

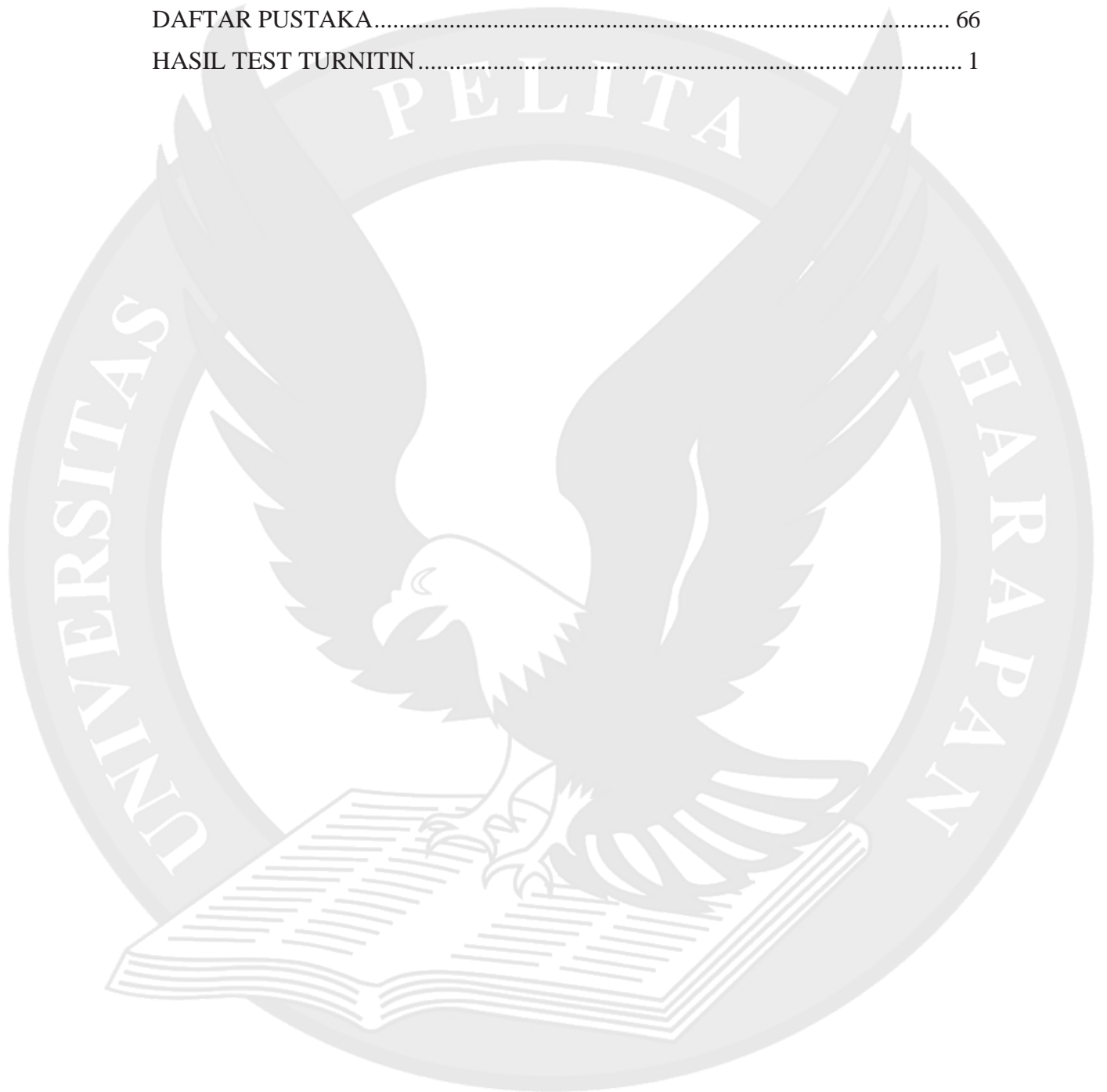
DAFTAR ISI

	halaman
PERNYATAAN KEASLIAN KARYA TUGAS AKHIR	viii
PERSETUJUAN DOSEN PEMBIMBING TUGAS AKHIR	ix
PERSETUJUAN TIM PENGUJI TUGAS AKHIR	x
.....	
ABSTRAK	xii
DAFTAR ISI	xiv
DAFTAR GAMBAR.....	xviii
DAFTAR TABEL	xx
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian.....	1
1.2 Rumusan Masalah Penelitian.....	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	4
1.4.1 Manfaat Teoritis	4
1.4.2 Manfaat Praktis.....	4
1.5 Sistematika Penulisan	5
BAB II LANDASAN TEORI	6
2.1 Data Mining dan Konsep Variabel	6
2.1.1 Data Mining.....	6
2.1.1.1 Business Understanding.....	6
2.1.1.2 Data Understanding	6
2.1.1.3 Data Preparation	6
2.1.1.4 Modelling	7
2.1.1.5 Evaluation.....	7
2.1.1.6 Deployment	7
2.1.2 Supervised Learning	8
2.1.3 Machine Learning.....	9
2.1.4 Logistic Regression (LR)	9
2.1.5 Random Forest (RF).....	10

2.1.6	Customer Lifetime Value	10
2.1.7	Maturity Date.....	11
2.1.8	Cross-Sell	11
2.1.9	Gimmick.....	12
2.1.10	Variabel Independen dan Variabel dependen	12
2.1.10.1	Gender.....	13
2.1.10.2	Age.....	13
2.1.10.3	Marital Status	14
2.1.10.4	Income	15
2.1.10.5	No Children.....	15
2.1.10.6	Occupation	16
2.1.10.7	Vintage.....	16
2.1.10.8	Nilai Premi Tahunan	17
2.1.10.9	Nilai Vouchers	17
2.1.10.10	Comms Media.....	17
2.1.10.11	Previously Claims	18
2.1.10.12	Complaints	18
2.1.10.13	Policy Ownership.....	19
2.1.10.14	Family in Force.....	19
2.1.10.15	Payment Type	20
2.1.10.16	Referring	20
2.1.10.17	Payment Frequency.....	21
2.1.10.18	Orphan Policy	21
2.1.10.19	Own Other Insurance	22
2.1.10.20	Premium Income Source.....	22
2.1.10.21	Currency.....	23
2.2	Model Penelitian.....	24
2.2.1	Model Logistic Regression.....	24
2.2.2	Model Random Forest.....	24
BAB III METODOLOGI PENELITIAN		25
3.1	Business Understanding.....	25
3.2	Data Understanding	26
3.3	Data Preparation	28
3.4	Modelling	33

BAB IV ANALISIS DAN PEMBAHASAN	36
4.1 Profil Nasabah	36
4.1.1 Deskripsi Variabel Independen	36
4.1.1.1 Variabel Independen – <i>numerical</i>	36
4.1.1.2 Variabel Independen – <i>categorical</i>	38
4.1.2 Deskripsi Variabel Dependen.....	40
4.2 Analisa Data Nasabah.....	40
4.2.1 Uji Duplikasi Data.....	40
4.2.2 Uji Multikolinearitas.....	41
4.2.3 Uji Korelasi	43
4.2.4 Uji Outliers	44
4.3 Hasil Penelitian.....	44
4.3.1 Logistic Regression (LR)	44
4.3.1.1 Uji Duplikasi	45
4.3.1.2 Uji Multikolinearitas.....	46
4.3.1.3 Uji Korelasi	46
4.3.1.4 Coefficient Interpretation Value.....	47
4.3.1.5 Rank Ordering	48
4.3.1.6 Confusion Matrix dan Performance dari model Logistic Regression	49
4.3.1.7 Nilai ROC dan AUC dari model Logistic Regression.....	50
4.3.2 Random Forest	51
4.3.2.1 Pohon (=tree) dari model Random Forest (RF).....	51
4.3.2.2 Feature Importance dari model Random Forest	51
4.3.2.3 Classification Report dan Confusion Matrix dari model Random Forest (RF).....	52
4.3.2.4 Nilai ROC dan AUC dari model Random Forest (RF).....	53
4.4 Pembahasan	53
4.4.1 Hasil Model Logistic Regression	53
4.4.2 Hasil Model Random Forest.....	58
4.4.3 Perbandingan Model Logistic Regression dan Random Forest	59
4.5 Managerial Implications	60
4.5.1 Next Step Action Jangka Pendek.....	60
4.5.2 Next Step Action Jangka Menengah/Panjang	61

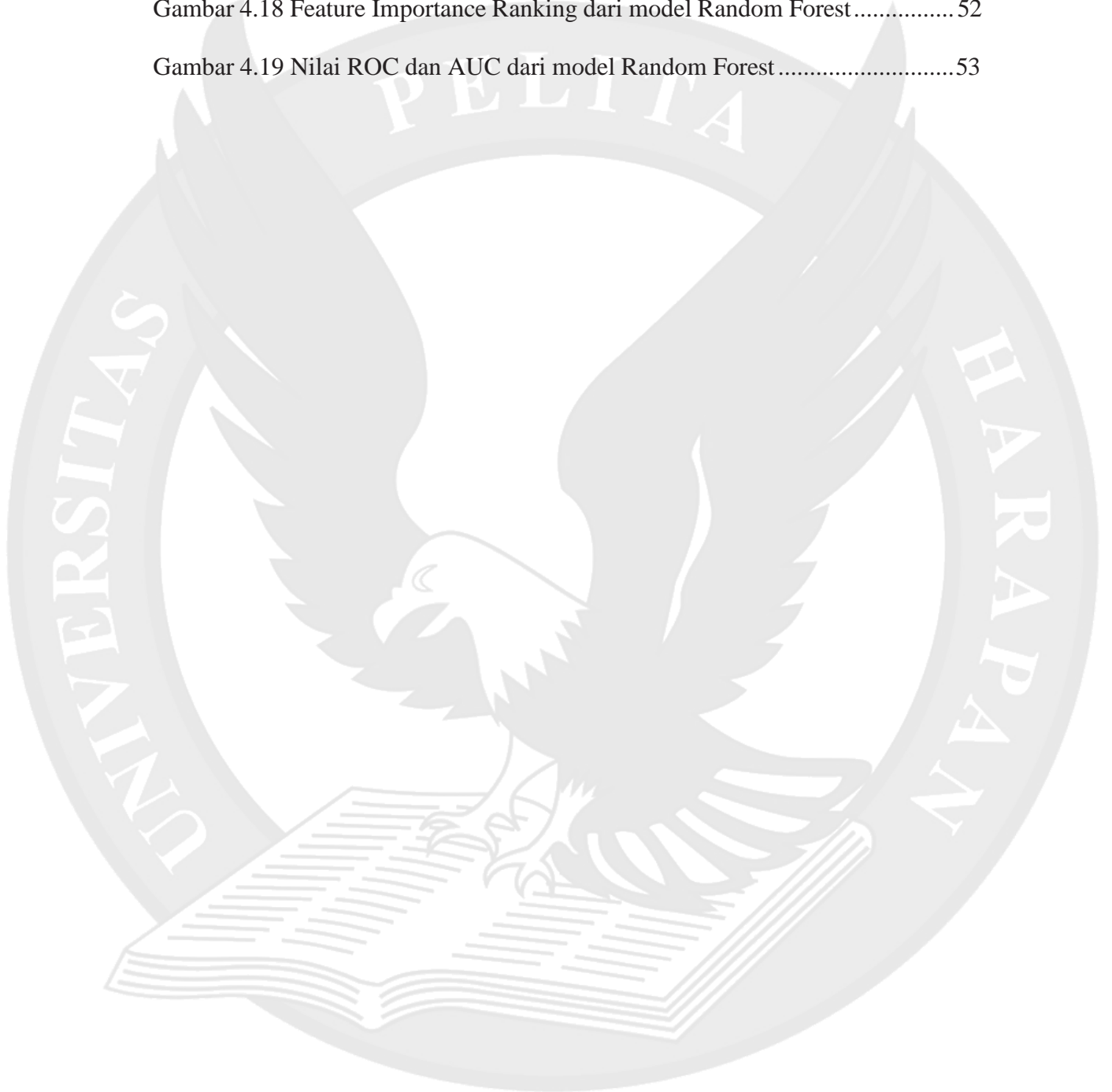
BAB V KESIMPULAN DAN SARAN	62
5.1 Kesimpulan.....	62
5.2 Saran.....	64
DAFTAR PUSTAKA.....	66
HASIL TEST TURNITIN.....	1



DAFTAR GAMBAR

	halaman
Gambar 2.1 Metode <i>Data Mining</i> CRISP-DM.....	8
Gambar 2.2 Langkah-langkah di metode <i>Data Mining</i> CRISP-DM	8
Gambar 2.3 Ilustrasi Model Logistic Regression (LR)	24
Gambar 2.4 Ilustrasi Model Random Forest (RF).....	24
Gambar 3.1 Proyeksi IQR pada populasi berdistribusi normal.....	32
Gambar 4.1 Boxplot Age dan Income.....	36
Gambar 4.2 Boxplot No Children dan Nilai Premi Tahunan	37
Gambar 4.3 Bar Chart Number of Children	37
Gambar 4.4 Bar Chart Gender dan Marital Status	38
Gambar 4.5 Bar Occupation dan Umur Polis	38
Gambar 4.6. Nilai Vouchers dan Communication Media	38
Gambar 4.7 Previously Claims dan Complaints	38
Gambar 4.8 Policy Ownership dan Family in Force.....	39
Gambar 4.9 Payment Type dan Referring	39
Gambar 4.10 Payment Frequency dan Orphan Policy	39
Gambar 4.11 Own Other Insurance dan Premium Income Source.....	39
Gambar 4.12 Currency	40
Gambar 4.13 Respond	40
Gambar 4.14 Association Matrix Heat Map Tahap Pertama.....	43
Gambar 4.15 Association Matrix Heat Map Tahap Kedua	46

Gambar 4.16 Nilai ROC dan AUC dari model Logistic Regression	50
Gambar 4.17 Salah satu Pohon (=tree) dari model Random Forest	51
Gambar 4.18 Feature Importance Ranking dari model Random Forest.....	52
Gambar 4.19 Nilai ROC dan AUC dari model Random Forest	53



DAFTAR TABEL

	halaman
Tabel 3.1 Variabel independen dan dependen dari obyek penelitian	28
Tabel 4.1 Statistik Variabel Independen – Data Numerical	36
Tabel 4.2 Nilai VIF dari semua variabel	41
Tabel 4.3 Nilai VIF dari Variabel setelah proses Cleansing Data	41
Tabel 4.4 Perubahan Penyebutan Variabel setelah proses Cleansing VIF Data ...	42
Tabel 4.5 Output Logistic Regression Tahap Pertama.....	44
Tabel 4.6 Perubahan Penyebutan Variabel setelah proses Cleansing P z 	45
Tabel 4.7 Nilai VIF setelah proses Uji Multikolinearitas	46
Tabel 4.8 Output Logistic Regression Tahap 2.....	47
Tabel 4.9 Rank Ordering Model Logistic Regression.....	49
Tabel 4.10 Prioritas Variabel setelah proses Rank Ordering	49
Tabel 4.11. Classification Report dan Confusion Matrix dari model Logistic Regression.....	49
Tabel 4.12 Classification Report dan Confusion Matrix dari Model Random Forest.....	52
Tabel 4.13 Dampak dan Rank Ordering Variabel X ke Variabel Y	54
Tabel 4.14 Range Nilai Tukar Rupiah terhadap USD	55
Tabel 4.15 Feature Importance Ranking dari model Random Forest	58
Tabel 4.16 Perbandingan performance model Logistic Regression dan Random Forest.....	59
Tabel 4.17 Perbandingan variabel dominan model Logistic Regression dan Random Forest.....	60