A COMPARATIVE ANALYSIS ON PREDICTING FOOTBALL MATCHES USING MACHINE LEARNING. (A CASE STUDY OF SPANISH LEAGUE)

BY

UTOMI AMAKA MICHELLE

PSC1707594

DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF PHYSICAL SCIENCE
UNIVERSITY OF BENIN
BENIN CITY

DECEMBER, 2022.

PREDICTION OF FOOTBALL MATCHES USING MACHINE LEARNING ALGORITHMS: (A CASE STUDY OF SPANISH LEAGUE)

BY

UTOMI AMAKA MICHELLE

PSC1707594

A PROJECT SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF PHYSICAL SCIENCES, UNIVERSITY OF BENIN, BENIN CITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR AWARD OF BACHELOR OF SCIENCE (B.Sc.) IN COMPTER SCIENCE

DECEMBER, 2022

CERTICATION

This is to certify that UTOMI AMAKA MICHELLE carried out this project work with Matriculation number PSC1707594 of Computer Science, University of Benin, Benin City.

_____ _____

MR. E. C. IGODAN DATE

(PROJECT SUPERVISOR)

APPROVAL

This project is hereby approved in partial fulfilment of the requirement for the award of Bachelor of Science (B.Sc.) Degree in Computer Science at the University of Benin. Benin City.

_____ _____

MR. E. C. IGODAN DATE

(Project Supervisor)

_____ _____

Prof. Mrs.A.O EGWALI DATE

(Head of Department)

DEDICATION

This project study is foremost dedicated to the Almighty God for His strength, wisdom, understanding, love, protection, inspiration and guidance who guided me throughout my course of study. In addition, to my mother Mrs. Utomi, for giving me the foundation upon which this academic journey is made possible, my siblings – Kimberly, Benson, Susan and Michael – for their endless support so far.

ACKNOWLEDEMENT

I would like to take this opportunity to thank all those who gave me the possibility to complete this project. Firstly, I would like to thank the Almighty God for providing me with the strength and resource to carry out this project.

I would like to offer my sincerest gratitude to my supervisor Mr. E. C. Igodan for his supervision, patience, knowledge and great support during the period of this project.

It is with radiant sentiment that I place on record, my best regard and deepest sense of gratitude to all lecturers of the department of Computer Science Department: Prof. Frank I. Amadin, Prof. Mrs. A.O. Egwali, Prof. Ekuobase, Prof. (Mrs.) F. A. Egbokhare, Dr. (Mrs.) G. O. Aziken, Dr. F. Chete, Dr. (Mrs.) V. I. Osunbor, Dr. E. P. Ebietomere, Mr. J. Odetayo, Mr. E. E. Obasohan, Mr. S.O.P. Oliomogbe, and Dr. (Mrs.) R. O. Osasere, Dr. F. O. Oliha, just to mention a few who made my university education a success.

I am entirely grateful to my parent Mrs. Ngozi Utomi, there is no doubt in my mind that without her continued support and counsel, I could never have come this far.

To my siblings The Famous five for their undying love and support from my travels from SA to my journey to UNIBEN, I love you all so much.

Lastly, I like to thank all my wonderful friends and course mates. To mention all would be an impossible task as my memory may fail; in an attempt to do that, many people would be omitted. Thank you all for your support throughout my stay in this university.

ABSTRACT

Football appears to be the most popular sports the world over, making it a game of betting for money making among other thing. This business of betting, over the years has gown making it a difficult and complex task in predicting correctly the outcome of football matches. This is as a result of the numerous number of factors that are considered but cannot be quantitatively valued or modeled. The aim of the project is to develop a machine learning algorithms for the prediction of football matches. The classification algorithms adopted in this project includes: K-Nearest Neighbor (KNN), support vector machines (SVM), Gaussian naïve Bayes (GNB), decision tree (DT) and Logistic Regression (LR) techniques. The dataset used was gathered from football-data-co.uk.

The models was built using python programming language environment. The comparative analysis carried out in this project support that machine learning algorithms perform well and shows room for future improvement.

Contents

1

CHAPTER ONE
INTRODUCTION

1.1 Background to the study

Football continues to be popular with 32 nations currently playing themselves to see who comes out victorious at the largest football competition in the world as the 2022 FIFA World Cup is in action. This year Edition of the 2022, FIFA World Cup forecasted to pull in 5 billion viewers according to FIFA President Gianni Infantino (MARCA, 2022). Very few events can even come close in terms of viewership, attendance and pricing. It makes sense that we would all be curious to learn which team will win the World Championship. The sport with the most television viewers,
the most costly television rights, the highest-paid athletes and events, and the most popular teams on social media called "The Beautiful Game," as the Brazilian player Pelé memorably described it

(Highlights et al., 2019). The area of influence of this sport is global and with the population of over 4 billion people that follow the sport to some degree (Sawe, World Atlas., 2019).
From World Cups to Nations Cups to Continental Championship Cup to league Championship, Football is the world's popular sports. Millions of people worldwide watch numerous leagues of football. As a result, there are more sponsorship agreements, betting companies operating internationally, sales of uniforms and other merchandise, match tickets, and commercials every day. The growth of the football business has brought together clubs looking for substantial financial assistance, which has significantly raised competitiveness amongst clubs.
One of the reasons for football being the most popular sport in the planet is its unpredictability. Numerous factors may be at play when the results of football matches are predicted, but curiosity and practical considerations are frequently the reasons behind soccer fans' predictions(Igea, 2019).2
The desire for financial gain might occasionally drive soccer fans to place bets online or at brick walls betting establishments. Other times, attempting to forecast the outcomes of the games just reflects an effort to gather data that will allow for a stimulating discussion with friends or family after the weekend.
It is challenging to quantify human precision when determining the outcome of a football game. Predicted theories influence precision. The accuracy of the assumptions for various leagues and tournaments varies, and they forecast the results for more leagues and tournaments than individuals
did. Comparing the network's accuracy to that of humans, this makes it challenging. One of the clever methods for demonstrating promising outcomes in the fields of categorization and prediction is machine learning (ML).
Machine learning (ML) algorithms are used to predict the outcomes of soccer matches. The Machine Language algorithm is trained using a data set consisting of a significant number of data
points (matches). For each match this training data includes features relevant to the games such as
the players, the number of corners, penalties, yellow and red cards, etc. of the match. The results of each match are also part of this training data. Prediction models may be built using the training
data. The models produced may be assessed comparing the results predicted by the model for a testing set of matches with the actual results of those matches (Igea, 2019).
Around the middle of the 20th century, the first effective efforts at using quantitative approaches to forecast football game outcomes were made. Using statistics, Moroney (1956) outlines a method
for predicting soccer results. Considering the limited computer power available at the time, fixing
the issue was a highly arduous and tied process.
It is becoming more and more common to predict soccer game results using data science approaches. In an effort to obtain a competitive advantage, football teams now hire data analysts. Each football game involves hundreds of player activities, but according to Ian Graham, director of research at Liverpool, the research division can only evaluate the ones that are collected from the football stat sheets and forms. The data that the research division manages is quite restricted, and Graham I. notes that he is striving to enhance the mathematical model of the games using video tracking. Improved simulations of a soccer match is created as more features are added to

the model. Better-detailed models will result in more accurate forecasts (Nytimes.com, 2019). Dhar (2013), prediction is the heart of remarkable disciplines in science and that is the reason why
the philosophy of prediction is employed in many companies around the world today to enable them forecast and make reasonable future plans. Machine learning, which an area of artificial intelligence is used in this project to predict football matches outcomes. The predictions would be
in four classes for each game: win, draw, loss or and cancel.

1.2 Research Motivation

This project aims to develop and assess a Machine Learning model able to predict results of football matches.

The project also provides answers to the following research questions:

a. To enhance the prediction capability.

b. To factor out which algorithm is more accurate.

1.3 Research Aim and Objectives

In this project, the researcher intend to implement a predictive model using machine learning approaches and deep learning methods for sports prediction. The specific objectives of this project
are to:

a. Extract relevant features from the football datasets using feature selection methods;

34

b. Design a predictive model for the prediction of the Spanish Primera Division;

c. Implement a predictive model using (a) and (b); and

d. Evaluate the performance of the predictive models in (c) based on standard metrics.

1.4 Research Methodology

This section describes the dataset, classification techniques and performance analysis. Seven Algorithms- long Short-term memory network (LSTM), k-nearest neighbors (KNN), support vector machine (SVM), logistic regression, decision tree and naïve Bayes are used in the experiment using the MLP model.

1.4.1 Dataset

The dataset used for the implementation was the Spanish La Liga; league of 2019/2020 to 2021/2022 seasons (football-data, 2022). The league consist of 20 teams played both home and away matches equaled to 380 matches per season and 1,520 matches for these four (4) years.

1.4.2 Performance Analysis

The performance of these classification algorithms was measured based on the accuracy. Accuracy
shows the rate at which the classifier meets the correct target class, that is, it determines the instances of data correctly classified (Rosli et al, 2018).

Accuracy = Total number of correctly predicted match result

---------- (1)

Total number of matches

The total number of correctly predicted Home Win, Away Win and Lose match results is equivalent to the total number of correctly predicted match results.5

1.5 scope of study

The scope of this project work is the Spanish Primera league (La Liga).

1.6 significance of study

With this study, we analyze the raw data, which includes previous fixture matches of the seasons and input values gathered from those matches. After analyzing, we expound the performance measures that can directly affect the results that will benefit teams and individuals who are engaged
in the betting scene.

CHAPTER TWO

LITERATURE REVIEW

2.0 INTRODUCTION

Until now, predictive analysist or researchers have tackled the predictions of football outcomes in
different approaches. Predicting a sporting event's outcome using technology has become increasingly important, as both the sports, betting industry and technology have expanded significantly. In actuality, when processing a large amount of data, humans have some limitations.
However, this problem can be overcome or solved using artificial intelligence techniques.

2.1 TYPES OF ARTIFICIAL INTELLIGENCE—WEAK AI VS. STRONG AI

I. Weak AI: Also, called Narrow AI or Artificial Narrow Intelligence (ANI)—is AI trained and focused to perform specific tasks. Weak AI drives most of the AI that surrounds us today. 'Narrow'
might be a more accurate descriptor for this type of AI as it is anything but weak; it enables some very robust applications, such as Apple's Siri, Amazon's Alexa, IBM Watson, and autonomous vehicles.
II. Strong AI: Is made up of Artificial General Intelligence (AGI) and Artificial Super Intelligence
(ASI). Artificial general intelligence (AGI), or general AI, is a theoretical form of AI where a machine would have an intelligence equaled to humans; it would have a self-aware consciousness
that has the ability to solve problems, learn, and plan. Artificial Super Intelligence (ASI)—also known as superintelligence —would surpass the intelligence and ability of the human brain.
III. While strong AI: is still entirely theoretical with no practical examples in use today that does
not mean AI researchers are not also exploring its development. In the meantime, the best examples
6of ASI might be from science fiction, such as HAL, the superhuman, rogue computer assistant in
2001: A Space Odyssey (IBM, 2020).

2.2 ARTIFICIAL INTELLIGENCE

The term intelligence would be the capacity to acquire and apply relevant skills to resolve issues and accomplish objectives in a context-specific manner in a world that is always changing and unpredictable. A manufacturing robot that has been fully preprogrammed is flexible, accurate, and
reliable but not intelligent (HAI, 2020).
John McCarthy, an emeritus Stanford professor, first used the phrase artificial intelligence (AI) in
1955. He later described it as "the science and engineering of creating machine intelligence". In a lot of research, people have taught robots to play chess or act cleverly, but today we focus on teaching machines to learn at least somewhat similarly to humans (HAI, 2020).

Artificial intelligence is a topic that, in its most basic form, combines computer science and substantial datasets to facilitate problem solving. Additionally, it includes the branches of artificial
intelligence known as deep learning and machine learning, which are commonly addressed together. These fields use AI algorithms to build expert systems that make predictions or categorize
information based on incoming data (IBM, 2020).
Artificial intelligence (AI) is the capacity of a digital computer or robot operated by a computer to
carry out actions frequently performed by intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from experience. It has been proven that computers can be programmed to perform extremely complicated tasks—like, for example, finding proofs for mathematical theorems or playing chess—with remarkable
7proficiency ever since the development of the digital computer in the 1940s. Nevertheless, despite
ongoing improvements in computer processing speed and memory space, there are currently no programs that can match human adaptability across a larger range of activities or those needing a substantial amount of background knowledge. On the other hand, some programs have reached the
performance levels of human experts and professionals in carrying out some specific tasks, so artificial intelligence in this constrained sense is present in a variety of applications, including voice or handwriting recognition, computer search engines, and medical diagnosis (Copeland, 2022).
AI applications include advanced web search engines (e.g., Google), recommendation systems (used by YouTube, Amazon and Netflix), understanding human speech (such as Siri and Alexa), self-driving cars (e.g., Tesla), automated decision-making and competing at the highest level in strategic game systems (such as chess and Go). (AlphaGo, 2017).
The AI effect is a phenomenon where actions once thought to require "intelligence" are frequently
taken out of the definition of AI, as machines grow more and more capable (McCorduck, 2004). For instance, optical character recognition is frequently left out of the category of AI-related technologies (Hackernoon, 2019), having become a routine technology (Schank, 1991).
Artificial intelligence originally established as an academic discipline in 1956 and has had multiple
waves of optimism since then, followed by disappointment and the loss of funding (known as an "AI winter") (Crevier, 1993), followed by new approaches, success and renewed funding.
Since its inception, AI research has experimented with and abandoned a wide range of methodologies, including modeling human problem solving, formal logic, extensive knowledge bases, and animal behavior imitation. Machine learning that is heavily based in mathematics and
8statistics has dominated the subject in the first two decades of the twenty-first century. This approach has been very effective in solving difficult issues in both industry and academics (Clark, 2015b). Traditional objectives of AI research include reasoning, knowledge representation, planning, learning, natural language processing, perception, and the ability to move and manipulate objects (Russell and Norvig, 2003),
One of the long-term objectives of the area is general intelligence, or the capacity to solve any

problem (Robert, 2016). Artificial intelligence (AI) researchers have integrated and modified a wide range of problem-solving techniques, including as formal logic, artificial neural networks, search and mathematical optimization, as well as approaches from statistics, probability, and economics, to address these issues. AI (Pennachin & Goertzel 2007) also influences computer science, psychology, linguistics, philosophy, and many other disciplines.

The field was founded on the assumption that human intelligence "can be so precisely described that a machine can be made to simulate it. This raised philosophical arguments about the mind and
the ethical consequences of creating artificial beings endowed with human-like intelligence; these
issues have previously been explored by myth, fiction and philosophy since antiquity (Newquist, 1994).

From assisting referee decisions, to improving physical performance and predicting future harm, to producing and implementing tactics, artificial intelligence is pervading more and more aspects of the footballing world.

Most recently, FIFA and UEFA have approved the use of Semi-automated Offside Technology (SAOT) at the 2022 FIFA World Cup in Qatar as well as for the 2022, UEFA Super Cup match and during the group stage matches of the 2022-2023 UEFA Champions League. However, there 9may be more to artificial intelligence technology than meets the eye, as well as its applications and
algorithms (LawInSport, 2022).

2.3 ARTIFICIAL INTELLIGENCE APPLICATIONS

Applications of artificial intelligence (AI) have been employed in business and academics to solve
specific issues. Like electricity or computers, artificial intelligence is a general-purpose technology
with a wide range of uses. It has been used in fields of language translation, image recognition, credit scoring, e-commerce and other domains (Erik and Tom, 2017).

There are numerous, real-world applications of AI systems today. Below are some of the most common examples:

I. Speech recognition: It is also known as automatic speech recognition (ASR), computer speech
recognition, or speech-to-text, and it is a capability, which uses natural language processing (NLP)
to process human speech into a written format. Many mobile devices incorporate speech recognition into their systems to conduct voice search—e.g. Siri—or provide more accessibility around texting.

II. Customer service: Online virtual agents are replacing human agents along the customer journey. They answer frequently asked questions (FAQs) around topics, like shipping, or provide personalized advice, cross-selling products or suggesting sizes for users, changing the way we think about customer engagement across websites and social media platforms. Examples include messaging bots on e-commerce sites with virtual agents, messaging apps, such as Slack and Facebook Messenger, and tasks usually done by virtual assistants and voice assistants.

III. Computer vision: This AI technology enables computers and systems to derive meaningful information from digital images, videos and other visual inputs, and based on those inputs, it can

10take action. This ability to provide recommendations distinguishes it from image recognition tasks.
Powered by convolutional neural networks, computer vision has applications within photo tagging
in social media, radiology imaging in healthcare, and self-driving cars within the automotive industry.
IV. Recommendation engines: AI algorithms can assist in identifying data trends that can be used
to create more efficient cross-selling tactics by using historical consumption behavior data. This is
used to make relevant add-on recommendations to customers during the checkout process for online retailers.
V. Automated stock trading: Designed to optimize stock portfolios, AI-driven high-frequency trading platforms make thousands or even millions of trades per day without human intervention (IBM, 2020).
1112
Figure 2.1. AI functional applications (WIPO, 2019)Figure 2.2 AI applications field (WIPO, 2022).

2.4 MACHINE LEARNING

The automatic recognition of deliberate patterns in data is referred to as machine learning. It has evolved into a typical tool in the recent years for almost every activity requiring data extraction from huge information collections. Machine learning is utilized in the current world to simplify human life in areas like search engines, anti-spam software and computer codes, digital cameras, smart phones, accident interference systems, and scientific fields including bioinformatics, 13medicine, physical science, and computer engineering. For instance, after each query, search engines learn how to present the simplest results.
A software system can safeguard credit card transactions by training anti-spam software systems to screen email communications. Computer programs can also learn how to spot frauds. Smartphones may achieve voice recognition commands and face recognition because digital cameras learn to detect and analyze faces. Accident prevention systems are already standard in modern vehicles, protecting both the driver and passengers from potential collisions. As is known,
a wide range of fields and specialties use machine learning. Due to this circumstance, different researchers have created numerous definitions. It is describe as the systematic use of computations
and factual models by computer frameworks to carry out a given activity successfully without the
need for clear illumination, relying instead on designs and inductive inference.

2.4.1 DEFINITION OF MACHINE LEARNING

According to Alpaydn, (2014), machine learning is the process of programming computers to optimize an execution model using case data or previous experience. Learning is the application of a computer program that optimizes a model we have with a few parameters. The model could be predictive to set predictions for the future, visual to gather data from data, or both. Since drawing conclusions from a sample is the primary objective of machine learning, the insights hypothesis is used while creating numerical models.

Machine learning is an ever-evolving field of technology whose algorithms are designed to mimic
human insights by absorbing information from the surrounding environment. Various fields, such
as aerospace engineering, pattern recognition, computer vision, finance, entertainment, and medical applications, have successfully adopted machine-learning techniques. (Murphy and Naqa,
2015)
14According to (Bishop, 2016) machine learning and pattern recognition are concerned with automatically identifying patterns in data by utilizing computer algorithms, as well as with acting on these patterns. Bishop further noted that in order to carry out the responsibilities of prediction and decision-making, machine learning algorithms build a model based on training data.
Machine learning is the study of using computers to mimic human learning, according to (Tang et
al.,) additionally, machine learning is a field of study that uncovers new skills and knowledge, separates it from already known information, and consistently enhances performance. The fundamental learning methods used by machine learning include rote learning, inductive reasoning, analogical learning, and deductive learning.
Rote learning represents the memory that stores and retrieves the information without reasoning and calculation. Inductive reasoning represents the common knowledge of specific instances, extracting the common regulations of data and reasoning from specific to general. Analogical learning is comparing similar actions and trying to find the relations and possible solutions of these
actions. Deductive learning represents the result of certain explications and clarified process of examples from general to specific.

2.4.2 HISTORY OF MACHINE LEARNING

When we examine the history of machine-learning field, in early, 1940's Warreen McCulloch and
Walter Pitts wrote a scientific paper (McCulloch and Pitts) about artificial neurons and this research is the first study of neural networks in the literature. Human brain data transmission mechanism and electrical circuits inspire proposed model. In 1950, Alan Turing published a manuscript (Turing, 2009) on computing machinery and intelligence. In this paper, Alan Turing develops the Turing Test to specify a case that a computer has real intelligence. In order to pass the test, a computer must be able to trick a human into believing it is human. In 1952, Arthur
1516
Samuel developed a computer program that plays checkers, and the program improves itself after each play with itself. In 1957, Frank Rosenblatt published a paper (Rosenblatt, 1958) on perceptron, which is a probabilistic model for information storage and organization in the brain. He defines the perceptron structure consisting of neurons and aiming to recognize patterns. Moreover, this scientific study is crucial for the invention of developing feed forward neural networks model and back propagation technique, which are the fundamental methods of the modern artificial neural network algorithms in early 70's. In 1967, Pelillo published a paper (Pelillo, 2014), which includes the nearest neighbor rule.
In 1980, Fukushima (Fukushima, 1980) proposed Neocognitron, which models artificial neural networks in a different perspective, being accepted as the ancestor of the convolutional neural networks model. (Hopfield, 1982) invented recurrent neural networks.

Recurrent structures are separated from feed forward structures because they use their outputs as inputs in the next process. Consequently, recurrent networks have memory. (Watkins, 1989) published his PhD thesis including the development of q learning that improves the usability of reinforcement learning. Reinforcement learning is the machine learning that consists of agents in a dynamic learning environment, which uses punishments and reward mechanisms. (Tesauro, 1992) developed a computer backgammon program, called TD-Gammon, which uses supervised learning and the multilayer neural network algorithm. The program was designed to play expert level backgammon. A researched paper that described random forest algorithm published by (Ho, 1995). In addition, (Vapnik and Cortes) published a researched paper in the same year that was described support vector machine algorithm. Both researches are new scientific discoveries and they contributed to the literature.

2.5 HOW DOES MACHINE LEARNING WORK?

Machine Learning is, undoubtedly, one of the most exciting subsets of Artificial Intelligence. It completes the task of learning from data with specific inputs to the machine. It is important to understand what makes Machine Learning work and, thus, how it can be used in the future.

The Machine Learning process starts with inputting training data into the selected algorithm. Training data being known or unknown data to develop the final Machine Learning algorithm. The
type of training data input does affect the algorithm, and that concept will be covered further shortly.

New input data is fed into the machine-learning algorithm to test whether the algorithm works correctly. The prediction and results are then checked against each other.

If the prediction and results do not match, the algorithm is re-trained multiple times until the data scientist gets the desired outcome. This enables the machine-learning algorithm to continually learn on its own and produce the optimal answer, gradually increasing in accuracy over time.

The next section discusses the three types of and use of machine learning.

2.6 TYPES OF MACHINE LEARNING

Machine learning is a subset of AI, which enables the machine to automatically learn from data, improve performance from experiences, and make predictions**.** Machine learning contains a set of
algorithms that work on a huge amount of data. Data is fed to these algorithms to train them, and based on training; they build the model & perform a specific task.

2.6.1 SUPERVISED LEARNING

As its name suggests, supervised machine learning is based on supervision. It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output. Here, the labelled data specifies that some of the 17inputs be already mapped to the output. More preciously, we can say; first, we train the machine
with the input and corresponding output, and then we ask the machine to predict the output using the test dataset.

2.6.2 CATEGORIES OF SUPERVISED MACHINE LEARNING

Supervised machine learning can be classified into two types of problems, which are given below:

I. Classification

Classification algorithms are used to solve the classification problems in which the output variable

is categorical, such as "Yes" or No, Male or Female, Red or Blue, etc. The classification algorithms
predict the categories present in the dataset. Classification deals with text categorization, natural language processing, informatics, fraud detection, face recognition, marketing, optical character recognition (Cimen, 2019). Some real-world examples of classification algorithms are Spam Detection, Email filtering, etc.
Figure 2.3. Classification example
18Some popular classification algorithms are listed below:
a) ANN Algorithm
b) Naïve Bayes Algorithms
c) K-NN Algorithm
d) Support Vector Machine Algorithm

II. Regression

Regression is the linear relationship between two or more variables. Unlike the classification, predictions of the regression is about the numeric outputs in order to measure the relationship between two or more variables. If analysis is performed using a single variable, it is called univariate regression, and if more than one variable is used, it is called multivariate regression analysis. Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.
19Figure 2.4. Example of Linear Regression
Some popular regression algorithms are listed below:
a) Simple Linear Regression Algorithm
b) Multivariate Regression Algorithm
c) Decision Tree Algorithm
d) Lasso Regression

2.6.3 UNSUPERVISED MACHINE LEARNING

Unsupervised learning is different from the supervised learning technique; as its name suggests, there is no need for supervision. It means, in unsupervised machine learning, the machine is trained
using the unlabeled dataset, and the machine predicts the output without any supervision.
2021
In unsupervised learning, the models are trained with the data that is neither classified nor labelled,
and the model acts on that data without any supervision.
The main aim of the unsupervised learning algorithm is to group or categories the unsorted dataset
according to the similarities, patterns, and differences. Machines are instructed to find the hidden patterns from the input dataset.

2.6.4 CATEGORIES OF UNSUPERVISED MACHINE LEARNING

Unsupervised Learning can be further classified into two types, which are given below:

I.

Clustering

The clustering technique is used when we want to find the inherent groups from the data. It is a way to group the objects into a cluster such that the objects with the most similarities remain in one group and have fewer or no similarities with the objects of other groups. An example of the

clustering algorithm is grouping the customers by their purchasing behavior.22
Figure 2.5 Example of clustering
Some of the popular clustering algorithms are given below:
a) K-Means Clustering algorithm
b) Mean-shift algorithm
c) DBSCAN Algorithm
d) Principal Component Analysis
e) Independent Component Analysis

II. Association

Association rule learning is an unsupervised learning technique, which finds interesting relations among variables within a large dataset. The main aim of this learning algorithm is to find the dependency of one data item on another data item and map those variables accordingly so that it can generate maximum profit. This algorithm is mainly applied in Market Basket analysis, Web usage mining, continuous production**,** etc.
Some popular algorithms of Association rule learning are Apriori Algorithm, Eclat, and FP-growth
algorithm.

2.6.5 REINFORCEMENT

Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance. Agent is rewarded for each good action and are punished for each bad action; hence, the goal of reinforcement learning agent is to maximize the rewards.
In reinforcement learning, there is no labelled data like supervised learning, and agents learn from
their experiences only.
The reinforcement learning process is similar to a human being; for example, a child learns various
things by experiences in his day-to-day life. An example of reinforcement learning is to play a game, where the Game is the environment, moves of an agent at each step define states, and the goal of the agent is to get a high score. Agent receives feedback in terms of punishment and rewards.
Due to its way of working, reinforcement learning is employed in different fields such as Game theory, Operation Research, Information theory, multi-agent systems.

232.6.7 CATEGORIES OF REINFORCEMENT LEARNING

Reinforcement learning is categorized mainly into two types of methods/algorithms:
POSITIVE REINFORCEMENT LEARNING: Positive reinforcement learning specifies increasing the tendency that the required behavior would occur again by adding something. It enhances the strength of the behavior of the agent and positively affects it.
NEGATIVE REINFORCEMENT LEARNING: Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behavior would occur again by avoiding the negative condition.

2.7 FEATURE SELECTION

The input variables that we give to our machine learning models are called features. Feature Selection is the method of reducing the input variable to your model by using only relevant data and getting rid of noise in data (simpliLearn, 2022). It is the process of automatically choosing

relevant features for your machine-learning model based on the type of problem you are trying to solve. We do this by including or excluding important features without changing them. It helps in
cutting down the noise in our data and reducing the size of our input data.
2425
Figure 2.6 Feature selection

2.7.1 OBJECTIVES OF FEATURE SELECTION

1. It eliminates irrelevant and noisy features by keeping the ones with minimum redundancy and maximum relevance to the target variable.
2. It reduces the computational time and complexity of training and testing a classifier, so it results
in more cost-effective models.
3. It improves learning algorithms' performance, avoids overfitting, and helps to create better general models (Towards Data Science, 2021).

2.8 TYPES OF FEATURE SELECTION METHODS

There are three categories of feature selection methods, depending on how they interact with the classifier, namely, filter, wrapper, and embedded methods.

I.

Filter methods are scalable (up to very high-dimensional data) and perform *fast* feature selection before classification so that the bias of a learning algorithm does not interact with 26 the bias of the feature selection algorithm.

They mainly act as rankers, ordering the features from best to worst.

The ranking of features depends on the intrinsic properties of the data, for example, variance, consistency, distance, information, correlation, etc.

There exist many filter methods, and new ones are developed regularly, each of them uses a different criterion to measure the relevance of the data.

One general definition for relevance is that a feature can be regarded as irrelevant if it is conditionally independent of the class labels or it does not influence the class labels; in these cases, it can be discarded.

II.

Wrapper methods use a machine-learning algorithm as a *black box evaluator* to find the best subsets of features, and so, they are dependent on the classifier.

Practically any combination of the search strategy and modeling algorithm can be used as a wrapper.

When a wrapper is performed in a dataset with plenty of features, it consumes exceptional computational resources and time to run.

Finally, these methods are simple to implement and can model feature dependencies.

III.

Embedded methods:

bridge the gap
between filters and wrappers.

To begin with, they fuse measurable and statistical criteria like a filter to choose some features, and then using a machine-learning algorithm, they pick the subset with the best classification performance.

They reduce the computational complexity of wrappers without re-classifying the subsets in each iteration and can model feature dependencies.

They
do not perform iterations.27
Feature selection is performed in the learning phase, meaning that these methods achieve both
model fitting and feature selection at the same time.
One disadvantage is their dependency on the classifier (Towards Data Science, 2021).

2.9 SUPPORT VECTOR MACHINE

Data Mining could be a pioneering and engaging analysis space thanks to its vast application areas
and task primitives. Support Vector Machine (SVM) is enjoying a decisive role because it provides
techniques that are particularly compatible to get ends up in an economical method and with a good level of quality. The use of support vector machine in various applications makes this tool inevitable for the development of products, which have implications for the society. Support vector
machines being computationally powerful tools for supervised learning, are widely employed in classification, clump, and regression problems. Support vector machines have a good performance
when implementing various kind of problems like face recognition, text categorization, bioinformatics, computer-science related topics, civil engineering, and electric electronics, etc. Support Vector Machine algorithm, first developed by Vladimir Vapnik in the fields of applied mathematics, learning theory and structural risk reduction, have demonstrated to figure successfully on various prediction and classification issues (Vapnik, 1998).
Many pattern recognition and regression problems involving estimate and prediction employ support vector machines. Support vector machines are able to capture a wide range of features thanks to the structural risk minimization theory's generalization concept. Support vector machines
fall within the category of supervised learning models, which examine data and identify patterns for the purposes of regression and classification. The features of the support vector machine 28 algorithm can be categorized as kernels, margin, duality, sparseness, and convexity (diggdata, 2014).
Figure 2.7 SVM Classification hyperplane

2.10 K NEAREST NEIGHBORS ALGORITHM

K nearest neighborhood algorithm as a non-parametric decision method that can be utilized for classification and regression operations (Hart et al., 2000). In classification, k nearest neighbor algorithm categorizes and labels the test samples regarding the feature sets with the k value and finding the closest values to the k value in every class on the feature set. Euclidian distance, Manhattan distance and Minkowski distance are the popular distance vector functions when finding the closest values 29
Figure 2.8 nearest Neighbor Algorithm Usage Example

Distance functions

Euclidean $\sqrt{\sum (xi - yi)}$
2
$k\ i=1$
2.1
Manhattan $\sum_{k\ i=1} |xi - yi|$ **2.2**
Minkowski $[\sum (|xi - yi|)$

q
$k\ i=1$
$]^{1/q}$ **2.3**
No a. Figure limit equationsFigure 2.8.1 (sayad, 2022) represents the Euclidian, Manhattan and Minkowski distance functions
formulas. As can be seen on the figure 2.8.1, in both Euclidian, Manhattan and Minkowski formula, Xi and Yi represent the points on the coordinate axis and the main aim is to find the distance between Xi and Yi (Chomboon et al., 2015). (Han et al., 2012) defines Minkowski distance formula that can be considered as a generalization of the both Euclidian and Manhattan distances.

2.11 ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) are brain-inspired algorithms that are used to forecast problems
and model complex patterns. The idea of biological neural networks in the human brain gave rise to the Artificial Neural Network (ANN), a deep learning technique. Artificial neural network refers
to a biologically inspired sub-field of artificial intelligence modeled after the brain. An effort to simulate how the human brain functions led to the creation of ANN. Although they are not exactly
the same, the operations of ANN and biological neural networks are very similar. The ANN algorithm accepts only structured and numeric data (Vidhya, 2021). ANN can model as the original
neurons of the human brain; hence, ANN processing parts are called Artificial Neurons.
ANN consist of a large number of interconnected neurons that are inspired by the working of a brain. These neurons have the capabilities to learn, generalize the training data and derive results from complicated data. Artificial neural networks neurons are linked to each other in various layers
of the networks. These neurons are known as nodes. These networks are used in the areas of classification & prediction, pattern & trend identifications, optimization problems, etc. ANN learns
from the training data (input and target output known) without any programming.
The learned neural network is called an expert system with the capability to analyze information and answer the questions of a specific field.
30Figure 2.9. Artificial Neural Networks
The units of calculation are neutrons. The artificial neurons are connected by synapses, which are really just weighted values. Artificial neural network consist of three layers:
31**Figure 2.10.** Layers of Artificial neural network
Input Layer: As its name implies, it accepts inputs supplied by the programmer in a variety of various formats.
Hidden Layer: The hidden layer presents in-between input and output layers. It performs all the calculations to find hidden features and patterns.
Output Layer: Using the hidden layer, the input is transformed over time to produce an output that is then communicated using this layer.
A bias is added as part of the weighted sum of the inputs that the artificial neural network computes
after receiving input. A transfer function serves as the visual representation of this calculation.
In order to produce the output, it passes the weighted total as an input to an activation function.

$\sum_{i=1}^{n} Wi * xi + b$ 2.4

32Activation functions choose whether a node should fire or not. Only those who are fired make it
to the output layer. Depending on the type of task we are completing, many activation functions can be used.

I NEURAL NETWORK

This section explains the difference between the three types of neural networks and cover the basics of deep neural networks. Here one will cover the following neural network types:

1. Multi-Layer perceptron (MLP)
2. Convolutional Neural networks (CNN)
3. Recurrent Neural Networks (RNN)

II. WHAT IS A DEEP NEURAL NETWORK?

A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between
the input and output layers. Deep neural network models have become a powerful tool for machine
learning and artificial intelligence.

Neural networks target brain-like functionality and are based on a simple artificial neuron: a nonlinear function (such as max (0, value)) of a weighted sum of the inputs. These pseudo neurons
are collected into layers, and the outputs of one layer becoming the inputs of the next in the sequence.

III. WHAT MAKES A NEURAL NETWORK "DEEP"?

Deep neural networks employ deep architectures in neural networks. "Deep" refers to functions with higher complexity in the number of layers and units in a single layer. The ability to manage large datasets in the cloud made it possible to build models that are more accurate by using
3334
additional and larger layers to capture higher levels of patterns. The two key phases of neural networks are called training (or learning) and inference (or prediction), and they refer to the development phase versus production or application. When creating the architecture of deep network systems, the developer chooses the number of layers and the type of neural network, and training determines the weights.

2.12 MULTILAYER PERCEPTRON (MLP)

A multilayer perceptron (MLP) is a class of a feedforward artificial neural network (ANN). MLPs
models are the most basic neural network, which is composed of a series of fully connected layers.

Today, MLP machine learning methods can be used to overcome the requirement of high computing power required by modern deep learning architectures. Each new layer is a set of nonlinear functions of a weighted sum of all outputs (fully connected) from the prior one.

Figure 2.11. Multilayer perceptrons35

Figure 2.4 multiple layer perceptron activation (medium, 2018)

Where x = inputs
w = weights
b = bias
f = activation function

2.13 CONVOLUTIONAL NEURAL NETWORK

A convolutional neural network (CNN) is another class of deep neural networks. CNNs are most commonly employed in computer vision. Given a series of images or videos from the real world, with the utilization of CNN, the AI system learns to automatically extract the features of these inputs to complete a specific task, e.g., image classification, face authentication, and image semantic segmentation. Different from fully connected layers in MLPs, in CNN models, one or multiple convolution layers extract the simple features from input by executing convolution operations. Each layer is a set of nonlinear functions of weighted sums at different coordinates of spatially nearby subsets of outputs from the prior layer, which allows the weights to be reused.

Figure 2.13. Basic architecture of CNN LeNet

Applying various convolutional filters, CNN machine learning models can capture the high-level representation of the input data, making CNN techniques widely popular in computer vision tasks.

Convolutional neural network example applications include image classification (e.g., AlexNet, VGG network, ResNet, MobileNet) and object detection (e.g., Fast R-CNN, Mask R-CNN, YOLO, SSD).

2.14 RECURRENT NEURAL NETWORK (RNN)

A recurrent neural network (RNN) is another class of artificial neural networks that use sequential

data feeding. RNNs have been developed to address the time-series problem of sequential input data. The input of RNN consists of the current input and the previous samples. Therefore, the connections between nodes form a directed graph along a temporal sequence. Furthermore, each 3637

neuron in an RNN owns an internal memory that keeps the information of the computation from the previous samples.

Figure 2.14 Recurrent neural network

RNN models are widely used in Natural Language Processing (NLP) due to the superiority of processing the data with an input length that is not fixed. The task of the AI here is to build a system that can comprehend natural language spoken by humans, e.g., natural language modeling,

word embedding, and machine translation. In RNNs, each subsequent layer is a collection of nonlinear functions of weighted sums of outputs and the previous state. Thus, the basic unit of RNN is called “cell”, and each cell consists of layers and a series of cells that enables the sequential

processing of recurrent neural network models (viso.ai, 2022).38

2.15 NAÏVE BAYES

The Nave Bayes algorithm is a supervised learning method for classification problems that is based

on the Bayes theorem (JavaTpoint, 2021).

It is mostly employed in text categorization with a large training dataset. The Naive Bayes Classifier is one of the most straightforward and efficient classification algorithms available today.

It aids in the development of efficient machine learning models capable of making accurate predictions.

Being a probabilistic classifier, it makes predictions based on the likelihood that an object will

occur. Spam filtration, Sentimental analysis, and article classification are a few examples of Naive
Bayes algorithms that are frequently used (JavaTpoint, 2021).
The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:
I.
Naïve: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence, each feature individually contributes to identify that it is an apple without depending on each other.
II.
Bayes: It is called Bayes because it depends on the principle of Bayes' Theorem
Bayes' Theorem:
Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.39
The formula for Bayes' theorem is given as:

$$P(A|B) = P(B|A)P(A)$$
$$P(B)$$

------- (2.5)
Where,
P (A|B) is Posterior probability: Probability of hypothesis A on the observed event B.
P (B|A) is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.
During the training process that the algorithm undertakes, the Naive Bayes Classifiers defines Parameters looking at each feature individually and it collects for each feature simple per-class Statistics. The Naive Bayes classifiers are applicable for continuous data and store the average values as well as the standard deviations of each feature for each class (Muller and Guido, 2018). The algorithm collects statistics of the mean and standard deviations for shots that resulted in a goal and for the shots that did not result in a goal. Hence, the Naive Bayes Classifier algorithm produces a model able to predict if a shot from a defined distance and angle results in a goal or not
(Igea, A, 2019).40

2.15.1 TYPES OF BAYES NAÏVE

Figure 2.15 Types of Bayes naïve (Medium, 2020)
(https://medium.com/@chaudhurysrijani/what-is-so-naive-about-naive-bayes-and-how-does-it deal-with-time-and-space-ea19826f5011)
I.
Gaussian: It is employed in classification and relies on the concept that feature data is distributed normally.
II.
Multinomial: Discrete counts are handled by multinomial. Let's take the issue of text classification as an example. We can take Bernoulli trials into consideration here, which is

one step further. Instead of counting the number of times a word appears in a document, we have "count how many times a word appears in a document." You can think of this as the "number of times outcome number x I is observed over the n trials."
III.
Bernoulli: If your feature vectors are binary, the binomial model is beneficial (i.e. zeros and ones). One use would be text classification using a "bag of words" model, where 1s and 0s represent words that occur in documents and those that do not (Medium, 2022).41

2.15.2 GAUSSIAN NAÏVE BAYES CLASSIFIER

In Gaussian Naïve Bayes, continuous values connected to each feature are presumed to be distributed using a Gaussian distribution (Normal distribution). Plotting it results in the bell-shaped
curve below, which is symmetric about the mean of the feature values (KDnuggets, 2022).
Figure 2.16 (KDnuggets, 2022)
(https://www.kdnuggets.com/2020/06/naive-bayes-algorithm-everything.html)
The likelihood of the features is assumed Gaussian; hence, conditional probability is given by:

$$P(x) = \sqrt{2\pi\sigma}\,1\,2\,e - (x - 2\sigma\,\mu\,2)\,2$$

--------- (2.6)
Now, what if any feature contains numerical values instead of categories i.e. Gaussian distribution.
Before constructing frequency tables, one alternative is to convert the numerical values to their 42
categorical equivalents. The third choice, as demonstrated above, could be to estimate the frequency by using the distribution of the numerical variable. Assuming normal or Gaussian distributions for numerical variables is one popular approach, for instance (KDnuggets, 2022). Two parameters define the probability density function for the normal distribution (mean and standard deviation).

$$\mu = \frac{1}{n}\sum_{i=1}^{n} xi$$

(2.7)

$$\sigma = \left[\frac{1}{n-1}\sum_{i=1}^{n} (xi - \mu)^2\right]^{0.5}$$

(2.8)

$$f(x) = \sqrt{2}\,1\,\pi\sigma\,e - (x\,2 - \sigma\,\mu\,2)\,2$$

(2.9)
(KDnuggets, 2022)

2.16 DECISION TREE

A decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. It has a hierarchical, tree structure, which consists of a root node, branches, internal nodes and leaf nodes (IBM, 2022).

The tree has nodes that split for the different values of a specific feature. The tree also has edges (Branches) that are outcomes of a split and go to the next feature or nodes. The root is the node Where the first split is performed and the tree also has leaves that are the terminal nodes which Predict the outputs of the problem (Igea. A, 2019).43

Figure 2.17. Decision Tree

(IBM, 2022)

(https://www.ibm.com/topics/decision trees#:~:text=A%20decision%20tree%20is%20a,internal%20nodes%20and%20leaf%20nodes.)

The root node of a decision tree, which lacks any incoming branches, is shown in the diagram above. The internal nodes, sometimes referred to as decision nodes, are fed by the root node's outgoing branches. Both node types undertake assessments based on the available attributes to create homogenous subsets, which are represented by leaf nodes or terminal nodes. The leaf nodes

represent all of the outcomes within the dataset (IBM, 2022).**2.17 LOGISTIC REGRESSION**

Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables (JavaTpoint, 2022).

Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome

must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.

Logistic Regression is much similar to the Linear Regression except that how they are used. Linear

Regression is used for solving Regression problems, whereas Logistic regression is used for solving the classification problems.

In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1). The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.

Logistic Regression is a significant machine-learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.

Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification (JavaTpoint, 2022).

4445

Figure 2.18 logistic regression (JavaTpoint, 2022)

https://www.javatpoint.com/logistic-regression-in-machine-learning

The mathematical steps to get Logistic Regression equations are given below:

The equation of the straight line can be written as:

$$y = b0 + b1x1 + b2x2 + b3x3 + \cdots + bnxn \quad (2.10)$$

In Logistic Regression y can be between 0 and 1 only, so for this let us divide the above equation by (1-y):

$$\frac{y}{1-y}; 0\ for\ y = 0, and\ infinity\ for\ y = 1 \quad (2.11)$$

But we need range between -[infinity] to +[infinity], then take logarithm of the equation it will become:

$$\log\left[\frac{y}{1-y}\right] = b0 + b1x1 + b2x2 + b3x3 + \cdots + bnxn \quad (2.12)$$

The above equation is the final equation for Logistic Regression (JavaTpoint, 2022).4647

LITERATURE REVIEW

Author's Name / year	Motivation	Objective(s)	Methodology	Contribution to Knowledge	Limitation(s)
Xu, (2012)	Predict sports performance accurately	To show how well the models and algorithm were implemented.	Using BP neural, Genetic Algorithm.	The system achieved as small as 26.637 on the training set and 32.3334 on the test set.	Limitations of data
Borycki, (2011)	Optimize betting profits and gains	To propose a model to predict sports outcome	Using Nearest Neighbors Search and Genetic		

algorithm This showed a profitable strategy so to make investing in sports betting is possible			
There was no investigation into the time complexity. New algorithms can be.	48		
Ulmer and Fernandez (2014)			
Due to the shame of being outdone by an octopus in predicting world cup games.			
Predictions in the results of soccer matches in the English premier league (EPL).			
Artificial neural network, Baseline, Naïve Bayes, Hidden Markov Model, Support vector Machine, Random forest, one vs-all			
Stochastic Gradient Limitation of DataDescent in correctly predicting the results of			

2012-2013
EPL Season,
showing
Manchester
United as the
winner.

49CHAPTER THREE
SYSTEM ANALYSIS AND DESIGN

3.1 INTRODUCTION

System analysis and design is a step-by-step process for developing high-quality information system. An informative system combines technology, people and data to provide support for business function such as order processing, inventory control, human resources and many more (Rosenblatt, 2013).

3.2 ANALYSIS OF STUDY

Due to the immense success of football, a great deal of research has been conducted to enhance prediction. One of the primary ways to engage people is to provide them with the platform that makes prediction on the go for their favorite matches.it provides an opportunity for viewers to be connected.

Most viewers read reviews on matches and make predictions before the match. For a single match,
reviews are posted in hundreds and often times even in thousands. Thus, it is difficult for any viewer to go through all reviews before making prediction. Hence, it is essential to classify these reviews and the filtration of these reviews can be done based on the categories: home win, away win, a draw with machine-learning algorithm.

3.3 EXISTING SYSTEM

Handling of large amount of data is a big challenge not just on the end of the viewers but also to the business owners to go through all the users reviews and make important predictions for the most desired matches. Existing method involved checking game head to head for the past 5 weeks
and making predictions from it.

5051

3.3.1 CONSTRAINTS OF EXISTING OF SYSTEM

Many problems opposing the current method were identified as the research and investigation exercise was being carried out. Such problem includes:

I.
Organization had always been concerned about the success of a project and how the public perceives the application. This concern results from a variety of motivations. They are
faced with the problem of assigning someone to manually compile game statistics from previous years.

II.
Physically search and input games standings.

III.
Difficulty in gathering user's interest.

IV.
Time consuming to read through since handling of large amount of data can be challenging.

3.4 JUSTIFICATION OF THE EXISITING

Due to the increasing variety of challenges encountered when attempting to gather a significant number of reviews on a certain product across several internet sites, it is necessary to find a possible solution to the constraints of existing system. To resolve this problem as stated in section
3.3.1. In bid to resolve this problem we would employ the use of artificial neural network technique
multilayer perceptron to classify the game statistics and overview of this classification.

3.4.1 MERITS OF THE PROPOSED WORK

Multilayer perceptron poses many advantages over physically reading the reviews. Multilayer perceptron is not only for classification problems, where you want to just find out continuous prediction instead it provides a platform to obtain discrete outcome. It is easy to implement and it can be used as a performance baseline. It does not require too many computational resources; it is
highly interpretable.it does not require ant tuning. It is easy to regularize and its output well calibrated predicted probabilities. The classification are divided into home win, away win and draw. The percentage of home win, away win and draw is easily obtained and it is accessible by the developer's future development of a model.

5253
INPUTS
Data collection
Data cleaning
Corpus
FEATURE SELECTION
PHASE
Classification of features
Feature selection
Corpus
Training
Multilayer
perceptron
KNN
CNN
RNN
SVM
Decision Tree,
Naïve Bayes,
and logistic
regression
TEST
Performance Evaluation
OUTPUT
Classification
Classification
Home
Away Win
Draw**3.4.1 RELEVANCE OF THE PROPOSED MODEL**

Using neural network technique of supervised teeing to classification requires clearly stated

methods. In the proposed method as shown in the architectural framework as shown in figure 3.1, definite and established procedures would be required, understanding their various relevance is necessary and would be discussed below:

1. INPUT COMPONENT: The above structure illustrates the proposed classification of reviews online platforms. To accomplish our goal, we analyze a dataset of football statistics gotten from data.org.

I.

Data collection: The dataset required for the proposed model is obtained. It is necessary because it provides the quality information to enable and improve the process of the model. It measures information on variables of interest that is required for evaluation.

II.

Data cleaning. The proposed model intends to conduct an analysis on the dataset obtained but it is necessary to remove invalid data points from the dataset, such as extra spaces. Removing duplicates and deleting formatting. Data cleaning is important and necessary because it helps improve data quality, increase the overall productivity and performance of the model.

III.

Corpus: This is the databank of the preprocessed reviews, the result obtained after data cleaning.

2. FEATURE SELECTION

Feature selection is also known as attribute selection. It is the selection of attributes most relevant

to the proposed model. It helps in creating an accurate predictive model, features that would provide higher chances of accuracy and requiring less data.55

3. A TRAINING, TESTING AND VALIDATION

Deploying a supervised machine learning, the model which is all about making a program generalized the input samples that has never been encountered. The dataset is split into two parts: training set and test set. Training set is used to make the model examine data and learn from its mistake while Validation is to ensure that the model is to be trained and the periodically be evaluated. Test set corresponds to the final evaluation and enables the model to be tested for its Generalization.

4. OUTPUT

The required classification after the testing, validation and testing is the output of the model.

3.5 DESIGN APPROACH

In machine learning, classification is used to classify the given content into a precise set based on a training set of data containing observations whose category is known in advance. To visually represent use model process and requirements and for the purpose of clarity we would be employing the use of data flow diagram which is mainly in 2 levels which are given as 0-level and

1-level as given in fig 3.2 and fig 3.3 respectively

1. 0-LEVEL DATA FLOW DIGRAM: This is a context diagram. It is designed to be an abstraction view, displaying the system as a singular process with the relationship to external entity

in this case data.org showing the input, the required process and output.Figure 3.2 0-level Data Flow Diagram (DFD)

2. 1-LEVEL DATA FLOW DIAGRAM

In 1-level DFD, it shows more details on the processes. In this level the main functionality of the system and a breakdown of the high
level into sub-processes is given.
Fig3.3 shows 1- level data flow diagram
Data.org
DATA COLLECTION (CSV FORMAT)
DATA CLEANING
FEATURE SELECTION
CLASSIFIER TRAINING
CLASSIFICATION
HOME
AWAY
DRAW
FOOTBALL stat (dataset)
Python
librariesSplit Dataset (using
sklearn model
selection)
80% training
20% testing
DATA PREPROCESSING
Minmax scaling for preprocessing the
reviews attributes such that the values
ranges from 0 - 1
Training Dataset
Label
Home = 0
Away = 1
Draw = 2
Learn SVM, LSTM,
MLP, NAÏVE BAYES,
LOGISTIC
REGRESSION,
DECISION TREE, KNN
Output Classifier
Classification of the
dataset into home,
away and draw
Data collection
(CSV format)
Data cleaning using
python panda library
function

3.5.1 MULTILAYER PERCEPTRON

Multilayer perceptron is an artificial neural network technique for analyzing a dataset in which there are one or more independent variables that determine an outcome. The outcome is measured
with a three variables which there are only three possible outcomes.

1. Data (reviews) collection

To provide an exhaustive study of neural network algorithms, the experiment based on analyzing the sentiment value of the standard dataset. We use the original data set of the android applications
review to test our methods of reviews classification. The dataset is available and collected through
(football-data), and this dataset consists of length 1280, and are uniform in 551 home wins, 373 away wins and 356 draws.

Table 3.2 data count for proposed model

HOME WIN
551
AWAY WIN
373
DRAW
356
TOTAL
1280

1. DATA PARTIONING

Figure 3.4: Data partition (Training and test set).

2. Data partitioning also known as data splitting is an approach used to divide the entire dataset into subsets of data. In an ideal situation, more independent data sets can be collected or simply and inexpensively repeat an experiment to collect new ones. Also independent datasets can be used for learning, model selection and even an assessment of the prediction performance. In such situations there exist no reason to split any dataset. However, in situations when only one dataset is available and newer ones cannot be collected then there is need for some strategy to perform prediction task effectively based on the dataset collected. In this section, several data splitting strategies were reviewed and predominantly the chosen strategy for this work. The dataset building
our model in a regression task can be divided into subsets of train and test data. Machine-Learning
deals with program learning from datasets and as a result the training data is the data which the machine learning programs learn to perform correlational tasks (classify, cluster, learn the attributes) while the testing data is the data, whose outcome is already known and is used to determine the accuracy of the machine learning algorithm, based on the training data (how effectively the leaning happened Figure 3.1 shows how a single dataset is divided into instances of train and test set as given to the machine learning algorithm.

3. Feature selection

Feature selection is one of the most important technique before data partitioning. It involves the extraction of the most and unused columns. Moreover, this technique is useful because of the greater dimensionality of text features and existence of noisy features and selects an optimal set of
features and removes the irrelevant feature in order to improve the classification performance. In this work, filter method of feature selection and ranker search technique were used to assign ranks
to attributes retrieved after carrying out data preprocessing task.

5960

4. Performance Evaluations

After training, the next step is to predict the output of the model on the testing dataset and a confusion matrix generated which classifies the review as positive, neutral or negative alongside precision, recall and F1 score, where precision is the ratio; $(tp\ tp + fp)$
where tp is number of true positives and fp is number of false positives Precision is intuitively the
ability of the classifier not to label as away a prediction that is home, and recall is the ability of the
classifier to find all the home predictions. F1 score is weighted average of the precision and recall,
where an F1 score reaches its best value at 1 and worst score at 0. It can be calculated as

$$F1 = 2 * precision\ (precision + * recall\ recall)$$

(3.1)

A confusion matrix is a summary of prediction results on a classification problem. The number of
correct and incorrect predictions are summarized with count values and broken down by class

Table 3.2: A typical structure of a confusion matrix

Class 1 predicted
Class 2 predicted
Class 3 predicted
Class 1 actual TP
FN
FN
Class 2 actual
FP
TP
FN
Class 3 actual
FP
FN
TP61

The results involve the following attributes:

a) True positive: True Positive in the testing data, which are correctly classified by the model as positive (P)

b) False Positive: False Positive in the testing data, which are incorrectly classified by the model as Positive (P)

c) True Negative: True Negative in the testing data, which are correctly classified by the model as Negative (N)

d) False Negative: False Negative in the testing data, which are incorrectly classified by the model as Negative (N)

Using Gaussian Naïve Bayes, one was able to predict the possible outcomes using the classifier to generate the results. Taking the some of the features, betting odds, which contains WHCD, WHCA, PSCA etc. and the full time results. After training the model, the confusion matrix from Scikit learning and pass it to the test data: the true values first, then predictions.

The rows represent the true label and the columns shows the predicted labels. Class sorts both rows and columns, so the first row shows all the samples that have a true label of draw, the middle row shows the label of wins and the last row show all the class that have a true label of

loss. Of course, it is quite hard to interpret this matrix in this raw format. So typically, one would plot it as a bar chart. This can be found in chapter four of this study.**CHAPTER FOUR**

IMPLEMENTATION AND DOCUMENTATION

4.1 OVERVIEW

System implementation refers to the construction and testing of the model itself. This section's goals are to construct the suggested model, record the procedure, and provide the findings.

4.2 SYSTEM REQUIREMENTS

System requirements are the conditions that a system must meet in order to operate smoothly, effectively, efficiently, and predictably, with the knowledge that doing so could affect performance
or result in problems.

4.2.1 HARDWARE REQUIREMENTS

To run the model conveniently, a decent laptop with the following hardware is required:

a. Processor: Intel 13, 15, 17, and above

b. RAM: 12GB or more

c. Memory: above 120GB of space

d. Processor Speed: 2.2GHZ and above

e. GPU: Because the model uses several cores for processing during its training stages, a GPU is necessary to increase the model's performance.

f. TPU: Because the model uses several cores for processing during its training stages, a TPU is necessary to increase the model's performance.

g. Internet: A stable internet connection from a reliable Internet Service Provider (ISP) because the software used 1s required to run online.

624.2.2 SOFTWARE REQUIREMENTS

A computer's software is an intangible Component of the computer. To run the model, the following software is required:

a. Operating System: either a Windows 7, 8, 10, or Linux operating system should be sufficient but Linux is recommended,

b. Integrated Development Environment (IDE): Google Colaboratory (Google Colab)

4.3 TOOLS FOR MODEL DEVELOPMENT

Software developers use these programs to create, debug, and maintain other programs and Applications. The next section lists the most important software tools utilized in the creation of The proposed model.

4.3.1 PROGRAMMING LANGUAGES USED

In the development of a model, the decision of which programming language to use is critical Since it helps the programmer to convey his or her ideas in a convenient manner.

A. Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built-in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a Scripting or glue language to connect existing components together. Python is simple, easy to learn
syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form. The proposed model utilizes some of python’s extensive library, which includes:

6364
I.
Numpy is a python library that allows provides its users with an array-processing package. It allows for the use of high-performance multidimensional array object.
II.
Pandas is used as a data manipulation tool that makes use of a data frame in storing its data in a tabular form.
III.
Seaborn is used for data visualization. It is a high-level interface based on matplotlib (another popular data Visualization library). It can be used in all sorts of data analysis tasks that require visualization of data and inferring information from it.
IV.
Matplotlib is a library used in visualizing data by providing the ability to create graph plots.

B. Scikit-Learn

Another powerful library used in the proposed model is Scikit-Learn. Scikit-learn is a library In Python, that provides many unsupervised and supervised learning algorithms. It is built Upon some of the familiar technologies like NumPy, pandas, and Matplotlib. The functionality That Scikit-learn provides includes:
a. Regression, including Linear and Logistic Regression
b. Classification, including K-Nearest Neighbors
C. Clustering, including K-Means and K-Means++
d. Model selection
e. Preprocessing, including Min-Max Normalization

4.4 SYSTEM TESTING

The effectiveness of the model is evaluated through system testing. Model performance testing often entails applying the models to the test dataset and assessing the results in terms of recall, accuracy, and precision.**4.5 RESULTS**

This section includes images of some of the model's intriguing feature graphs, as well as the suggested model's outcome at the end.

4.5.1 RAW DATASET

This is the available dataset obtained from the data collection phase (football-data.co.uk).The dataset contains some columns related to betting statistics and it is stored in a comma separated values (CSV) format.

4.5.2 PREPROCESSED DATA

The image table 4.1 shows the output derived from a pre-processed data, which has been normalize, the main aim of reducing and possible elimination of redundancy in the data and face the large variance of data.
Table 4.1 Preprocessed Data Ath. Madrid 2019
65Table 4.2 Preprocessed Data for Ath. Madrid 2020
Table 4.3 Preprocessed Data Barcelona 2019
66Table 4.4 Preprocessed Data Barcelona 2020
Table 4.5 Preprocessed Data Celta 2019
67Table 4.6 Preprocessed Data Celta 2020
Table 4.7 Preprocessed Data Real Madrid 2019
68Table 4.8 Preprocessed Data Real Madrid 2020
Table 4.9 Preprocessed Data Sevilla 2019

69Table 4.10 Preprocessed Data Sevilla 2020
Table 4.11 Preprocessed Data Valencia 2019
7071
Table 4.12 Preprocessed Data Valencia 2020

4.7 SPLITING, LABELLING OF DATASET AND PERFORMANCE EVALUATION

4.7.1 CLASSIFICATION ALGORITHM

As earlier stated in chapter 3, we would be splitting the available dataset into 80% training and 20% testing. The main aim of the training data is to ensure the system learns on this data since the
output is known in order to generalize on other data later on. While the test data is used to evaluate
the performance of the model using some performance metrics accuracy, specificity and sensitivity.

4.7.2 PERFORMANCE EVALUATION TABLE

MODELS
ACCURACY VALUES
SPECIFICITY
SENSITIVITY
PRECISION
FSCORE
SVM
0.569554
0.527288
0.565385
0.563218
0.773895
DECISION TREE
0.506562
0.462184
0.607692
0.478788
0.740089
MULTILAYER
PERCEPTRON
0.611549
0.563242
0.6
0.586466
0.792558
KNN
0.598425
0.567106
0.584615
0.575758
0.790823
NAÏVE BAYES
0.543307
0.549809
0.55
0.510714
0.778683
LOGISTIC REGRESSION
0.619423
0.577049
0.580769
0.599206
0.801319

4.8 MODEL EVALUATION

I evaluated my model using a performance evaluation table to get the highest accuracy and the highest precision in other to calculate the accuracy, recall and precision as shown below:
Highest Accuracy (x) = Precision value x 100%
Highest Accuracy (x) = 0.619423 x 100% = 61.94%
ERROR (X) = 1 – ACCURACY (X)

ERROR (X) = 1 – 0.619423 = 0.380577 = 38%
This model has a precision of 0.619, which means that when the model predicts the result is correct
probably 61.9% of the time.

4.9 ANALYSIS OF THE FACTORS AFFECTING FOOTBALL PREDICTION

How much of a factor is home advantage?

In the Spanish league season, each team plays 19 home games and 19 away games. According to our analysis, teams playing at home had a better chance of winning than teams playing away.

Steps Taken:

1. All the data of the past 12 years was combined and grouped according to the Full Time Result (FTR).
2. The length of each group was evaluated to get Home wins, Away wins and Draws.

Results: The following table and graph illustrates the number of Home wins, Away wins and Draws in a particular Spanish season.

7374

Table 4.2 Game Statistics

Fig 4.1 Graphical illustration75

Table 4.3 Game Aggregate

Home Wins	**Away Wins**	**Draws**	**Overall**
551	373	356	

Fig 4.2 and Fig 4.3 illustrates the total number of home wins, away wins and draws in the past 3 years. The Aggregate win percentage for the home side as well as the Away side has been represented as a pie chart in Fig 4.2 and bar chart in Fig 4.3 respectively.

Fig 4.2 Pie Chart

Fig 4.3 Bar chart**Conclusions:**

The information above makes it evident that during each season of the Spanish football league, the
home team has won more games than the visiting team. These factors could be to blame for this
a) Football is a team sport, and crowd energy has an impact on the team's overall mood. When a team is playing at home, the audience supports them and this improves their performance.
b) The host team may perform better due to familiarity with the field and the surrounding weather.
c) Home teams do not have to travel far to get to the stadiums, whereas visiting teams might have
to travel the full length and breadth of the nation, wearing them out.

Finally:

1. Home teams have a definite advantage over Away teams. On aggregate, home team win 46.65%
matches compared to 27.72% matches won by the away teams.
2. Using the different algorithms and methods to compare the data and results which showed that

logistic regression algorithm was better in predicting accurate result with an 80% mark. Although,
these predictions should not be over looked because the Spanish league can be upsetting where any team is capable of defeating any other team.

76**CHAPTER FIVE**

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 SUMMARY

The aim of this project was to develop and design a predictive model using machine learning approaches and deep learning methods for sports prediction. This was done by taking the Spanish
football League as a case study, using the Feature selection technique and LSTM, KNN, SVM, LOGISTIC REGRESSION, NAÏVE BAYES and DECISION TREE as the algorithm for the model.

The prediction model was developed after datasets were extracted online and preprocessed, trained
and tested with the preprocessed data

The model was implemented using Python with its corresponding libraries such as SK Learn, Numpy etc. The model had an accuracy of 61.94% after testing it.

The chapters give briefs of how the design was carried out, the structure program and the prediction
analysis.

5.2 CONCLUSION

In conclusion, the project's aim was achieved, which was design of a predictive model. The model
was tested with an accuracy of 61.94%. The accuracy is good, seeing that existing works hardly use the features used in this project such as Home Team Yellow Card (HY) and Away Team Yellow Card (AY).

7778

5.3 RECOMMENDATION

Since the accuracy of the prediction was only 61.94%, there is still much work to be done to increase its accuracy. The types of features selected go a long way in making the model more accurate. I recommend that the algorithms chosen must be what would affect the predicting accuracy and more algorithms should be added to get a higher prediction.**REFERENCE**

"AlphaGo – Google DeepMind". https://wildoftech.com/alphago-google-deepmind/ Archived from the original on 20 October 2021.

, https://www.javatpoint.com/artificial-neural-network

“A simplest introduction to Support Vector Machine,” 07-Aug-2014” [Online]. Available:https://diggdata.in/post/94066544971/support-vector-machine- without-tears.

“Artificial neutral network,” Analytics Vidhya, javatpoint, [Online]. Available: https://www.analyticsvidhya.com/blog/2021/09/introduction-to-artificial-neural-networks/

“Deep neural network or artificial neural network,” Viso.ai, 2022[online]. Available: https://viso.ai/deep-learning/deep-neural-network-three-popular-types/

“Feature selection,” simpliLearn [online]: https://www.simplilearn.com/tutorials/machine learning-tutorial/feature-selection-in-machine-learning#:~:text=ISTRegister%20Now-,What%20is%20Feature%20Selection%3F,you%20are%20trying%20to%20solve.

“K-nearest neighbors algorithm,” Wikipedia, 21-Jul-2019. [Online]. Available:

https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm.
A. Cutler, D. R. Cutler, and J. R. Stevens, “Ensemble Machine Learning,” Ensemble Mach. Learn.,
no. January, 2012, doi: 10.1007/978-1-4419-9326-7. [1]
Ashok83 (10 September 2019). "How AI Is Getting Groundbreaking Changes In Talent Management And HR Tech". Hackernoon. https://hackernoon.com/how-ai-is-getting groundbreaking-changes-in-talent-management-and-hr-tech-d24ty3zzd Archived from the original on 11 September 2019.
7980
Brynjolfsson, Erik; Mitchell, Tom (22 December 2017). "What can machine learning do? Workforce implications". Science. 358 (6370): 1530–1534. Bib code: 2017Sci...358.1530B. doi:10.1126/science.aap8062. PMID 29269459. S2CID 4036151.
C. Cortes and V. Vapnik, “Support-Vector Networks,” Machine Learning. [Online]. Available: https://dl.acm.org/citation.cfm?id=218929.
C. J. C. H. Watkins, “Learning from delayed rewards,” Master’s thesis, 1989.
C. M. Bishop, “Pattern Recognition and Machine Learning,” Springer-Verlag New York, 2016.
C. M. F. Che Mohd Rosli, M. Z. Saringat, N. Razali, and A. Mustapha, “A Comparative Study of Data Mining Techniques on Football Match Prediction,” J. Phys. Conf. Ser., vol. 1020, no. 1, 2018,
doi: 10.1088/1742-6596/1020/1/012003.
Clark, Jack (2015b). "Why 2015 Was a Breakthrough Year in Artificial Intelligence". Bloomberg.com. Archived from the original on 23 November 2016.
Copeland, B. (2022, November 11). Artificial intelligence. Encyclopedia Britannica. https://www.britannica.com/technology/artificial-intelligence
Crevier, Daniel (1993, pp. 161–162, 197–203, 211, 240). AI: The Tumultuous Search for Artificial
Intelligence. New York, NY: Basic Books. ISBN 0-465-02997-3.
Dhar, V., (2013). Data Science and prediction. Communications of the ACM, 56, 12 PP. 64-73
E. Alpaydin, “Introduction to machine learning,” Cambridge, MA: The MIT Press, 2014.
El Naqa I., Murphy M.J, “What Is Machine Learning?” Machine Learning in Radiation Oncology.
Springer, Cham, 2015F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in
the brain.” Psychological Review, vol. 65, no. 6, pp. 386–408, 1958.
Figure 2.1 and 2.2 [online] Available:
https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf
Football-data (2022) [Online] Available: https://www.football-data.co.uk/spainm.php [Accessed 8 December 2022].
G. Tesauro, “Temporal difference learning and TD-Gammon,” Communications of the ACM, vol.
38, no. 3, pp. 58–68, 1995.
Highlights, M., Cup, F., Kits, F., Kits, 2., 1, F., Updates, F., Updates, M., Open, A., Open, F., Open, U., 2019, S., Cowboys, D., Eagles, P., Rams, L., United, M., Barcelona, F., Madrid, R., Saint-Germain, P., Munich, B., Milan, A., Roma, A., City, L., United, N., Money, S., Sports, O., SPORTS, M., 2018, F., Warriors, G., Lakers, L. and Cavaliers, C. (2019). 25 World's Most Popular

Sports (Ranked by 13 factors). [online] TOTAL SPORTEK. Available at: https://www.marca.com/en/world-cup/2022/12/07/639087e922601df07f8b45dd.html (MARCA, 2022)
https://www.totalsportek.com/most-popular-sports/ [Accessed 7 December 2022].
https://www.worldatlas.com/articles/what-are-the-most-popular-sports-in-the-world.html.
IBM cloud education; https://www.ibm.com/cloud/learn/what-is-artificial-intelligence
J. Han, M. Kamber, and J. Pei, "Data mining: concepts and techniques," Amsterdam: Morgan Kaufmann, 2012.
81J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," PNAS, 01-Apr-1982. [Online]. Available: https://www.pnas.org/content/79/8/2554.abstract.
Joachims. T, "Learning to Classify Text Using Support Vector Machines," Springer. [Online]. Available: https://www.springer.com/la/book/9780792376798.
K. Chomboon, P. Chujai, P. Teerarassammee, K. Kerdprasop, and N. Kerdprasop, "An Empirical Study of Distance Metrics for k-Nearest Neighbor Algorithm," The Proceedings of the 2nd International Conference on Industrial Application Engineering 2015, Jan. 2015.
K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern
recognition unaffected by shift in position," Biological Cybernetics, vol. 36, no. 4, pp. 193–202, 1980.
LawInSport, (August 2022) [online] available: https://www.lawinsport.com/topics/item/artificial
intelligence-in-football-current-uses-contracting-tips-for-clubs
M. Pelillo, "Alhazen and the nearest neighbor rule," Pattern Recognition Letters," vol. 38, pp. 34–
37, 2014.
M. Turing, "Computing Machinery and Intelligence," Parsing the Turing Test, pp. 23–65, 2009.
McCorduck, Pamela (2004), Machines Who Think (2nd ed.), Natick, MA: A. K. Peters, Ltd., ISBN
1-56881-205-1.
Moroney M. (1956). Facts from figures, 3rd edition. Penguin: London.
Muller, A. and Guido, S. (2018). Introduction to Machine Learning with Python. O'Reilly Media.
8283
Nervous activity," Springer Link. [Online]. Available: https://link.springer
.com/article/10.1007/BF02478259.
Newquist, HP (1994). The Brain Makers: Genius, Ego, and Greed in the Quest for Machines That
Think. New York: Macmillan/SAMS. ISBN 978-0-672-30412-5.
Nytimes.com. (2019). How Data (and Some Breathtaking Soccer) Brought Liverpool to the Cusp of Glory. [online] Available at: https://www.nytimes.com/2019/05/22/magazine/soccer dataliverpool. html [Accessed 7 December 2022].
R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification," Wiley- Interscience, 2000.
Roberts, Jacob (2016). "Thinking Machines: The Search for Artificial Intelligence". Distillations. Vol. 2, no. 2. pp. 14–23.
Russell, Stuart J.; Norvig, Peter (2003), Artificial Intelligence: A Modern Approach (2nd ed.),

Upper Saddle River, New Jersey: Prentice Hall, ISBN 0-13-790395-2
S. Sayad, “K nearest Neighbors - Classification,” KNN Classification. [Online]. Available: https://www.saedsayad.com/k_nearest_neighbors.htm.
Sawe, B. (2019). The Most Popular Sports in the World. [online] WorldAtlas. Available at:
Schank, Roger C. (1991). "Where's the AI". AI magazine. Vol. 12, no. 4.
Stanford University- HAI pdf - https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions
HAI.pdfT. K. Ho, "Random decision forests," Proceedings of 3rd International Conference on Document
Analysis and Recognition, Montreal, Quebec, Canada, 1995, pp.278-282vol.1. doi: 10.1109/ICDAR.1995.598994
V. N. Vapnik, “Statistical learning theory,” New York: Wiley, 1998.
W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in
84