

**ANALISIS KOMPARATIF KETEPATAN HASIL DETEKSI KANKER
PROSTAT MENGGUNAKAN METODE *RANDOM FOREST* DAN *K-
NEAREST NEIGHBOR (KNN)***

Diajukan Sebagai Syarat Untuk Menyelesaikan

Pendidikan Program Strata-1 Pada

Jurusan Teknik Informatika



Oleh :

Retno Tri Aprillia

NIM : 09021382025160

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2023

LEMBAR PENGESAHAN SKRIPSI

ANALISIS KOMPARATIF KETEPATAN HASIL DETEKSI
KANKER PROSTAT MENGGUNAKAN METODE RANDOM
FOREST DAN K-NEAREST NEIGHBOR

Oleh:

Retno Tri Aprillia
NIM : 09021382025160

Palembang, 20 Oktober 2023

Pembimbing I,

Pembimbing II,



Dian Palupi Rini, M.Kom., Ph.D.
NIP. 19789223296942992



Annisa Darmawahyuni, M.Kom
NIP. 199006302023212044

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari Jumat tanggal 21 Oktober 2023 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Retno Tri Aprillia
NIM : 09021302025160
Judul : Analisis Komparatif Ketepatan Hasil Deteksi Kanker Prostat Menggunakan Metode *Random Forest* dan *K-Nearest Neighbor* (KNN)

Dan dinyatakan **LULUS**.

1. Penguji I




Kanda Januar Miraswan, M.T.
NIP. 199001092019031012

2. Pembimbing I



Dian Palupi Rini, M.Kom., Ph.D.
NIP. 19789223296942992

3. Pembimbing II



Annisa Darmawahyuni, M.Kom.
NIP. 199006302023212044

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
197812222006042003

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Retno Tri Aprillia
NIM : 09021382025160
Jurusan : Teknik Informatika
Program Studi : Teknik Informatika Bilingual
Fakultas : Ilmu Komputer
Angkatan : 2020

Dengan ini menyatakan bahwa:

1. Dalam penyusunan/penulisan proyek akhir harus bersifat orisinal dan tidak melakukan plagiarisme baik produk software/hardware..
2. Dalam penyelesaian proyek akhir dilaksanakan di Laboratorium Fakultas Ilmu Komputer Universitas Stiwijaya dan tidak diselesaikan atau dikerjakan oleh pihak lain diluar civitas akademik Fakultas Ilmu Komputer Universitas Stiwijaya

Demikian Pernyataan ini saya buat sebenar-benarnya dan saya bersedia diberikan sanksi apabila dikemudian hari pernyataan saya ini terbukti tidak benar yaitu:

1. Tidak mengikuti ujian Komprehensif atau tidak lulus ujian Komprehensif
2. Bersedia mengganti judul atau topik proyek akhir setelah mendapat persetujuan. dari pembimbing proyek akhir



Palembang, 20 Oktober 2023

1a,


Retno Tri Aprillia

09021382025160

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Retno Tri Aprillia
NIM : 09021302025160
Judul : Analisis Komparatif Ketepatan Hasil Deteksi Kanker
Prostat Menggunakan Metode *Random Forest* dan *K-Nearest Neighbor* (KNN)

Hasil Pengecekan *Software* Turnitin : 8%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya saya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, 20 Oktober 2023



Retno Tri Aprillia
NIM 09021382025160

MOTO DAN PERSEMBAHAN

*"Setidaknya kamu memiliki satu alasan kenapa kamu bertahan atas segalanya,
cari bila kamu tidak menemukannya"*

*"Dan Kami tidak membebani seseorang melainkan sesuai dengan
kesanggupannya."*

(QS. Al-Baqarah: 286)

*"Maka ketahuilah, tidak ada sesuatu yang harus dihormati di sisi Allah selain
orang-orang yang berilmu." (QS. Fatir: 28)*

Skripsi ini kapersambahkan kepada :

- Allah Subhanahu Wa Ta'ala
- Keluarga Saya
- Sahabat dan Teman-Teman Saya
- Dosen Pembimbing Akademik dan Skripsi
- Fakultas Ilmu Komputer
- Universitas Sriwijaya

ABSTRACT

Prostate cancer is a malignant disease that typically affects men and occurs in the prostate gland located beneath the bladder. Prostate cancer ranks as the second most common cancer in the United States and was the fifth in Indonesia with 13,563 cases in 2020. While the exact cause remains unknown, early detection of prostate cancer can be achieved through the use of machine learning and data mining techniques. Data were sourced from Kaggle, incorporating ten features. The Random Forest and K-Nearest Neighbor (KNN) methods were employed for classification, with PCA used to select eight components corresponding to the number of features in the training data. The research findings reveal that KNN outperformed Random Forest in classification performance. In KNN, the optimal parameter K was identified as 19, achieving an accuracy, precision, recall, and F1 score of 100%. In contrast, Random Forest attained an accuracy of 75%, with precision and recall at 85% and 75%, and an F1 score of 75%. These results indicate that KNN can classify data with higher accuracy. The findings demonstrate the efficacy of both methods in classifying prostate cancer patients.

Keywords : Prostate Cancer, Classification, K-Nearest Neighbor (KNN), Random Forest, PCA

Palembang, 20 Oktober 2023

Pembimbing I,

Pembimbing II,



Dian Palupi Rini, M.Kom., Ph.D.

Annisa Darmawahyuni, M.Kom

NIP. 19789223296942992

NIP. 1671147006900002

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.

NIP: 197812222006042003

ABSTRAK

Kanker prostat merupakan penyakit ganas yang umumnya menyerang pria dan terjadi pada kelenjar prostat di bawah kandung kemih. Kanker prostat menjadi penyakit kedua paling umum di Amerika Serikat dan berada di urutan kelima dengan jumlah kasus 13.563 pada tahun 2020 di Indonesia. Meskipun penyebab pasti belum diketahui, deteksi dini dari kanker prostat dapat dilakukan dengan menggunakan mesin pembelajaran dan *data mining*. Data diambil dari Kaggle dengan mengambil 10 fitur. Metode *Random Forest* dan *K-Nearest Neighbour* (KNN) digunakan untuk klasifikasi, dan PCA digunakan dengan memilih 8 komponen sesuai dengan jumlah fitur pada data latih. Hasil penelitian menunjukkan bahwa KNN menghasilkan kinerja klasifikasi yang lebih baik daripada *Random Forest*. Dalam KNN, parameter K optimal adalah 19, mencapai akurasi, presisi, *recall*, dan skor F1 sebesar 100%. Sementara itu, *Random Forest* mencapai akurasi 75%, presisi 85%, *recall* 75%, dan skor F1 75%. Hasil ini menunjukkan bahwa KNN mampu mengklasifikasikan data dengan tingkat keakuratan yang lebih tinggi. Hasil yang didapatkan menunjukkan bahwa kedua metode mampu mengklasifikasikan penderita penyakit kanker prostat.

Kata Kunci : Kanker Prostat, Klasifikasi, *K-Nearest Neighbor* (KNN), *Random Forest*, PCA

Palembang, 20 Oktober 2023

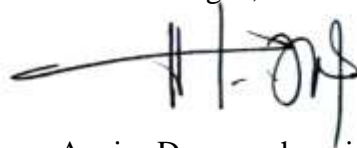
Pembimbing I,



Dian Palupi Rini, M.Kom., Ph.D.

NIP. 19789223296942992

Pembimbing II,



Annisa Darmawahyuni, M.Kom

NIP. 1671147006900002

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

KATA PENGANTAR

Puji syukur kepada Allah Subanahu Wa Ta'ala atas Rahmat dan karunia-Nya, penyusunan tugas akhir dengan judul “Analisis Komparatif Ketepatan Hasil Deteksi Kanker Prostat Menggunakan Metode *Random Forest* dan *K-Nearest Neighbor* (KNN)” dapat diselesaikan dengan baik. Tugas akhir ini disusun sebagai syarat dalam menyelesaikan studi Strata-1 program studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis ingin mengucapkan terima kasih kepada semua pihak yang telah membantu dan memberikan dukungan, bimbingan, kerjasama dan motivasi kepada penulis dalam menyelesaikan tugas akhir ini. Adapun penulis ingin mengucapkan terima kasih kepada:

1. Ayah penulis Alm Petrimaisyah yang selalu mendidik, dan mendoakan penulis walaupun sudah berpulang
2. Ibu Dian Palupi Rini, M.Kom., Ph.D. dan ibu Annisa Darmawahyuni, M.Kom., selaku dosen pembimbing tugas akhir yang telah memberikan arahan, bimbingan, serta meluangkan waktu, tenaga, dan pikirannya selama proses penyusunan tugas akhir ini.
3. Tamir Gading Hasibuan selaku orang yang suka memarahi penulis apabila penulis tidak makan, begadang, dan tidak menggunakan *skincare* selama pengerjaan tugas akhir.
4. Ibu penulis Sri Susdianah, B.Sc., mamas penulis Prayitno Fidiantoro, S.T., dan mbak penulis dr. Dian Destriyanah, Sp.N., yang selalu mendoakan dan memberikan dukungan kepada penulis.
5. Hanny Putri Gayatri, Febiyanti, Anissa Arrayyan, Silvi Agustina, Adrian Sutansaty, Aldo Haliansyah, dan semua teman-teman di kelas Bilingual B yang telah memberikan dukungan, hiburan, dan telah berproses selama perkuliahan.
6. Naqiyyah Nurrosyadah selaku sahabat penulis sejak SMA yang mau memberikan dukungan kepada penulis, walaupun dia setelah wisuda melupakan kepada penulis.

7. Gandi, Euis, Citra, Emak selaku sahabat penulis yang suka berkeluh kesah mengenai perkuliahan dan telah memberikan dukungan, serta memberikan semangat kepada penulis sejak semester III.
8. Pegawai Fakultas Ilmu Komputer Universitas Sriwijaya yang telah membantu penulis selama perkuliahan luring berlangsung.
9. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu selama masa perkuliahan dan penyusunan tugas akhir ini di Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari dalam penyusunan tugas akhir ini jauh dari sempurna baik dari segi penyusunan, bahasan, ataupun penulisannya. Oleh karena itu, penulis mengharapkan kritik dan saran yang sifatnya membantu. Semoga tugas akhir ini bermanfaat bagi pengembangan wawasan dan peningkatan ilmu pengetahuan di kalangan masyarakat luas serta dapat dijadikan kajian untuk penelitian lainnya.

Palembang, 20 Oktober 2023



Retno Tri Aprillia

DAFTAR ISI

| | |
|--|------|
| HALAMAN JUDUL..... | i |
| HALAMAN PENGESAHAN..... | ii |
| TANDA LULUS UJIAN SIDANG SKRIPSI | iii |
| HALAMAN PERNYATAAN BEBAS PLAGIAT | iv |
| HALAMAN MOTO DAN PERSEMBAHAN | v |
| ABSTRACT..... | vi |
| ABSTRAK | vii |
| KATA PENGANTAR | viii |
| DAFTAR ISI..... | x |
| DAFTAR TABEL..... | xiv |
| DAFTAR GAMBAR | xvii |
| DAFTAR LAMPIRAN..... | xxi |
| | |
| BAB I PENDAHULUAN | I-1 |
| 1.1 Pendahuluan | I-1 |
| 1.2 Latar Belakang Masalah | I-1 |
| 1.3 Rumusan Masalah | I-3 |
| 1.4 Tujuan Penelitian | I-3 |
| 1.5 Manfaat Penelitian | I-3 |
| 1.6 Batasan Masalah | I-4 |
| 1.7 Sistematika Penulisan | I-4 |
| 1.8 Kesimpulan..... | I-5 |
| | |
| BAB II KAJIAN LITERATUR | II-1 |
| 2.1 Pendahuluan | II-1 |
| 2.2 Landasan Teori | II-1 |
| 2.2.1 Kanker Prostat | II-1 |
| 2.2.2 Klasifikasi | II-2 |
| 2.2.3 <i>Decision Tree</i> | II-3 |

| | |
|--|--------|
| 2.2.4 <i>Random Forest</i> | II-5 |
| 2.2.5 <i>K-Nearest Neighbor</i> | II-7 |
| 2.2.6 <i>Principal Component Analysis(PCA)</i> | II-8 |
| 2.2.7 Evaluasi Model..... | II-9 |
| 2.3 Penelitian Terkait..... | II-10 |
| 2.4 Kesimpulan..... | II-12 |
| | |
| BAB III METODOLOGI PENELITIAN..... | III-1 |
| 3.1 Pendahuluan | III-1 |
| 3.2 Pengumpulan Data..... | III-1 |
| 3.2.1 Jenis dan Sumber Data | III-1 |
| 3.2.2 Metode Pengumpulan Data | III-2 |
| 3.3 Tahapan Penelitian | III-2 |
| 3.3.1 Kerangka Kerja | III-3 |
| 3.3.2 Pengumpulan Data | III-4 |
| 3.3.3 Kriteria Pengujian | III-4 |
| 3.3.4 Format Data Pengujian..... | III-4 |
| 3.3.5 Alat Bantu Penelitian | III-5 |
| 3.3.6 Pengujian Penelitian..... | III-5 |
| 3.3.7 Analisis Hasil Pengujian dan Kesimpulan | III-6 |
| 3.4 Metode Pengembangan Perangkat Lunak | III-7 |
| 3.4.1 Fase Insepsi | III-7 |
| 3.4.2 Fase Elaborasi | III-7 |
| 3.4.3 Fase Konstruksi..... | III-7 |
| 3.4.4 Fase Transisi | III-8 |
| 3.5 Manajemen Proyek Penelitian..... | III-8 |
| 3.6 Kesimpulan..... | III-11 |
| | |
| BAB IV PENGEMBANGAN PERANGKAT LUNAK | IV-1 |
| 4.1 Pendahuluan | IV-1 |
| 4.2 Fase Insepsi | IV-1 |

| | | |
|--|---|-------|
| 4.2.1 | Pemodelan Bisnis | IV-1 |
| 4.2.2 | Kebutuhan Sistem | IV-2 |
| 4.2.3 | Analisis dan Desain..... | IV-3 |
| 4.2.3.1 | Analisis dan Kebutuhan Perangkat Lunak | IV-3 |
| 4.2.3.2 | Analisis Data | IV-4 |
| 4.2.3.3 | Analisis <i>Random Forest</i> | IV-4 |
| 4.2.3.4 | Analisis <i>K-Nearest Neighbor</i> (KNN)..... | IV-11 |
| 4.2.3.5 | Desain Perangkat Lunak..... | IV-16 |
| 4.3 | Fase Elaborasi..... | IV-23 |
| 4.3.1 | Pemodelan Bisnis | IV-23 |
| 4.3.1.1 | Perancangan Data..... | IV-23 |
| 4.3.1.2 | Perancangan Antar Muka..... | IV-24 |
| 4.3.2 | Kebutuhan Sistem | IV-26 |
| 4.3.3 | Diagram..... | IV-26 |
| 4.3.3.1 | Diagram Aktivitas | IV-26 |
| 4.3.3.2 | Diagram Alur | IV-28 |
| 4.4 | Fase Kontruksi..... | IV-31 |
| 4.4.1 | Kebutuhan Sistem | IV-31 |
| 4.4.2 | Diagram Kelas..... | IV-31 |
| 4.4.3 | Implementasi | IV-32 |
| 4.4.3.1 | Implementasi Kelas..... | IV-32 |
| 4.4.3.2 | Implementasi Antar Muka..... | IV-34 |
| 4.5 | Fase Transisi | IV-35 |
| 4.5.1 | Pemodelan Bisnis | IV-35 |
| 4.5.2 | Rencana Pengujian | IV-35 |
| 4.5.3 | Implementasi | IV-37 |
| 4.6 | Kesimpulan..... | IV-38 |
| BAB V HASIL DAN ANALISIS PENELITIAN..... | | V-1 |
| 5.1 | Pendahuluan | V-1 |
| 5.2 | Data Hasil Percobaan..... | V-1 |

| | |
|--|----------|
| 5.2.1 Konfigurasi Percobaan | V-1 |
| 5.2.2 Data Hasil Konfigurasi..... | V-2 |
| 5.2.2.1 Hasil Percobaan Parameter <i>Random Forest</i> | V-2 |
| 5.2.2.2 Hasil Percobaan Parameter KNN..... | V-2 |
| 5.3 Analisis Hasil Pengujian..... | V-3 |
| 5.3.1 Analisis Hasil Pengujian Metode <i>Random Forest</i> | V-3 |
| 5.3.2 Analisis Hasil Pengujian Metode KNN | V-4 |
| 5.4 Kesimpulan..... | V-6 |
| BAB VI KESIMPULAN DAN SARAN | VI-1 |
| 6.1 Pendahuluan | VI-1 |
| 6.2 Kesimpulan..... | VI-1 |
| 6.3 Saran | VI-2 |

DAFTAR TABEL

| | |
|---|--------|
| Tabel II-1 <i>Confusion Matrix</i> | II-10 |
| Tabel III-1 Rancangan Tabel Hasil Pengujian Metode <i>Random Forest</i> | III-4 |
| Tabel III-2 Rancangan Tabel Hasil Pengujian Metode KNN | III-5 |
| Tabel III-3 Tabel <i>Work Breakdown Structure</i> (WBS) | III-10 |
| Tabel IV-1 Tabel Kebutuhan Fungsional..... | IV-2 |
| Tabel IV-2 Tabel Kebutuhan Non-Fungsional | IV-3 |
| Tabel IV-3 Tabel <i>Data Sample Random Forest</i> | IV-5 |
| Tabel IV-4 Tabel <i>Data Sample Threshold Random Forest</i> | IV-6 |
| Tabel IV-5 Tabel Rincian Total <i>Random Forest</i> | IV-7 |
| Tabel IV-6 Tabel Perhitungan Entropy dan Gain Pohon-1 | IV-7 |
| Tabel IV-7 Tabel Perhitungan Entropy Total Pohon-1 | IV-8 |
| Tabel IV-8 Tabel Perhitungan Entropy dan Gain Pohon-2..... | IV-9 |
| Tabel IV-9 Tabel Perhitungan Entropy Total Pohon-2..... | IV-9 |
| Tabel IV-10 Tabel Perhitungan Entropy dan Gain Pohon-3..... | IV-10 |
| Tabel IV-11 Tabel Data Uji <i>Random Forest</i> | IV-11 |
| Tabel IV-12 Tabel Data <i>Sample K-Nearest Neighbor</i> (KNN)..... | IV-12 |
| Tabel IV-13 Normalisasi Data <i>K-Nearest Neighbor</i> (KNN) | IV-13 |
| Tabel IV-14 Tabel Data Uji <i>K-Nearest Neighbor</i> (KNN) | IV-13 |
| Tabel IV-15 Tabel Perhitungan Jarak Data <i>Sample</i> dan Data Uji | IV-14 |
| Tabel IV-16 Tabel Perhitungan Jarak <i>Euclidian</i> | IV-15 |
| Tabel IV-17 Tabel Pengurutan Data | IV-15 |
| Tabel IV-18 Tabel Hasil Klasifikasi | IV-16 |
| Tabel IV-19 Tabel Definisi Aktor..... | IV-17 |
| Tabel IV-20 Tabel Definisi <i>Use Case</i> | IV-17 |
| Tabel IV-21 Tabel Skenario Menampilkan Hasil <i>Testing</i> | IV-18 |
| Tabel IV-22 Tabel Skenario <i>Load Data</i> | IV-20 |
| Tabel IV-23 Tabel Skenario Menampilkan Hasil Klasifikasi..... | IV-21 |

| | |
|---|-------|
| Tabel IV-24 Tabel Implementasi Kelas | IV-32 |
| Tabel IV-25 Rencana Pengujian <i>Use Case</i> Melakukan <i>Load Data</i> | IV-36 |
| Tabel IV-26 Rencana Pengujian <i>Use Case</i> Menampilkan Prediksi..... | IV-36 |
| Tabel IV-27 Rencana Pengujian <i>Use Case</i> Menampilkan Klasifikasi..... | IV-36 |
| Tabel IV-28 Tabel Pengujian <i>Use Case Load Data</i> | IV-37 |
| Tabel IV-29 Tabel Pengujian <i>Use Case</i> Hasil Prediksi | IV-37 |
| Tabel IV-30 Tabel Pengujian <i>Use Case</i> Hasil Klasifikasi dan Komparasi.... | IV-37 |
| Tabel V-1 Tabel Hasil Perhitungan Klasifikasi <i>Random Forest</i> | V-2 |
| Tabel V-2 Tabel Hasil Perhitungan Klasifikasi KNN..... | V-2 |

DAFTAR GAMBAR

| | |
|---|-------|
| Gambar II-1 Model <i>Decision Tree</i> | II-3 |
| Gambar II-2 Proses <i>Random Forest</i> | II-6 |
| Gambar III-1 Diagram Tahapan Penelitian..... | III-2 |
| Gambar III-2 Kerangka Kerja Penelitian | III-3 |
| Gambar III-3 Tahapan Pengujian Penelitian..... | III-6 |
| Gambar IV-1 Pohon-1 | IV-8 |
| Gambar IV-2 Pohon-2..... | IV-9 |
| Gambar IV-3 Pohon-3..... | IV-10 |
| Gambar IV-15 <i>Use Case Diagram</i> | IV-16 |
| Gambar IV-16 Rancangan Antarmuka Halaman Testing | IV-24 |
| Gambar IV-17 Rancangan Antarmuka Halaman <i>Load Data</i> | IV-25 |
| Gambar IV-18 Rancangan Antarmuka Halaman Klasifikasi..... | IV-25 |
| Gambar IV-19 Diagram Aktivitas Menampilkan Hasil Testing..... | IV-27 |
| Gambar IV-20 Diagram Aktivitas <i>Load Data</i> | IV-27 |
| Gambar IV-21 Diagram Aktivitas Menampilkan Hasil Komparasi | IV-28 |
| Gambar IV-22 <i>Sequence Diagram</i> Menampilkan Hasil Testing..... | IV-29 |
| Gambar IV-23 <i>Sequence Diagram</i> Memuat Data | IV-29 |
| Gambar IV-24 <i>Sequence Diagram</i> Menampilkan Hasil Komparasi | IV-30 |
| Gambar IV-25 Diagram Kelas | IV-32 |
| Gambar IV-26 Implementasi Antarmuka Halaman testing | IV-34 |
| Gambar IV-27 Implementasi Antarmuka Halaman <i>Load Data</i> | IV-34 |
| Gambar IV-28 Implementasi Antarmuka Halaman Komparasi..... | IV-35 |
| Gambar V-1 Grafik Hasil Klasifikasi Metode <i>Random Forest</i> | V-4 |
| Gambar V-2 Grafik Hasil Klasifikasi KNN..... | V-5 |
| Gambar V-3 Komparasi Hasil Klasifikasi <i>Random Forest</i> dan KNN | V-5 |

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini mendeskripsikan dasar dari penelitian yang akan dilakukan mengenai komparasi metode *Random Forest* dan *K-Nearest Neighbors* pada penyakit Kanker Prostat. Bab ini membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah yang diterapkan untuk memfokuskan cakupan penelitian pada area yang relevan dan mencegah terjadinya penyimpangan.

1.2 Latar Belakang Masalah

Kanker prostat merupakan penyakit keganasan sistem urogenital yang terjadi pada pria. Kanker ini terjadi pada prostat, kelenjar kecil di bawah kandung kemih yang menghasilkan cairan semen. Kanker prostat menjadi jenis kanker kedua yang paling sering terdiagnosis di Amerika Serikat setelah kanker paru-paru (Siegel, et al., 2020). Menurut Global Cancer Statistics, kanker prostat di Indonesia berada di urutan kelima dengan jumlah kasus sebanyak 13.563 pada tahun 2020. Angka tersebut menunjukkan bahwa kanker prostat juga menjadi masalah kesehatan yang cukup signifikan di Indonesia. Kanker ini terjadi pada prostat, kelenjar kecil di bawah kandung kemih yang menghasilkan cairan semen. Adapun penyebab dari kanker prostat belum ditemukan, namun kanker ini dapat dilakukan deteksi dini.

Kanker prostat dapat dideteksi dengan tes PSA (*Prostate Specific Antigen*) dan biopsi prostat, namun kedua metode ini memiliki kekurangan dan tingkat akurasi yang rendah. Dalam hal ini, beberapa penelitian telah dilakukan untuk mengembangkan metode pendeteksian kanker prostat. Salah satu penelitian yang dilakukan oleh (Ngatimin, et.al., 2018) yang menggunakan metode *decision Tree* dan *Naïve Bayes* dalam mendeteksi kanker prostat. Hasil penelitian menunjukkan bahwa akurasi tertinggi dihasilkan oleh *Naïve Bayes* sebesar 80,00% dan *Decision Tree* sebesar 76,67%.

Penelitian lain juga dilakukan oleh (Niza, et.al., 2019) yang membandingkan kinerja metode *Logistic Regression* dan *Artificial Neural Network* (ANN) dalam mendeteksi kanker prostat. Data yang digunakan meliputi umur pasien, hasil tes PSA, dan hasil pemeriksaan fisik lainnya. Hasil penelitian menunjukkan bahwa nilai akurasi ANN sebesar 88,57% lebih tinggi daripada *Logistic Regression* sebesar 85,71%. Penelitian lain oleh Jurnal IEEE Access (2021) menggunakan algoritma SVM untuk mendeteksi kanker prostat dari data citra MRI multimodal. Pada penelitian ini dihasilkan bahwa algoritma SVM menghasilkan akurasi sebesar 96,7% dalam membedakan antara kanker prostat dan jaringan sehat pada citra MRI multimodal.

Penelitian lain yang membandingkan kinerja KNN, *Naïve Bayes*, dan *Random Forest* dilakukan oleh (Rahmat, Yulianto., 2018) dalam mengklasifikasikan data pembeli kendaraan. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memberikan hasil yang lebih baik dalam klasifikasi data pembeli kendaraan dengan akurasi sebesar 90,9% dibandingkan dengan algoritma KNN dengan akurasi 85,3%, dan algoritma *Naïve Bayes* dengan akurasi 81,8%.

Berdasarkan pembahasan di atas, dapat dikatakan bahwa penggunaan metode *Random Forest* mampu menghasilkan nilai akurasi yang lebih baik. Metode *Random Forest* dikenal sebagai salah satu teknik klasifikasi yang optimal untuk mengatasi data dalam jumlah yang besar dan mampu mengatasi *overfitting* (Yuniarti et al, 2021). Selain itu, meskipun metode KNN mendapatkan nilai akurasi yang lebih rendah daripada metode *Random Forest*, KNN merupakan metode yang efektif untuk digunakan pada *data training* yang besar dan mampu memberikan hasil yang baik bahkan dengan jumlah *data training* yang terbatas (Wibowo et al., 2020).

Berdasarkan kelebihan dari *Random Forest* dan *K-Nearest Neighbor* (KNN), maka pada penelitian ini dibandingkan hasil deteksi kanker prostat menggunakan metode *Random Forest* dan *K-Nearest Neighbor* (KNN). Pada penelitian ini juga akan dilakukan pemilihan fitur yang tepat pada masing-masing metode menggunakan *Principal Component Analysis* (PCA). Hal ini dikarenakan bahwa seleksi fitur dengan PCA dapat meningkatkan nilai akurasi prediksi pada

kedua metode klasifikasi tersebut (Mubarok et.,al 2020). PCA memiliki kemampuan untuk mengurangi dimensi data tanpa menghilangkan kolom data yang penting, sehingga dapat mempercepat waktu komputasi dan meningkatkan akurasi klasifikasi. Kemudian, dilakukan evaluasi dan perbandingan akurasi hasil deteksi antara kedua metode tersebut untuk menentukan metode mana yang lebih efektif dan efisien dalam mendeteksi kanker prostat.

1.3 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana penerapan pemilihan fitur yang tepat pada metode Random Forest dan KNN dapat meningkatkan akurasi deteksi kanker prostat?
2. Bagaimana penerapan dan hasil perbandingan dari akurasi hasil deteksi kanker prostat menggunakan metode *Random Forest* dan *K-Nearest Neighbor* (KNN)?

1.4 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengimplementasikan penerapan metode *Random Forest* dan *K-Nearest Neighbor* (KNN) untuk mendeteksi kanker prostat.
2. Mengetahui hasil analisis secara komperatif oleh metode *Random Forest* dan *K-Nearest Neighbor* (KNN) untuk mendeteksi kanker prostat.

1.5 Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Sistem yang dihasilkan dapat berguna bagi bidang kedokteran untuk mendeteksi penderita yang mengalami penyakit kanker prostat.
2. Hasil penelitian dapat dijadikan sumber rujukan bagi peneliti lain untuk melanjutkan penelitian mengenai deteksi kanker prostat.

1.6 Batasan Masalah

Adapun batasan masalah penelitian ini adalah sebagai berikut:

1. Data yang digunakan pada penelitian ini adalah *Prostate Cancer* yang diambil dari Kaggle.
2. Metode yang digunakan untuk perhitungan jarak pada *K-Nearest Neighbor* (KNN) menggunakan *Euclidian Distance*.
3. Pada pemilihan fitur digunakan *Principal Component Analysis (PCA)* untuk dilakukan tahap *pre-processing* pada data.

1.7 Sistematika Penulisan

Sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini akan membahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah, serta sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan membahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi metode yang digunakan dan penelitian terkait yang relevan pada penelitian ini.

BAB III. METODELOGI PENELITIAN

Pada bab ini akan menjelaskan tahapan yang akan diterapkan pada penelitian menggunakan metode *Random Forest* dan *K-Nearest Neighbor* (KNN). Setiap rencana tahapan penelitian di deskripsikan dengan rinci yang mengacu pada suatu kerangka kerja. Selanjutnya akan dilakukan perancangan manajemen proyek pengembangan dalam pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan membahas mengenai tahapan yang dilakukan pada proses pengembangan perangkat lunak yang merupakan metode *Random Forest* dan *K-Nearest Neighbor* (KNN).

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini membahas hasil pengujian penelitian berdasarkan format pengujian yang sudah direncanakan. Selain itu, pada bab ini dibahas hasil analisis deteksi kanker prostat sebagai dasar kesimpulan dalam penelitian ini.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan penelitian dan juga saran yang diharapkan agar dapat menjadi sumber referensi untuk pengembangan penelitian lain dengan metode yang sama.

1.8 Sistematika Penulisan

Berdasarkan penjelasan di atas, maka dapat disimpulkan dalam penelitian ini telah dibahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

DAFTAR PUSTAKA

- Aditya. (2022). Machine learning models for data classification. *International Journal of Computer Science and Mobile Computing*, 11(2), 150-157.
- Arrieumerlou, Y. (2021). Principal Component Analysis For Feature Selection in Machine Learning: A review. *Artificial Intelligence Review*, 54(3), 2201-2237. doi: 10.1007/s10462-020-09822-8.
- Alimin, & Riyadi, S. (2020). Sistem Pengembangan Deteksi Kanker Prostat Berbasis Image Processing dengan Metode Convolutional Neural Network. *Jurnal Teknologi dan Sistem Komputer*, 8(2), 143-148.
- Darmawan, D. (2020). Klasifikasi Kanker Prostat Berdasarkan Tingkat Keganasan dan Stadium. *Jurnal Kedokteran dan Kesehatan Indonesia*, 11(2), 105-111. doi: 10.24853/jkk.2020.11.2.105-111
- Hanifah, N., & Arifin, A. Z. (2019). Decision Tree untuk Analisis Data Mining. *Jurnal SIMETRIS*, 10(1), 349-356.
- Hasan, I. K., Resmawan, & Ibrahim, J. (2022). Perbandingan K-Nearest Neighbor dan Random Forest dengan Seleksi Fitur Information Gain untuk Klasifikasi Lama Studi Mahasiswa. *Indonesian Journal of Applied Statistics*, 5(1), 58-66. <https://doi.org/10.13057/ijas.v5i1.58056>
- Huljanah, M., Rustam, Z., Utama, S., & Siswantining, T. (2019). Feature Selection using Random Forest Classifier for Predicting Prostate Cancer. *Journal of Physics: Conference Series*, 1280(1), 012027. doi: 10.1088/1742-6596/1280/1/012027
- IEEE Access. (2021). Support Vector Machine-based Classification of Multimodal Magnetic Resonance Imaging for Prostate Cancer Diagnosis. *IEEE Access*, 9, 53268-53280. doi: 10.1109/access.2021.3060315

- Kim, S., Lee, J., & Park, J. (2019). Anatomy of the prostate gland: histological features and their correlation with imaging findings. *Ultrasonography*, 38(2), 81-89. <https://doi.org/10.14366/usg.18035>
- Larasati, S., & Jaya, A. F. (2022). Analysis of classification models in machine learning for predicting heart disease. *Journal of Physics: Conference Series*, 1927(1), 012020. <https://doi.org/10.1088/1742-6596/1927/1/012020>
- Lee, J., Kim, S., Park, J., & Han, K. (2022). Stem cell markers in prostate cancer: an immunohistochemical study. *Pathology & Oncology Research*, 28(3), 1863-1869. <https://doi.org/10.1007/s12253-021-01212-w>
- Loupe, G. (2022). Understanding Random Forest. *Journal of Machine Learning Research*, 23(1), 1-18. <https://jmlr.org/papers/v23/21-818.html>
- Mubarok, H., Widyastuti, H., & Kurniawan, E. (2020). Penerapan *Principal Component Analysis* (PCA) Pada Seleksi Fitur Data Profil Resiko Kredit. *Jurnal Sains dan Teknologi Informasi*, 5(1), 23-28.
- Ngatimin, A. S., Sudiby, Y., & Arifin, A. Z. (2018). Comparison of decision tree and naive bayes method for prostate cancer detection. *Journal of Physics: Conference Series*, 971(1), 012006. doi: 10.1088/1742-6596/971/1/012006
- Niza, C. M., Suherman, S., & Setiawan, N. A. (2019). Performance Comparison of Logistic Regression and Artificial Neural Network for Prostate Cancer Detection. *Journal of Physics: Conference Series*, 1204(1), 012031. doi: 10.1088/1742-6596/1204/1/012031
- Pratama, A. M. (2021). A comparative study of decision tree and Naïve Bayes for classification problem. *Journal of Physics: Conference Series*, 1818(1), 012030. <https://doi.org/10.1088/1742-6596/1818/1/012030>
- Pratiwi, C. A., & Irhamah. (2021). Klasifikasi Microarray “Prostate Cancer” Menggunakan Metode Fuzzy Support Vector Machine (FSVM)-Genetic

- Algorithm. *Jurnal Ilmiah Matematika, Statistika, dan Komputasi*, 7(1), 23-34. <https://doi.org/10.1017/jimstk.2021.17>
- Rahmat, D., & Yulianto, E. (2018). Classification of Vehicle Purchasing Data Using K-Nearest Neighbor, Naïve Bayes and Random Forest. *Journal of Physics: Conference Series*, 1028(1), 012073. doi: 10.1088/1742-6596/1028/1/012073
- Sari, R. K., & Sari, M. (2019). Comparison of classification accuracy using decision tree and naive bayes algorithm in determining employee promotion. *Journal of Physics: Conference Series*, 1317(1), 012029. doi: 10.1088/1742-6596/1317/1/012029
- Safri, M., Supriatna, I., & Khairuddin, N. (2020). Anatomy of the prostate gland: A review. *Journal of Medical Science and Clinical Research*, 8(6), 784-788. <https://doi.org/10.18535/jmscr/v8i6.84>
- Siegel, R.L., Miller, K.D., Fuchs, H.E., & Jemal, A. (2020). Cancer Statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1), 7-30. doi: 10.3322/caac.21590
- Suyanto. (2020). Metode Evaluasi Klasifikasi. *Jurnal Sistem Informasi Bisnis*, 4(1), 51-62.
- Tanaka, M. (2017). The histology of the human prostate gland: a review. *Japanese Journal of Clinical Oncology*, 47(9), 773-777. <https://doi.org/10.1093/jjco/hyx089>
- Wibowo, A.T., Saikhu, R., & Soelaiman, R. (2016). Implementation Of Image Contained Spam Detection Algorithm Using SVM And Random Forest.
- Widodo, D. S. (2022). Data mining classification technique for customer classification based on purchase behavior. *International Journal of Innovative Technology and Exploring Engineering*, 11(1), 129-134. <https://doi.org/10.35940/ijitee.J1022.1186S21>