

**IMPLEMENTASI *BOOTSTRAP AGGREGATING*
PADA METODE *DECISION TREE* DAN REGRESI LOGISTIK
UNTUK KLASIFIKASI KANKER SERVIKS**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Jurusan Matematika pada Fakultas MIPA**

Oleh:
SITI HASMA WATI
08011381823076



**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SRIWIJAYA
2022**

HALAMAN PENGESAHAN

IMPLEMENTASI *BOOTSTRAP AGGREGATING* PADA METODE *DECISION TREE* DAN REGRESI LOGISTIK UNTUK KLASIFIKASI KANKER SERVIKS

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
di Jurusan Matematika pada Fakultas MIPA

Oleh:

SITI HASMA WATI
08011381823076

Indralaya, Maret 2022
Pembimbing Utama

Pembimbing Kedua



Endang Sri K. M.Si.
NIP.197702082002122003



Dr. Yulia Resti, M.Si.
NIP. 197307191997022001

Mengetahui,



PERNYATAAN KEASLIAN KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama mahasiswa : Siti Hasma Wati
NIM : 08011381823076
Fakultas/Jurusan : MIPA/Matematika

Menyatakan bahwa skripsi ini adalah hasil karya saya sendiri dan karya ilmiah ini belum pernah diajukan sebagai pemenuhan persyaratan untuk memperoleh gelar kesarjanaan strata satu (S1) dari Universitas Sriwijaya maupun perguruan tinggi lain.

Semua informasi yang dimuat dalam skripsi ini yang berasal dari penulis lain baik yang dipublikasikan atau tidak telah diberikan penghargaan dengan mengutip nama sumber penulis secara benar. Semua isi dari skripsi ini sepenuhnya menjadi tanggung jawab saya sebagai penulis.

Demikianlah surat pernyataan ini saya buat dengan sebenarnya.

Indralaya, 29 Maret 2022



Penulis

HALAMAN PERSEMBAHAN

MOTTO

“Hidup itu seperti mengendarai sepeda.

Untuk menjaga keseimbanganmu, kamu harus terus bergerak.”

- Albert Einstein -

Saya rasa memang betul, kita harus terus bergerak untuk menjaga
keseimbangan dan tidak jatuh, sehingga kita bisa sampai
ke tempat tujuan yang hendak kita capai...

Bergerak untuk maju..

Bergerak untuk berusaha..

Bergerak untuk belajar..

Bergerak untuk bersyukur..

Bergerak untuk meraih cita-cita..

Bergerak untuk membahagiakan orang yang dicintai..

Dan bergerak untuk menuju sang Pencipta dengan mendapatkan Ridha-Nya..

" Ayo *take action*. Skripsi tidak akan selesai apabila kamu hanya
memikirkannya. Jangan menjadi pemalas dan penakut karena ada
harapan besar keluarga di atas pundakmu."

Skripsi ini kupersembahkan kepada:

- 1. Allah SWT**
- 2. Kedua Orangtua dan (Almh) Nenekku**
- 3. Saudaraku**
- 4. Keluarga Besarku**
- 5. Semua Dosen dan Guruku**
- 6. Sahabatku**
- 7. Almamaterku**

KATA PENGANTAR

Assalamu ’alaikum Warahmatullahi Wabarakatuh

Puji syukur kehadirat Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Implementasi Bootstrap Aggregating pada Metode Decision Tree dan Regresi Logistik untuk Klasifikasi Kanker Serviks**”. Shalawat beserta salam semoga selalu tercurahkan kepada baginda kita Nabi Muhammad SAW beserta keluarga, sahabat dan para pengikutnya hingga akhir zaman.

Penulis menyadari bahwa dalam proses penyusunan skripsi ini banyak rintangan dan tantangan yang dihadapi. Namun dengan kesabaran dan ketekunan yang dilandasi rasa tanggung jawab, penulis dapat menyelesaikan skripsi ini meskipun masih banyak kekurangannya. Penulis juga menyadari bahwa penulisan ini tidak dapat terselesaikan tanpa dukungan dari berbagai pihak. Dengan segala hormat dan penuh kerendahan hati, penulis mengucapkan terima kasih kepada kedua orang tua tercinta, Bapak **Nasarudin** dan Ibu **Siti Rofa** atas segala do'a terbaik, perhatian, semangat serta nasihat yang selalu diberikan kepada penulis.

Penulis juga mengucapkan terima kasih kepada Ibu **Dr. Yulia Resti, M.Si.** selaku Dosen Pembimbing Utama dan Ibu **Endang Sri Kresnawati, M.Si.** selaku Dosen Pembimbing Kedua sekaligus Dosen Pembimbing Akademik yang telah bersedia meluangkan waktu untuk memberikan bimbingan, saran, nasihat, motivasi, serta pengetahuan yang sangat berarti bagi penulis.

Penulis juga mengucapkan terimakasih kepada:

1. Bapak **Hermansyah, Ph.D.** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
2. Bapak **Drs. Sugandi Yahdin, M.M.** selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
3. Ibu **Dr. Dian Cahyawati, M.Si.** selaku Sekretaris Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.
4. Ibu **Dr. Ir. Herlina Hanum, M.Si.** dan Ibu **Des Alwine Zayanti, M.Si.** selaku Dosen Pembahas Skripsi yang telah memberikan waktu serta saran dan tanggapan yang bermanfaat bagi penulis dalam penggerjaan skripsi ini.
5. Ibu **Oki Dwipurwani, M.Si.** selaku Ketua Seminar dan Bapak **Drs. Endro Setyo Cahyono, M.Si.** selaku Sekretaris Seminar yang telah membantu pelaksanaan seminar dan sidang sarjana penulis.
6. **Seluruh Dosen Jurusan Matematika FMIPA UNSRI** atas ilmu serta didikan yang telah diberikan kepada penulis selama menempuh pendidikan di Jurusan Matematika FMIPA UNSRI.
7. Bapak **Irwansyah** dan Ibu **Hamida** yang telah banyak membantu penulis dalam hal administrasi di Jurusan Matematika FMIPA UNSRI.
8. Keluarga besarku terutama (almh) **Nenekku** serta saudaraku **Muhammad Nasir Sulas, S.T.** dan **Muhammad Saing, S.Kom.** atas do'a, motivasi, dan dukungan yang diberikan.

9. Sahabat-sahabatku, **Ica, Mifta, Irma, Anggi, Rara, Desi, Santi, Sukma, Fasa, Aina, dan Natasya** serta seluruh teman-teman Angkatan **2018** atas dukungan, semangat, bantuan, dan kebersamaannya.
10. Kakak tingkat Angkatan **2016** dan **2017** yang telah memberikan ilmu dan pengalamannya serta adik tingkat Angkatan **2019** dan **2020** yang telah memberikan dukungan dan doa serta semangat.
11. **Semua Pihak** yang tidak dapat penulis sebutkan satu persatu. Semoga segala kebaikan yang diberikan mendapat balasan dari Allah SWT.

Penulis berharap agar skripsi ini dapat bermanfaat bagi semua pihak yang membutuhkan terutama mahasiswa/mahasiswi Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sriwijaya.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Indralaya, Maret 2022

Penulis

**IMPLEMENTATION OF BOOTSTRAP AGGREGATING
ON DECISION TREE AND LOGISTIC REGRESSION METHODS
FOR CLASSIFICATION OF CERVIC CANCER**

**By:
SITI HASMA WATI
08011381823076**

ABSTRACT

Cervical cancer is one of the most common diseases suffered by women in the world. Global Burden Cancer recorded the incidence of cervical cancer worldwide in 2020 as many as 604,127 cases while the death rate from cervical cancer was recorded at 341,831 cases. The high mortality rate from cervical cancer is also related to the delay in diagnosis of the disease. Therefore, it is necessary to have a study that discusses the classification process to diagnose cervical cancer patients with high accuracy results. The purpose of this study is to classify cervical cancer based on pap smear cell image extraction using the decision tree method and logistic regression with the implementation of bootstrap aggregating (bagging) and without the implementation of bagging. The data used in this study is a dataset of Pap smear cell image extraction of cervical cancer 7 classes originating from Herlev University Hospital. The results of this study indicate that the implementation of bagging can improve the performance of a single method for cervical cancer classification. Classification using the decision tree method resulted in accuracy, precision, recall, and specificity of 85.87%, 56.09%, 54.88%, and 91.48%, respectively. While the classification using the decision tree method with bagging resulted in accuracy, precision, recall, and specificity of 87.89%, 61.85%, 61.59%, and 92.61%, respectively. Classification using logistic regression method resulted in accuracy, precision, recall, and specificity of 89.75%, 68.61%, 68.45%, and 93.75%, respectively. While the classification using the logistic regression method with bagging produces the best accuracy, precision, recall, and specificity, which are 90.53%, 70.13%, 70.70%, and 94.25%, respectively.

Keywords: Classification, Cervical Cancer, Bootstrap Aggregating, Decision Tree, Logistic Regression

**IMPLEMENTASI *BOOTSTRAP AGGREGATING*
PADA METODE *DECISION TREE* DAN REGRESI LOGISTIK
UNTUK KLASIFIKASI KANKER SERVIKS**

Oleh:

SITI HASMA WATI

08011381823076

ABSTRAK

Kanker serviks merupakan salah satu penyakit yang paling banyak diderita oleh perempuan di dunia. *Global Burden Cancer* mencatat angka kejadian kanker serviks di seluruh dunia pada tahun 2020 sebanyak 604.127 kasus sedangkan angka kematian akibat kanker serviks tercatat sebanyak 341.831 kasus. Tingginya angka kematian akibat kanker serviks juga berkaitan dengan keterlambatan diagnosis penyakit tersebut. Oleh karena itu, diperlukan adanya suatu penelitian yang membahas proses klasifikasi untuk melakukan diagnosis pasien penderita kanker serviks dengan hasil akurasi tinggi. Tujuan penelitian ini yaitu mengklasifikasi kanker serviks berdasarkan ekstraksi citra sel *pap smear* menggunakan metode *decision tree* dan regresi logistik dengan implementasi *bootstrap aggregating (bagging)* dan tanpa implementasi *bagging*. Data yang digunakan pada penelitian ini yaitu *dataset* hasil ekstraksi citra sel *pap smear* kanker serviks 7 kelas yang berasal dari Rumah Sakit Universitas Herlev. Hasil penelitian ini menunjukkan bahwa implementasi *bagging* dapat meningkatkan kinerja metode tunggal untuk klasifikasi kanker serviks. Klasifikasi menggunakan metode *decision tree* menghasilkan akurasi, presisi, *recall*, dan spesifitas masing-masing sebesar 85,87%, 56,09%, 54,88%, dan 91,48%. Sedangkan klasifikasi menggunakan metode *decision tree* dengan *bagging* menghasilkan akurasi, presisi, *recall*, dan spesifitas masing-masing sebesar 87,89%, 61,85%, 61,59%, dan 92,61%. Klasifikasi menggunakan metode regresi logistik menghasilkan akurasi, presisi, *recall*, dan spesifitas masing-masing sebesar 89,75%, 68,61%, 68,45%, dan 93,75%. Sedangkan klasifikasi menggunakan metode regresi logistik dengan *bagging* menghasilkan akurasi, presisi, *recall*, dan spesifitas terbaik yaitu masing-masing sebesar 90,53%, 70,13%, 70,70%, dan 94,25%.

Kata Kunci: Klasifikasi, Kanker Serviks, *Bootstrap Aggregating*, *Decision Tree*, Regresi Logistik

DAFTAR ISI

| | |
|--|-------------|
| HALAMAN JUDUL | i |
| HALAMAN PENGESAHAN | ii |
| HALAMAN PERSEMBAHAN | iii |
| KATA PENGANTAR..... | iv |
| ABSTRACT..... | vii |
| ABSTRAK | viii |
| DAFTAR ISI | ix |
| DAFTAR TABEL..... | xii |
| DAFTAR GAMBAR..... | xiii |
| DAFTAR LAMPIRAN | xiv |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Perumusan Masalah | 5 |
| 1.3 Batasan Masalah..... | 5 |
| 1.4 Tujuan Penelitian..... | 6 |
| 1.5 Manfaat Penelitian | 6 |
| BAB II TINJAUAN PUSTAKA..... | 7 |
| 2.1 Kanker Serviks | 8 |
| 2.2 Data Mining | 8 |
| 2.3 <i>Machine Learning</i> | 9 |
| 2.4 Klasifikasi | 10 |
| 2.5 Diskritisasi Data | 10 |
| 2.6 <i>Bootstrap Aggregating</i> | 11 |
| 2.7 <i>Decision Tree</i> | 14 |
| 2.8 Tabel Kontingensi | 17 |
| 2.9 Analisis Regresi..... | 17 |
| 2.10 Regresi Logistik | 18 |
| 2.11 Regresi Logistik Multinomial..... | 18 |
| 2.11.1 Estimasi Parameter | 20 |

| | |
|---|-----------|
| 2.11.2 Pengujian Parameter | 25 |
| 2.11.3 Uji Kesesuaian Model | 26 |
| 2.11.4 <i>Odds Ratio</i> | 27 |
| 2.12 <i>Confusion Matrix</i> | 28 |
| BAB III METODOLOGI PENELITIAN | 31 |
| 3.1 Waktu dan Tempat Penelitian | 31 |
| 3.1.1 Waktu Penelitian | 31 |
| 3.1.2 Tempat Penelitian | 31 |
| 3.2 Data Penelitian | 31 |
| 3.3 Metode Penelitian..... | 33 |
| BAB IV HASIL DAN PEMBAHASAN | 37 |
| 4.1 Deskripsi Data | 37 |
| 4.2 Diskritisasi Data | 39 |
| 4.3 Partisi Data..... | 40 |
| 4.4 <i>Decision Tree</i> | 41 |
| 4.5 <i>Decision Tree</i> dengan Implementasi <i>Bootstrap Aggregating</i> | 50 |
| 4.6 Regresi Logistik | 58 |
| 4.6.1 Estimasi Parameter | 58 |
| 4.6.2 Pengujian Parameter | 63 |
| 4.6.3 Uji Kesesuaian Model | 65 |
| 4.6.4 <i>Odds Ratio</i> | 66 |
| 4.6.5 Model Regresi Logistik Multinomial | 68 |
| 4.7 Regresi Logistik dengan Implementasi <i>Bootstrap Aggregating</i> | 71 |
| 4.7.1 Estimasi Parameter | 71 |
| 4.7.2 Pengujian Parameter | 72 |
| 4.7.3 Uji Kesesuaian Model | 73 |
| 4.7.4 <i>Odds Ratio</i> | 75 |
| 4.7.5 Model Regresi Logistik Multinomial | 76 |
| 4.8 Analisis Hasil..... | 80 |
| BAB V KESIMPULAN DAN SARAN | 83 |
| 5.1 Kesimpulan | 83 |

| | |
|-----------------------------|-----------|
| 5.2 Saran..... | 84 |
| DAFTAR PUSTAKA | 85 |
| LAMPIRAN | 90 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1 Tabel kontingensi | 17 |
| Tabel 2.2 <i>Confusion matrix</i> untuk n kelas | 28 |
| Tabel 2.3 Kategori nilai kinerja model | 30 |
| Tabel 3.1 Karakteristik sel <i>pap smear</i> | 31 |
| Tabel 3.2 Variabel prediktor penelitian | 32 |
| Tabel 4.1 <i>Dataset</i> sel <i>pap smear</i> | 37 |
| Tabel 4.2 Tabulasi data variabel X1 terhadap variabel Y | 38 |
| Tabel 4.3 Nilai maksimum, minimum, dan <i>range</i> setiap variabel prediktor | 39 |
| Tabel 4.4 Hasil diskritisasi data..... | 40 |
| Tabel 4.5 Data latih | 40 |
| Tabel 4.6 Data uji | 41 |
| Tabel 4.7 Nilai <i>entropy</i> dan <i>gain</i> untuk penentuan <i>root node</i> | 43 |
| Tabel 4.8 Nilai <i>entropy</i> dan <i>gain</i> untuk penentuan <i>internal node</i> pertama dari variabel K/C (<i>node 1.1</i>) | 45 |
| Tabel 4.9 Nilai <i>entropy</i> untuk penentuan <i>internal node</i> pertama dari variabel KerneShort (<i>node 1.1.1</i>) | 46 |
| Tabel 4.10 <i>Confusion matrix</i> metode <i>decision tree</i> | 48 |
| Tabel 4.11 Nilai TP, TN, FP, dan FN metode <i>decision tree</i> | 49 |
| Tabel 4.12 Data hasil <i>bootstrap</i> pertama | 51 |
| Tabel 4.13 Nilai <i>entropy</i> untuk penentuan <i>root node</i> | 51 |
| Tabel 4.14 Nilai <i>entropy</i> untuk penentuan <i>internal node</i> pertama dari variabel K/C (<i>node 1.1</i>) metode <i>decision tree</i> dengan <i>bagging</i> | 53 |
| Tabel 4.15 Nilai <i>entropy</i> untuk penentuan <i>internal node</i> pertama dari variabel KerneShort (<i>node 1.1.1</i>) metode <i>decision tree</i> dengan <i>bagging</i> | 54 |
| Tabel 4.16 <i>Confusion matrix</i> metode <i>decision tree</i> dengan <i>bagging</i> | 57 |
| Tabel 4.17 Nilai TP, TN, FP, dan FN metode <i>decision tree</i> dengan <i>bagging</i> | 57 |
| Tabel 4.18 Uji simultan..... | 64 |
| Tabel 4.19 Uji Hosmer dan Lemeshow | 66 |
| Tabel 4.20 <i>Confusion matrix</i> metode regresi logistik | 69 |
| Tabel 4.21 Nilai TP, TN, FP, dan FN metode regresi logistik | 70 |
| Tabel 4.22 Uji simultan metode regresi logistik dengan <i>bagging</i> | 72 |
| Tabel 4.23 Uji Hosmer dan Lemeshow metode regresi logistik dengan <i>bagging</i> | 74 |
| Tabel 4.24 <i>Confusion matrix</i> metode regresi logistik dengan <i>bagging</i> | 79 |
| Tabel 4.25 Nilai TP, TN, FP, dan FN metode regresi logistik dengan <i>bagging</i> ... | 79 |
| Tabel 4.26 Perbandingan tingkat ketepatan klasifikasi | 81 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1 <i>Bootstrap sampling</i> | 13 |
| Gambar 2.2 Alur klasifikasi menggunakan metode <i>decision tree</i> dengan implementasi <i>bagging</i> | 13 |
| Gambar 2.3 Alur klasifikasi menggunakan metode regresi logistik dengan implementasi <i>bagging</i> | 14 |
| Gambar 2.4 Struktur pohon keputusan | 15 |
| Gambar 4.1 Sebaran frekuensi kelas kanker serviks | 38 |
| Gambar 4.2 Pohon keputusan <i>root node</i> | 44 |
| Gambar 4.3 Pohon keputusan <i>node 1.1</i> | 45 |
| Gambar 4.4 Pohon keputusan <i>node 1.1.1</i> | 47 |
| Gambar 4.5 Pohon keputusan <i>root node</i> menggunakan metode <i>decision tree</i> dengan <i>bagging</i> | 52 |
| Gambar 4.6 Pohon keputusan <i>node 1.1</i> menggunakan metode <i>decision tree</i> dengan <i>bagging</i> | 53 |
| Gambar 4.7 Pohon keputusan <i>node 1.1.1</i> menggunakan metode <i>decision tree</i> dengan <i>bagging</i> | 55 |

DAFTAR LAMPIRAN

| | |
|--|-----|
| Lampiran 1 Tabulasi data setiap variabel prediktor terhadap variabel respon | 90 |
| Lampiran 2 Interval dari setiap variabel prediktor untuk diskritisasi data | 96 |
| Lampiran 3 Perhitungan <i>entropy</i> dan <i>gain</i> untuk penentuan <i>node</i> 1.1 | 98 |
| Lampiran 4 Perhitungan <i>entropy</i> dan <i>gain</i> untuk penentuan <i>node</i> 1.1.1 | 102 |
| Lampiran 5 Model pohon keputusan..... | 106 |
| Lampiran 6 Perhitungan <i>entropy</i> dan <i>gain</i> untuk penentuan <i>node</i> 1.1 metode <i>decision tree</i> dengan <i>bagging</i> | 118 |
| Lampiran 7 Perhitungan <i>entropy</i> dan <i>gain</i> untuk penentuan <i>node</i> 1.1.1 metode <i>decision tree</i> dengan <i>bagging</i> | 122 |
| Lampiran 8 Model pohon keputusan metode <i>decision tree</i> dengan <i>bagging</i> | 126 |
| Lampiran 9 Hasil klasifikasi menggunakan metode <i>decision tree</i> dengan <i>bagging</i> | 133 |
| Lampiran 10 Iterasi <i>Newton Raphson</i> metode regresi logistik..... | 133 |
| Lampiran 11 Hasil estimasi parameter, statistik uji wald, dan <i>odds ratio</i> metode regresi logistik | 134 |
| Lampiran 12 Hasil estimasi parameter, statistik uji wald, dan <i>odds ratio</i> metode regresi logistik dengan <i>bagging</i> | 141 |
| Lampiran 13 Hasil klasifikasi menggunakan metode regresi logistik dengan <i>bagging</i> | 149 |
| Lampiran 14 Tabel <i>chi square</i> | 150 |
| Lampiran 15 Tabel distribusi normal..... | 151 |

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker serviks merupakan salah satu penyakit yang paling banyak diderita oleh perempuan di dunia. *Global Burden Cancer* (Globocan) mencatat angka kejadian kanker serviks di seluruh dunia pada tahun 2020 sebanyak 604.127 kasus sedangkan angka kematian akibat kanker serviks tercatat sebanyak 341.831 kasus (Global Burden Cancer, 2021). Data yang bersumber dari Globocan menunjukkan angka kejadian kanker serviks di Indonesia pada tahun 2018 menempati urutan ke-2 dengan persentase sebesar 10,7% dan angka kematian menempati urutan ketiga sebesar 10,3% (Yulia, 2019). Data yang bersumber dari Rumah Sakit Kanker Dharmais pada tahun 2018 menunjukkan bahwa kasus kanker terbanyak adalah kanker payudara sebesar 19,18%, kanker serviks sebesar 10,69%, dan kanker paru-paru sebesar 9,89% (Pangribowo, 2019).

Tingginya angka kematian akibat kanker serviks juga berkaitan dengan keterlambatan diagnosis dan pemeriksaan penyakit tersebut, karena kanker serviks sering kali tidak teridentifikasi sampai penyakit itu berkembang lebih lanjut (World Health Organization, 2022). Ada beberapa tes untuk deteksi dini kanker serviks, namun yang paling populer di Indonesia ada 2 yakni tes *Pap Smear* dan Inspeksi Visual Asam Asetat (IVA). *Pap Smear* merupakan tes untuk mendeteksi ada atau tidaknya perubahan sel-sel rahim yang beresiko menjadi kanker. Sel-sel rahim tersebut diamati menggunakan mikroskop dan didiagnosis oleh ahli secara manual. Analisis hasil pemeriksaan *pap smear* secara manual memiliki kelemahan

yaitu membutuhkan banyak waktu dan rentan terhadap kesalahan (Hussain *et al.*, 2020). Suatu sistem diagnosis yang akurat untuk hasil pemeriksaan *pap smear* akan sangat bermanfaat untuk mengatasi kelemahan tersebut. Maka dari itu diperlukan adanya suatu penelitian yang membahas proses klasifikasi untuk melakukan diagnosis pasien penderita kanker serviks dengan hasil akurasi tinggi sehingga diharapkan dapat menyamakan hasil diagnosis dari tenaga medis.

Terdapat banyak metode dalam melakukan klasifikasi data yang memberikan hasil yang akurat dan baik, salah satunya yaitu metode *Decision Tree*. *Decision tree* merupakan metode klasifikasi yang menggunakan struktur pohon keputusan untuk mengklasifikasi suatu data. Hasil klasifikasi menggunakan *decision tree* sangat mudah dipahami karena menyajikan informasi dengan cara yang sederhana yaitu divisualisasikan dalam bentuk pohon (Patel *et al.*, 2012). Metode *decision tree* bersifat non-parametrik sehingga tidak harus memenuhi asumsi-asumsi klasik metode klasifikasi seperti normalitas dan multikolinearitas (Hastuti, 2016).

Metode lainnya yang dapat digunakan untuk melakukan klasifikasi yaitu regresi logistik. Regresi logistik merupakan suatu metode analisis yang digunakan untuk mengetahui hubungan antara satu atau lebih variabel prediktor dengan variabel respon yang bertipe data kategori (Hosmer *et al.*, 2013). Regresi logistik juga dapat digunakan untuk mengklasifikasi data menggunakan model yang telah diperoleh (Yudissanta & Ratna, 2012).

Decision tree memiliki kelemahan yaitu ketidakstabilan model yang dihasilkan, artinya jika terdapat perubahan kecil dalam data latih maka berubah pula struktur pohnnya (Khamidah *et al.*, 2018). Begitupula dengan regresi logistik,

metode pengklasifikasian dengan regresi logistik memberikan pendugaan parameter yang tidak stabil, artinya jika terdapat perubahan dalam data latih maka menyebabkan perubahan yang signifikan pada model (Aqiqi, 2013). Oleh karena itu, diperlukan suatu metode untuk memperbaiki stabilitas tanpa menurunkan ketepatan klasifikasi yaitu *bootstrap aggregating (bagging)*.

Bagging merupakan salah satu metode yang menggunakan prinsip *ensemble method* yaitu metode yang menggunakan kombinasi dari beberapa model (Fitriah *et al.*, 2012). Pada pengklasifikasian *bagging*, model dibentuk dari beberapa *dataset* baru yang diperoleh dengan pengambilan sampel secara acak (*bootstrap*). Hasil klasifikasi akhir diperoleh dengan *voting* dari hasil klasifikasi masing-masing model (Quinlan, 2006). *Bagging* dirancang untuk memperbaiki stabilitas dan meningkatkan ketepatan klasifikasi yang dihasilkan oleh metode tunggal, *bagging* pertama kali dikenalkan oleh Breiman pada tahun 1994 (Hasibuan *et al.*, 2019).

Terdapat beberapa penelitian sebelumnya yang menggunakan *bagging* diantaranya yaitu, Suniantara *et al.* (2019) melakukan penelitian untuk mengklasifikasikan berat badan bayi lahir dengan menerapkan *bagging* pada metode regresi logistik ordinal. Pada penelitian tersebut, nilai akurasi yang diperoleh dengan implementasi *bagging* meningkat sebesar 13,3% dari akurasi yang diperoleh hanya dengan metode regresi logistik ordinal sebesar 63,37%. Pada penelitiannya, Mirqotussa'adah *et al.* (2017) menggunakan *bagging* pada algoritma C4.5 untuk mendiagnosis diabetes. Dari hasil penelitian tersebut, dengan menerapkan *bagging* untuk prediksi berbasis *ensemble* pada algoritma C4.5 dapat meningkatkan akurasi sebesar 1,18% dari akurasi awal 68,61%.

Subasi *et al.* (2019) melakukan penelitian untuk mengklasifikasi aktivitas manusia berbasis *smartphone* dengan menerapkan *bagging* pada metode *random tree*. Hasil dari penelitian tersebut menunjukkan bahwa pengklasifikasian dengan menerapkan *bagging* dapat meningkatkan akurasi sebesar 9,81%. Pada penelitiannya, Gaikwad & Thool (2015) menggunakan *bagging* pada metode *Partial Decision Tree* (PART) untuk mendeteksi intrusi. Dari hasil penelitian tersebut, penerapan *bagging* dapat meningkatkan akurasi sebesar 0,6% dari akurasi awal 77,7901%. Penelitian sebelumnya menunjukkan bahwa implementasi *bagging* dapat meningkatkan kinerja suatu metode tunggal.

Penelitian mengenai klasifikasi citra sel *pap smear* kanker serviks menggunakan *dataset* Herlev pernah dilakukan oleh beberapa peneliti sebelumnya diantaranya yaitu, Dewi *et al.*, (2020) melakukan pengklasifikasian menggunakan metode *Naïve Bayes*, *Sample Bootstrapping*, dan *Weighted PCA*. Pada penelitian tersebut diperoleh akurasi sebesar 97,3% untuk 2 kelas dan 85,7% untuk 7 kelas. Sholik (2017) melakukan pengklasifikasian menggunakan *dataset* yang sama menggunakan metode *Fuzzy k-Nearest Neighbor* menghasilkan akurasi sebesar 91,59% untuk 2 kelas dan 67,89% untuk 7 kelas. Pada penelitiannya, Hemalatha & Rani (2016) melakukan klasifikasi citra sel *pap smear* kanker serviks menggunakan metode *Multi Layer Perceptron*. Dari hasil penelitiannya diperoleh nilai akurasi, presisi, dan *recall* masing-masing sebesar 85,05%, 60,72%, dan 78,94%. Penelitian sebelumnya menunjukkan bahwa tingkat ketepatan klasifikasi dalam pengklasifikasian citra sel *pap smear* kanker serviks 7 kelas lebih rendah dibandingkan 2 kelas.

Berdasarkan penelitian-penelitian sebelumnya, penulis tertarik melakukan penelitian mengenai implementasi *bagging* yang diharapkan dapat meningkatkan kinerja metode *decision tree* dan regresi logistik untuk klasifikasi kanker serviks. Hasil kinerja diukur berdasarkan tingkat ketepatan klasifikasi yang pada penelitian ini dibatasi pada akurasi, presisi, *recall*, dan spesifisitas.

1.2 Perumusan Masalah

Rumusan masalah pada penelitian ini yaitu sebagai berikut.

1. Bagaimana pengklasifikasian kanker serviks menggunakan metode *decision tree* dan regresi logistik?
2. Bagaimana pengklasifikasian kanker serviks menggunakan metode *decision tree* dan regresi logistik dengan implementasi *bagging*?
3. Bagaimana tingkat ketepatan dalam klasifikasi kanker serviks menggunakan metode *decision tree* dan regresi logistik?
4. Bagaimana tingkat ketepatan dalam klasifikasi kanker serviks menggunakan metode *decision tree* dan regresi logistik dengan implementasi *bagging*?
5. Apakah *bagging* dapat meningkatkan kinerja metode *decision tree* dan regresi logistik pada klasifikasi kanker serviks?

1.3 Batasan Masalah

Batasan masalah penelitian ini yaitu sebagai berikut.

1. Penelitian ini menggunakan *dataset* hasil ekstraksi citra sel *pap smear* kanker serviks yang berasal dari Rumah Sakit Universitas Herlev.

2. Data dipartisi menjadi 80% data latih (733 citra) dan 20% data uji (184 citra).
3. Tingkat ketepatan klasifikasi dibatasi pada akurasi, presisi, *recall*, dan spesifisitas.

1.4 Tujuan Penelitian

Tujuan penelitian ini yaitu sebagai berikut.

1. Mengetahui pengklasifikasian kanker serviks menggunakan metode *decision tree* dan regresi logistik.
2. Mengetahui pengklasifikasian kanker serviks menggunakan metode *decision tree* dan regresi logistik dengan implementasi *bagging*.
3. Menghitung tingkat ketepatan dalam klasifikasi kanker serviks menggunakan metode *decision tree* dan regresi logistik.
4. Menghitung tingkat ketepatan dalam klasifikasi kanker serviks menggunakan metode *decision tree* dan regresi logistik dengan implementasi *bagging*.
5. Mengetahui apakah ada peningkatan kinerja metode *decision tree* dan regresi logistik dengan implementasi *bagging* pada klasifikasi kanker serviks.

1.5 Manfaat Penelitian

Manfaat penelitian ini yaitu sebagai berikut.

1. Dapat membantu tenaga medis terkait informasi dalam melakukan klasifikasi diagnosis pada pasien penderita kanker serviks.

2. Sebagai bahan referensi untuk penelitian selanjutnya.
3. Diharapkan dapat meningkatkan ilmu pengetahuan terutama bagi para pembaca dan masyarakat mengenai pengklasifikasian kanker serviks.
4. Diharapkan dapat meningkatkan wawasan dan kemampuan peneliti dalam menyelesaikan karya tulis ilmiah.

DAFTAR PUSTAKA

- Andriani, A. (2013). Sistem Pendukung Keputusan Berbasis Decision Tree Dalam Pemberian Beasiswa Studi Kasus : Amik “ BSI Yogyakarta .” *Seminar Nasional Teknologi Informasi Dan Komunikasi 2013 (SENTIKA 2013)*, 163–168.
https://repository.bsi.ac.id/index.php/unduh/item/48930/Sentika_2013Anik-Andriani.pdf
- Aqiqi, N. (2013). *Aplikasi Bagging untuk Meningkatkan Ketepatan Klasifikasi Pda Regresi Multinomial Logistik*. Universitas Jember.
- Bahety, A. (2014). *Extension and Evaluation of ID3-Decision Tree Algorithm*. 2(1), 1–8.
- Bhardwaj, R., & Vatta, S. (2013). Implementation of ID3 Algorithm. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(6), 845–851. www.ijarcsse.com
- Bland, J. M., & Altman, D. G. (2000). The odds ratio. *Bmj*, 320(7247), 1468. <https://doi.org/10.1136/bmj.320.7247.1468>
- Cancer Council Victoria. (2020). *Cervical Cancer*. Melbourne: Cancer Council Victoria. https://www.cancervic.org.au/cancer-information/types-of-cancer/cervical_cancer/cervical-cancer-overview.html
- Charisa, T. J. (2021). Penerapan Metode Ensemble untuk Klasifikasi Kanker Serviks Menggunakan Algoritma Support Vector Machine, Multilayer Perceptron, dan K-Nearest Neighbor. In *Repository.Unsri.Ac.Id*. Universitas Sriwijaya.
- Dash, R., Paramguru, R. L., & Dash, R. (2011). Comparative Analysis of Supervised and Unsupervised Discretization Techniques. *International Journal of Advances in Science and Technology*, 2(3), 29–37. https://www.researchgate.net/profile/Rajashree_Dash/publication/266058863_Comparative_Analysis_of_Supervised_and_Unsupervised_Discretization_Techniques/links/55727c6b08aeacff1ffacde8.pdf
- Defiyanti, S. (2017). Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik. *CITEE*, 256–261.
- Dewi, Y. N., Rianto, H., Riana, D., & Siregar, J. (2020). Integrasi Metode Sample Bootstrapping dan Weighted Principal Component Analisys (PCA) Untuk Meningkatkan Performa Naïve Bayes Pada Citra Tunggal Pap Smear. *Inti Nusa Mandiri*, 14(2), 133–138.
- Dewi, Y. N., & Sariyati, F. A. (2019). Metode Sample Bootstrapping Untuk Meningkatkan Performa Algoritma Naive Bayes Pada Citra Tunggal Pap Smear. *Jurnal Teknik Informatika*, 12(1), 1–10. <https://doi.org/10.15408/jti.v12i1.11031>

- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap* (D. R. Cox, D. V. Hinkley, N. Reid, D. B. RUBIN, & B. W. Silverman (eds.); 1st ed.). Springer Science Business Media, B.V.
- Fajri, M. (2015). Splitting Rule Dan Penerapan Bagging Pada Pohon Klasifikasi. In *Fakultas Matematika Dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember*.
- Fara, N. J. A. (2020). *Klasifikasi Nodul Payudara (Mammae) Berbasis Ciri Tepi Pda Citra Hasil Ultrasonografi (USG) di RSUD Tugurejo Semarang Menggunakan Scilab* [Universitas Islam Negeri Walisongo Semarang]. https://www.slideshare.net/maryamkazemi3/stability-of-colloids%0Ahttps://barnard.edu/sites/default/files/inline/student_user_guide_for_spss.pdf%0Ahttp://www.ibm.com/support%0Ahttp://www.spss.com/sites/dm-book/legacy/ProgDataMgmt_SPSS17.pdf%0Ahttps://www.n
- Fitriah, W. W., Mashuri, M., & Irhamah. (2012). Faktor-Faktor yang Mempengaruhi Keparahan Korban Kecelakaan Lalu Lintas di Kota Surabaya dengan Pendekatan Bagging Regresi Logistik Ordinal. *Jurnal Sains Dan Seni ITS*, 1(1), 253–258.
- Gaikwad, D. P., & Thool, R. C. (2015). Intrusion Detection System using Bagging with Partial Decision Tree Base Classifier. *Procedia Computer Science*, 49(1), 92–98. <https://doi.org/10.1016/j.procs.2015.04.231>
- Garillo-Manliguez, C. A. (2016). Generalized Confusion Matrix for Multiple Classes. *ResearchGate*, 5–7. <https://doi.org/10.13140/RG.2.2.31150.51523>
- Global Burden Cancer. (2021). Cervix uteri. In *The Global Cancer Observatory*.
- Gogtay, N. J., Deshpande, S. P., & Thatte, U. M. (2017). Principles of Regression Analysis. *Journal of Association of Physicians of India*, 65, 48–52. <https://doi.org/10.2307/2984136>
- Hasibuan, M. J. A., Rusgiyono, A., & Safitri, D. (2019). Pemodelan Produk Domestik Regional Bruto (PDRB) Di Provinsi Jawa Tengah Menggunakan Bootstrap Aggregating Multivariate Adaptive Regression Splines (Bagging Mars). *Jurnal Gaussian*, 8(1), 139–148. <https://doi.org/10.14710/j.gauss.v8i1.26628>
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. In *Springer Series in Statistics* (Second). https://doi.org/10.1111/j.1751-5823.2009.00095_18.x
- Hastuti, Y. (2016). KLASIFIKASI KARAKTERISTIK MAHASISWA UNIVERSITAS COKROAMINOTO PALOPO MENGGUNAKAN METODE NAÏVE BAYES DAN DECISION TREE. *Jurnal Dinamika*, 07(02), 34–41.
- Hemalatha, K., & Rani, D. K. U. (2016). Improvement of Multi Layer Perceptron Classification on Cervical Pap smear data with Feature Extraction. *International Journal of Innovative Research in Science, Engineering, and*

- Technology*, 5(12), 20419–20424. <https://doi.org/10.15680/IJIRSET.2016.0512024>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression Second Edition* (2nd ed.). John Wiley & Sons.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression Third Edition* (3rd ed.). John Wiley & Sons. <https://doi.org/10.2307/2532419>
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme Learning Machine: Theory and Applications. *Neurocomputing*, 70(1–3), 489–501. <https://doi.org/https://doi.org/10.1016/j.neucom.2005.12.126>
- Hussain, E., Mahanta, L. B., Das, C. R., & Talukdar, R. K. (2020). A comprehensive study on the multi-class cervical cancer diagnostic prediction on pap smear images using a fusion-based decision from ensemble deep convolutional neural network. *Tissue and Cell*, 65. <https://www.sciencedirect.com/science/article/abs/pii/S0040816619304872?via%3Dihub>
- Jantzen, J., Norup, J., Dounias, G., & Bjerregaard, B. (2005). Pap-smear Benchmark Data For Pattern Classification. *Proc. NiSIS 2005, Albufeira, Portugal*, 1–9.
- Junaedi, H., Budianto, H., Maryati, I., & Melani, Y. (2011). Data Transformation Pada Data Mining. *Prosiding Konferensi Nasional Inovasi Dalam Desain Dan Teknologi-IDeaTech*, 7, 93–99.
- Kastawan, P. W., Wiharta, D. M., & Sudarma, I. M. (2018). Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil. *Majalah Ilmiah Teknologi Elektro*, 17(3), 371–376. <https://doi.org/10.24843/mite.2018.v17i03.p11>
- Khamidah, F. S. N., Hapsari, D. P., & Nugroho, H. (2018). Implementasi Fuzzy Decision Tree Untuk Prediksi Gagal Ginjal Kronis. *INTEGER: Journal of Information Technology*, 3(1), 19–28. <https://doi.org/10.31284/j.integer.2018.v3i1.155>
- Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling with Applications in R. In *Springer* (Vol. 26). http://appliedpredictivemodeling.com/s/Applied_Predictive_Modeling_in_R.pdf
- Larose, D. T. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. In *Wiley Interscience* (Vol. 9780470908). John Wiley and Sons. <https://doi.org/10.1002/9781118874059>
- Lutfi, M. (2019). Implementasi Metode K-Nearest Neighbor dan Bagging Untuk Klasifikasi Mutu Produksi Jagung. *Agromix*, 10(2), 130–137. <https://doi.org/10.35891/agx.v10i2.1636>
- Mirqotussa'adah, Muslim, M. A., Sugiharti, E., Prasetyo, B., & Alimah, S. (2017). Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa

- Diabetes. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 8(2), 135–143. <https://doi.org/10.24843/lkjiti.2017.v08.i02.p07>
- Naqa, I. El, & Murphy, M. J. (2015). *What Is Machine Learning?* Cham Springer. https://doi.org/https://doi.org/10.1007/978-3-319-18305-3_1
- Pangribowo, S. (2019). Beban Kanker di Indonesia. *Pusat Data Dan Informasi Kementerian Kesehatan RI*, 1–16.
- Patel, B. N., Prajapati, S. G., & Lakhtaria, D. K. I. (2012). Efficient Classification of Data Using Decision Tree. *Bonfring International Journal of Data Mining*, 2(1), 6–12. <https://doi.org/10.9756/bijdm.1098>
- Prasetyo, R. T., & Pratiwi. (2015). Penerapan Teknik Bagging Pada Algoritma Klasifikasi Untuk Mengatasi Ketidakseimbangan Kelas Dataset Medis. *Jurnal Informatika*, II(2), 395–403. <https://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/118>
- Puspitasari, A. M., Ratnawati, D. E., & Widodo, A. W. (2018). Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(2), 802–810. <http://j-ptiik.ub.ac.id>
- Quinlan, J. R. (2006). Bagging, Boosting, and C4.5. *Proceedings of the National Conference on Artificial Intelligence*.
- Riana, D., Widyantoro, D. H., Latifah, T., & Mengko, R. (2013). Ekstraksi dan Klasifikasi Tekstur Citra Sel Nukleus Pap Smear. *Jurnal TICOM*, 1(3), 62–70.
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review Paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Setyawan, D. A., & Fathicah, C. (2018). Pengembangan Metode Decision Tree dengan Diskritisasi Data dan Splitting Atribut Menggunakan Hierarchical Clustering dan Dispersion Ratio. *JUTI: Jurnal Ilmiah Teknologi Informasi*, 18(2), 179–187.
- Sholik, M. (2017). *Klasifikasi Sel Serviks Pada Citra Pap Smear Berdasarkan Fitur Bentuk Deskriptor Regional dan Fitur Tekstur Uniform Rotated Local Binary Pattern*. Institut Teknologi Sepuluh November.
- Softia, D., Yuniarti, D., & Goejantoro, R. (2015). Analisis Regresi Eksponensial. In *Jurnal Eksponensial* (Vol. 6, Issue 2, pp. 57–64).
- Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing and Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>

- Somantri, O., Sasmito, G. W., & Sungkar, M. S. (2015). Optimalisasi Neural Network dengan Bootstrap Aggregating (Bagging) untuk Penentuan Prediksi Harga Listrik. *Scientific Journal of Informatics*, 1(2), 185–192. <https://doi.org/10.15294/sji.v1i2.4025>
- Subasi, A., Fllatah, A., Alzobidi, K., Brahimi, T., & Sarirete, A. (2019). Smartphone-Based Human Activity Recognition Using Bagging and Boosting. *Procedia Computer Science*, 163, 54–61. <https://doi.org/10.1016/j.procs.2019.12.086>
- Suniantara, I. K. P., Putra, I. G. E. W., & Suwardika, G. (2019). Peningkatan Ketepatan Klasifikasi dengan Metode Bootstrap Aggregating pada Regresi Logistik Ordinal. *INTENSIF: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi*, 3(1), 32–42. <https://doi.org/10.29407/intensif.v3i1.12587>
- Sykes, A. O. (1993). An Introduction to Regression Analysis. *Coase-Sandor Working Paper Series in Law and Economics*, 20. <https://doi.org/10.1002/9781118267912.ch6>
- Turban, E., Aronson, J. E., & Liang, T.-P. (2005). Decision Support Systems and Intelligent Systems. In *Prentice-Hall of India Private Limited* (7th ed.). Asoke K. Ghosh. <https://doi.org/10.1002/9780470755891.ch11>
- Williams, J. G. (2013). Bootstrap Sampling is With Replacement: A Comment on Ayillon et al. (2011). *River Research and Applications*, 29(3), 399–401. <https://doi.org/10.1002/rra.1605>
- World Health Organization. (2022). *Cervical Cancer*. [https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-\(hpv\)-and-cervical-cancer](https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer)
- Yudissanta, A., & Ratna, M. (2012). Analisis Pemakaian Kemoterapi pada Kasus Kanker Payudara dengan Menggunakan Metode Regresi Logistik Multinomial (Studi Kasus Pasien di Rumah Sakit “X” Surabaya). *Jurnal Sains Dan Seni ITS*, 1(1), 112–117. http://ejurnal.its.ac.id/index.php/sains_seni/article/view/1269
- Yulia, D. (2019). Capaian Pelaksanaan Program Deteksi Dini Kanker Serviks Melalui Pemeriksaan IVA Tes di Kota Jambi Tahun 2015-2018. *UGM Public Health Symposium*, 7659.