Table of Contents

Abstract .............................................................................................................................. 00

Acknowledgment ................................................................................................................ 00

Declaration .......................................................................................................................... 00

CHAPTER I – INTRODUCTION .............................................................................................. 00

1.1 Background ................................................................................................................... 00

1.2 Problem Statement ....................................................................................................... 00

1.3 Research aim ................................................................................................................. 00

1.4 Research Questions ....................................................................................................... 00

1.5 Research Objectives ...................................................................................................... 00

1.6 Project Plan ................................................................................................................... 00

CHAPTER 2: LITERATURE REVIEW ....................................................................................... 00

2.1 Introduction .................................................................................................................. 00

2.1.1 Hotel Reservation Cancellation ......................................................................... 00

2.1.2 Possible Reasons for Reservation Cancellation …................................................ 00

2.1.3 Challenges Due to Hotel Reservation Cancellation ............................................. 00

2.2 What is One Hot Encoding ……………………....................................................................... 00

2.3 Data Imbalance ………………………………………………............................................................. 00

2.3.1 How Data imbalance will impact the performance of model …………………………… 00

2.3.2 Solutions for Data Imbalance ………………………………………………………………………….. 00

2.3.2.1 Over Sampling ……………………………….………………………………………………………. 00

2.3.2.2 Under Sampling ……………………………………………………………………………………… 00

2.4 Standardization …………………………………………………………………………………………………………. 00

2.4.1 What is Standardization …………………………………………………………………………………… 00

2.4.2 Standardization Procedure ………………………………………………………………………………. 00

2.4.3 Benefits of Standardization ……………………………………………………………………………… 00

2.5 Grid Search Cross Validation ……………………………………………………………………………………… 00

2.5.1 What is Grid Search Cross Validation ………………..……………………………………………… 00

2.5.2 Operation of Grid Search Cross Validation ……………………………………………………….. 00

2.5.3 Advantages and Disadvantages of Grid Search CV ……………………………………………. 00

2.6 Overview of Machine Learning ....................................................................................... 00

2.6.1 Types of Machine Learning ................................................................................... 00

2.6.1.1 Supervised Learning ………............................................................................ 00

2.6.1.2 Unsupervised Learning ................................................................................ 00

2.6.1.3 Reinforcement Learning ………………………………………………………………………….. 00

2.7 Classification Algorithms ………………….….…....................................................................... 00

2.7.1 Logistic Regression ……………………………………………………………………..…..………………… 00

2.7.2 K-Nearest Neighbours ............................................................................................ 00

2.7.3 Decision Trees …..................................................................................................... 00

2.7.4 Random Forest ....................................................................................................... 00

2.7.4.1 Random Forest Feature Importance ……………………………………………………… 00

2.7.5 Challenges of Classification Algorithms …............................................................... 00

2.7.6 Other Classification Algorithms ………….................................................................. 00

2.8 Performance Metrics ........................................................................................................ 00

2.8.1 Accuracy ……............................................................................................................ 00

2.8.2 Confusion Matrix .................................................................................................... 00

2.9 Research Gap .................................................................................................................... 00

CHAPTER 3: RESEARCH METHODOLOGY ................................................................................. 00

3.1 Introduction ........................................................................................................................ 00

3.2 Hotel Reservations Dataset ……........................................................................................... 00

3.3 Data Pre-processing ............................................................................................................ 00

3.3.1 Data Loading ............................................................................................................. 00

3.3.2 Data Cleaning ............................................................................................................ 00

3.4 One-Hot Encoding ………………………………………………………………………………………………………….. 00

3.5 Data Visualization ………….………………………………………………………………………………………………. 00

3.6 Training and Testing Data Separation …………………………………………………………………………….. 00

3.7 Feature Standardization …………………................................................................................... 00

3.8 Oversampling using SMOTE ……………………………………………………………………………………………. 00

3.9 Grid Search on ML Algorithms on Over-Sampled Data ……………………………………………………. 00

3.10 Model Building with Best Params of Algorithms on Over-Sampled Data .......................... 00

3.11 Grid Search on ML Algorithms on Under-Sampled Data ….……………………………………………. 00

3.12 Model Building with Best Params of Algorithms on Under-Sampled Data ........................ 00

3.13 Performance Metrics …....................................................................................................... 00

3.13.1 Accuracy ................................................................................................................... 00

3.13.2 Confusion Matrix ….…………………………………………………………………………………………….. 00

3.14 Summary ............................................................................................................................ 00

CHAPTER 4: RESULTS AND DISCUSSION .................................................................................... 00

4.1 Experimentation setup ........................................................................................................ 00

4.2 Data Pre – Processing ......................................................................................................... 00

4.2.1 Checking for missing value ........................................................................................ 00

4.2.2 Data Separation Results ............................................................................................ 00

4.3 Data Visualization Results ………………………………………………………………………………………………. 00

4.4 Machine Algorithms Results on Over-Sampled Data ……..................................................... 00

4.4.1 Logistic Regression ……………………………………………………………………………………………….. 00

4.4.2 K-Nearest Neighbours …………………………………………………………………………………………… 00

4.4.3 Decision Tree ……………………………………….................................................................... 00

4.4.4 Random Forest ……………………………………………………………………………………………………… 00

4.4.4.1 Feature Importance Results …………................................................................. 00

4.4.5 Combined Results of All Algorithms on Over-Sampled Data …………………………………. 00

4.5 Machine Algorithms Results on Under-Sampled Data …..................................................... 00

4.5.1 Logistic Regression ……………………………………………………………………………………………….. 00

4.5.2 K-Nearest Neighbours …………………………………………………………………………………………… 00

4.5.3 Decision Tree ……………………………………….................................................................... 00

4.5.4 Random Forest ……………………………………………………………………………………………………… 00

4.5.4.1 Feature Importance Results …………................................................................. 00

4.5.5 Combined Results of All Algorithms on Under-Sampled Data ………………………………. 00

4.6 Comparison of Algorithms Results on Over-Sampled and Under-Sampled Data ................ 00

CHAPTER 5: CONCLUSION AND FUTURE WORK ........................................................................ 00

REFERENCES .............................................................................................................................. 00

APPENDICIES ............................................................................................................................. 00