

**APLIKASI KLASIFIKASI KANKER SERVIKS  
MENGUNAKAN DATA *MICRORNA*  
DENGAN ALGORITMA *RANDOM FOREST***



**SKRIPSI**

**Disusun Sebagai Salah Satu Syarat  
untuk Memperoleh Gelar Sarjana Komputer  
pada Departemen Ilmu Komputer/ Informatika**

**Disusun Oleh:**

**EKO ABDUL AZIZ**

**24010315120020**

**DEPARTEMEN ILMU KOMPUTER/ INFORMATIKA  
FAKULTAS SAINS DAN MATEMATIKA  
UNIVERSITAS DIPONEGORO**

**2019**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Eko Abdul Aziz

NIM : 24010315120020

Judul : Aplikasi Klasifikasi Kanker Serviks Menggunakan Data *MicroRNA* Dengan Algoritma *Random Forest*

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 24 Juni 2019

  
METERAI  
TEMPEL  
6000  
Rp. ENCIK RUPIAH  
Eko Abdul Aziz  
24010315120020

## HALAMAN PENGESAHAN

Judul : Aplikasi Klasifikasi Kanker Serviks Menggunakan Data *MicroRNA* Dengan  
Algoritma *Random Forest*  
Nama : Eko Abdul Aziz  
NIM : 24010315120020

Telah diujikan pada sidang skripsi pada tanggal 29 Mei 2019 dan dinyatakan lulus pada  
tanggal 29 Mei 2019.

Semarang, 24 Juni 2019

Mengetahui,

Ketua Departemen Ilmu Komputer/ Informatika  
FSM Undip

Panitia Penguji Skripsi  
Ketua,



Dr. Retno Kusumaningrum, S.Si, M.Kom  
NIP. 198104202005012001



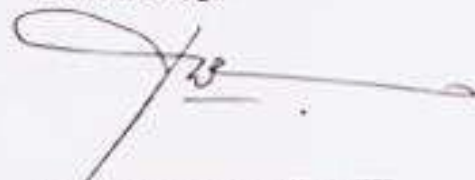
Beta Noranita, S.Si, M.Kom  
NIP. 197308291998022001

## HALAMAN PENGESAHAN

Judul : Aplikasi Klasifikasi Kanker Serviks Menggunakan Data *MicroRNA* Dengan  
Algoritma *Random Forest*  
Nama : Eko Abdul Aziz  
NIM : 24010315120020

Telah diujikan pada sidang skripsi pada tanggal 29 Mei 2019 dan dinyatakan lulus pada tanggal 29 Mei 2019.

Semarang, 24 Juni 2019  
Pembimbing,



Indra Waspada, ST, MTI  
NIP. 197902122008121002

## ABSTRAK

Kanker serviks merupakan salah satu penyakit paling mematikan bagi kaum wanita di dunia. Kematian akibat kanker serviks terbesar dialami negara-negara berpenghasilan rendah atau menengah, termasuk Indonesia. Salah satu upaya pencegahan untuk menanggulangi kematian akibat kanker serviks adalah melakukan deteksi dini kanker serviks. Teknologi terbaru deteksi dini kanker adalah dengan menggunakan pembelajaran mesin untuk klasifikasi yang menghasilkan prediksi kondisi pasien, sehingga bisa digunakan untuk pedoman awal pengobatan kanker. Tetapi, para peneliti biomedis tidak semuanya bisa melakukan klasifikasi dan memerlukan sebuah alat yang bisa memfasilitasi mereka dalam melakukan klasifikasi. Dengan demikian, dibuat aplikasi menggunakan *framework* Django untuk klasifikasi kanker serviks menggunakan data *microRNA* dengan algoritma *Random Forest*. Dalam kasus kanker, *microRNA* bisa digunakan sebagai biomarker diagnostik untuk deteksi dini kanker sehingga bisa membantu perencanaan perawatan penderita kanker. Algoritma *Random Forest* dipilih karena menghasilkan visualisasi model berupa pohon keputusan yang mudah dipahami oleh peneliti biomedis dan bisa melakukan perhitungan *feature importance* untuk rekomendasi biomarker. Aplikasi ini dibangun menggunakan model proses *waterfall* yang pada fase pengumpulan persyaratan aplikasinya didasarkan pada analisis metodologi CRISP-DM. Hasil pengujian aplikasi ini menyatakan bahwa aplikasi ini layak digunakan sebagai solusi baru untuk memfasilitasi peneliti biomedis pada domain kanker serviks.

**Kata Kunci :** Aplikasi, Klasifikasi, Kanker Serviks, *MicroRNA*, *Random Forest*, Model Proses *Waterfall*, CRISP-DM, Django

## ABSTRACT

Cervical cancer is one of the deadliest diseases for women in the world. Deaths from cervical cancer are greatest in low or middle income countries, including Indonesia. One effort to tackle deaths from cervical cancer is early detection of cervical cancer. The latest technology for early detection of cancer is to use machine learning for classification that produces patient predictions, so that it can be used to guide the initial treatment of cancer. However, not all biomedical researchers can classify and ask for a tool that can facilitate them in classifying. Thus, applications made using the Django framework for classification of cervical cancer using microRNA data with the Random Forest algorithm. In the case of cancer, the microRNA can be used as a biomarker for early detection of cancer so that it can help plan the treatment of cancer patients. Random Forest Algorithms are chosen because they produce a visualization model that forms decision trees that are easily approved by biomedical researchers and can calculate important features for biomarker assessment. This application was built using the waterfall process model which at the stage of fulfilling the application requirements was based on the analysis carried out by CRISP-DM. The results of testing this application state that this application is worthy of being used as a new solution to facilitate biomedical researchers in the cervical cancer domain.

**Keywords:** Application, Classification, Cervical Cancer, MicroRNA, Random Forest, Waterfall Process Model, CRISP-DM, Django.

## KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “Aplikasi Klasifikasi Kanker Serviks Menggunakan Data *MicroRNA* Dengan Algoritma *Random Forest*”.

Dalam penyusunan laporan skripsi ini penulis mendapat banyak bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih kepada:

1. Ibu Dr. Retno Kusumaningrum, S.Si., M.Kom, selaku Ketua Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang.
2. Bapak Panji Wisnu Wirawan, ST, MT, selaku Koordinator Skripsi Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro, Semarang.
3. Bapak Indra Waspada, S.T, MTI, selaku dosen pembimbing skripsi yang telah membantu dalam membimbing dan mengarahkan penulis hingga selesainya skripsi ini.
4. Orang tua, keluarga, teman dekat dan sahabat yang telah mendukung, membantu, serta memberikan semangat kepada penulis dalam menyelesaikan skripsi ini.
5. Semua pihak yang telah membantu kelancaran dalam penyusunan skripsi, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dokumen skripsi ini masih jauh dari sempurna. Oleh karena itu, saran dan kritik yang membangun sangat penulis harapkan. Akhir kata, semoga skripsi ini dapat bermanfaat bagi semua pihak

Semarang, 24 Juni 2019

Eko Abdul Aziz

## HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Eko Abdul Aziz  
NIM : 24010315120020  
Program Studi : Ilmu Komputer/ Informatika  
Jenis Karya : Skripsi

Demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*)** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

Aplikasi Klasifikasi Kanker Serviks Menggunakan Data *MicroRNA* Dengan Algoritma *Random Forest*

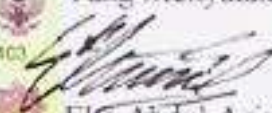
Beserta perangkat yang ada (jika diperlukan). Dengan Hak bebas Royalti Non-eksklusif ini Universitas Diponegoro berhak menyimpan, mengalihmedia/ formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya

Semarang, 24 Juni 2019

Yang Menyatakan



  
Eko Abdul Aziz  
24010315120020

## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PENGESAHAN .....	iv
ABSTRAK .....	v
ABSTRACT .....	vi
KATA PENGANTAR.....	vii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS .....	viii
DAFTAR ISI .....	ix
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL .....	xvii
DAFTAR LAMPIRAN .....	xix
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan dan Manfaat .....	3
1.4. Ruang Lingkup .....	3
1.5. Sistematika Penulisan .....	4
BAB II LANDASAN TEORI.....	5
2.1. Perkembangan Penelitian Aplikasi Klasifikasi Kanker Menggunakan Data <i>MicroRNA</i> .....	5
2.2. Kanker Serviks.....	5
2.3. <i>MicroRNA</i> sebagai <i>Biomarker</i> Kanker .....	6
2.4. Aplikasi Web .....	7
2.5. Klasifikasi .....	7
2.6. Pembelajaran Mesin.....	7
2.7. Penambangan Data .....	8

2.8. <i>Random Forest</i> .....	8
2.8.1. Algoritma <i>Random Forest</i> .....	9
2.8.2. Fungsi <i>Feature Importance</i> pada <i>Random Forest</i> .....	13
2.9. <i>K-Fold Cross Validation</i> .....	13
2.10. <i>Confusion Matrix</i> .....	14
2.11. CRISP-DM .....	15
2.12. Model Proses <i>Waterfall</i> .....	18
2.13. <i>Object-Oriented Analysis and Design (OOAD)</i> .....	20
2.13.1. Fase Analisa Sistem .....	21
2.13.1.1. Analisa Kebutuhan Sistem .....	21
2.13.1.2. Analisa <i>Use Case</i> .....	22
2.13.1.3 <i>Class Diagram</i> Analisis.....	24
2.13.2. Fase Desain Sistem.....	25
2.13.2.1. <i>Sequence Diagram</i> .....	25
2.13.2.2 <i>Class Diagram</i> Desain.....	26
2.14. Django .....	27
2.15. Pengujian <i>Black Box</i> .....	27
BAB III METODOLOGI PENELITIAN.....	28
3.1. Persyaratan Aplikasi .....	28
3.1.1. Implementasi Metodologi CRISP-DM.....	29
3.1.1.1. Fase Pemahaman Bisnis .....	29
3.1.1.2. Fase Pemahaman Data.....	29
3.1.1.3. Fase Pengolahan Data.....	30
3.1.1.4. Fase Pemodelan .....	30
3.1.1.4.1. <i>Select Modeling Technique</i> .....	30
3.1.1.4.2. <i>Generate Test Design</i> .....	30
3.1.1.4.3. <i>Build Model</i> .....	33
3.1.1.4.4. <i>Asses Model</i> .....	45

3.1.1.5. Fase Evaluasi .....	46
3.1.2. Identifikasi Persyaratan Aplikasi.....	46
3.2. Analisa Aplikasi.....	46
3.2.1. Analisa <i>Use Case</i> .....	47
3.2.2. <i>Class Diagram</i> Analisis.....	47
3.3. Desain Sistem .....	47
3.3.1. <i>Sequence Diagram</i> .....	47
3.3.2. <i>Class Diagram</i> Desain.....	47
3.3.3. Skema Basis Data .....	47
3.3.4. Sketsa Antarmuka .....	47
3.4. Spesifikasi Aplikasi .....	48
3.5. Implementasi Aplikasi .....	48
3.5.1. Implementasi Basis Data .....	48
3.5.2. Implementasi Kode Program .....	48
3.5.3. Implementasi Antarmuka.....	48
3.6. Pengujian Aplikasi.....	48
BAB IV HASIL DAN PEMBAHASAN.....	49
4.1. Persyaratan Aplikasi .....	49
4.1.1. Implementasi Metodologi CRISP-DM.....	49
4.1.1.1. Fase Pemahaman Bisnis .....	49
4.1.1.2. Fase Pemahaman Data.....	49
4.1.1.2.1. <i>Collect Initial Data</i> .....	50
4.1.1.2.2. <i>Describe Data</i> .....	50
4.1.1.2.3. <i>Explore Data</i> .....	51
4.1.1.2.4. <i>Verify Data Quality</i> .....	51
4.1.1.3. Fase Pengolahan Data.....	51
4.1.1.3.1. <i>Select Data</i> .....	51
4.1.1.3.2. <i>Clean Data</i> .....	51

4.1.1.3.3. <i>Construct Data</i> .....	52
4.1.1.4. Fase Pemodelan .....	53
4.1.1.4.1. <i>Select Modeling Technique</i> .....	53
4.1.1.4.2. <i>Generate Test Design</i> .....	53
4.1.1.4.3. <i>Build Model</i> .....	53
4.1.1.4.4. <i>Asses Model</i> .....	56
4.1.1.5. Fase Evaluasi .....	56
4.1.2. Identifikasi Persyaratan Aplikasi.....	58
4.2. Analisa Aplikasi.....	60
4.2.1. Analisa <i>Use Case</i> .....	60
4.2.2.1 Daftar Aktor.....	60
4.2.2.2 Daftar Use Case .....	61
4.2.2.3 Use Case Diagram .....	61
4.2.2. <i>Use Case Detail</i> .....	63
4.2.3. <i>Class Diagram Analisis</i> .....	68
4.3. Desain Sistem .....	73
4.3.1. <i>Sequence Diagram</i> .....	73
3.3.5. <i>Class Diagram Desain</i> .....	82
3.3.6. Skema Basis Data .....	83
3.3.7. Sketsa Antarmuka .....	85
4.4. Spesifikasi Perangkat.....	91
4.5. Implementasi Aplikasi .....	92
4.5.1. Implementasi Basis Data .....	92
4.5.2. Implementasi Kode Program .....	95
4.5.3. Implementasi Antarmuka.....	96
4.6. Pengujian Aplikasi.....	103
4.6.1. Pengujian <i>Black Box</i> .....	103
4.6.1.1. Rencana Pengujian Aplikasi.....	103

4.6.1.2. Hasil Pengujian Aplikasi .....	103
4.6.2. Skenario Pengujian Aplikasi.....	106
BAB V PENUTUP .....	107
5.1 Kesimpulan.....	107
5.2 Saran .....	107
DAFTAR PUSTAKA.....	108
LAMPIRAN – LAMPIRAN .....	111

## DAFTAR GAMBAR

Gambar 2.1 Ilustrasi <i>Ensemble Learning</i> (Raschka and Mirjalili 2017).....	8
Gambar 2.2 <i>K-Fold Cross Validation</i> Dengan $K = 10$ (Raschka and Mirjalili 2017).....	14
Gambar 2.3 <i>Confusion Matrix</i> (Raschka and Mirjalili 2017).....	14
Gambar 2.4 Siklus Hidup CRISP-DM (Chapman et al. 2000).....	16
Gambar 2.5 Penjabaran Elemen Untuk Setiap Fase CRISP-DM (Chapman et al. 2000) ...	18
Gambar 2.6 Model Proses <i>Waterfall</i> (O’Docherty 2005).....	19
Gambar 2.7 <i>Use Case Diagram</i> (Ramnath and Dathan 2011) .....	23
Gambar 2.8 Tabel <i>Use Case</i> Menambah Buku (Ramnath and Dathan 2011).....	23
Gambar 2.9 <i>Class Diagram</i> Tingkat Analisis (Ramnath and Dathan 2011).....	25
Gambar 2.10 <i>Sequence Diagram</i> Menambah Buku (Ramnath and Dathan 2011).....	26
Gambar 2.11 <i>Class Diagram</i> Tingkat Desain (Ramnath and Dathan 2011).....	27
Gambar 3.1 <i>Split Data</i> Untuk Fitur miR-221 Dengan Nilai <i>Threshold</i> 62 .....	36
Gambar 3.2 <i>Split Data</i> Untuk Fitur miR-203 Dengan Nilai <i>Threshold</i> 3 .....	40
Gambar 3.3 Hasil Pohon ke-1 .....	43
Gambar 3.4 Hasil Pohon ke-2 .....	43
Gambar 4.1 <i>Use Case Diagram</i> Aplikasi .....	62
Gambar 4.2 <i>Class Diagram</i> Tingkat Analisis .....	69
Gambar 4.3 <i>Sequence Diagram</i> Eksplorasi Dataset.....	73
Gambar 4.4 <i>Sequence Diagram</i> Menampilkan Semua Data Perhitungan RF.....	77
Gambar 4.5 <i>Sequence Diagram</i> Menambah Data Perhitungan RF.....	78
Gambar 4.6 <i>Sequence Diagram</i> Menghapus Data Perhitungan RF .....	79
Gambar 4.7 <i>Sequence Diagram</i> Melihat Detail Data Perhitungan RF.....	80
Gambar 4.8 <i>Sequence Diagram</i> Melihat Hasil Evaluasi Model dan <i>Feature Importance</i> ..	80
Gambar 4.9 <i>Sequence Diagram</i> Melihat Hasil Pohon.....	81
Gambar 4.10 <i>Class Diagram</i> Tingkat Desain .....	82
Gambar 4.11 Sketsa Antarmuka Eksplorasi Dataset.....	86
Gambar 4.12 Sketsa Antarmuka Melihat Semua Perhitungan RF .....	86
Gambar 4.13 Sketsa Antarmuka Menambah Perhitungan RF.....	88
Gambar 4.14 Sketsa Antarmuka Menghapus Perhitungan RF .....	88
Gambar 4.15 Sketsa Antarmuka Melihat Detail Perhitungan RF .....	89

Gambar 4.16 Sketsa Antarmuka Melihat Hasil Evaluasi Model dan Sketsa Antarmuka Melihat Hasil <i>Feature Importance</i> .....	90
Gambar 4.17 Sketsa Antarmuka Melihat Hasil Pohon.....	90
Gambar 4.18 Implementasi Tabel Dataset .....	93
Gambar 4.19 Implementasi Tabel SetLabel .....	93
Gambar 4.20 Implementasi Tabel SetFitur.....	94
Gambar 4.21 Implementasi Tabel SetHyperparameterRF .....	94
Gambar 4.22 Implementasi Tabel RandomForest.....	95
Gambar 4.23 Implementasi Antarmuka Eksplorasi Data .....	96
Gambar 4.24 Implementasi Antarmuka Eksplorasi Data .....	97
Gambar 4.25 Implementasi Antarmuka Melihat Semua Perhitungan RF .....	98
Gambar 4.26 Implementasi Antarmuka Menambah Perhitungan RF .....	99
Gambar 4.27 Implementasi Antarmuka Menghapus Perhitungan RF.....	99
Gambar 4.28 Implementasi Antarmuka Melihat Detail Perhitungan RF .....	100
Gambar 4.29 Implementasi Antarmuka Melihat Detail Perhitungan RF .....	100
Gambar 4.30 Implementasi Antarmuka Melihat Hasil Evaluasi Model .....	101
Gambar 4.31 Implementasi Antarmuka Melihat Hasil <i>Feature Importance</i> .....	101
Gambar 4.32 Implementasi Antarmuka Melihat Hasil Pohon .....	102
Gambar 4.33 Implementasi Antarmuka Melihat Hasil Pohon .....	102
Gambar 6.1 Hasil Pohon ke-1 .....	112
Gambar 6.2 Hasil Pohon ke-2 .....	113
Gambar 6.3 Hasil Pohon ke-3 .....	114
Gambar 6.4 Hasil Pohon ke-4 .....	115
Gambar 6.5 Hasil Pohon ke-5 .....	115
Gambar 6.6 Hasil Pohon ke-6 .....	116
Gambar 6.7 Hasil Pohon ke-7 .....	116
Gambar 6.8 Hasil Pohon ke-8 .....	117
Gambar 6.9 Hasil Pohon ke-9 .....	117
Gambar 6.10 Hasil Pohon ke-10 .....	118
Gambar 6.11 Memasukkan Data <i>MicroRNA</i> Kanker Serviks .....	131
Gambar 6.12 Melakukan Eksplorasi Singkat Dataset .....	132
Gambar 6.13 Melakukan Pengaturan Set Label Dataset .....	132
Gambar 6.14 Melakukan Pengaturan Set Fitur Dataset Dengan Memilih Semua Fitur ...	133

Gambar 6.15 Melakukan Pengaturan Set Hiperparameter Algoritma RF.....	133
Gambar 6.16 Melakukan Perhitungan RF Untuk Membuat Model Klasifikasi Menggunakan Semua Fitur .....	134
Gambar 6.17 Melihat Hasil Evaluasi Model Klasifikasi Menggunakan Semua Fitur .....	134
Gambar 6.18 Melihat Hasil <i>Feature Importance</i> .....	135
Gambar 6.19 Melakukan Pengaturan Set Fitur Dataset Dengan Memilih Enam Fitur .....	135
Gambar 6.20 Melakukan Perhitungan RF Untuk Membuat Model Klasifikasi Menggunakan Enam Fitur .....	136
Gambar 6.21 Melihat Hasil Evaluasi Model Klasifikasi Menggunakan Enam Fitur .....	136

## DAFTAR TABEL

Tabel 2.1 Perkembangan Penelitian Aplikasi Klasifikasi Kanker dengan Data <i>MicroRNA</i> .	5
Tabel 3.1 Data Label Hasil Uji.....	31
Tabel 3.2 Tabel Hasil <i>Confusion Matrix</i> .....	31
Tabel 3.3 Tabel Dataset untuk Membangun Model Klasifikasi.....	34
Tabel 3.4 Tabel <i>Bootstrap Sample</i> Pohon ke-1 .....	35
Tabel 3.5 Tabel <i>Information Gain Child</i> .....	39
Tabel 3.6 Tabel <i>Bootstrap Sample</i> Pohon ke-2 .....	42
Tabel 4.1 Tabel Data <i>MicroRNA</i> Kanker Serviks .....	50
Tabel 4.2 Tabel Data <i>MicroRNA</i> Kanker Serviks Saat Proses Pemetaan Nilai Label .....	52
Tabel 4.3 Tabel Data <i>MicroRNA</i> Kanker Serviks Setelah Proses Pembersihan dan Pemetaan Nilai Label .....	52
Tabel 4.4 Tabel Hasil Perhitungan <i>Feature Importance</i> .....	54
Tabel 4.5 Tabel Hasil Evaluasi Model Klasifikasi Dengan Semua Fitur Setelah Proses Pembersihan .....	56
Tabel 4.6 Analisa Persyaratan Fungsional APKKS Berdasarkan Tahapan CRISP-DM.....	60
Tabel 4.7 Daftar Aktor Aplikasi.....	61
Tabel 4.8 Daftar <i>Use Case</i> Aplikasi .....	61
Tabel 4.9 Tabel <i>Use Case</i> Eksplorasi Dataset.....	63
Tabel 4.10 Tabel <i>Use Case</i> Melihat Semua Perhitungan RF .....	64
Tabel 4.11 Tabel <i>Use Case</i> Melakukan Perhitungan RF.....	64
Tabel 4.12 Tabel <i>Use Case</i> Menghapus Perhitungan RF .....	65
Tabel 4.13 Tabel <i>Use Case</i> Melihat Detail Perhitungan RF .....	66
Tabel 4.14 Tabel <i>Use Case</i> Melihat Hasil Evaluasi Model.....	66
Tabel 4.15 Tabel <i>Use Case</i> Melihat Hasil <i>Feature Importance</i> .....	67
Tabel 4.16 Tabel <i>Use Case</i> Melihat Hasil Pohon.....	67
Tabel 4.17 Identifikasi <i>Class Analysis</i> .....	70
Tabel 4.18 Identifikasi Atribut <i>Class Analysis</i> .....	72
Tabel 4.19 Rancangan Tabel Dataset .....	83
Tabel 4.20 Rancangan Tabel SetLabel .....	83
Tabel 4.21 Rancangan Tabel SetFitur .....	84

Tabel 4.22 Rancangan Tabel SetHyperparameterRF .....	84
Tabel 4.23 Rancangan Tabel RandomForest.....	85
Tabel 4.24 Implementasi Kode Program Menambah Perhitungan RF.....	96
Tabel 4.25 Hasil Pengujian <i>Black Box</i> .....	103
Tabel 6.1 Rencana Skenario Pengujian Aplikasi.....	127
Tabel 6.2 Hasil Evaluasi Model Klasifikasi Dengan Semua Fitur .....	128
Tabel 6.3 Hasil Perhitungan <i>Feature Importance</i> .....	129
Tabel 6.4 Tabel Hasil Evaluasi Model Klasifikasi Dengan Enam Fitur .....	130
Tabel 6.5 <i>Test Case</i> Pengujian <i>Black Box</i> .....	137

## DAFTAR LAMPIRAN

Lampiran 1. Hasil Pohon Keputusan.....	112
Lampiran 2. Potongan Kode Program Menambah Perhitungan RF.....	119
Lampiran 3. Skenario Pengujian APKKS .....	127
Lampiran 4. Rencana Pengujian <i>Black Box</i> APKKS .....	137

# BAB I

## PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan skripsi mengenai aplikasi klasifikasi kanker serviks menggunakan data *microRNA* dengan Algoritma *Random Forest*.

### 1.1. Latar Belakang

Kanker atau tumor ganas adalah penyakit yang disebabkan pertumbuhan sel-sel abnormal yang tumbuh di luar batas. Kanker akan menyerang organ-organ tubuh yang bersebelahan hingga menyebar ke organ-organ lainnya dan akhirnya menyebabkan kematian (WHO 2018). Secara global kanker menduduki peringkat dua dunia untuk kasus penyebab kematian pada manusia (WHO 2018). Salah satu jenis kanker yang mematikan adalah kanker serviks. Kanker serviks muncul pada leher rahim wanita yang berfungsi sebagai pintu masuk menuju rahim dari vagina (Kemenkes RI 2017). Kemenkes RI (2017) menyebutkan bahwa kanker serviks merupakan salah satu penyakit paling mematikan bagi kaum wanita Indonesia, setidaknya tidak kurang dari 15.000 kasus setiap tahunnya. Kanker serviks menduduki peringkat empat dunia untuk kasus penyakit kanker pada wanita, tercatat 570.000 kasus baru pada tahun 2018 dengan persentase 90% kematian terjadi di negara berpenghasilan rendah dan menengah (WHO 2019).

Kematian akibat kanker dapat dikurangi jika kanker terdeteksi dan diobati sejak dini (WHO 2018). Teknologi terbaru deteksi dini kanker adalah dengan menggunakan pembelajaran mesin untuk klasifikasi yang menghasilkan prediksi kondisi pasien, sehingga bisa digunakan untuk pedoman awal pengobatan kanker (Kourou et al. 2015). Penelitian Diaz-Uriarte (2007) menyebutkan bahwa para peneliti biomedis memerlukan sebuah alat yang bisa memfasilitasi mereka dalam melakukan klasifikasi, sehingga dibuatlah aplikasi GeneSrF berbasis web khusus data *microarray* yang bisa menerapkan konteks klasifikasi pasien.

Tiga tahun terakhir, penelitian terkait pembuatan aplikasi klasifikasi sebenarnya sudah dilakukan, tetapi dikhususkan untuk data *microRNA*. *MicroRNA* bisa digunakan sebagai biomarker diagnostik untuk deteksi dini kanker sehingga bisa membantu perencanaan perawatan penderita kanker (Madhavan et al. 2013). Penelitian pembuatan

aplikasi klasifikasi kanker yang dimaksud adalah pada penelitian Anjarwati (2017) yang menghasilkan aplikasi berbasis web dalam melakukan klasifikasi profil fitur kanker ovarium menggunakan data *microRNA* dengan algoritma *Backpropagation*. Kemudian, penelitian Setianto (2018) juga menghasilkan aplikasi berbasis web dalam melakukan klasifikasi kanker payudara menggunakan data profil fitur *microRNA* dengan algoritma *Backpropagation* yang dioptimasi dengan *Stimulus-Sampling* dan Nguyen-Widrow. Dan penelitian Ningrum (2018) yang menghasilkan aplikasi berbasis web dalam melakukan klasifikasi kanker payudara menggunakan data *microRNA* dengan algoritma *Deep Neural Network* yang dioptimasi dengan *Batch Normalization*.

Tiga penelitian di atas yakni penelitian Anjarwati (2017), penelitian Setianto (2018) dan penelitian Ningrum (2018) menghasilkan aplikasi klasifikasi kanker, tetapi aplikasi yang dihasilkan masih memiliki kekurangan dalam hal visualisasi model klasifikasi yang dihasilkan. Hal ini dikarenakan, tiga penelitian tersebut lebih berfokus untuk peningkatan kinerja performa algoritma yang digunakan. Oleh karena itu, berdasarkan permasalahan tersebut akan dibangun sebuah aplikasi klasifikasi kanker serviks menggunakan data *microRNA* dengan algoritma *Random Forest* sebagai solusi baru untuk memfasilitasi peneliti biomedis pada domain kanker serviks. Algoritma *Random Forest* dipilih karena menghasilkan visualisasi model klasifikasi berupa pohon keputusan yang mudah dipahami peneliti biomedis non IT dan bisa menghitung *feature importance* yang bisa digunakan sebagai rekomendasi biomarker *microRNA* kanker serviks.

Aplikasi klasifikasi kanker serviks menggunakan data *microRNA* dengan algoritma *Random Forest* ini akan dibangun dengan model proses *waterfall*. Model proses *waterfall* lebih mudah digunakan (Sommerville 2011). Pada prinsipnya, model proses *waterfall* hanya boleh digunakan ketika persyaratannya ditentukan atau dipahami dengan baik dan tidak mungkin berubah secara radikal selama pengembangan sistem (Sommerville 2011). Dengan demikian, pada fase pengumpulan persyaratan aplikasi didasarkan pada analisis metodologi CRISP-DM yang teruji secara ilmiah sebagai model proses penambangan data yang handal, sehingga diperoleh persyaratan aplikasi yang spesifik dan tidak berubah-ubah. CRISP-DM menyediakan standar proses umum untuk pemecahan masalah dari bisnis atau unit penelitian dengan strategi penambangan data dan sekaligus berfungsi sebagai peta jalan yang bermanfaat untuk bergerak menuju pemenuhan tujuan penelitian (Larose and Larose 2014).

## 1.2. Rumusan Masalah

Berdasarkan latar belakang di atas maka dapat dirumuskan suatu permasalahan yaitu bagaimana membuat sebuah aplikasi untuk klasifikasi kanker serviks menggunakan data *microRNA* dengan algoritma *Random Forest* yang dibangun berdasarkan model proses *waterfall* dan pada fase pengumpulan persyaratan aplikasinya didasarkan pada analisis metodologi CRISP-DM.

## 1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai dalam penelitian ini adalah solusi baru untuk memfasilitasi peneliti biomedis pada domain kanker serviks dengan sebuah aplikasi. Aplikasi tersebut adalah aplikasi klasifikasi kanker serviks menggunakan data *microRNA* dengan algoritma *Random Forest Forest* yang dibangun berdasarkan model proses *waterfall* dan pada fase pengumpulan persyaratan aplikasinya didasarkan pada analisis metodologi CRISP-DM.

Manfaat dalam pembuatan aplikasi ini adalah terdapat visualisasi model klasifikasi dalam bentuk pohon keputusan, sehingga lebih mudah dipahami untuk peneliti biomedis yang tidak memiliki pemahaman dibidang IT. Selain itu, juga terdapat perhitungan *feature importance* yang bisa digunakan sebagai rekomendasi biomarker kepada para peneliti biomedis.

## 1.4. Ruang Lingkup

Dalam penyusunan skripsi ini perlu adanya batasan-batasan agar pembahasan lebih terarah dan tidak melebihi target yang diteliti. Batasan-batasan tersebut adalah:

1. Data yang digunakan sebagai dasar dalam melakukan klasifikasi adalah nilai ekspresi gen data *microRNA* yang diperoleh dari *paper* (Witten et al. 2010).
2. Aplikasi yang akan dibuat berbasis *web* dengan menggunakan HTML, CSS, JavaScript, bahasa pemrograman python dengan *framework* Django dan menggunakan sistem manajemen basis data MySQL dengan *tools* phpMyAdmin
3. Aplikasi ini dibangun dengan model proses *Waterfall* yang akan dilakukan sampai tahap pengujian.
4. Fase pengumpulan persyaratan akan didasarkan analisis metodologi CRISP-DM
5. Implementasi metodologi CRISP-DM dilakukan hingga fase evaluasi.
6. Fase analisa dan desain aplikasi akan dikembangkan dengan pendekatan *Object-Oriented Analysis and Design* (OOAD)

7. Klasifikasi hanya bisa dilakukan menggunakan data *microRNA* yang disimpan dalam format file csv.
8. Dalam membuat model klasifikasi pada aplikasi hanya menggunakan Algoritma *Random Forest*.

### **1.5. Sistematika Penulisan**

Sistematika penulisan yang digunakan dalam skripsi ini terbagi dalam beberapa pokok bahasan, yaitu:

#### **BAB I      PENDAHULUAN**

Bab ini memaparkan tentang latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan dalam penyusunan skripsi.

#### **BAB II     LANDASAN TEORI**

Bab ini memaparkan tinjauan pustaka dan dasar teori yang digunakan untuk penyelesaian masalah, pelaksanaan dan penyusunan skripsi.

#### **BAB III    METODOLOGI PENELITIAN**

Bab ini memaparkan tentang garis besar penyelesaian masalah skripsi, penjelasan singkat dari langkah-langkah yang akan dilakukan dan perhitungan manual dari algoritma yang akan digunakan.

#### **BAB IV    HASIL DAN PEMBAHASAN**

Bab ini menyajikan hasil dan pembahasan dari langkah-langkah yang dilalui berdasarkan metodologi penelitian dalam menyelesaikan permasalahan skripsi.

#### **BAB V     PENUTUP**

Bab ini menyajikan tentang kesimpulan dari pemaparan yang telah dijabarkan pada bab-bab sebelumnya dan saran pengembangan penelitian kedepan.