

Table of Contents

PERNYATAAN KEASLIAN KARYA TUGAS AKHIR	ii
PERSETUJUAN DOSEN PEMBIMBING TUGAS AKHIR	iii
PERSETUJUAN TIM PENGUJI TUGAS AKHIR.....	iv
ABSTRACT.....	v
ABSTRAK.....	vi
ACKNOWLEDGEMENTS	vii
Table of Contents	ix
Table of Figures	xii
Table of Tables.....	xiv
Chapter 1: INTRODUCTION.....	1
1.1 Background	1
1.2 Problem Formulation	2
1.3 Problem Limitation	2
1.4 Research Purposes.....	3
1.5 Methodology	3
1.6 Writing System	3
Chapter 2: UNDERLYING THEORY.....	5
2.1 Customer Churn	5

2.2 Data Mining	6
2.3 Classification Algorithm	9
2.4 Machine Learning Models	10
2.4.1 Decision Tree	11
2.4.2 Random Forest Classifier	13
2.4.3 XGBoost Classifier	14
Chapter 3: SYSTEM DESIGN.....	17
3.1 Data Source and Type	17
3.2 Calculating Churn	19
3.3 Research Variable	19
3.4 Data Analysis Method.....	20
3.5 Research Steps	20
3.6 Building the Model with Orange Data Mining	23
3.7 Evaluation Method	24
Chapter 4: RESULT AND ANALYSIS	27
4.1 Tools.....	27
4.2 Dataset.....	27
4.3 Data Pre-Processing	29
4.3.1 Training Data and Testing Data Proportions	29
4.3.2 Data Preparation in Orange	30
4.4 Using Random Forest Model	32

4.5 Using XGBoost Model.....	33
4.6 Choosing the Best Model.....	34
4.6.1 Result Interpretation.....	36
4.7 Churn Rate Analysis	40
Chapter 5: CONCLUSIONS AND SUGGESTIONS	66
5.1 Conclusions.....	66
5.2 Suggestions	67
References.....	68
Appenddix A Churn Dataset	A-0
Appenddix B Churn Rate Graph	B-1
Appenddix C Box Plot of Churn Rate %	C-1

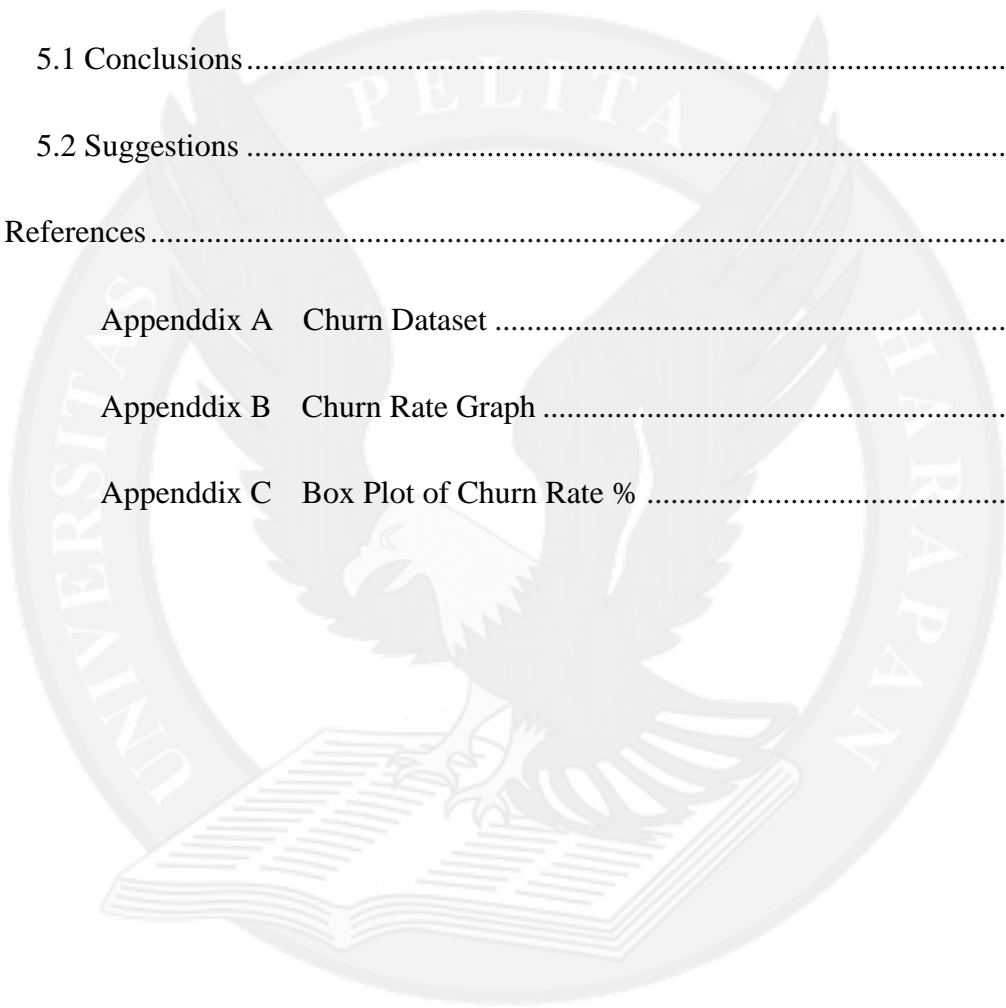


Table of Figures

Figure 2.1 Steps of Data Mining (Ramageri, 2010).....	9
Figure 2.2 Random Forest Scheme (Kulkarni & Sinha, 2013).....	13
Figure 2.3 XGBoost Scheme (Chen & Guestrin, 2016).....	15
Figure 3.1 Research Flow Chart.....	22
Figure 4.1 Loading Dataset into Orange	31
Figure 4.2 Imputing File in Orange.....	31
Figure 4.3 Sampling Data in Orange.....	32
Figure 4.4 Random Forest Widget in Orange	32
Figure 4.5 Random Forest Parameters in Orange	33
Figure 4.6 Gradient Boosting Widget in Orange	33
Figure 4.7 Gradient Boosting Parameters in Orange	34
Figure 4.8 Test and Score Widget in Orange	34
Figure 4.9 Loading Testing Data in Orange.....	35
Figure 4.10 Prediction Widget in Orange	35
Figure 4.11 ROC Curve of The Random Forest Model.....	38
Figure 4.12 Churn Rate Ganjil 2019	44
Figure 4.13 Churn Rate Genap 2019.....	45
Figure 4.14 Churn Rate Ganjil 2020	46
Figure 4.15 Churn Rate Genap 2020.....	47
Figure 4.16 Churn Rate Ganjil 2021	48
Figure 4.17 Churn Rate Teknik Informatika - S1	49
Figure 4.18 Box Plot Teknik Informatika - S1	50
Figure 4.19 Churn Rate Sistem Informasi - S1	51

Figure 4.20	Box Plot Sistem Informasi - S1	52
Figure 4.21	Churn Rate Informatika (Kampus Kota Medan) - S1.....	53
Figure 4.22	Box Plot Informatika (Kampus Kota Medan) - S1	54
Figure 4.23	Churn Rate Sistem Informasi (Kampus Kota Medan) - S1	55
Figure 4.24	Box Plot Sistem Informasi (Kampus Kota Medan) - S1	56
Figure 4.25	Churn Rate vs Accreditation Plot Ganjil 2019	58
Figure 4.26	Churn Rate vs Accreditation Plot Genap 2019.....	59
Figure 4.27	Churn Rate vs Accreditation Plot Ganjil 2020	60
Figure 4.28	Churn Rate vs Accreditation Plot Genap 2020.....	61
Figure 4.29	Churn Rate vs Accreditation Plot Ganjil 2021	62

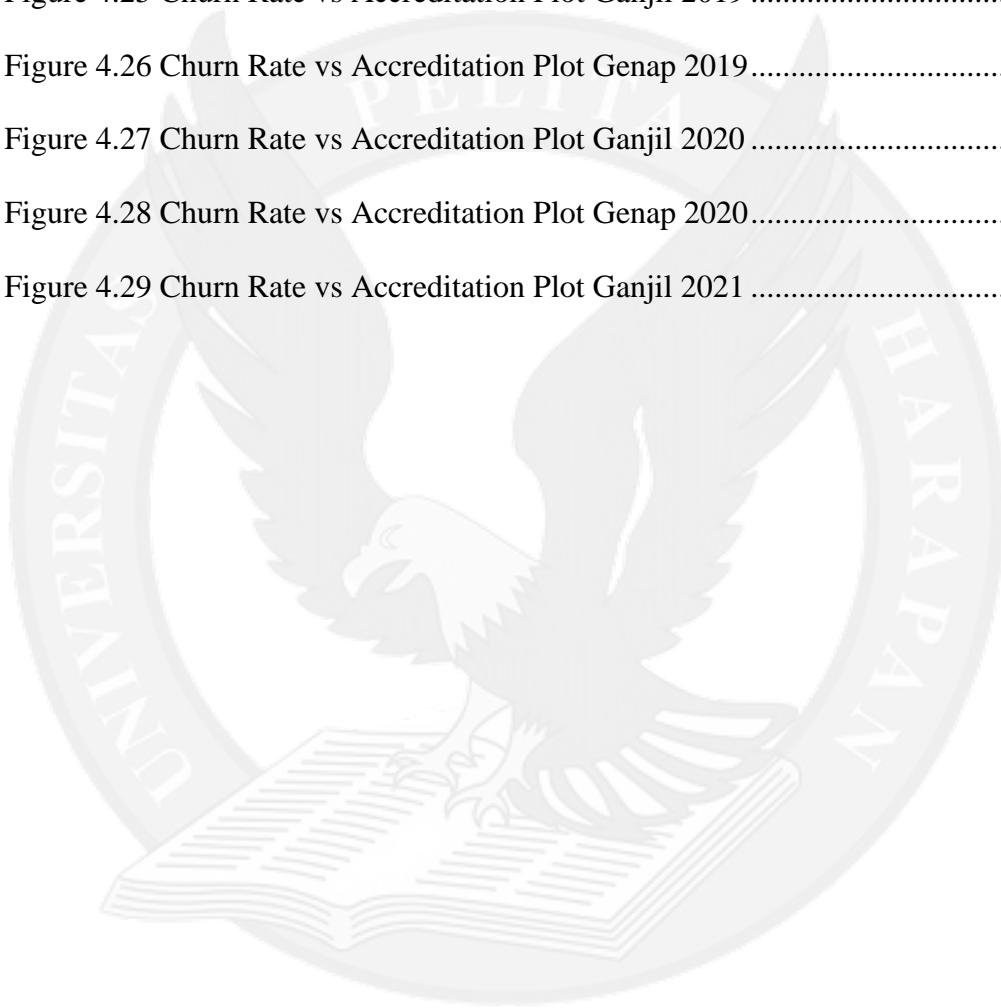


Table of Tables

Table 3.1 List of Variables.....	19
Table 3.2 Confusion Matrix	24
Table 3.3 AUC Score Categories	26
Table 4.1 Dataset Sample.....	28
Table 4.2 Data Sample from data_churn.csv	29
Table 4.3 Data Proportion.....	30
Table 4.4 RMSE and MSE Score.....	36
Table 4.5 Confusion Matrix of The Random Forest Model.....	38

