

PREDICTING EMPLOYEE RETENTION BASED ON SOCIO-ECONOMIC FACTORS: A CASE OF AN SRI LANKAN APPAREL ORGANIZATION

H.M.D.A.B Attanayake

(179075T)

Degree of Master of Science in Business Statistics

Department of Mathematics

University of Moratuwa,
Sri Lanka

March 2023

PREDICTING EMPLOYEE RETENTION BASED ON SOCIO-ECONOMIC FACTORS: A CASE OF AN SRI LANKAN APPAREL ORGANIZATION

H M D A B Attanayake

(179075T)

Dissertation submitted in partial fulfilment of the requirements for the Master of Science in
Business Statistics

Department of Mathematics

University of Moratuwa
Sri Lanka

March 2023

Declaration

I declare that this is my own work, and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 3/10/2023

The above candidate has carried out research for the Masters Dissertation under my supervision.

Supervisor: Dr. A.A Sunethra
Senior Lecturer,
Department of Statistics,
University of Colombo.

Signature:

Date: 20-03-2023

Abstract

Employee attrition is considered as one of the biggest challenges faced by the apparel manufacturing organizations in Sri Lanka and it's an industry which highly depends on its human workforce. Nearly 60% of the terminated employees are terminated during the first six months of their time at the organization considered in this study which makes it difficult to ensure a smooth production process. Currently a scientific method is not in place to identify the first six months termination risk associated with the employees in advance and if so, the management can take necessary actions to mitigate those risk factors. This study is aimed to build a predictive model for one of the leading apparel manufacturing organizations in Sri Lanka to identify the first six months retention risk of the newly recruits one month after recruitment based on socio-demographic factors and their earnings. Data obtained through the employee interview process and payroll process were used for this study and a scalable end-to-end machine learning tree boosting algorithm called XG-Boost was used to build the predictive model. The built model went through a parameter tuning process to improve its accuracy and after that the accuracy of the pruned model is around 86% and the model suggested that employee earnings, BMI (Body Mass Index), Education level, Age, Family opinion on the job are the most prominent factors affecting employee decision to stay in the organization for more than six months or not. The developed model has been implemented at one of the manufacturing facilities in the organization, and insights given by the model on each employee's retention status being shared with all stakeholders through a dashboard.

Acknowledgement

I would like to extend my sincere gratitude to my supervisor, Dr. A.A. Sunethra for the immense support given by contributing her valuable time to discuss on the progress of the research and by pointing out areas of further improvement. She personally went through each of my work to correct and notify me if any necessary actions had to be taken. It was a great privilege to have had such a dedicated supervisor to make this research a success.

I would also like to take this an opportunity to thank to the course coordinator of the MSc. in Business Statistics degree program, Ms. Ravindi Jayasundara for her consistent support and guidance given throughout the period which made it possible to complete and make this research a success.

I would also like to thank all the lecturers of the Department of Mathematics at University of Moratuwa for the guidance given throughout my period at the university

Table of Contents

Declaration.....	i
Abstract.....	ii
Acknowledgement	iii
Table of Contents	iv
List of Figures.....	vii
List of Tables	viii
Chapter 1 Introduction.....	9
1.1 Background of the study.....	9
1.1.1 Expectation Mismatch.....	10
1.1.2 Variability in Income	10
1.1.3 Lack of Career Progression.....	11
1.2 Significance of the study scope	12
1.3 Significance of the study to the apparel sector.....	12
1.4 Objectives	14
1.5 Organization of the thesis.....	14
Chapter 2 Literature Review	15
2.1 Introduction to Employee Attrition	15
2.2 Impact of Employee Attrition on Company Performance, Productivity and Profitability	15
2.3 Prediction Models for Employee Attrition.....	16
2.4 Prediction Models for Employee Attrition in the Apparel Industry.....	17
2.5 Employee Turnover Prediction Recent Research Trends	18
2.6 Limitations of Employee Turnover Predictive Algorithms.....	19

2.7	Significance of the Study	19
Chapter 3 Theory and Methodology		21
3.1	Phases of the Study	21
3.2	Data Preparation	22
3.3	Model Specification	24
3.3.1	Working Model of XG-Boost Algorithm.....	25
3.3.2	Working Model of Logistic Regression Algorithm	28
3.3.2.1	Logistic Regression Algorithm Objective Function	28
3.4	Model Fitting	29
3.4.1	Tools and Platforms	29
3.4.2	Train & Test Data Set Split.....	30
3.5	Model Evaluation	30
3.6	Model Improvement	31
3.6.1	Hyperparameter Tuning using Grid Search	32
3.6.2	Hyperparameter Tuning in XG-Boost Algorithm.....	32
3.7	Prediction	34
Chapter 4 Preliminary Analysis		35
4.1	Characteristics of the Categorical Variables	35
4.2	Characteristics of the Continuous Variables	40
Chapter 5 Advanced Analysis Results		43
5.1	Model Accuracy	43
5.2	Hyper Parameter Values.....	43
5.3	Confusion Matrix of the Trained Model	44
5.4	Feature Importance of the Trained Model.....	45
Chapter 6 Discussion		48
6.1	Summary of Findings.....	48
6.2	Model Adaptation.....	48

6.3	Future Work	50
	References	51
	Appendix.....	54

List of Figures

Figure 3.1: XG-Boost Algorithm Function.....	26
Figure 3.2: Confusion Matrix	31
Figure 4.1: Histogram for Survival Status vs Gender.....	36
Figure 4.2: Histogram for Survival Status by Residence.....	36
Figure 4.3: Histogram for Survival Status by Previous Job.....	36
Figure 4.4: Histogram for Survival Status by Civil Status	37
Figure 4.5: Histogram for Survival Status by Family Opinion on Job	37
Figure 4.6: Histogram for Survival Status by Distance to Residence.....	38
Figure 4.7: Histogram for Survival Status by Job Expectation	38
Figure 4.8: Histogram for Survival Status by Referral Method.....	39
Figure 4.9: Histogram for Survival Status by Contribution to Family	39
Figure 4.10: Histogram for Survival Status by Reason to Choose Apparel	39
Figure 4.11: Histogram for Survival Status by Spouse Occupation	40
Figure 4.12: Box Plot for Retention Status vs Youngest Child Age	41
Figure 4.13: Box Plot for Retention Status vs No of Children	41
Figure 4.14: Box Plot for Retention Status vs No of Children (Less than 5 Years).....	41
Figure 4.15: Box Plot for Retention Status vs 2nd Month Gross Salary	41
Figure 4.16: Box Plot for Retention Status vs Age.....	41
Figure 4.17: Box Plot for Retention Status vs Experience (Years)	41
Figure 4.18: Box Plot for Retention Status vs BMI.....	41
Figure 5.1: Confusion Matrix	44
Figure 6.1: Digitized Application	49
Figure 6.2: Power BI Dashboard	50

List of Tables

Table 5.1: Hyperparameter Values	43
Table 5.2: Model Performance Indicators	44