

**KLASIFIKASI *IMAGE* TUMBUHAN OBAT SIRIH HIJAU DAN
SIRIH MERAH MENGGUNAKAN METODE *DECISION TREE***

Skripsi

Oleh

**DEWI LESTARI
1817051041**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

ABSTRACT

KLASIFIKASI *IMAGE* TUMBUHAN OBAT SIRIH HIJAU DAN SIRIH MERAH MENGGUNAKAN METODE *DECISION TREE*

By

DEWI LESTARI

Medicinal plants or plants that grow wild in the yard can be used as substitutes for drugs, which are typically obtained from a pharmacy. There are many types of medicinal plants, including betel leaf. With the increasing use of technology, it is possible to classify the leaves of green and red betel plants based on their features. The data used in this study are 1000 green betel images and 1000 red betel images. The total number of images used in the study is 2000. The Decision Tree classifier is based on the use of six texture features: dissimilarity, correlation, homogeneity, contrast, ASM, and energy. The percentage of training data used in a prediction model is 60%:40%, 65%:35%, 70%:30%, 75%:25%, and 80%:20%. Confusion metric is used to measure the results of the classification test. The Decision Tree method achieved an average accuracy of 98%, precision 98.3%, recall 97.1%, and f1-score is 97.5%. The Naïve Bayes method yields an average accuracy 92.1%, precision 87.5%, recall 97.8%, and F1 score 92.3%. The Decision Tree Method is more accurate than the Naïve Bayes Method.

Keywords: Classification, Medicinal plants, GLCM, Decision Tree, Confusion Metric.

ABSTRAK

KLASIFIKASI *IMAGE* TUMBUHAN OBAT SIRIH HIJAU DAN SIRIH MERAH MENGGUNAKAN METODE *DECISION TREE*

Oleh

DEWI LESTARI

Tumbuhan obat atau lebih disingkat dengan sebutan TOGA merupakan tanaman hasil budidaya rumahan maupun tanaman yang tumbuh secara liar pada pekarangan rumah yang berguna sebagai pengganti obat-obatan kimia. Daun sirih merupakan salah satu jenis tumbuhan obat yang memiliki banyak jenis. Dengan perkembangan teknologi yang semakin pesat, tentunya dapat digunakan untuk mengklasifikasikan daun tumbuhan obat sirih hijau dan sirih merah. Data yang digunakan yaitu 1000 data image sirih hijau dan 1000 data image sirih merah dengan total keseluruhan data sebanyak 2000 data image. Model klasifikasi Decision Tree dilakukan dengan menggunakan 6 parameter feature ekstraksi tekstur Gray Level Co-Occurance Matric (GLCM) yaitu dissimilarity, correlation, homogeneity, contrast, ASM, dan energy. Perbandingan persentasi data latih dan data uji yang digunakan ialah 60%:40%, 65%:35%, 70%:30%, 75%:25%, dan 80%:20%. Confusion matriks digunakan untuk mengukur hasil pengujian klasifikasi. Hasil pengujian dengan metode Decision Tree diperoleh nilai rata-rata akurasi sebesar 98%, presisi 98,3%. recall 97,1% dan f1-score 97,5%. Adapun hasil dari pengujian dengan metode Naïve Bayes didapatkan hasil rata-rata accuracy sebesar 92,1%, precission 87,5%. recall 97,8% dan F1 Score 92,3%. Dari hasil perbandingan tersebut dapat disimpulkan bahwa metode Decision Tree lebih unggul dibandingkan dengan metode Naïve Bayes.

Kata Kunci: Klasifikasi, Tumbuhan Obat, GLCM, Decision Tree, Confusion Matriks

**KLASIFIKASI *IMAGE* TUMBUHAN OBAT SIRIH HIJAU DAN
SIRIH MERAH MENGGUNAKAN METODE *DECISION TREE***

Oleh

DEWI LESTARI

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA KOMPUTER**

Pada

**Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

Judul Skripsi : **KLASIFIKASI *IMAGE* TUMBUHAN OBAT SIRIH HIJAU DAN SIRIH MERAH MENGGUNAKAN METODE *DECISION TREE***

Nama Mahasiswa : **Dewi Lestari**

Nomor Pokok Mahasiswa : 1817051041

Program Studi : S1 Ilmu Komputer

Fakultas : Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing



Rizky Prabowo, M.Kom.
NIP 19880807 201903 1 011

2. Ketua Jurusan Ilmu Komputer

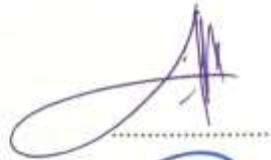


Didik Kurniawan, S.Si., M.T.
NIP 19800419 200501 1 004

MENGESAHKAN

1. Tim Penguji

Ketua : Rizky Prabowo, M.Kom.



Penguji 1 : Prof. Admi Syarif, Ph.D.




Penguji 2 : Bambang Hermanto, S.Kom., M.Cs.



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam




Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T.
NIP. 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 21 Juni 2022

PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul **"Klasifikasi Image Tumbuhan Obat Sirih Hijau dan Sirih Merah Menggunakan Metode Decision Tree"** merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, 21 Juni 2022



DEWI LESTARI
NPM.1817051041

RIWAYAT HIDUP



Penulis dilahirkan di Negeri Ratu, Sungkai Utara, Lampung Utara pada tanggal 23 Desember 2000, sebagai anak pertama dari dua bersaudara dari pasangan Bapak Wawan Gunawan dan Ibu Lisnawati. Penulis menempuh pendidikan formal pertama kali di Pendidikan Taman Kanak-kanak (TK) An-Nur diselesaikan pada tahun 2006, kemudian melanjutkan Sekolah Dasar (SD) diselesaikan di SD Negeri Ratu pada tahun 2012.

Penulis melanjutkan pendidikannya di Sekolah Menengah Pertama diselesaikan di SMP Negeri 1 Sungkai Utara pada tahun 2015, dan Sekolah Menengah Atas (SMA) diselesaikan di SMA Negeri 1 Sungkai Utara pada tahun 2018.

Pada Tahun 2018, penulis terdaftar sebagai mahasiswa Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur Penerimaan Mahasiswa Perluasan Akses Pendidikan (PMPAP). Selama menjadi mahasiswa, penulis melakukan beberapa kegiatan di antaranya:

1. Melaksanakan Karya Wisata Ilmiah (KWI) di Desa Way Bungur, Lampung Selatan, pada bulan Desember tahun 2018.
2. Penulis aktif di organisasi mahasiswa ROIS FMIPA sebagai anggota bidang Kajian dan Keumatan pada tahun 2018
3. Penulis aktif di Organisasi Himpunan Mahasiswa Komputer (HIMAKOM) sebagai anggota bidang Badan Khusus pada tahun 2020.

4. Penulis juga aktif dalam kepanitian yang diadakan dilingkungan kampus FMIPA UNILA seperti Penerimaan Mahasiswa Baru Ilmu Komputer (PRINTER) 2019, Penerimaan Mahasiswa Baru Tingkat Fakultas pada tahun 2020, Pekan Raya Jurusan (PRJ) HIMAKOM 2020.
5. Pada tahun 2021 penulis melakukan Kerja Praktik di Radar Lampung TV
6. Pada tahun 2021 penulis juga melakukan Kuliah Kerja Nyata di Desa Padang Ratu, Kecamatan Sungkai Utara, Kabupaten Lampung Utara.

PERSEMBAHAN

Puji dan syukur saya ucapkan kepada Allah Subhanahu Wa Ta'ala atas segala rahmat dan ridho-Nya yang selalu memberikan keyakinan, kekuatan, kesabaran serta kelancaran kepadaku sehingga skripsi ini dapat diselesaikan.

Kupersembahkan karya ini untuk:

Mama dan Papa yang selalu memberikan segalanya kepadaku. Terima kasih atas doa, kasih sayang, perhatian, dukungan, pengorbanan, serta hal lainnya yang kalian berikan dan tak akan terbalaskan.

Kepada adikku serta keluarga besar yang selalu memberikan doa dan dukungan semangat kepadaku.

Sahabat-sahabatku, terima kasih telah menemaniku, mendukungku, dan selalu memberikan kebahagiaan dalam hidupku.

Keluarga Besar Ilmu Komputer 2018

Almamater Tercinta Universitas Lampung

MOTTO

"Bekerjalah untuk akhiratmu seolah-olah kamu akan mati esok hari, dan bekerjalah untuk kehidupan duniamu seolah-olah kamu akan hidup selamanya."

-Ali bin Abi Thalib -

"Kamu tidak harus menjadi hebat untuk memulai, tetapi kamu harus mulai untuk menjadi hebat."

- Zig Ziglar -

"Ketika segalanya menjadi sulit, berhentilah sejenak dan lihat ke belakang dan lihat seberapa jauh kamu telah melangkah. Jangan lupa betapa berharganya itu. Kamu adalah bunga yang paling indah, lebih dari siapapun di dunia ini."

- Kim Taehyung -

"Hidup itu sulit, dan segala sesuatunya tidak selalu berjalan dengan baik, tetapi kita harus berani dan melanjutkan hidup kita."

- Suga -

"Ambition is the first step to success, the second step is action"

-Anonim-

SANWACANA

Puji syukur penulis ucapkan kehadirat Allah SWT, karena atas rahmat dan hidayah-Nya skripsi ini dapat diselesaikan. Skripsi ini dengan judul “*Klasifikasi Image Tunbuan Obat Sirih Hijau dan Sirih Merah Menggunakan Metode Decision Tree*” adalah salah satu syarat untuk memperoleh gelar Sarjana Komputer di Universitas Lampung.

Sebenarnya skripsi ini tidak terlepas dari dukungan banyaknya pihak, sehingga pada kesempatan ini terima kasih disampaikan kepada semua pihak yang telah membantu dan membimbing selama berlangsungnya proses pengerjaan hingga terselesainya skripsi ini.

Ungkapan terima kasih ini disampaikan kepada :

1. Kedua orang tuaku tercinta, Papah Wawan Gunawan dan Mamah Lisnawati yang senantiasa mengiringi langkahku dengan segala dukungan dan doa, serta tiada hentinya memberikan nasihat, material, bimbingan serta curahan kasih sayang.
2. Adekku tersayang Devi Darmawan yang selalu memberikan bantuan, doa, perhatian dan dukungan selama ini, semoga kita bisa menjadi putri yang membanggakan kedua orangtua.
3. Seluruh keluarga besar yang telah membantu, mendukung dalam menyelesaikan studi S1 di Ilmu Komputer Unila, telah memberikan doa, semangat, perhatian dan kasih sayang kepadaku.
4. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T. selaku dekan FMIPA Unila;
5. Bapak Didik Kurniawan, S.Si., M.T. selaku Ketua Jurusan Ilmu Komputer;

6. Bapak Rizky Prabowo, M.Kom. selaku Dosen Pembimbing Utama atas kesediaannya untuk memberikan bimbingan, bantuan, saran dan kritik dalam proses penyelesaian skripsi ini:
7. Bapak Prof. Admi Syarif, Ph.D. selaku Dosen Pembahas I dalam penelitian skripsi ini. Terima kasih untuk masukan dan saran-saran dalam proses penyelesaian skripsi ini;
8. Bapak Bambang Hermanto, S,Kom., M.cs. selaku Dosen Pembahas II dalam penelitian skripsi ini. Terima kasih untuk masukan dan saran-saran dalam proses penyelesaian skripsi ini;
9. Bapak Didik Kurniawan, S.Si., M.T. selaku Dosen Pembimbing Akademik;
10. Bapak dan Ibu dosen tanpa terkecuali yang telah memberikan ilmu pengetahuan, wawasan selama penulis menyelesaikan pendidikan di jurusan ilmu komputer.
11. Bapak dan Ibu Staf administrasi Jurusan Ilmu Komputer serta staf administrasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung yang telah membantu penulis dalam menyelesaikan administrasi akademik.
12. Keluarga, teman, sahabat seperjuangan di Ilmu Komputer, Intania Rahmadhilla, Ratih Indah Wardani, Rahmadila Nurjannah, Suci Hikmawati, yang sudah menghibur, menolong, memberikan semangat untuk penulis selama proses skripsi berlangsung bahkan dari awal perkuliahan di Ilmu Komputer.
13. Keluarga Besar Mas Pikri Mba Intan, yaitu Aulia Ahmad Nabil, Fikri Mulya Permana yang sudah menghibur, menolong, memberikan semangat untuk penulis selama proses skripsi berlangsung bahkan dari awal perkuliahan di Ilmu Komputer.
14. Windy Desty Ariany, selaku teman pertama di Ilmu Komputer sekaligus rekan dalam menyelesaikan skripsi ini yang sudah bersama-sama memberikan semangat dan bertukar pikiran serta saling membantu satu sama lain.
15. Arafia Isnayu Akaf, selaku teman dekatku yang telah membantu, memberikan dukungan, semangat serta menghibur, menjadi pendengar keluh kesah selama penulis menyelesaikan studi S1 di Ilmu Komputer.

16. Ester Novi Andriani, Wilda Rahmadona. Teman seperjuanganku sedari SMA yang memberikan motivasi untuk melanjutkan pendidikan S1, bersama-sama berjuang daftar masuk perguruan tinggi sampai sekarang. Sukses selalu untuk kita.
17. Keluarga besar Duabelas IPA 3 yang telah memberikan dukungan serta hiburan kepada penulis dalam menyelesaikan skripsi.
18. Kepada Dewi Lestari, diriku sendiri. Terima kasih sudah kuat dan bertahan dalam proses penyelesaian skripsi ini. Terima kasih untuk malam-malam panjang yang sudah dihabiskan untuk menuangkan segala fikiran dalam menyelesaikan skripsi ini. Terima kasih untuk kaki yang selalu kuat untuk melangkah, tangan yang selalu kuat untuk menghapus peluh serta air mata. Terima kasih untuk tidak menyerah walau sering salah. Ini langkah awal, mari bertahan lagi dengan tantangan kehidupan yang baru.

DAFTAR ISI

Halaman

| | |
|--|------------|
| DAFTAR ISI | i |
| DAFTAR TABEL | iii |
| DAFTAR GAMBAR | iv |
| | |
| I. PENDAHULUAN | 1 |
| 1.1. Latar Belakang | 1 |
| 1.2. Rumusan Masalah | 3 |
| 1.3. Batasan Masalah | 3 |
| 1.4. Tujuan Penelitian | 3 |
| 1.5. Manfaat Penelitian | 3 |
| | |
| II. TINJAUAN PUSTAKA | 4 |
| 2.1. Penelitian Terdahulu | 4 |
| 2.2. Klasifikasi Citra (<i>Image</i>)..... | 8 |
| 2.3. Tumbuhan Obat | 9 |
| 2.4. Daun Sirih | 9 |
| 2.5. <i>Decision Tree</i> | 10 |
| 2.6. <i>Feature Extraction</i> | 15 |
| 2.7. <i>Confusion Matrix</i> | 18 |
| | |
| III. METODOLOGI PENELITIAN | 22 |

| | |
|--|-----------|
| 3.1. Tempat dan Waktu Penelitian | 22 |
| 3.2. Data dan Alat Pendukung | 23 |
| 3.3. Alur Kerja Penelitian | 28 |
| | |
| V. SIMPULAN DAN SARAN..... | 56 |
| 5.1. Simpulan | 56 |
| 5.2. Saran | 57 |
| | |
| DAFTAR PUSTAKA | 58 |

DAFTAR TABEL

| Tabel | Halaman |
|---|---------|
| 1. Penelitian Terdahulu | 4 |
| 2. Data <i>Image</i> Daun Sirih Hijau | 24 |
| 3. Data <i>Image</i> Daun Sirih Merah | 24 |

DAFTAR GAMBAR

| Gambar | Halaman |
|---|---------|
| 1. Struktur <i>Decision Tree</i> | 10 |
| 2. Representasi <i>Confusion Matrix</i> | 19 |
| 3. Pengambilan Data Image. | 23 |
| 4. Diagram Alur Penelitian | 28 |

I. PENDAHULUAN

1.1. Latar Belakang

Klasifikasi citra (*image*) artinya suatu proses pengelompokan semua *pixel* di suatu gambar ke dalam kelompok yang bisa diinterpretasikan sebagai suatu *property* yang khusus. Pada pengklasifikasian citra (*image*) ada dua metode yang dapat dipergunakan untuk mengidentifikasi dan mengklasifikasikan sebuah citra satelit yaitu klasifikasi terbimbing serta klasifikasi tak terbimbing yang berkaitan dengan pendekatan dalam pengenalan pola (Chang & Ren, 2000). Klasifikasi terbimbing atau *Supervised Classification* adalah proses pengelompokan pixel-pixel suatu metode klasifikasi yang dilakukan dengan teknik analisis yang mana pada kriteria kelas pengelompokannya ditetapkan berdasarkan penciri kelas yang diperoleh melalui pembuatan area atau *training area*. Sedangkan klasifikasi tidak terbimbing atau *Unsupervised Classification* adalah proses pengelompokan pixel-pixel pada image menjadi beberapa kelas dengan menggunakan analisis *cluster*. Pada pembentukan kelas Sebagian besar dikerjakan oleh computer yang dikelompokkan berdasarkan kemiripan dari data itu sendiri (Purwanto & Lukiawan, 2019).

Tumbuhan obat atau lebih disingkat dengan sebutan TOGA merupakan tanaman hasil budidaya rumahan maupun tanaman yang tumbuh secara liar pada pekarangan rumah yang berguna sebagai pengganti obat kimia (Hariyanti, 2019). Penggunaan tanaman obat sebagai pengganti obat-obatan kimia telah dipraktekkan semenjak jaman Mesir kuno sekitar 2500 tahun sebelum masehi. Indonesia

mempunyai 30.000 jenis tanaman dari total 40.000 jenis tanaman yang ada di seluruh dunia. Terdapat sekitar 940 jenis tanaman di Indonesia yang dapat digunakan untuk pengobatan sebagai alternatif dari penggunaan obat-obat kimia (Hariyanti, 2019).

Dengan perkembangan teknologi yang semakin pesat, penggunaan komputer dapat digunakan untuk mengklasifikasikan berbagai jenis objek salah satunya klasifikasi citra (*image*). Klasifikasi citra (*image*) dengan menerapkan metode *machine learning* sudah banyak dilakukan untuk menghasilkan suatu tingkat akurasi serta kesimpulan terhadap objek tertentu. Salah satu penelitian sebelumnya yang dilakukan oleh (Pamungkas et al., 2020) yaitu Klasifikasi Gambar Gigitan Ular Menggunakan *Regionprops* dan Algoritma *Decision Tree*, penelitian tersebut bertujuan untuk mengklasifikasikan gambar bekas gigitan ular apakah termasuk ular berbisa atau tidak berbisa. Klasifikasi gambar menggunakan ekstraksi ciri *Regionprops*. Algoritma *Decision Tree* digunakan dalam penelitian ini karena cocok digunakan untuk mendukung data biner *classification* serta memperoleh hasil yang maksimal walaupun dengan data latih yang tidak terlalu banyak. Hasil dari model *Decision Tree* yang digunakan dalam penelitian ini menghasilkan nilai akurasi 97.41%, *precision* 92,85%, *recall* 91,42%, dan *F1 score* 92,06%.

Setelah mengidentifikasi dari berbagai jenis penelitian sebelumnya, maka penulis mengusulkan penelitian skripsi dengan judul “**Klasifikasi Image Tumbuhan Obat Sirih Hijau dan Sirih Merah Menggunakan Metode Decision Tree**”. Penelitian ini akan berfokus dalam mengklasifikasikan daun tumbuhan obat sirih hijau dan sirih merah dengan menggunakan ekstraksi fitur tekstur sebagai parameter yang membantu dalam proses klasifikasi *image* daun tumbuhan obat sirih hijau dan sirih merah menggunakan model klasifikasi *machine learning Decision Tree*.

1.2. Rumusan Masalah

Berdasarkan pemaparan latar belakang diatas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana mengklasifikasikan daun tumbuhan obat sirih hijau dan sirih merah menggunakan metode klasifikasi *Decision Tree*.
2. Bagaimana performance metode *Decision Tree* jika dibandingkan dengan metode klasifikasi *Naïve Bayes*.

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Penelitian ini menggunakan metode klasifikasi *Decision Tree* dalam mengklasifikasikan daun tumbuhan obat sirih hijau dan sirih merah.
2. Data *image* daun sirih hijau dan sirih merah didapatkan dari pengambilan gambar dengan kamera ponsel sebanyak 1000 data *image* daun sirih hijau dan 1000 data *image* daun sirih merah dengan total sebanyak 2000 data *image*.

1.4. Tujuan Penelitian

Tujuan penelitian ini untuk menerapkan metode klasifikasi *Decision Tree* dan membandingkan dengan metode klasifikasi *Naïve Bayes* dalam mengklasifikasikan tumbuhan obat daun sirih hijau dan sirih merah.

1.5. Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Sebagai wawasan dan pengetahuan mengenai klasifikasi daun tumbuhan obat sirih hijau dan sirih merah menggunakan metode klasifikasi *Decision Tree*.
2. Mampu mengetahui tingkat keakuratan metode klasifikasi *Decision Tree* dalam mengklasifikasikan jenis tumbuhan obat sirih hijau dan sirih merah sehingga dapat dijadikan sebagai informasi bagi penelitian selanjutnya.

II. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian ini tidak terlepas dari beberapa penelitian terdahulu yang telah dilakukan. Berikut merupakan beberapa ringkasan dari penelitian terdahulu.

Tabel 1. Penelitian Terdahulu

| No | Nama Peneliti | Judul | Data | Metode | Hasil |
|----|------------------------|--|--|---------------------------------------|---|
| 1 | (Batitis et al., 2020) | Image Classification of Abnormal Red Blood Cells Using Decision Tree Algorithm | Data sel darah merah abnormal didapatkan dari rumah sakit, buku, dan jurnal, terdapat 40 gambar yang terdiri dari 600 sel sampel | Metode Algoritma <i>Decision Tree</i> | Metode <i>decision tree</i> mampu digunakan untuk mengklasifikasikan sel darah merah abnormal. Namun kesalahan dalam klasifikasi adalah hasil perbedaan yang kecil antara atribut yang digunakan. Tingkat |

| | | | | | |
|---|--------------------------|--|---|-----------------------------|---|
| | | | jenis sel darah merah abnormal yang berbeda. | | akurasi terbaik yang didapatkan rata-rata 89,31% sedangkan tingkat akurasi kesalahan sebesar 10,96%. Tingkat kesalahan tersebut dikarenakan deteksi kualitas dari gambar sel darah merah yang diperoleh bukan dari algoritma yang digunakan. |
| 2 | (Pamungkas et al., 2020) | Klasifikasi Gambar Gigitan Ular Menggunakan Regionprops dan Algoritma <i>Decision Tree</i> | Dataset gambar luka gigitan ular yang terdiri dari 34 gambar gigitan dengan 16 gambar gigitan ular berbisa dan 18 gambar gigitan ular tidak berbisa | Metode <i>Decision Tree</i> | Hasil menunjukkan proses klasifikasi gambar gigitan ular menunjukkan bahwa kombinasi parameter <i>distance</i> dengan luas area atau dengan <i>eccentricity</i> ketika menggunakan <i>7-fold validation</i> menghasilkan nilai yang paling tinggi |

| | | | | | |
|---|-------------------------------|---|--|------------------------|---|
| | | | | | dengan akurasi 97,41%, <i>precision</i> 92,85%, <i>recall</i> 91,42%, dan <i>F1 score</i> 92,06%. |
| 3 | (Charbuty & Abdulazeez, 2021) | Classification Based on Decision Tree Algorithm for Machine Learning. | Data yang digunakan 12 data Penelitian mengenai data mining machine learning dari berbagai bidang. | <i>Decision Tree</i> . | Penelitian ini membahas penggunaan metode klasifikasi DT dalam melakukan klasifikasi data mining yang dilakukan oleh 12 peneliti. Dari penelitian tersebut didapatkan tingkat akurasi terbaik sebesar 99,93% artinya algoritma DT andal dalam melakukan klasifikasi machine learning. |

Penelitian terdahulu digunakan sebagai acuan dan perbandingan dalam penelitian ini. Beberapa penelitian terdahulu yang digunakan adalah sebagai berikut:

2.1.1. Image Classification of Abnormal Red Blood Cells Using Decision Tree Algorithm (2020).

Penelitian ini dilakukan oleh (Batitis et al., 2020). Penelitian ini membahas tentang klasifikasi *image* sel darah merah abnormal dengan menggunakan data *image* sel darah merah abnormal yang didapatkan

dari rumah sakit, buku, dan jurnal, terdapat 40 gambar yang terdiri dari 600 sel sampel jenis sel darah merah yang berbeda-beda. Klasifikasi terbagi menjadi 10 jenis sel darah merah yaitu *spherocyte*, *codocyte*, *stomatocyte*, *ovalocyte*, *elliptocyte*, *degmacyte*, *drepanocyte*, *dacrocytes*, *acanthocyte* dan *echinocyte* menggunakan algoritma klasifikasi *Decision Tree*. Hasil menunjukkan bahwa metode *Decision Tree* mampu digunakan untuk mengklasifikasikan sel darah merah abnormal. Namun kesalahan dalam klasifikasi adalah hasil perbedaan yang kecil antara atribut yang digunakan. Tingkat akurasi terbaik yang didapatkan rata-rata 89,31% sedangkan tingkat akurasi kesalahan sebesar 10,96%. Tingkat kesalahan tersebut dikarenakan deteksi kualitas dari gambar sel darah merah yang diperoleh bukan dari algoritma yang digunakan.

2.1.2. Klasifikasi Gambar Gigitan Ular Menggunakan *Regionprops* dan Algoritma *Decision Tree* (2020)

Penelitian ini dilakukan oleh (Pamungkas et al., 2020). Penelitian ini membahas tentang klasifikasi gambar gigitan ular yang menggunakan dataset gambar luka gigitan ular yang terdiri dari 34 gambar gigitan dengan 16 gambar gigitan ular berbisa dan 18 gambar gigitan ular tidak berbisa. Klasifikasi gambar ular ini menggunakan ekstraksi ciri *regionprops* yang terdapat beberapa fitur yang digunakan untuk membantu proses ekstraksi ciri antara lain *centroid*, luas area, *distance*, dan *eccentricity*. Hasil menunjukkan proses klasifikasi gambar gigitan ular menunjukkan bahwa kombinasi parameter *distance* dengan luas area atau dengan *eccentricity* ketika menggunakan *7-fold validation* menghasilkan nilai yang paling tinggi dengan akurasi 97,41%, *precision* 92,85%, *recall* 91,42%, dan *F1 score* 92,06%.

2.1.3. Classification Based on Decision Tree Algorithm for Machine Learning (2021)

Penelitian ini dilakukan oleh (Charbuty & Abdulazeez, 2021). Peneliti mengulas mengenai data mining *machine learning* dari berbagai bidang seperti analisis penyakit medis, klasifikasi teks, klasifikasi penggunaan *smartphone*, *processing image* dan lain sebagainya menggunakan algoritma DT. Dari bidang-bidang tersebut peneliti memaparkan bagaimana menghasilkan sebuah klasifikasi akurasi dengan menggunakan algoritma DT. Terdapat 12 penelitian yang menerapkan algoritma DT salah satunya penelitian yang dilakukan oleh (Zou, et al., 2018) dalam penelitian tersebut membahas bagaimana memprediksi penyakit diabetes militus dengan menggunakan perbandingan algoritma machine learning yaitu DT, *random forest*, *convolutional neural network*. Data yang digunakan yaitu data pasien diabetes militus di salah satu rumah sakit di Luzhou China, dari hasil klasifikasi tersebut didapatkan hasil akurasi terbaik sebesar 0,8084. Dari berbagai bidang penelitian yang telah dilakukan dengan menggunakan algoritma DT didapatkan sebuah kesimpulan akurasi terbaik yang dicapai algoritma DT ini sebesar 99,93% hasil ini menunjukkan bahwa kinerja algoritma DT sangat baik ketika digunakan sebagai repositori klasifikasi machine learning.

2.2. Klasifikasi Citra (Image)

Klasifikasi merupakan suatu aktivitas pengelompokan yang berguna untuk mempermudah dalam mencari sesuatu yang diperlukan jika terkelompokkan dengan baik (Lestari, 2016). Citra (*image*) ialah gambar pada bidang dwimatra (dua dimensi) yang ditinjau dari sudut pandang matematis citra merupakan fungsi menerus (*continue*) dan intensitas cahaya pada bidang dwimatra (Arsy, 2016). Adapun klasifikasi *image* yaitu suatu proses pengelompokan seluruh *pixel* pada

suatu gambar ke dalam kelompok yang dapat diinterpretasikan menjadi suatu *property* yang spesifik (Chang & Ren, 2000).

2.3. Tumbuhan Obat

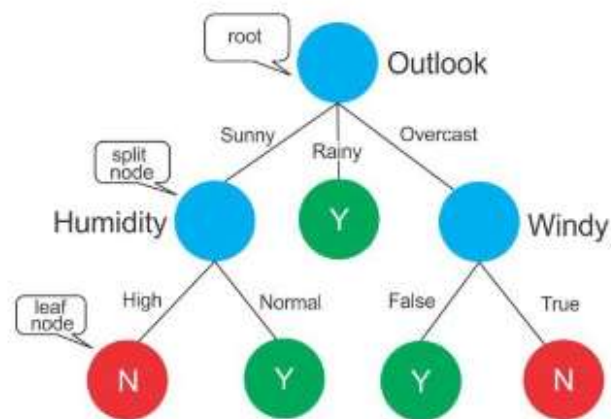
Tumbuhan obat atau lebih disingkat dengan sebutan TOGA ialah tanaman hasil budidaya rumahan maupun tanaman yang tumbuh secara liar di pekarangan rumah yang memiliki kegunaan menjadi pengganti obat kimia (Hariyanti, 2019). Penggunaan tanaman obat sudah di praktekkan sejak jaman Mesir kuno kurang lebih 2500 tahun sebelum masehi. Indonesia memiliki 30.000 jenis tanaman dari total 40.000 jenis tanaman yang ada di seluruh dunia. Terdapat kurang lebih 940 jenis tanaman di Indonesia yang dapat dipergunakan untuk pengobatan sebagai alternatif dari penggunaan obat-obat kimia (Hariyanti, 2019).

2.4. Daun Sirih

Daun sirih atau dengan nama latin *Piper betle L* merupakan salah satu tumbuhan herbal berasal dari Indonesia yang tumbuh merambat pada batang pohon lain. Tumbuhan sirih memiliki senyawa seperti *saponin*, *flavonoid*, *polifenol* dan *minyak atsiri*. Tumbuhan yang memiliki bau yang khas ini tentunya dapat digunakan sebagai obat alternatif, manfaat daun sirih sendiri dapat menurunkan kadar gula dalam darah bagi penderita diabetes, menurunkan kolestrol, pengobatan sakit gigi dan lain sebagainya. Daun sirih juga memiliki banyak jenis seperti sirih hijau, sirih merah, sirih gading, sirih bulu, sirih irian dan sirih hitam. Dari jenis daun sirih tersebut tentunya memiliki khasiat yang baik buat kesehatan terutama dalam pengganti obat-obatan kimia dapat menggunakan daun tumbuhan herbal (Putri et al., 2020).

2.5. Decision Tree

Decision Tree merupakan salah satu metode klasifikasi dengan memodelkan prediksi menggunakan struktur pohon atau struktur hirarki (Iykra, 2018). *Decision Tree* memiliki konsep mengubah data menjadi suatu pohon keputusan dan aturan-aturan keputusan. Metode yang digunakan dalam penerapan metode *Decision Tree* dapat mem-breakdown proses pengambilan keputusan yang kompleks menjadi lebih simple, sehingga pengambilan keputusan akan lebih menginterpretasikan solusi dari permasalahan (Iykra, 2018). Pohon keputusan dimulai dengan sebuah root node (titik awal) yang digunakan untuk pengambilan tindakan awal berdasarkan root node tersebut diteruskan dengan memecahkan leaf node sesuai dengan algoritma *Decision Tree* (Nurzahputra et al., 2016).



Gambar 1. Struktur *Decision Tree* (Nurzahputra et al., 2016).

Pada Gambar 1 tersebut merupakan salah satu contoh pohon keputusan untuk bermain tenis. Pohon keputusan tersebut memiliki 1 root, 3 split node yaitu keadaan cuaca yang terdiri dari cerah, hujan, dan mendung serta memiliki 2 leaf node yaitu high dan normal. Rules atau aturan dalam pengambilan keputusan di mulai dari outlook (root). Keputusan dapat disimpulkan melalui leaf node. Pada Gambar 1 tersebut keputusan yang dapat diambil yaitu jika cuaca cerah dengan kelembapan normal maka dapat bermain tenis dan lain sebagainya.

Struktur dari *Decision Tree* dibentuk dari tiga simpul yaitu (Nurzahputra et al., 2016).

- Simpul *root* adalah titik awal dari metode *Decision Tree*
- Simpul perantara adalah simpul yang digunakan sebagai perantara berhubungan dengan suatu pernyataan atau pengujian.
- Simpul *leaf* merupakan simpul terakhir yang memuat suatu keputusan akhir atau kelas target untuk suatu pohon keputusan.

Dalam pembentukan pohon keputusan banyak algoritma yang dapat dipakai yaitu ID3, C4.5 dan *Classification and Regression Tree (CART)*.

2.5.1. Algoritma Iterative Dichotomiser 3 (ID3)

Iterative Dichotomiser 3 merupakan algoritma yang dikembangkan oleh J. Ross Quinlan. Algoritma ID3 ialah salah satu metode klasifikasi *Decision Tree* atau algoritma pengambilan keputusan dengan pohon keputusan yang paling dasar. Algoritma ID3 melakukan pencarian atau analisis secara menyeluruh pada semua kemungkinan yang akan terjadi pada sebuah pohon keputusan. Dalam membentuk pohon keputusan ID3 tidak jauh berbeda dengan algoritma C45 yaitu dengan menghitung nilai entropy dan menghitung nilai information gain dari setiap variable (Tyasti, Ispriyanti, & Hoyyi, 2015)

2.5.2. Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma yang dapat digunakan untuk mengimplementasikan pohon keputusan. Algoritma C4.5 digunakan untuk klasifikasi data yang memiliki atribut dengan tipe data diskrit dan kontinu (Elisa, 2017). Algoritma C4.5 merupakan pengembangan dari algoritma *Iterative Dichotomiser (ID3)* yang merupakan algoritma paling dasar untuk membuat pohon keputusan. Namun, algoritma ID3 tidak mampu mengatasi atribut dengan data kosong sehingga diperlukan

algoritma C4.5 yang dapat digunakan untuk mengatasi percaharian atribut dengan data kosong serta dapat melakukan pemangkasan cabang yang sebelumnya tidak terdapat pada algoritma ID3 (Setio 2020). Menurut (Elisa, 2017) beberapa langkah dalam membangun pohon keputusan algoritma C4.5 adalah sebagai berikut.

1. Langkah awal yaitu menentukan atribut yang akan digunakan sebagai akar, untuk memilih atribut sebagai akar didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.
2. Membuat cabang untuk tiap-tiap nilai dari atribut.
3. Bagi kasus dalam cabang dengan
4. Ulangi proses untuk setiap cabang sampai semua kasus memiliki kelas yang sama dan didapatkan hasil keputusan pada cabang terbawah.

Menurut (Laksmi et al., 2013) adapun tahapan dari proses algoritma C4.5 dapat diuraikan sebagai berikut.

1. Mempersiapkan Data Training.

Data yang digunakan merupakan data tabel yang memiliki atribut dan kelas atribut yang digunakan sebagai penentuan keputusan.

2. Menghitung Nilai *Entropy*

Entropy merupakan nilai informasi yang menyatakan ukuran ketidakpastian, yaitu perbedaan keputusan terhadap nilai atribut tertentu. Semakin tinggi nilai *entropy* maka semakin tinggi pula ketidakpastian dari atribut tersebut. Nilai *entropy* dapat dihitung dengan rumus persamaan 1 sebagai berikut.

$$\text{Entropy (S)} = - \sum_{a=1}^n p_i \times \log_2 p_i \dots\dots\dots(1)$$

Adapun keterangan sebagai berikut.

S : himpunan kasus

a : atribut

n : jumlah partisi atribut

p_i : probabilitas yang diperoleh dari S dibagi dengan total kasus

3. Menghitung Nilai *Gain*

Gain merupakan atribut yang digunakan untuk memilih tes atribut pada setiap simpul pohon keputusan. Dengan demikian *gain* sebagai penentu tingkat pengaruh suatu atribut terhadap keputusan atau ukuran efektifitas suatu variable dalam mengklasifikasikan data. Nilai *gain* dapat dihitung dengan rumus persamaan 2 sebagai berikut.

$$Gain(S, A) = entropy(S) - \sum_{a=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \dots \dots (2)$$

Adapun keterangan sebagai berikut

S : himpunan kasus

a : atribut

n : jumlah partisi atribut

$|S_i|$: jumlah kasus pada partisi ke-i

$|S|$: jumlah kasus dalam S

4. Mengulangi proses ke-2 hingga semua cabang memiliki kelas yang sama. Pada proses percabangan akan berhenti apabila.

- Semua kasus dalam simpul n sudah mendapatkan kelas yang sama
- Tidak ada variabel independen pada kasus yang dipartisi lagi
- Tidak ada kasus di dalam cabang yang kosong.

2.5.3. Algoritma CART (*Classification and Regression Trees*).

Classification and Regression Trees atau CART merupakan salah satu algoritma yang dapat digunakan untuk proses klasifikasi menggunakan pohon keputusan. Algoritma CART akan membagi data menjadi dua (biner) yang kemudian menghasilkan sebuah pohon keputusan yang menggambarkan hasil pengelompokan. Algoritma CART bertujuan untuk mendapatkan suatu kelompok data yang akurat sebagai pencari dari suatu pengklasifikasian (Astuti, 2018). Pembentukan pohon keputusan di bentuk dengan membagi secara biner (*binery recursive partitioning*). Pembagian dilakukan dengan membagi data menjadi simpul kiri dan simpul kanan. Pemilihan dilakukan pada setiap simpul sampai dihasilkan simpul akhir. Variable yang dipilih pada simpul pertama merupakan variable yang penting. Langkah-langkah algoritma CART adalah sebagai berikut (Larose, 2014).

1. Mempersiapkan data yang akan diklasifikasikan
2. Menentukan variable *predictor* yang berguna sebagai variabel dasar pengelompokan berdasarkan variabel target.
3. Membagi data menjadi ruas kanan dan ruas kiri.
4. Mengukur goodness (kesesuaian) dari calon masing-masing cabang s pada simpul keputusan t yang dihitung dengan rumus.

$$\Phi (s | t) = 2 P_L P_R \sum_{j=1}^{jumlah\ kategori} |P (j| t_L) - P (j|t_R)| \dots \dots \dots (3)$$

Keterangan

t_L : calon cabang kiri dari node keputusan t

t_R : calon cabang kanan dari node keputusan t

P_L : jumlah data cabang kiri / jumlah data dari node keputusan t

P_R : jumlah data cabang kanan / jumlah data dari node keputusan t

$P_j | t_L$: jumlah data berkategori j pada calon cabang kiri dibagi jumlah data node keputusan t

$P_j | t_R$: jumlah data berkategori j pada calon cabang kanan dibagi jumlah data mode keputusan j.

5. Tentukan calon cabang untuk node keputusan dengan memilih nilai tersebsar sampai cabang ini tidak dihitung lagi selanjutnya
6. Gambarkan cabang node keputusan dan node kejadian terminasi
7. Ulangi Langkah 4 sampai tidak terdapat lagi cabang node keputusan.

2.6. *Feature Extraction*

Feature Extraction adalah teknik pengambilan ciri dari suatu citra yang memiliki nilai yang didapatkan untuk di analisis ke tahap selanjutnya (Religia, 2019). *Feature Extraction* bertujuan untuk melakukan perhitungan serta perbandingan yang dapat digunakan untuk proses klasifikasi suatu citra berdasarkan ciri histogram yang dimiliki (Religia., 2019) . Ekstraksi ciri adalah tahapan mengekstrak informasi atau ciri yang ingin dikenali dan dibedakan antara objek satu dengan objek yang lain dari data citra (Agustien et al., 2018). Secara umum ekstraksi ciri terbagi menjadi 5 yaitu ekstraksi ciri warna, bentuk, tekstur, ukuran dan geometri.

2.6.1. Ekstraksi Fitur Warna

Gambar tersusun dari banyak *pixel* yang memiliki intensitas warna masing-masing. Gambar tersusun atas nilai warna RGB (Rai et al., 2017). Pada tahapan ekstraksi fitur warna dapat dilakukan dengan mengkonversikan nilai RGB menjadi HSV (*Hue, Saturation, Value*) (Robianto dkk, 2021). Penggunaan model warna HSV pertama kali diterapkan oleh A.R Smith 1978. Adapun penjelasan mengenai HSV adalah sebagai berikut (Robianto et al., 2021).

- *Hue* merupakan warna asli seperti kuning, merah sebagai penentu kemerahan dan kehijauan

$$H = \begin{cases} 0^\circ & \Delta = 0 \\ 60^\circ \times \left(\frac{G' - B'}{\Delta} \bmod 6\right) & , C_{max} = R' \\ 60^\circ \times \left(\frac{B' - R'}{\Delta} + 2\right) & , C_{max} = G' \\ 60^\circ \times \left(\frac{R' - G'}{\Delta} + 4\right) & , C_{max} = B' \end{cases}$$

- *Saturation* merupakan kemurnian dari warna yang menyatakan sebagai kekuatan warna. Perhitungan nilai *Saturation*:

$$S = \begin{cases} 0 & , C_{max} = 0 \\ \frac{\Delta}{C_{max}} & , C_{max} \neq 0 \end{cases}$$

- *Value* adalah kecerahan asli warna. Angka menyatakan *value* yaitu antara 0-100% angka 0 menyatakan warna hitam dan apabila semakin tinggi nilainya maka semakin terang dan citra memiliki variasi warna terbaru. Untuk mendapatkan derajat Hue [0, 360] menggunakan persamaan berikut.

Perhitungan nilai *Value*:

$$V = C_{max}$$

2.6.2. Ekstraksi Fitur Bentuk.

Fitur bentuk merupakan karakter dari suatu objek yang merupakan konfigurasi oleh garis dan kontur. Fitur bentuk dapat dikategorikan berdasarkan batas (*boundary-based*) dan berdasarkan *region-based* (Rai et al., 2017). Teknik berdasarkan batas (*boundary-based*) menggambarkan karakteristik eksternal contohnya adalah *pixel* sepanjang batas objek. Sedangkan teknik berdasarkan *region-based* memiliki parameter *area*, *centroid*, *boundingBox*, *eccentricity*, *equivDiameter* (Agustien et al., 2018).

2.6.3. Ekstraksi Fitur Tekstur

Ekstraksi fitur tekstur merupakan ciri dalam sebuah gambar yang merupakan susunan struktur dari permukaan suatu gambar (Sugiarta dkk, 2017). Ekstraksi fitur tekstur dapat menggunakan salah satu teknik *Gray Level Co-Occurance Matrix* (GLCM) yang berperan sebagai matrik pengambilan nilai keabuan dari sebuah gambar. Tahapan yang dilakukan untuk melakukan ekstraksi fitur tekstur (Rai et al., 2017). Citra warna harus dirubah terlebih dahulu menjadi citra grayscale.

- Nilai citra gambar RGB dirubah menjadi abu-abu menggunakan rumus keabuan = $0.2989 * R + 0.5870 * G + 0.1140 * B$
- Segmentasi nilai warna ke dalam 16 bin
- Menghitung nilai *co-occurance* matrik dengan 4 arah yaitu 0° , 45° , 90° , dan 135° .
- Menghitung informasi ciri tekstur yaitu *contrast*, *correlation*, *energy*, *homogeneity*, dan *entropy*. Adapun rumus persamaannya sebagai berikut.

Contrast, dihitung dengan persamaan 4 berikut:

$$\sum_k k^2 [\sum_i \sum_j p(i, j)] \dots\dots\dots(4)$$

Correlation, dihitung dengan persamaan 5 berikut:

$$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i\sigma_j} \dots\dots\dots(5)$$

Energy dihitung dengan persamaan 6 berikut:

$$\sum_{i,j} p(i, j)^2 \dots\dots\dots(6)$$

Homogeneity dihitung dengan persamaan 7 berikut:

$$\sum_{i,j} \frac{p(i,j)}{1+|i-j|} \dots\dots\dots(7)$$

2.6.4. Ekstraksi Fitur Ukuran

Ekstraksi fitur ukuran berfungsi untuk membedakan antara ukuran objek satu dengan objek lainnya dengan menggunakan parameter ukuran yaitu luas. Luas dalam ekstraksi ini adalah banyaknya suatu *pixel* yang membentuk suatu objek, sedangkan keliling merupakan banyaknya piksel yang mengelilingi suatu objek (Saputro & Sari, 2019).

2.6.5. Ekstraksi Fitur Geometri

Ekstraksi fitur geometri didasarkan pada hubungan antara dua buah titik, garis, atau bidang pada citra digital. Jarak antara 2 buah titik dalam satuan piksel dapat ditentukan dengan persamaan *Euclidean* ataupun *Manhattan* (Widianto, 2019). Adapun rumus menghitungnya dapat di lihat pada persamaan 9 dan 10 sebagai berikut (Religia, 2019)

Euclidean

$$\sqrt{d1(x, y) = \sum_{i=1}^n |f_i(x) - f_j(y)|^2} \dots\dots\dots(8)$$

Manhattan

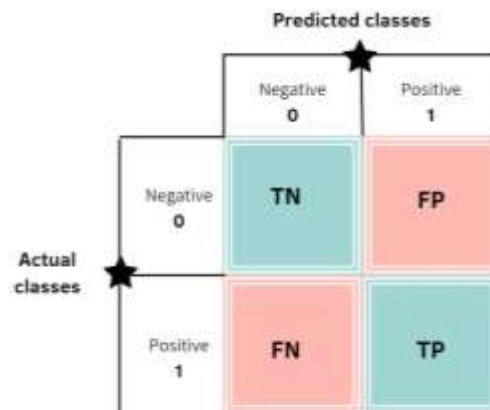
$$d2(x, y) = \sum_{i=1}^n |f_i(x) - f_j(y)|^2 \dots\dots\dots(9)$$

2.7. *Confusion Matrix*

Confusion matrix merupakan salah satu metode yang digunakan untuk mengukur dan mengevaluasi performa kinerja dari metode klasifikasi (*supervised learning*) pada machine learning dimana keluarannya dapat berupa dua kelas atau lebih (Karsito, 2019). *Confusion matrix* berbentuk tabel matriks yang menggambarkan kinerja dari suatu model klasifikasi pada rangkaian data uji yang menyatakan data uji yang dinyatakan terklasifikasi dengan benar dan data uji yang dinyatakan

terklasifikasi salah (Nugroho, 2019). Terdapat empat istilah dalam merepresentasikan hasil proses klasifikasi pada *confusion matrix* yaitu sebagai berikut.

- True Positive (TP) : adalah jumlah data dengan nilai hasil positif dan nilai hasil prediksi positif.
- False Positive (FP) : adalah jumlah data dengan nilai hasil negatif dan nilai hasil prediksi positif.
- True Negative (TN) : adalah jumlah data dengan nilai hasil negatif dan nilai hasil prediksi negatif.
- False Negative (FN) : adalah jumlah data dengan nilai hasil positif dan nilai hasil prediksi negative



Gambar 2. Representasi *Confusion Matrix* (Nugroho, 2019)

Representasi nilai dari *True Positive* dan *True Negative* akan memberikan hasil prediksi data yang bernilai benar, sedangkan untuk nilai dari *False Positive* dan *False Negative* akan memberikan hasil prediksi data yang bernilai salah. Positif dan negative merupakan suatu hasil prediksi dari model (Nugroho, 2019). Nilai dari *True Positive*, *True Negative*, *False Positive* dan *False Negative* adalah nilai yang digunakan untuk menghitung evaluasi *performance metrics* yang

berguna untuk mengukur kinerja model yang telah dibuat. Terdapat beberapa *performance metrics* yang biasa digunakan yaitu sebagai berikut:

1. Accuracy

Accuracy adalah jumlah presentasi yang menggambarkan seberapa akurat model dapat mengklasifikasikan data dengan benar dari keseluruhan data. Nilai *accuracy* ini merupakan nilai yang mengolah data yang terprediksi dengan tepat ditambah dengan data yang terprediksi tidak tepat dibagi dengan jumlah keseluruhan data yang ada (Nugroho, 2019). Untuk menghitung tingkat akurasi dapat digunakan rumus dalam persamaan 10 sebagai berikut.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \dots \dots \dots (10)$$

2. Precision (Positive Predictive Value)

Precision dapat digunakan untuk mengukur kinerja sistem dengan menghitung data yang terklasifikasi dengan tepat dan data yang ditemukan terklasifikasi tidak tepat. Data yang terklasifikasi dengan tepat dijadikan data acuan, kemudian dibagi dengan hasil prediksi *False Positive* atau data yang terprediksi tidak tepat maka akan diketahui kecocokan antara data acuan dengan data prediksi. Sehingga semakin banyak data acuan yang terprediksi tidak tepat ditemukan pada proses klasifikasi maka nilai dari *precision* akan semakin kecil (Nugroho, 2019). Untuk menghitung nilai *precision* dapat digunakan rumus dalam persamaan 11 sebagai berikut.

$$precision = \frac{TP}{TP + FP} \dots \dots \dots (11)$$

3. Recall atau Sensitivity

Recall merupakan model untuk mengukur evaluasi kinerja sistem dalam menemukan sebuah informasi kembali. Dengan demikian rasio prediksi data yang terklasifikasi benar akan dibandingkan dengan keseluruhan data yang benar positif (Nugroho, 2019). Semakin banyak data yang tepat tidak ditemukan maka nilai recall akan semakin kecil. Untuk menghitung nilai *recall* dapat digunakan rumus dalam persamaan 12 sebagai berikut.

$$recall = \frac{TP}{TP + FN} \dots \dots \dots (12)$$

4. F1-Score

F1-Score merupakan perbandingan rata-rata precision dan recall. Untuk menghitung nilai formula *f1-score* dapat digunakan rumus dalam persamaan 13 sebagai berikut (Saputro & Sari, 2019).

$$F1 - Score = 2x \frac{Recall \times Precision}{Recall + Precision} \dots \dots \dots (13)$$

III. METODOLOGI PENELITIAN

3.1. Tempat dan Waktu Penelitian

Penelitian dilakukan di Gedung Ilmu Komputer (GIK) lantai 2 Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Universitas Lampung. Adapun Timeline penelitian dapat dilihat pada Tabel 2.

Tabel 2. Timeline Penelitian

| Tahapan | Kegiatan | 2021 | | 2022 | | | | |
|---------------------|--|------|-----|------|-----|-----|-----|-----|
| | | Nov | Des | Jan | Feb | Mar | Apr | Mei |
| Penelitian Awal | Studi Literatur | ■ | | | | | | |
| | Penentuan Tema | ■ | | | | | | |
| | Pengambilan Data <i>Image</i> | ■ | | | | | | |
| | Penyusunan draft (Bab I-III) | ■ | ■ | ■ | | | | |
| Penelitian Lanjutan | Pre-processing | | ■ | ■ | | | | |
| | Ekstraksi ciri | | | ■ | ■ | | | |
| | Metode klasifikasi dengan menggunakan <i>Decision Tree</i> | | | | ■ | ■ | | |
| | Evaluasi dan Hasil Akurasi | | | | | ■ | | |
| | Penyusunan draft hasil (Bab IV-V) | | | | | ■ | ■ | |
| Evaluasi | Revisi Skripsi | | | | | | | ■ |

3.2. Data dan Alat Pendukung

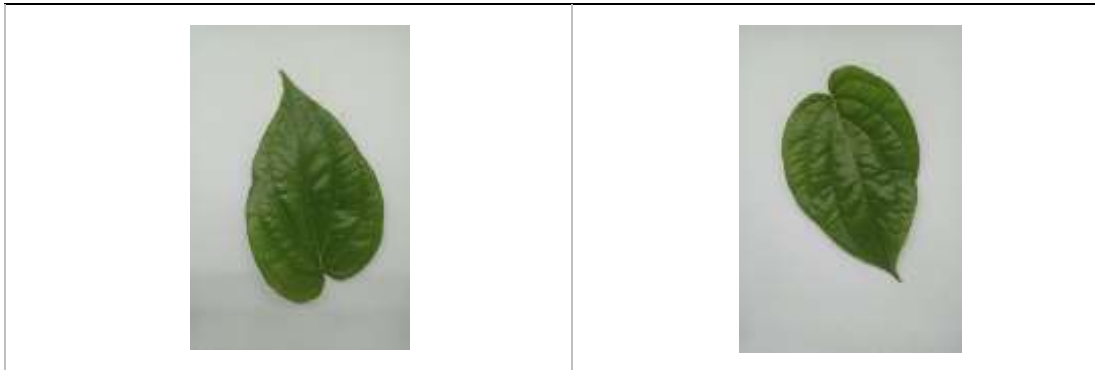
Adapun data dan alat pendukung yang digunakan untuk melakukan penelitian ini adalah sebagai berikut:

3.2.1. Data

Data yang digunakan yaitu data primer yang diambil peneliti dari observasi langsung dengan melakukan pengambilan gambar sample daun sirih hijau sebanyak 1000 data *image* dan sirih merah sebanyak 1000 data *image* sehingga total keseluruhan data *image* daun sirih hijau dan sirih merah yaitu 2000 data *image* dengan format JPG/JPEG (*Joint Photographic Experts Group*). Pengambilan gambar dengan menggunakan 1 sample daun sirih merah dilakukan sebanyak 24 kali searah jarum jam. Untuk mendapatkan 1000 data *image* daun sirih merah maka sample daun yang digunakan sebanyak 42 sample dan begitu pula dengan data *image* daun sirih hijau menggunakan 42 sample daun sirih hijau yang dilakukan dengan pengambilan gambar sebanyak 24 kali sehingga diperoleh 1000 data *image* daun sirih hijau. Sehingga total keseluruhan sample daun sirih yang digunakan yaitu 84 sample daun.



Gambar 3. Pengambilan Data *Image*.

Table 2. Data *Image* Daun Sirih Hijau

Pada tabel 2 diatas merupakan sample data *image* daun sirih hijau. Pengambilan gambar menggunakan 1 daun sirih hijau menghasilkan 24 gambar dengan rotasi yang berbeda. Total keseluruhan sampel daun yang digunakan sebanyak 42 sampel daun sirih hijau. Pada penelitian ini menggunakan 1000 data *image* daun sirih hijau.

Table 3. Data *Image* Daun Sirih Merah

Pada tabel 3 diatas merupakan sample data *image* daun sirih merah. Pengambilan gambar menggunakan 1 daun sirih merah menghasilkan 24 gambar dengan rotasi yang berbeda. Total keseluruhan sampel daun yang digunakan sebanyak 42 sampel daun sirih merah. Pada penelitian ini menggunakan 1000 data *image* daun sirih merah.

3.2.2. Alat Pendukung

Alat pendukung dalam penelitian ini terbagi menjadi dua jenis yaitu sebagai berikut:

3.2.2.1. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan pada penelitian ini yaitu:

- a. Processor: Intel Core i3 7020U, dual core (4 thread) clock up to 2.3 GHz
- b. Installed RAM: 4.00 GB
- c. Mini Studio ukuran 40 x 30 x 30
- d. Tripod 3110
- e. Handphone Android 3GB/32GBs
- f. Kamera Belakang: 13 MP + 8 MP + 2 MP dan kamera depan 8 MP.

3.2.2.2. Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan pada penelitian ini yaitu:

- a. Sistem Operasi: Windows 10 Pro 64-Bit.
- b. *Tools*
 - Anaconda Versi 3
Anaconda dikembangkan oleh Peter Wang dan Travis Oliphant pada tahun 2012. Anaconda adalah sebuah platform yang digunakan untuk menjalankan aplikasi dalam mengelola *packages* untuk menggunakan *library* dalam mengolah kode program yang dibutuhkan untuk *data learning* (Simon, 2020).
 - Jupyter Notebook 3.0.14
Jupyter merupakan perangkat lunak *opensource* yang didalamnya terdapat berbagai macam bahasa

program. Jupyter lab fleksibel digunakan dalam hal mendukung *workflow* untuk data *sains*, komputasi ilmiah dan machine learning (Simon, 2020).

- *Python Programming*

Bahasa pemrograman yang dikembangkan oleh Guido Van Rossum pada tahun 1991. Python merupakan Bahasa pemrograman *interpretative* yang dapat digunakan diberbagai platform dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode (Simon, 2020). Bahasa pemrograman python merupakan Bahasa pemrograman yang sangat populer karena dapat melakukan pengolahan *Data Science, Machine Learning dan Internet of Things (IoT)*

c. Library

Library atau package yang digunakan dalam penelitian ini adalah sebagai berikut.

- *Numpy*

Numpy merupakan library yang dimiliki python yang digunakan untuk menghitung operasi matematika pada array, membuat vektor matriks, mencari nilai max dan min dan lain sebagainya.

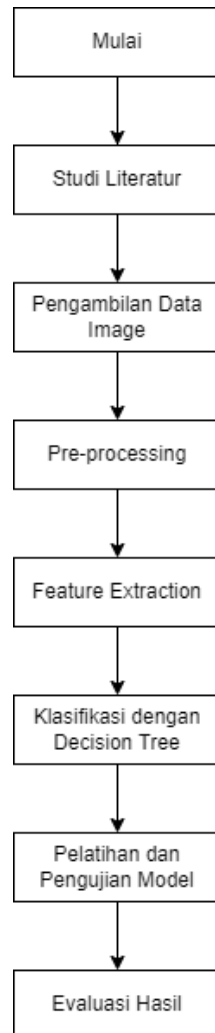
- *Pandas*

Pandas merupakan library yang berguna untuk melakukan *preprocessing* dan analisis data. Library *pandas* dapat mengolah data dalam bentuk csv, txt, excel dan sebagainya.

- *Scikit Learn*
Scikit Learn merupakan library yang dapat digunakan untuk permodelan *data science* seperti algoritma regresi, pengelompokkan, algoritma Naïve Bayes, *Decision Tree* dan lain sebagainya
- *Matplotlib*
Matplotlib merupakan library yang berguna untuk memvisualisasikan data menjadi bentuk grafik secara 2D maupun 3D didalam python.
- *Keras*
Keras merupakan *neural network* API level tinggi yang umumnya ditulis dalam kode program python. *Keras* banyak digunanakan karena *opensource* selain itu keras juga membuat *prototyping* lebih mudah dan cepat (Simon, 2020).
- *TensorFlow*
TensorFlow berfungsi untuk membantu dalam memperoleh tool yang tepat dalam riset, *coding*, maupun pengembangan *machine learning*.
- *Pydotplus*
Packages pydotplus digunakan untuk membuat *graph Decision Tree*. *Pydotplus* dapat digunakan apabila telah menginstal *package graphviz*.

3.3. Alur Kerja Penelitian

Penelitian ini melalui beberapa tahapan yang dapat dilihat pada Gambar 4.



Gambar 4. Diagram Alur Penelitian

Berikut ini penjelasan dari setiap tahapan pada Gambar 4 Diagram Alur Penelitian.

3.3.1. Studi Literatur

Pada tahap studi literatur dilakukan pencaharian untuk mendapatkan sumber teori valid yang mendukung penelitian ini. Sumber data maupun teori tersebut didapatkan melalui jurnal penelitian terdahulu, jurnal yang berguna untuk mengukuhkan teori yang dibuat, buku-buku, artikel

website yang mendukung makna mengenai penelitian. Tahap ini bertujuan untuk dijadikan sebagai referensi dan memperkuat penelitian yang sedang dilakukan.

3.3.2. Pengambilan Data *Image*

Data yang digunakan pada penelitian ini menggunakan data primer. Pengambilan data *image* dilakukan dengan cara mengambil gambar menggunakan ponsel dengan objek daun tumbuhan obat sirih hijau sebanyak 1000 data *image* dan sirih merah sebanyak 1000 data *image* sehingga total keseluruhan data *image* sirih hijau dan sirih merah berjumlah 2000 data *image* dengan format JPG.

3.3.3. *Pre-processing*

Tahap *pre-processing* adalah tahap yang dilakukan untuk menyiapkan data training sebelum diolah. *Pre-processing* berfungsi untuk menghasilkan data citra yang lebih baik sebelum diproses ke tahap berikutnya. Tahap *pre-processing* yang dilakukan yaitu proses *resize*. Pada tahap *resize* memastikan semua gambar memiliki rasio *pixel* yang sama. Hal ini dilakukan karena biasanya ukuran pada dimensi gambar yang dihasilkan tidak selalu sama dan untuk melakukan proses klasifikasi maka rasio dari semua gambar harus sama.

3.3.4. *Feature Extraction*

Setelah tahap *pre-processing* selesai kemudian dengan data gambar tersebut akan dilakukan tahap *Feature Extraction* yang bertujuan untuk menghasilkan nilai dari data gambar yang akan di analisis untuk proses selanjutnya. Pada penelitian ini menggunakan *Feature Extraction* tekstur *Grey Level Co-Occurance Matrix* (GLCM). Proses *feature extraction* dilakukan menggunakan *tools* Jupyter Notebook yang terdapat dalam platform Anaconda dalam bahasa pemrograman python.

3.3.5. Klasifikasi dengan Model *Decision Tree*

Setelah melakukan tahapan ekstraksi ciri, data tersebut dibagi menjadi data latih dan data uji. Dari hasil ekstraksi ciri tersebut dihasilkan nilai parameter dari fitur ekstraksi tekstur *Grey Level Co-Occurance Matrix* (GLCM). Proses klasifikasi tumbuhan obat sirih hijau dan sirih merah menggunakan metode *Decision Tree*.

3.3.6. Pengujian dan Pelatihan Model

Pada tahapan pengujian model, 1000 data daun tumbuhan obat sirih hijau dan sirih merah sebanyak 1000 akan dilakukan pembagian data kedalam data training dan data testing. Pengujian akan dilakukan sebanyak 5 kali dengan setiap kali pengujian akan membagi data training dan data testing dengan proporsi yaitu pengujian pertama 60%:40%, pengujian kedua 65%:35%, pengujian ketiga 70%:30%, pengujian keempat 75%:25%, dan pengujian kelima 80%:20%. Pembagian proporsi pengujian data tersebut berdasarkan jurnal rujukan yang digunakan.

3.3.7. Evaluasi

Hasil dari pengujian tersebut akan dihitung menggunakan model data yang diukur dengan membangun *confusion matrix*. Pada tahapan *confusion matrix* ini akan dihasilkan nilai akurasi, *precision*, *recall*, dan *F1 score*. Nilai akurasi hasil tersebut akan berguna untuk melihat dan menunjukkan tingkat kebenaran dari hasil klasifikasi *image* daun tumbuhan obat sirih merah sirih hijau dengan menggunakan metode *Decision Tree*. Hasil akurasi ini akan menunjukkan apakah metode klasifikasi yang digunakan dapat bekerja dengan baik.

V. SIMPULAN DAN SARAN

5.1. Simpulan

Pada penelitian Klasifikasi *Image* Tumbuhan Obat Sirih Hijau dan Sirih Merah Menggunakan Metode Decision Tree dapat disimpulkan sebagai berikut.

1. Penelitian ini sudah berhasil mengimplementasikan metode Decision Tree dalam mengklasifikasikan tumbuhan obat sirih hijau dan sirih merah. Perbandingan pembagian penggunaan data latih dan data uji sebesar 60%:40%, 65%:35%, 70%:30%, 75%:25%, dan 80%:20%. Permodelan dilakukan sebanyak 3 kali dengan masing-masing permodelan dilakukan percobaan 5 kali percobaan menggunakan proporsi pembagian data latih dan data uji.
2. Dari ketiga permodelan dalam 5 kali percobaan klasifikasi image sirih hijau dan sirih merah menggunakan metode Decision Tree didapatkan hasil nilai rata-rata accuracy sebesar 98%, precision 98,3%. recall 97,1% dan F1 Score 97,5%.
3. Penelitian Klasifikasi *Image* Tumbuhan Obat Sirih Hijau dan Sirih Merah Menggunakan Metode Decision Tree dibandingkan dengan penelitian Klasifikasi *Image* Tumbuhan Obat Sirih Hijau dan Sirih Merah Menggunakan Metode Naïve Bayes lebih unggul tingkat akurasi yang dihasilkan. Penelitian menggunakan metode Decision Tree menghasilkan tingkat akurasi sebesar 98% sedangkan metode Naïve Bayes sebesar

92,1%. Jadi dapat ditarik kesimpulan bahwa metode Decision Tree dapat bekerja dengan baik untuk proses klasifikasi tumbuhan obat sirih hijau dan sirih merah.

5.2. Saran

Adapun saran yang diberikan dalam penelitian ini adalah sebagai berikut.

1. Penelitian ini dapat dilakukan dengan menggunakan metode klasifikasi lainnya atau dapat menggunakan metode klasifikasi *deep learning* dengan harapan dapat menemukan hasil klasifikasi yang lebih baik sebagai bahan perbandingan.
2. Penelitian ini dapat ditambahkan dengan menggunakan data *image* tumbuhan obat lainnya dengan jumlah proporsi data *image* yang lebih banyak untuk penelitian lebih lanjut.

DAFTAR PUSTAKA

- Agustien, L., Armanto, H., Zaman, L., Informasi, T., Tinggi, S., Surabaya, T., & Person, K. (n.d.). *Seminar Nasional Teknologi dan Rekayasa (SENTRA) 2018 ISSN (Cetak) 2527-6042 eISSN (Online)*.
- Arsy, L., Nurhayati, O. D., & Martono, K. T. 2016, April. Aplikasi Pengolahan Citra Digital Meat Detection Dengan Metode Segmentasi K-Mean Clustering Berbasis OpenCV dan Eclipse. *Jurnal Teknologi dan Sistem Komputer*, 2.
- Astuti, R. 2018. DATA MINING UNTUK KLASIFIKASI DENGAN ALGORITMA CART (CLASSIFICATION AND REGRESSION TREES). *Media Informatika*, 17.
- Batitis, V. M. E., Caballes, M. J. G., Ciudad, A. A., Diaz, M. D., Flores, R. D., & Tolentin, E. R. E. (2020). Image Classification of Abnormal Red Blood Cells Using Decision Tree Algorithm. *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, 498–504. <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00093>
- Chang, C.-I., & Ren, H. 2000. An Experiment-Based Quantitative and Comparative Analysis of Target Detection and Image Classification Algorithms for Hyperspectral Imagery. In *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* (Vol. 38, Issue 2).
- Charbuty, B., & Abdulazeez, A. 2021. Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://doi.org/10.38094/jastt20165>

- Elisa, E. 2017, Juni. JOIN | Volume 2 No. 1 | Juni 2017ISSN 2527-9165Makalah dikirIm: 12Mei 2017; Revisi: 22 Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2, 6. Retrieved November 2021
- Hariyanti, S., & A'yun, N. Q. 2019. *Budidaya Toga*. Semarang, Jawa Tengah: CV.HarianJatengNetwork. Retrieved oktober 2021, from https://kkn.unnes.ac.id/lapkknunnes/21903_3374121001_7_Kelurahan%20Pakintelan_219030018_20191206_111143.pdf
- Iykra. 2018, july. Mengenal Decision Tree dan Manfaatnya. Retrieved Oktober 2021, from <https://medium.com/iykra/mengenal-decision-tree-dan-manfaatnya-b98cf3cf6a8d>
- Karsito, & Susanti, S. 2019, Maret. KLASIFIKASI KELAYAKAN PESERTA PENGAJUAN KREDIT RUMAH DENGAN ALGORITMA NAÏVE BAYES DI PERUMAHAN AZZURA RESIDENCIA. *SIGMA – Jurnal Teknologi Pelita Bangsa*, 9, 6.
- Laksmi, M., Martin, Begum, M., & Venkatesan, P. 2013, June. An Analysis on Performance of Decision Tree Algorithms using Student’s Qualitative Data. *I.J.Modern Education and Computer Science*, 5, 10. doi:10
- Larose, D. T., & Larose , C. D. 2014. *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining*. Canada: John Wiley & Sons, Inc., Hoboken, New Jersey.
- Lestari, I. D. 2016, Oktober. Klasifikasi online dan google. *Jurnal Iqra'*, 2.
- Nugroho, K. S. 2019, November. Confusion Matrix untuk Evaluasi Model pada Supervised Learning. Retrieved November 2021, from <https://ksnugroho.medium.com/confusion-matrix-untuk-evaluasi-model-pada-supervised-machine-learning-bc4b1ae9ae3f>
- Nurzahputra, A., Ratna Safitri, A., & Aziz Muslim, M. (n.d.). *Klasifikasi Pelanggan pada Customer Churn Prediction Menggunakan Decision Tree*.

- Pamungkas, Y. W., Adiwijaya, A., & Utama, D. Q. 2020. Klasifikasi Gambar Gigitan Ular Menggunakan Regionprops dan Algoritma Decision Tree. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 1(2), 69. <https://doi.org/10.30865/json.v1i2.1789>
- Purwanto, E. H., & Lukiawan, R. 2019. PARAMETER TEKNIS DALAM USULAN STANDAR PENGOLAHAN PENGINDERAAN JAUH: METODE KLASIFIKASI TERBIMBING. *Jurnal Standardisas*, 21.
- Putri, A. K., Satwika, E., Sulistyana, Y., Arindias, Z., Studi, M. P., Biologi, P., Keguruan, F., & Pendidikan, I. (n.d.). *Studi Morfologi Piper betle L. dan Pemanfaatannya dalam Kehidupan Sehari-Hari*.
- Rai, G., Sugiarta, A., Made, I., & Widyantara, O. (n.d.). Ekstraksi Fitur Warna, Tekstur dan Bentuk untuk Clustered-Based Retrieval of Images (CLUE). *Teknologi Elektro*, 16.
- Religia, Y. 2019, September. FEATURE EXTRACTION UNTUK KLASIFIKASI PENGENALAN WAJAH MENGGUNAKAN SUPPORT VECTOR MACHINE DAN K-NEAREST NEIGHBOR. *Jurnal Ilmiah Informatika, Arsitektur dan Lingkungan*, jurnal.pelitabangsa.ac.id.
- Robianto, Sitorus, S. H., & Ristian, U. 2021. PENERAPAN METODE DECISION TREE UNTUK MENGLASIFIKASIKAN MUTU BUAH JERUK BERDASARKAN FITUR WARNA DAN UKURAN. *Jurnal Komputer dan Aplikasi*, 09, 76-86.
- Setio, P. B., Saputro, D. R., & Winarno, B. 2020. Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5. *Journal Universitas Sebelas Maret (UNNES)*, 3, 8.
- Simon, H. S. 2020. Penentuan Posisi Objek Berbasis Image Processing Dengan Menggunakan Metode Convolutional Neural Network.
- Tyasti, A. E., Ispriyanti, D., & Hoyyi, A. 2015. ALGORITMA ITERATIVE DICHOTOMISER 3 (ID3) UNTUK MENGIDENTIFIKASI DATA REKAM MEDIS

(Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta). *Jurnal Gaussian*, 4, 237-246.

Widianto, M. H. 2019, Desember 23. Ekstraksi Ciri dalam dunia sinyal. Retrieved from <https://binus.ac.id/bandung/2019/12/ekstraksi-ciri-dalam-dunia-sinyal/>

Widhi Saputro, I., & Wulan Sari, B. (2019). Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction. *Citec Journal*, 6(1).

Zou, Q., Kaiyang, Q., Yamei, L., Dehui, Y., Ju, Y., & Tang, H. 2018, November. Predicting Dabetes Mellitus With Machine Learning Techniques. (S. I. Tao Huang, Ed.) *Frontiers in Genetics*, 9. doi:10.3389/fgene.2018.00515