

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PENGESAHAN .....	ii
HALAMAN PERNYATAAN .....	iii
HALAMAN PERSEMBAHAN .....	iv
HALAMAN MOTTO .....	v
PRAKATA .....	vi
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xi
INTISARI .....	xiii
ABSTRACT .....	xiv
<b>I. PENDAHULUAN .....</b>	<b>1</b>
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah .....	3
1.3. Tujuan dan Manfaat Penelitian .....	3
1.4. Tinjauan Pustaka .....	4
1.5. Metode Penelitian .....	7
1.6. Sistematika Penulisan .....	7
<b>II. DASAR TEORI .....</b>	<b>9</b>
2.1. <i>Machine Learning</i> .....	9
2.2. <i>Supervised Learning</i> .....	11
2.3. <i>Feature Selection</i> .....	12
2.3.1. <i>Information Gain</i> .....	12
2.3.2. <i>Gain Ratio</i> .....	13
2.3.3. <i>Chi-Square</i> .....	13
2.3.4. <i>Random Forest Importance</i> .....	13
2.4. <i>Decision Tree</i> .....	14
2.4.1. Terminologi <i>decision tree</i> .....	16

2.4.2. <i>Decision Tree</i> Berdasarkan Algoritma <i>CART</i> .....	17
2.5. <i>Bootstrap aggregating (bagging)</i> .....	19
2.6. <i>Imbalance Data</i> .....	20
2.7. <i>k-Fold Cross Validation</i> .....	21
2.8. Evaluasi Model Klasifikasi .....	22
2.9. <i>AUC (Area Under Curve)</i> .....	24
<b>III. <i>RANDOM OVER SAMPLING (ROS), RANDOM FOREST DAN DECISION TREE ALGORITMA C5.0</i></b> .....	<b>26</b>
3.1. <i>Random Over sampling (ROS)</i> .....	26
3.2. <i>Random Forest</i> .....	27
3.2.1. Algoritma <i>Random Forest</i> .....	28
3.2.2. <i>Out of Bag (OOB) Estimasi Error</i> .....	29
3.2.3. <i>Variable Importance</i> .....	30
3.3. <i>Decision Tree Algoritma C5.0</i> .....	30
<b>IV. <i>STUDI KASUS</i></b> .....	<b>35</b>
4.1. Penanganan Data Pada <i>Missing Value</i> .....	35
4.2. <i>Random Over Sampling</i> .....	37
4.3. Data Partisi .....	38
4.4. <i>Random Forest Pada Balance Data</i> .....	39
4.4.1. Penentuan mtry dan jumlah pohon (ntree) terbaik .....	39
4.4.2. Pembuatan model <i>Random Forest</i> Pada <i>Balance Data</i> ....	42
4.4.3. Evaluasi model klasifikasi <i>Random Forest</i> Pada <i>Balance Data</i> .....	51
4.4.4. <i>Variable importance Random Forest</i> Pada <i>Balance Data</i> ..	52
4.5. <i>Random Forest Pada Imbalance Data</i> .....	53
4.5.1. Penentuan mtry dan jumlah pohon (ntree) terbaik .....	54
4.5.2. Pembuatan model <i>Random Forest</i> Pada <i>Imbalance Data</i> ..	55
4.5.3. Evaluasi model klasifikasi <i>Random Forest</i> Pada <i>Imbalance Data</i> .....	56
4.5.4. <i>Variable importance Random Forest</i> Pada <i>Imbalance Data</i> ..	57

4.6.	<i>Decision Tree</i> Berdasarkan Algoritma C5.0 Pada <i>Balance Data</i> ...	58
4.6.1.	Pembuatan model <i>Decision Tree</i> Berdasarkan Algoritma C5.0 Pada <i>Balance Data</i> .....	59
4.6.2.	Evaluasi model klasifikasi <i>Decision Tree</i> Berdasarkan Algoritma C5.0 Pada <i>Balance Data</i> .....	66
4.7.	<i>Decision Tree</i> Berdasarkan Algoritma C5.0 Pada <i>Imblance Data</i> .....	67
4.7.1.	Pembuatan model <i>Decision Tree</i> Berdasarkan Algoritma C5.0 Pada <i>Balance Data</i> .....	67
4.7.2.	Evaluasi model klasifikasi <i>Decision Tree</i> Berdasarkan Algoritma C5.0 Pada <i>Balance Data</i> .....	68
<b>V.</b>	<b>KESIMPULAN</b> .....	<b>70</b>
5.1.	Kesimpulan .....	70
5.2.	Saran .....	71
	<b>DAFTAR PUSTAKA</b> .....	<b>72</b>
<b>A.</b>	<b>SKRIP PROGRAM R</b> .....	<b>75</b>
1.1.	Skrup program R <i>random over sampling</i> pada analisis klasifikasi <i>random forest</i> .....	75
1.2.	Skrup program R klasifikasi <i>random forest</i> .....	78
1.3.	Skrup program R <i>random over sampling</i> pada analisis klasifikasi <i>decision tree</i> algoritma C5.0 .....	80
1.4.	Skrup program R klasifikasi <i>decision tree</i> algoritma C5.0 .....	82
<b>B.</b>	<b>DATA STROKE</b> .....	<b>84</b>