

## DAFTAR ISI

<b>HALAMAN JUDUL</b>	<b>i</b>
<b>HALAMAN PENGESAHAN</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN</b>	<b>iii</b>
<b>HALAMAN PERSEMBAHAN</b>	<b>iv</b>
<b>HALAMAN MOTTO</b>	<b>v</b>
<b>PRAKATA</b>	<b>vi</b>
<b>DAFTAR ISI</b>	<b>viii</b>
<b>DAFTAR TABEL</b>	<b>xi</b>
<b>INTISARI</b>	<b>xiii</b>
<b>ABSTRACT</b>	<b>xiv</b>
<b>I PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang Masalah	1
1.2 Tujuan dan Manfaat Penelitian	3
1.3 Pembatasan Masalah	3
1.4 Tinjauan Pustaka	4
1.5 Metode Penelitian	6
1.6 Sistematika Penulisan	6
<b>II DASAR TEORI</b>	<b>8</b>
2.1 Machine Learning	8
2.1.1 Perbedaan Utama Pemodelan Statistik dan Machine Learning	9
2.1.2 Langkah-Langkah Pengembangan Model Machine Learning dan Penerapan	10
2.2 Supervised dan Unsupervised Learning	11
2.3 Feature Scaling	13
2.4 Decision Tree	13
2.4.1 Terminologi dan Istilah yang terkait dengan <i>Decision Tree</i>	14
2.4.2 <i>Decision Tree</i> Berdasarkan Algoritma CART	15
2.4.3 Kelebihan dan Kelemahan <i>Decision Tree</i>	17
2.5 Bagging ( <i>Bootstrap Aggregating</i> )	18
2.6 <i>Imbalanced Data</i>	19
2.7 k-Fold Cross Validation	20
2.8 Evaluasi Model Klasifikasi	21
2.9 <i>ROC Curves</i>	23

<b>III SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE), RANDOM FOREST DAN K-NEAREST NEIGHBORS</b>	<b>25</b>
3.1 SMOTE (Synthetic Minority Oversampling Technique)	25
3.1.1 Synthethic Minority Over-Sampling Technique Nominal Continous (SMOTE-NC)	27
3.2 Random Forest	28
3.2.1 Karakteristik Akurasi <i>Random Forest</i>	28
3.2.2 Algoritma <i>Random Forest</i>	33
3.2.3 <i>Out of Bag</i> (OOB) Estimasi Error	33
3.2.4 Variable Importance	34
3.3 K-Nearest Neighbors	34
3.3.1 Perbedaan metode metrik jarak	36
<b>IV STUDI KASUS</b>	<b>38</b>
4.1 Penanganan Data Pada <i>Missing Value</i>	38
4.2 <i>SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE)</i>	40
4.3 Data Partisi	45
4.4 <i>SMOTE Random Forest</i>	45
4.4.1 Penentuan mtry dan jumlah pohon (ntree) terbaik	46
4.4.2 Pembuatan model <i>smote random forest</i>	48
4.4.3 Evaluasi model klasifikasi <i>smote random forest</i>	57
4.4.4 <i>Variable importance smote random forest</i>	58
4.5 <i>Random Forest</i> pada <i>Imbalaced Data</i>	59
4.5.1 Penentuan mtry dan jumlah pohon (ntree) terbaik	60
4.5.2 Pembuatan model <i>random forest</i>	62
4.5.3 Evaluasi model klasifikasi <i>random forest</i>	63
4.5.4 <i>Variable importance random forest</i>	64
4.6 <i>SMOTE k-Nearest Neighbors</i>	65
4.6.1 Pembuatan model <i>smote k-nearest neighbors</i>	66
4.6.2 Evaluasi model klasifikasi <i>smote k-nearest neighbors</i>	68
4.7 <i>k-Nearest Neighbors</i> pada <i>Imbalanced Data</i>	69
4.7.1 Pembuatan model <i>k-nearest neighbors</i>	69
4.7.2 Evaluasi model klasifikasi <i>k-nearest neighbors</i>	72
4.8 Perbandingan Evaluasi Model Klasifikasi	73
<b>V KESIMPULAN</b>	<b>75</b>
5.1 Kesimpulan	75
5.2 Saran	76

<b>DAFTAR PUSTAKA</b>	<b>77</b>
<b>A SKRIP PROGRAM R</b>	<b>79</b>
1.1 skrip program r klasifikasi smote random forest	79
1.2 skrip program r klasifikasi random forest	80
1.3 skrip program r klasifikasi smote k-nearest neighbors	82
1.4 skrip program r klasifikasi k-nearest neighbors	84
<b>B CERVICAL CANCER RISK FACTORS DATA</b>	<b>86</b>