0.898	0.880	0.889	59.9160
QDA	0.7951	0.952	0.625	0.947	0.753	1.6390
NB	0.6481	0.857	0.341	0.881	0.492	0.4530

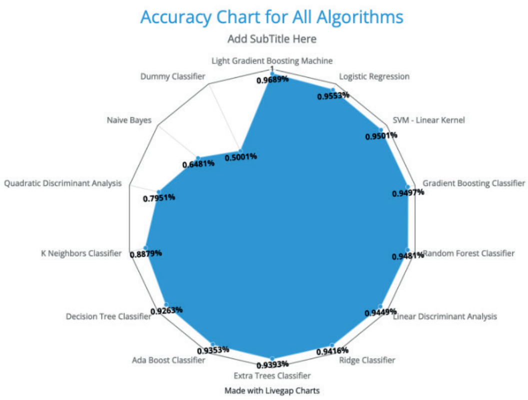


Figure 6. Radar Chart of Accuracy Ratio for Models

Accuracy ratio, precision, recall, and F1 scores of LR, SVM, and GBC algorithms also perform well compared to LightGBM. All implemented algorithms are successful except Quadratic Discriminant Analysis and Naive Bayes. In addition, compared to the running time of other algorithms, KNN, GBC, and RF are slower than others.

Table 3
LightGBM tune parameters and description

Parameter	Description	Value
boosting_type	Boosting type	gbdt
importance_type	Type of feature importance	split
learning_rate	Boosting learning rate	0.1
max_depth	maximum distance between the root node and leaf node	-1
n_estimators	Number of estimators	100
num_leaves	Control parameter for the complexity of the tree model	31
subsample	Sub-samples ratio	1.0

The majority of high accuracy ratios are ensemble learning models. As a result, ensemble learning algorithms outperform results in classification problems in gaming. Specifying the learning rate as small and the number of iterations as high in the LightGBM algorithm and having large training data are essential parameters for developing successful models. For this reason, the use of 'learning_rate' = 0.1, 'n_estimators' = 100 and "gbdt" as the enhancement type in our model has been effective in the performance success of the model.

5. CONCLUSION AND DISCUSSION

Table 4 lists research published on MOBA games over the past few years. It includes the year, author, game genre, and best accuracy. Since 2017, classification and clustering algorithms, mainly classification, have been implemented in MOBA games every year. According to the literature, Random Forest and Logistic Regression algorithms are high-performance in gaming. Research has been conducted with LoL (League of Legends) data with high accuracy scores. Among the reasons for this, as mentioned in the previous parts of the article, it is thought that there is a problem accessing data. In [30][33] Naive Bayes and Logistic Regression achieved a good level of success with 0.77 in previous studies but not as well as others. In addition to classification, clustering models were implemented on such gaming datasets [31].

However, among those implemented in Table 4, it has a lower accuracy ratio than others. In another study, statistics-based models do not perform much accuracy in AI applications in gaming. Moreover, ML algorithms that implemented ensemble learning have become life-saving models in data competitions and researchers. Here again, the best results are performed through tree-based ensemble learning algorithms.

Table 4
Related works

Year	Authors	Genre	Best Accuracy
2017	Almeida and et al. (Almeida et al., 2017)	MOBA	Naive Bayes (0.77)
2018	Mora-Cantallops and Sicilia (Mora-Cantallops & Sicilia, 2018)	MOBA	k-means
2019	Ani and et al. (Ani et al., 2019)	MOBA	Random Forest (0.99)
2019	Porokhnenko, etc.(Porokhnenko et al., 2019)	MOBA	Logistic Regression (0.70)
2021	Costa and et al.(Costa et al., 2021)	MOBA	TSSTN
2021	Yang and et al. (Yang et al., 2021)	MOBA	Logistic Regression and Random Forest (0.97)

Machine learning algorithms are generally implemented for classification problems in gaming, and deep learning is not needed much. Because deep learning works better on datasets with image, audio, video, and text content, it is also difficult to predict how well it will work, as it's a black-box model. For future works, there will be models with much larger data sets to work with sets with a large variety of attributes and missing data, which will increase the applicability of such models on a company basis.

ROC and Precision-Recall curves, features importance plots, number of predicted classes, confusion matrix, and algorithms abbreviation and title are attached in Appendix A. AUC plots explain how well the model can predict classes, as seen as ROC curves for the LightGBM classifier in Appendix A. The higher the AUC, the better the model is at predicting false as false and true as true. Accuracy increases as the curve gets closer to the top-left (Krzanowski & Hand, 2009). For example, the higher the AUC in our data dataset of winner and loser, the better the model performs in predicting between winners and losers. When the ROC plot of LightGBM algorithm is analyzed, the curve has an almost 100% accuracy ratio. In other words, the model performed very well in both positive and negative classes for problems. Looking at the confusion matrix, samples are labelled as a winner in data but classified as a loser in prediction (n=591) and labelled as a loser but classified as a winner (n=704). It is almost impossible to achieve 100% accuracy rate in gaming datasets. Because LoL is a game played with 5v5 and teammates' well performance can make the same team's bad players win the game. For this reason, the misclassification of 1295 samples is acceptable. When examining players, it can be interpreted both groups were equally classified correctly.

The ROC curve is a very crucial performance measure for classification problems. ROC is a probability curve and area under it, and it represents the degree or measure of separability. The ROC curve has FPR (False Positive Rate) on the X axis and TPR (True Positive Rate) on the Y axis. The higher the level under the curve, the higher the class discrimination performance. In research, successful results were achieved on negatives by examining ROC curves. As seen in the Features Importance plot, 'earned_gpm' (Earned Game Per Minutes) is the most important attribute. Then the list goes on to include 'earned_gold_shared', 'gold_spent', 'assist' and 'death' attributes. Checking the chart and interpreting that gold earned in-game is important for a player who has never played LoL. The game's main aim is not to kill rivals throughout the session but to develop quickly individually by killing the rivals with the resources in the session. In this respect, the importance of predicted features makes this research meaningful.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The author has no conflict of interest to declare.
Grant Support: The author declared that this study has received no financial support.

REFERENCES

Almeida, C. E. M., Correia, R. C. M., Eler, D. M., Olivete-Jr, C., Garci, R. E., Scabora, L. C., & Spadon, G. (2017). Prediction of winners in MOBA games. *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. https://doi.org/10.23919/CISTI.2017.7975774

Alpaydin, E. (2020). *Introduction to machine learning* (Fourth edition). The MIT Press.

Ani, R., Harikumar, V., Devan, A. K., & Deepa, O. S. (2019). Victory prediction in League of Legends using Feature Selection and Ensemble methods. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 74–77. https://doi.org/10.1109/ICCS45141.2019.9065758

Arık, K., Gezer, M., & Tayali, S. T. (2022). Bibliometric Analysis of Scientific Studies Published on Game Customer Churn Analysis Between 2008 and 2022. *Journal of Politics*, *5*(1), 21.

Beverly Peders. (2018). *Film vs. Video Games From A Screenwriter's Perspective*. https://www.wescreenplay.com/blog/film-vs-video-games-from-a-screenwriters-perspective/

Costa, L. M., Mantovani, R. G., Monteiro Souza, F. C., & Xexéo, G. (2021). Feature Analysis to League of Legends Victory Prediction on the Picks and Bans Phase. *2021 IEEE Conference on Games (CoG)*, 01–05. https://doi.org/10.1109/CoG52621.2021.9619019

Donovan, T. (2010). *Replay: The history of video games*. Yellow Ant.

Galehantomo, G. (2015). Platform Comparison Between Games Console, Mobile Games And PC Games. *SISFORMA*, *2*, 23. https://doi.org/10.24167/sisforma.v2i1.407

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (Second edition). O'Reilly Media, Inc.

Han, J., & Kamber, M. (2012). *Data mining: Concepts and techniques* (3rd ed). Elsevier.

IBM. (2021, August 17). *IBM Documentation*. https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview

Justesen, N., Bontrager, P., Togelius, J., & Risi, S. (2019). *Deep Learning for Video Game Playing* (arXiv:1708.07902). arXiv. https://doi.org/10.48550/arXiv.1708.07902

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (n.d.). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*. 9.

Kent, S. L. (2001). *The ultimate history of video games: From Pong to Pokémon and beyond: the story behind the craze that touched our lives and changed the world* (1st ed). Prima Pub.

Kleinbaum, D. G., & Klein, M. (2010). Introduction to Logistic Regression. In *Logistic Regression: A Self-Learning Text* (pp. 1–39). Springer. https://doi.org/10.1007/978-1-4419-1742-3_1

Krzanowski, W. J., & Hand, D. J. (2009). *ROC curves for continuous data*. Chapman & Hall/CRC.

League of Legends Live Player Count and Statistics. (2022, July 11). https://activeplayer.io/league-of-legends/

Lemarechal, C. (2012). Cauchy and the Gradient Method. *Documenta Mathematica*, 4.

Mason, L., Baxter, J., Bartlett, P. L., & Frean, M. R. (1999). *Boosting Algorithms as Gradient Descent*. *3*, 7.

Michael Donovan. (2014, February 3). *The MMO vs. The MOBA*. https://www.gamedeveloper.com/disciplines/the-mmo-vs-the-moba

Microsoft. (2022). *LightGBM Documentation*. https://lightgbm.readthedocs.io/en/v3.3.2/

Mora-Cantallops, M., & Sicilia, M.-Á. (2018). Player-centric networks in League of Legends. *Social Networks*, *55*, 149–159. https://doi.org/10.1016/j.socnet.2018.06.002

Nathan Reiff. (2022). *10 Biggest Entertainment Companies*. https://www.investopedia.com/articles/investing/020316/worlds-top-10-entertainment-companies-cmcsa-cbs.asp

Nestor Gilbert. (2022). *Number of Gamers Worldwide 2022/2023: Demographics, Statistics, and Predictions*. https://financesonline.com/number-of-gamers-worldwide/

Nilsson, N. J. (2010). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge University Press.

Oracle. (2020). *What is Customer Loyalty?* https://www.oracle.com/tr/cx/marketing/customer-loyalty/what-is-customer-loyalty/

Porokhnenko, I., Polezhaev, P., & Shukhman, A. (2019). Machine Learning Approaches to Choose Heroes in Dota 2. *2019 24th Conference of Open Innovations Association (FRUCT)*, 345–350. https://doi.org/10.23919/FRUCT.2019.8711985

Robson, J., & Meskin, A. (2016). Video Games as Self-Involving Interactive Fictions. *The Journal of Aesthetics and Art Criticism*, *74*(2), 165–177.

Schreier, J. (2017). *Blood, sweat, and pixels: The triumphant, turbulent stories behind how video games are made* (First edition). Harper Paperbacks.

Shajihan, N. (2020). *Classification of stages of Diabetic Retinopathy using Deep Learning*. https://doi.org/10.13140/RG.2.2.10503.62883

Simon Kemp. (2022). *Digital 2022: Global Overview Report* [Analysis]. https://datareportal.com/reports/digital-2022-global-overview-report

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*(1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7

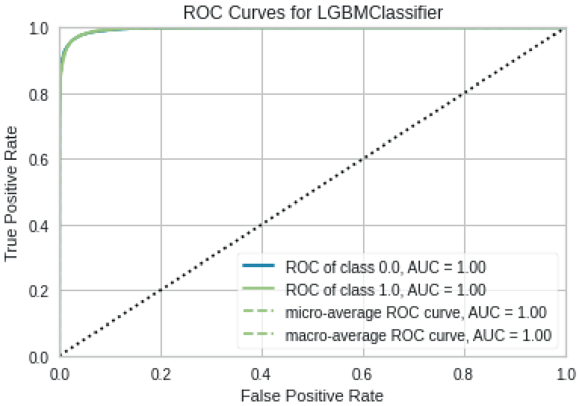
Tim Sevenhuysen. (n.d.). *Oracle's Elixir—LoL Esports Stats*. Retrieved September 15, 2022, from https://oracleselixir.com/tools/downloads

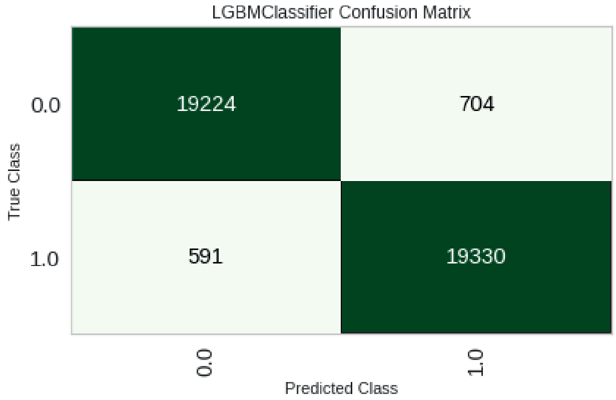
William Noble. (2006). *What is a support vector machine? | Nature Biotechnology*. https://www.nature.com/articles/nbt1206-1565

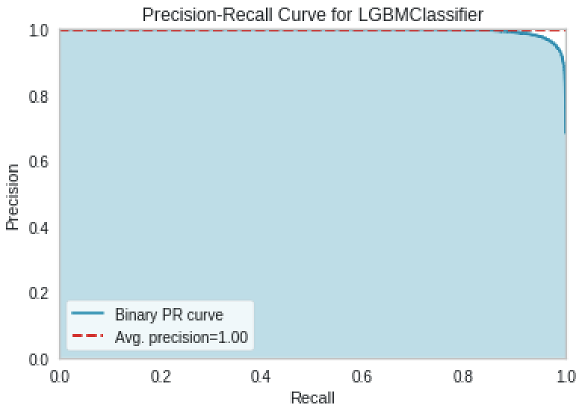
Wright, R. E. (1995). Logistic regression. In *Reading and understanding multivariate statistics* (pp. 217–244). American Psychological Association.

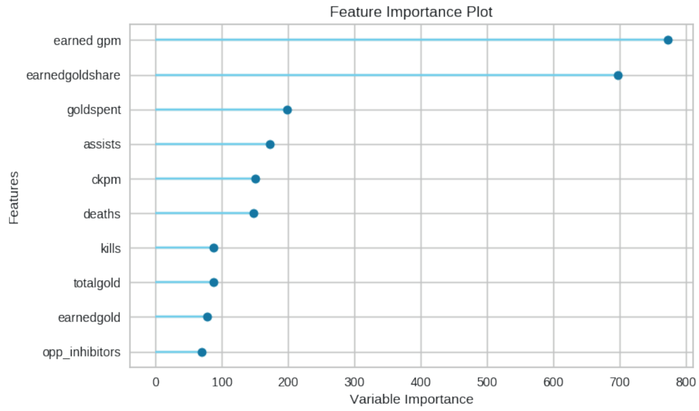
Yang, Z., Pan, Z., Wang, Y., Cai, D., Liu, X., Shi, S., & Huang, S.-L. (2021). *Interpretable Real-Time Win Prediction for Honor of Kings, a Popular Mobile MOBA Esport* (arXiv:2008.06313). arXiv. https://doi.org/10.48550/arXiv.2008.06313

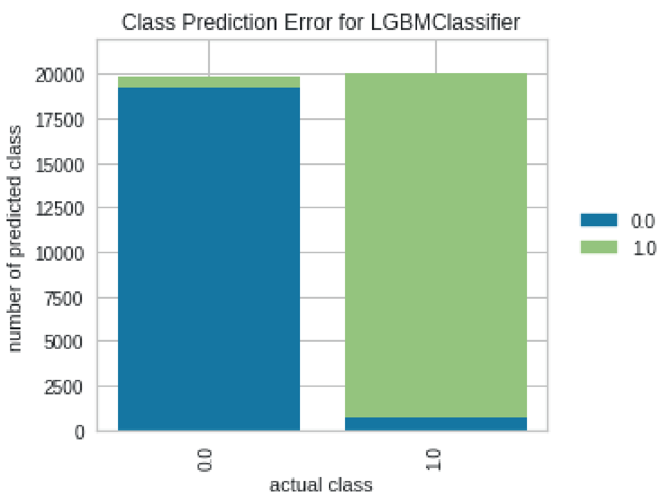
APPENDIX A











Abbreviation	Full Algorithm Title
LGBM	LightGBM
LR	Logistic Regression
SVM	Support Vector Machines
GBC	Gradient Boosting Classifier
RF	Random Forest
LDA	Linear Discriminant Analysis
RC	Ridge Classifier
ETC	Extra Trees Classifier
ADAB	Adaptive Boosting AdaBoost
DT	Decision Tree
KNN	K-Nearest Neighbor
QDA	Quadratic Discriminant Analysis
NB	Naive Bayes

APPENDIX B

Term	Description
A	Total assists
AGT	Average game time/duration, in minutes
APG	Assists per game
B%	Percentage of games in which the champion was banned (not tied to a specific role)
BLND%	Blind-pick rate: percentage of games in which this player/champion was picked before their lane opponent (not always available)
BN%	Baron control rate
CCPM	Crowd control dealt to champions per minute
Champion	Champion name
CKPM	Average combined kills per minute (team kills + opponent kills)
CS%P15	Average share of team's total CS post-15-minutes
CSD10	Average creep score difference at 10 minutes
CSD15	Average creep score difference at 15 minutes
CSD20	Average creep score difference at 20 minutes
CSPM	Average monsters + minions killed per minute
CTR%	Counter-pick rate: percentage of games in which this player/champion was picked after their lane opponent (not always available)
CWPM	Control wards purchased per minute
D	Total deaths
D%P15	Average share of team's damage to champions post-15-minutes
DMG%	Damage Share: average share of team’s total damage to champions
DMG%P15	Average share of team's damage to champions post-15-minutes
DPG	Deaths per game
DPM	Average damage to champions per minute
DRG%	Dragon control rate: percent of all Dragons killed that were taken by the team, reflecting only elemental drakes if ELD% is present
DTH%	Average share of team’s deaths
EGPM	Average earned gold per minute (excludes starting gold and inherent gold generation)
EGR	Early-Game Rating
ELD%	Elder dragon control rate
Event	Event name
F3T%	First-to-three-towers rate (percentage of games in which team was the first to 3 tower kills
FB%	First Blood rate -- for players/champions, percent of games earning a First Blood participation (kill or assist)
FBN%	First Baron rate
FBV%	First Blood Victim rate -- percent of games player/champion was killed for First Blood
FD%	First dragon rate
FT%	First tower rate
GD10	Average gold difference at 10 minutes
GD15	Average gold difference at 15 minutes
GD20	Average gold difference at 20 minutes
GOLD%	Gold Share: average share of team’s total gold earned (excludes starting gold and inherent gold generation)
GP	Games Played
GPM	Average gold per minute
GPR	Gold percent rating (average amount of game’s total gold held, relative to 50%)
GSPD	Average gold spent percentage difference
GXD10	Average gold+experience difference at 10 minutes
GXD15	Average gold+experience difference at 15 minutes
GXD20	Average gold+experience difference at 20 minutes
HLD%	Rift Herald control rate
IWC%	Average percentage of opponent’s invisible wards cleared
JNG%	Jungle Control: average share of game’s total jungle CS
K	Total kills
KD	Kill-to-Death Ratio
KDA	Total Kill/Death/Assist ratio
KP	Kill participation: percentage of team's kills in which player earned a Kill or Assist
KPG	Kills per game
KS%	Kill share: player's percentage of their team's total kills
L	Losses
LNE%	Lane Control: average share of game’s total lane CS
Losses	Total Losses
LP	Ladder Points
MLR	Mid/Late Rating
OE Rating	Oracle’s Elixir Performance Rating
OE Rtg	Oracle’s Elixir Performance Rating
P%	Percentage of games champion was picked in this role.
P+B%	Percentage of games in which the champion was either banned or picked in any role
Player	Player's in-game name
Pos	Position
PPG	Turret plates destroyed per game
Rank	Official Leaderboard Rank
STL	Neutral objectives stolen
STLPG	Neutral objectives stolen per game
StPG	Neutral objectives stolen per game
Team	Team name
VSPM	Vision score per minute
VWC%	Average percentage of opponent’s visible wards cleared
W	Wins
W%	Win percentage
WC%	Average percentage of opponent wards cleared
WCPM	Average wards cleared per minute
Wins	Total Wins
WPM	Average wards placed per minute
XPD10	Average experience difference at 10 minutes
XPD15	Average experience difference at 15 minutes
XPD20	Average experience difference at 20 minutes