CLASSIFICATION OF OUTCOMES OF ENGLISH PREMIER LEAGUE MATCHES USING MACHINE LEARNING MODELS

BY

IYIOLA, TOMILAYO PROMISE

(20PCD02190)

AUGUST, 2022

CLASSIFICATION OF OUTCOMES OF ENGLISH PREMIER LEAGUE MATCHES USING MACHINE LEARNING MODELS

BY

IYIOLA TOMILAYO PROMISE

(20PCD02190)

B.Sc Mathematics, Bowen University, Iwo.

A DISSERTATION SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE MASTER OF SCIENCE (M.Sc.) DEGREE IN INDUSTRIAL MATHEMATICS IN THE DEPARTMENT OF MATHEMATICS, COLLEGE OF SCIENCE AND TECHNOLOGY, COVENANT UNIVERSITY.

AUGUST, 2022

ACCEPTANCE

This is to attest that this dissertation is accepted in partial fulfilment of the requirements for the award of the degree of Master of Science in Industrial Mathematics in the Department of Mathematics, College of Science and Technology, Covenant University, Ota, Nigeria.

Mr. Taiwo B. Erewunmi
(Secretary, School of Postgraduate Studies) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DECLARATION

I, IYIOLA, TOMILAYO PROMISE (20PCD02190) declare that this research was carried out by me under the supervision of Dr. Hilary I. Okagbue of the Department of Mathematics, College of Science and Technology, Covenant University, Ota, Nigeria. I attest that the dissertation has not been presented either wholly or partially for the award of any degree elsewhere. All sources of data and scholarly information used in this dissertation are duly acknowledged.

IYIOLA, TOMILAYO PROMISE

(Student) **Signature and Date**

CERTIFICATION

We certify that this dissertation titled **"CLASSIFICATION OF OUTCOMES OF ENGLISH PREMIER LEAGUE MATCHES USING MACHINE LEARNING MODELS"** is an original research work carried out by **IYIOLA, TOMILAYO PROMISE (20PCD02190)** in the Department of Mathematics, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria under the supervision of Dr. H. I. Okagbue. We have examined and found this work acceptable as part of the requirements for the award of Master of Science in Industrial Mathematics.

Dr. Hilary I. Okagbue
(Supervisor) **Signature and Date**

Prof. Samuel A. Iyase
(Head of Department) **Signature and Date**

Prof. Kayode Ayinde
(External Examiner) **Signature and Date**

Prof. Akan B. Williams
(Dean, School of Postgraduate Studies) **Signature and Date**

DEDICATION

To the Almighty God for grace, opportunity, and divine wisdom, To my lovely parents and siblings.

ACKNOWLEDGEMENTS

I want to give glory and praise to the Almighty God for helping me through the course of this programme.

I appreciate the Chancellor, Dr. D. O. Oyedepo, for establishing this great institute of learning and providing an enabling environment for research.

I sincerely appreciate Prof. S. A. Iyase, Head of the Department of Mathematics, Covenant University, for his words of encouragement and wise counsel all through the programme.

My appreciation also goes to the post-graduate coordinator of the Department of Mathematics, Covenant University, Dr. S. O. Edeki for his contributions and guidance throughout this research work.

I acknowledge the intellectual tutelage of my supervisor, Dr. H. I. Okagbue and his immense support and guidance towards the commencement and completion of this dissertation.

I extend my appreciation to the post-graduate committee of the Department of Mathematics, Covenant University, Prof. S. A. Iyase, Prof. T. A Anake, Prof. A. M. Okedoye, Dr. M. C. Agarana, Dr. S. O. Edeki, Dr. A. A. Opanuga, Dr. J. G. Oghonyon, Dr. O. O. Agboola, Dr. O. A. Odetunmibi, Dr. G. O. Alao, Dr. H. I. Okagbue, Dr. A. F. Adedotun, and Dr. O. F. Imaga, for their contributions to the success of this work in one way or the other.

I am also thankful to the administrative officer of the Department of Mathematics, Covenant University, Mrs. T. A. Ajayi for her assistance during this research work. I will also like to appreciate the entire faculty and staff of the Department of Mathematics. My appreciation also goes to my wonderful course-mate, Lekan Igbekele, for his contribution to this academic journey.

The appreciation will not end without acknowledging my lovely parents, Dr. O. A. Iyiola and Mrs. M. A. Iyiola for their ever-present love, care, prayers, support and assistance, may God continue to bless and keep them in Jesus' name. I also extend my appreciation to my lovely siblings, Olumayowa Aina (Nee Iyiola), Ayobamiji Iyiola, and Ayanfeoluwa Iyiola, for their support, assistance and words of encouragement

during the course of this research work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

$\bar{x}$:. Arithmetic mean.

μ or μ_x :. Population mean.

$\sum$: Summation.

R_i : Sum rank of the i_{th} group.

τ : Threshold.

X : Discrete random variable

$N(t_i)$: No of times that $t(i)$ occurs.

$N(t)$: Total no of counts.

t : Set of events.

$F_T(x)$: Boosted classifier.

$F_{t-1}(x)$: Previous stage trained boosted classifier.

IG : Information gain.

N : No of data points in a data set.

y_i : Real value of y

$\emptyset$: Empty set.

Ω : Nonempty/sample space.

LR :Logistic regression.

ML : Machine learning.

AB : Adaptive boosting.

GB : Gradient boosting.

CA : Classification accuracy.

AUC : Area under curve.

ANN : Artificial neural network.

SVM : Support vector machine

ABSTRACT

Football remains an important sport in the world and it has a lot of followers. Researchers are often interested in the analysis of the results of football matches, which helps in the prediction or classification of outcomes (results) of football matches based on some variables. Most of the available models of prediction and classification of outcomes are based on a selected variable or a large number of variables. The use of a few variables can not predict accurately and the use of large variables leads to the problem of interpretation (Parsimony). This work used feature selection methods to reduce sixteen selected independent variables (football related) to six variables in the classification of the outcome variable (home win, away win, and draw) of five seasons of English premier league matches. As expected, a home win is a modal observation in all five seasons. The Kruskal Wallis test showed that the median outcome was not the same for the five seasons, while four machine learning models classified the outcome using the six best variables recommended via the feature selection. Furthermore, the result of the first half and second half was used to classify the final outcome. Five performance metrics attest that the ML models are good in the classification. Cross-Validation ensured that the issues of over-fitting were adequately addressed. Bookmakers may find this research interesting as some variables were identified as key to the classification of outcomes of football matches.

Keywords: Algorithm, Classification, Cross-Validation, English Premier League, Feature Selection, Frequency, Machine Learning, Kruskal Wallis.