

**KLASIFIKASI PENYAKIT PADA TANAMAN TEH
MENGUNAKAN ALGORITMA C4.5**

(Studi Kasus: PTPN IV UNIT KEBUN SIDAMANIK)

SKRIPSI

OLEH:

PHILIPUS ASTERIUS

(188160086)



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2023

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 16/1/24

Access From (repository.uma.ac.id)16/1/24

**KLASIFIKASI PENYAKIT PADA TANAMAN TEH
MENGUNAKAN ALGORITMA C4.5**

(Studi Kasus: PTPN IV UNIT KEBUN SIDAMANIK)

SKRIPSI

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana di Fakultas Teknik

Universitas Medan Area

Oleh:

PHILIPUS ASTERIUS

188160086

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

2023

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

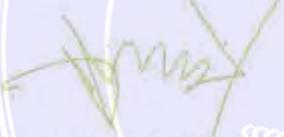
1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

HALAMAN PENGESAHAN

Judul Skripsi : KLASIFIKASI PENYAKIT PADA TANAMAN TEH
MENGUNAKAN ALGORITMA C4.5
(Studi Kasus: PTPN IV UNIT KEBUN SIDAMANIK)

Nama : Philipus Asterius
NPM : 188160086
Fakultas : Teknik

Disetujui Oleh ;
Komisi Pembimbing


Dr. Dian Novandri ST, M.Kom
Pembimbing I


Andre Hasudungan lubis, S.Ti.M.Sc
Pembimbing II


Dr. Ing. Supriatno, S.T, M.T
Dekan Fakultas Teknik


Rizki Muliono, S.Kom, M.Kom
K.a. Prodi

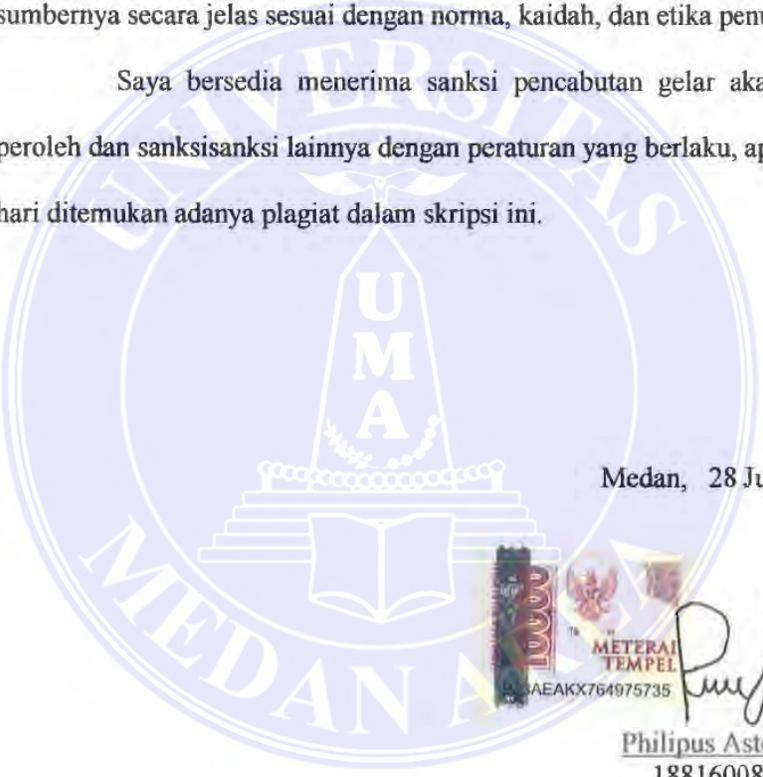
Tanggal lulus : 28 juli 2023

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksisanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 28 Juli 2023



Philipus Asterius
Philipus Asterius
188160086

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini :

Nama : Philipus Asterius
NPM : 188160086
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul : **Klasifikasi Penyakit pada Tanaman Teh Menggunakan Algoritma C4.5 (Studi Kasus : PTPN IV UNIT Kebun Sidamanik)**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan

Pada tanggal : 28 Juli 2023

Yang menyatakan



(Philipus Asterius)

v

RIWAYAT HIDUP

Penulis dilahirkan di Aek Bingke, pada tanggal 19 Juni 1999 dari ayah Delsan Purba dan ibu Erdimauli. Penulis merupakan putra ketiga dari 4 (empat) bersaudara. Penulis menyelesaikan pendidikan sekolah dasar di SDN 064 Aek Bingke pada tahun 2012. Pada tahun yang sama penulis melanjutkan pendidikan sekolah menengah pertama di SMP Swasta Berkat Aek Bingke selama 3 (tiga) tahun dan selesai pada tahun 2015. Penulis melanjutkan tingkat pendidikan selanjutnya di SMAN 1 Panyabungan Utara dan lulus pada tahun 2018. Tahun 2018 penulis juga terdaftar sebagai mahasiswa Teknik Informatika di Universitas Medan Area.

Selama mengikuti perkuliahan, Penulis melaksanakan Kerja Praktek (KP) di Kantor Desa Medan Estate.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Kuasa atas segala karuniaNya sehingga skripsi ini berhasil diselesaikan. Tema yang dipilih dalam penelitian ini ialah Deep Learning dengan judul “Klasifikasi penyakit pada tanaman teh menggunakan Algoritma C4.5

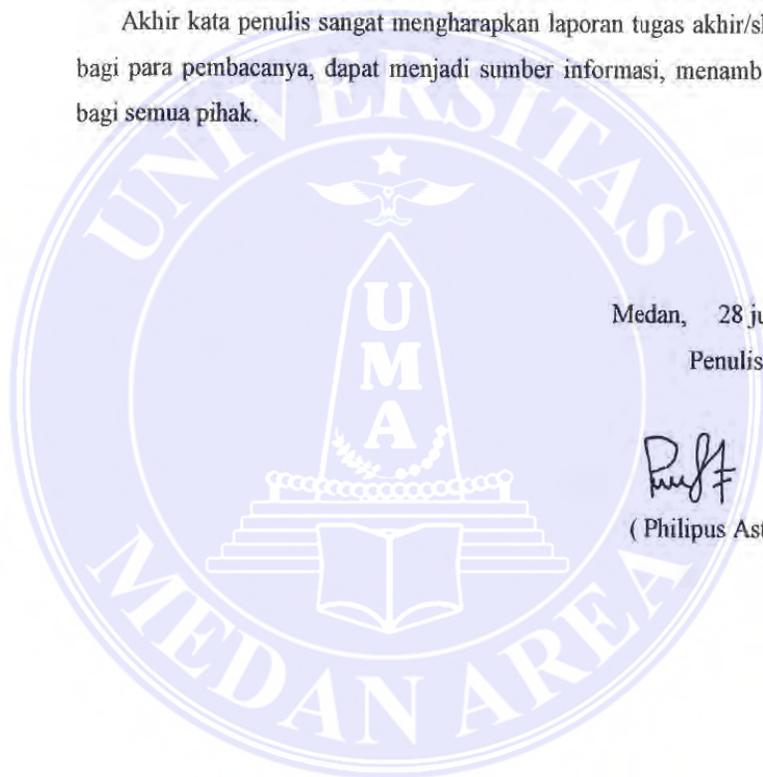
Skripsi ini merupakan salah satu syarat untuk menyelesaikan pendidikan untuk mencapai gelar sarjana di Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area. Pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc. selaku Rektor Universitas Medan Area.
2. Bapak Dr. Eng., suprianto, ST, M.T. selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom selaku Kepala Program Studi Teknik Informatika Universitas Medan Area.
4. Bapak Dr Dian Novandri, S.T, M.Kom dan Bapak Andre Hasudungan Lubis, S.Ti, M.Sc. selaku Komisi Pembimbing tugas akhir yang dengan sabar telah membimbing saya serta memberikan masukan-masukan yang berguna bagi saya.
5. Orang tua penulis yaitu Bapak Delsan Darius Purba dan Erdimauli Manalu, kaka saya Anna Magdalena Purba, Cristina Purba, adek saya Efren wasintor Purba yang telah mendoakan tiada henti dan memberikan semangat serta membantu penulis dalam segi materi dan moril sehingga penulis dapat menyelesaikan skripsi ini dengan sebaik baiknya.
6. Seluruh Dosen dan Staf Program Studi Teknik Informatika Universitas Medan Area.
7. Seluruh teman-teman yang sudah memberikan dukungannya selama penulisan proposal skripsi ini, khususnya teman-teman Teknik Informatika angkatan 2018.

8. Seluruh pihak yang tidak dapat disebutkan satu persatu yang membantu dalam penyelesaian skripsi ini

Penulis menyadari bahwa laporan tugas akhir/skripsi ini masih memiliki kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan demi kesempurnaan laporan tugas akhir/skripsi ini. Penulis berharap laporan tugas akhir/skripsi ini dapat bermanfaat baik untuk kalangan pendidikan maupun masyarakat.

Akhir kata penulis sangat mengharapkan laporan tugas akhir/skripsi ini dapat berguna bagi para pembacanya, dapat menjadi sumber informasi, menambah wawasan khususnya bagi semua pihak.



Medan, 28 juli 2023

Penulis

A handwritten signature in black ink, appearing to read 'Philipus Asterius'.

(Philipus Asterius)

ABSTRAK

Tanaman teh (*camellia sinensis*) merupakan tanaman penting dalam perekonomian Indonesia. Bukan hanya sebagai bahan konsumsi saja, tanaman ini juga dapat berperan dalam membantu perekonomian seperti menyediakan lapangan pekerjaan, sehingga mengurangi pengangguran, dan sangat berguna untuk menambah devisa negara. Perkebunan teh di Sidamanik memiliki lahan yang sangat luas yaitu sekitar 6.373 Ha. Petani disana kerap mengalami kesulitan dalam menjaga kualitas produksinya. Hal ini salah satunya dipengaruhi oleh sulitnya menentukan jenis penyakit yang menjangkit pada tanaman. Sehingga perlu mengklasifikasikan penyakit-penyakit pada tanaman teh tersebut. Dengan memanfaatkan teknologi informatika atau data mining Algoritma C4.5 adalah algoritma yang sudah banyak digunakan untuk klasifikasi data memiliki atribut-atribut numerik, kemampuan dari hasil klasifikasi yang dapat digunakan memprediksi nilai atribut bertipe dikret dari record yang baru. Hasil akurasi algoritma tersebut adalah 96% dengan jumlah data yang digunakan yaitu 400 *record* meggunakan proporsi data dan *training* 80% dan testing 20%. Berdasarkan hasil akurasi yang dicapai, dapat dikatakan bahwa Algoritma C4.5 baik digunakan dalam klasifikasi penyakit tanaman teh.

Kata Kunci: Penyakit tanaman teh, Algoritma C4.5, Klasifikasi, Data Mining.

ABSTRACT

The tea plant (camellia sinensis) plays an important role in Indonesian economy. This plant act as the consumption material and play contribute to providing jobs, reducing unemployment, and is very useful for increasing state division. The tea plantations in Sidamanik have a very large area of 6,373 ha. Farmers often felt difficulties in maintaining the quality of their production, including determining the type of disease that infects plants. Hence, it is necessary to classify the diseases in the tea plant by utilizing informatics technology or data mining. C4.5 Algorithm is tool that has been widely used for classifying data with numeric attributes, the ability of the classification results is used to predict the value of dict type attributes from new records. The results shows that the algorithm's accuracy is 96% with the amount of data used, namely 400 records using the proportion of 80% data training and 20% data testing. Based on classification, there are 43,67% plants are infected Blister Blight, 31,33% are infected Phytophthora infestans, and about 25% of plants are infected die back disease.

Keywords: tea plant disease, C4.5 algorithm, Classification, Data Mining.

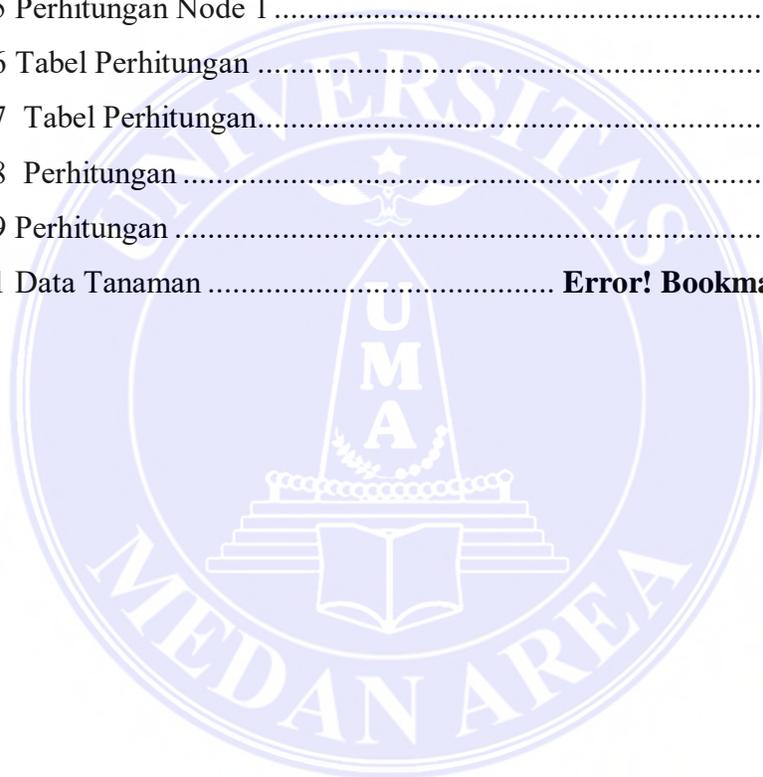
DAFTAR ISI

HALAMAN PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	iv
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS Error! Bookmark not defined.	
RIWAYAT HIDUP	vi
ABSTRAK	ix
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Perumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian.....	4
1.6. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1. Tanaman Teh.....	6
2.1.1. Penyakit Tanaman Teh.....	7
2.2. Data mining	8
2.3. Klasifikasi	9
2.4. Pohon Keputusan	9
2.5. Algoritma C4.5	10
2.6. <i>Confusion Matrix</i>	11
2.7. <i>Python</i>	12
2.8. <i>Google Colab</i>	12

2.9. Flowchart	13
2.10. Penelitian Terdahulu	14
BAB III METODOLOGI PENELITIAN	19
3.1. Pengumpulan Data.....	19
3.1.1. Jenis Data Penelitian	19
3.1.2. Metode Pengumpulan Data	21
3.2. Tahapan Penelitian.....	22
3.3. Kebutuhan Perangkat	23
3.4. Proses Penerapan Perhitungan Manual C4.5.....	24
3.5. Pilih Atribut Akar	24
BAB IV HASIL DAN PEMBAHASAN.....	44
4.1. Hasil.....	44
4.1.1 Data dan Visualisasi.....	44
4.1.2 Proses Encoder	46
4.1.3 Korelasi	47
4.1.4 Analisis 70: 30 Data	49
4.1.5 Analisis 80: 20 Data	51
4.2. Pembahasan.....	52
BAB V KESIMPULAN DAN SARAN	53
5.1. Kesimpulan.....	53
5.2. Saran	53
DAFTAR PUSTAKA.....	54
LAMPIRAN	57

DAFTAR TABEL

Tabel 2. 1 <i>Confusion Matrix</i>	12
Tabel 2. 2 Penelitian Terdahulu.....	15
Tabel 3. 1 Data Jenis Penyakit	19
Tabel 3. 2 Data Jenis Penyakit	20
Tabel 3. 3 Data Tanaman	21
Tabel 3. 4 Contoh Kasus	24
Tabel 3. 5 Perhitungan Node 1	28
Tabel 3. 6 Tabel Perhitungan	32
Tabel 3. 7 Tabel Perhitungan.....	36
Tabel 3. 8 Perhitungan	39
Tabel 3. 9 Perhitungan	42
Tabel 4. 1 Data Tanaman	Error! Bookmark not defined.



DAFTAR GAMBAR

Gambar 2. 1 Cacar daun.....	7
Gambar 2. 2 Busuk daun.....	7
Gambar 2. 3 Mati ujung	8
Gambar 2. 4 Flowchar.....	14
Gambar 3. 1 Tahapan Penelitian.....	22
Gambar 3. 2 Pohon keputusan.....	30
Gambar 3. 3 Pohon keputusan.....	33
Gambar 3. 4 Pohon keputusan.....	37
Gambar 3. 5 Pohon keputusan.....	40
Gambar 3. 6 Pohon keputusan.....	43
Gamabr 4. 1. Library yang digunakan	Error! Bookmark not defined.
Gamabr 4. 2 Membaca Data Excel	Error! Bookmark not defined.
Gamabr 4. 3 cek mising	Error! Bookmark not defined.
Gamabr 4. 4 klasifikai masing-masing penyakit	Error! Bookmark not defined.
Gamabr 4. 5 label <i>Encoder</i>	Error! Bookmark not defined.
Gamabr 4. 6 korelasi antar Variabel	Error! Bookmark not defined.
Gamabr 4. 7 Hasil akurasi dan confusion	Error! Bookmark not defined.

SSS

BAB I PENDAHULUAN

1.1. Latar Belakang

Perkembangan teknologi informasi yang semakin pesat saat ini menuntut manusia untuk mampu mengikuti perkembangannya. Pemanfaatan teknologi informasi tersebut sangat mendukung untuk mempermudah pekerjaan di berbagai aspek kehidupan manusia, baik untuk kebutuhan operasional, manajemen, maupun pengambilan keputusan (Mardi, 2018). Teknologi informasi telah banyak digunakan di sektor pertanian, yakni berperan penting dalam mempermudah petani dalam pengambilan keputusan. Salah satu manfaatnya adalah dapat memberikan informasi tentang penyakit pada tanaman, waktu penanaman, dan waktu panen. Dengan demikian, hal tersebut sangat penting guna mengurangi gagal panen dan kerugian bagi petani salah satunya di sektor pertanian teh (Burhan, 2018).

Tanaman teh (*camellia sinensis*) merupakan tanaman penting dalam perekonomian Indonesia. Bukan hanya sebagai bahan konsumsi saja, tanaman ini juga dapat berperan dalam membantu perekonomian seperti menyediakan lapangan pekerjaan, sehingga mengurangi pengangguran, dan sangat berguna untuk menambah devisa Negara (Burhan, 2018). Teh dikenal di Indonesia sejak tahun 1686 yaitu ketika seorang warga Belanda yang bernama Dr. Andreas Cleyer membawa ke Indonesia. Teh banyak disukai oleh masyarakat Indonesia karena rasanya yang sangat khas dan harum ketika diseduh atau diminum membuat banyak orang menyukainya, bahkan ada jenis teh yang dapat digunakan sebagai obat untuk penyembuhan penyakit

(Ma'rifah, 2020).

Masa budi daya tanaman teh, ada beberapa kendala yang menyebabkan turunnya produksi pada tanaman teh diantaranya masalah hama dan penyakit yang dapat menurunkan hasil produksi tanaman teh. Berdasarkan pengetahuan pakar hama yang sering menyerang pada tanaman teh antara lain Tungau Jingga (*Tenuipelpus Obovatus Donn*), Ulat Penggulung Daun (*Enarmonia Leuestoma*), dan berbagai hama lainnya (Fazri Ernawati Erlansari, 2019).

Provinsi Sumatera utara memiliki perkebunan teh terbesar ke dua di Indonesia yang letaknya berada di Kecamatan Sidamanik.

Proses pengklasifikasian penyakit tanaman teh masih dilakukan secara manual, yakni petani hanya melakukan pengamatan langsung pada tanaman tersebut. Sehingga, prosesnya sangat memakan waktu dan biaya. Sementara itu, terdapat solusi yang dapat dilakukan dalam menangani permasalahan tersebut, yakni data mining. Data mining adalah proses pengumpulan data besar dan pengambilan keputusan dan pola atau hubungan dalam data set, dalam Teh khas Sumatera utara ini dikenal dengan teh hitam yang sangat banyak diminati oleh negara Amerika Serikat dan negara-negara Eropa. Namun, produksinya belum menguntungkan pihak petani yang disebabkan oleh kualitasnya yang masih menurun (Damanik, Okprana, & Sormin, 2022). Perkebunan teh di Sidamanik ini memiliki lahan yang sangat luas yaitu sekitar 6.373 ha petani disana kerap mengalami kesulitan dalam menjaga kualitas produksinya. Hal ini salah satunya dipengaruhi oleh sulitnya menentukan jenis penyakit yang menjangkit pada tanaman. Sehingga perlu mengklasifikasikan penyakit-penyakit pada tanaman teh tersebut. data mining banyak algoritma yang

dapat digunakan dalam pembuatan pohon keputusan, sebagai berikut algoritma ID3, CART dan C4.5 (Swastina, 2018).

Penelitian sebelumnya yang dilakukan oleh (Sularno & Anggraini, 2017) yang menggunakan klasifikasi metode C4.5 untuk klasifikasi tingkat keganasan hama pada tanaman padi, dimana hasil penelitian yang didapat adalah Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi eksekutif dan sistem digunakan untuk menggambarkan proses yang terkait dengan pengklasifikasian keganasan hama tanaman padi. Sehingga, dapat dikatakan bahwa algoritma C4.5 dapat digunakan untuk mengklasifikasikan penyakit pada tanaman. Dalam penelitian ini Algoritma C4.5 digunakan untuk mengklasifikasi penyakit tanaman teh

1.2. Perumusan Masalah

Rumusan masalah pada penelitian ini bagaimana menerapkan algoritma C4.5 dalam mengklasifikasi penyakit pada tanaman teh?

1.3. Batasan Masalah

Adapun batasan masalah pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menganalisa data hasil rekaman PTPN IV Sidamanik.
2. Jenis penyakit yang akan diklasifikasi:
 - a. Penyakit Daun :
 - 1) *Blister Blight* penyakit cacar daun teh ini yang disebabkan oleh jenis jamur bernama *Exobasidium vexans*.
 - 2) Busuk Daun adalah penyakit yang disebabkan jamur *cylindrocladiumilicicola*.
 - 3) Mati Ujung disebabkan oleh jamur *Pestalolita Theae* yang menyerang tanaman terutama melalui luka atau bagian daun yang rusak.
3. Penentuan penyakit pada tanaman teh menggunakan metode algoritma C4.5

1.4. Tujuan Penelitian

Tujuan penelitian sebagai berikut:

1. Menerapkan algoritma C4.5 untuk mengklasifikasi penyakit tanaman teh.
2. Membantu pihak yang terkait atau perusahaan untuk mengklasifikasi penyakit pada tanaman teh.

1.5. Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Secara Teoritis
 - a. Penelitian ini untuk pengembangan ilmu pengetahuan, terkhusus ilmu Teknik Informatika dan Ilmu Komputer, serta dalam ilmu pertanian.
 - b. Menemukan cara pengimplementasian Algoritma C4.5 terhadap pengklasifikasian penyakit tanaman teh.
2. Secara Metodologi, penelitian ini sebagai rujukan untuk mengembangkan sistem yang lebih lanjut.
3. Secara Praktis
 - a. Bagi Peneliti
Memahami proses kerja Algoritma C4.5 ke bentuk implementasi.
 - b. Bagi Universitas
Sebagai bahan referensi penelitian yang akan datang dan sebagai bahan evaluasi bagi universitas dalam mengembangkan pengetahuan, yang berkaitan dengan Algoritma C4.5.

1.6. Sistematika Penulisan

Sistematika penulisan yang akan diajukan pada penyusunan skripsi ini antara lain:

BAB I: Pendahuluan

Pada bab ini menjelaskan secara singkat tentang latar belakang, rumusan masalah, tujuan serta manfaat, batasan masalah dan sistematika penulisan.

BAB II: Tinjauan Pustaka

Pada bab ini menerapkan tentang teori dasar yang berhubungan dengan program yang akan dirancang serta metode yang digunakan.

BAB III: Metodologi Penelitian

Pada bab ini menerapkan metodologi penelitian ataupun perancangan yang digunakan, setiap rencana dari tahapan penelitian dideskripsikan secara rinci.

BAB IV: Implementasi dan Pengujian Sistem

Pada bab ini membahas tentang langkah-langkah implementasi dan hasil program yang terdiri dari tampilan program, alur program dan penjelasan program.

BAB V: Kesimpulan dan Saran

Pada bab ini menjelaskan tentang kesimpulan dan saran dari penelitian yang telah dilakukan

BAB II TINJAUAN PUSTAKA

2.1. Tanaman Teh

Teh merupakan jenis tanaman dengan nama latin *Camellia Sinensis* yang kebanyakan hidup di wilayah tropis. Karakteristik tanaman teh adalah batangnya lebih tegak dan daunnya tunggal tumbuh berselang-seling, daun teh nya lonjong dan memiliki tulang runcing yang berbulu di ujungnya. Tepi daun teh tajam dengan gerigi halus. Dibandingkan dengan daun teh tua, daun teh muda warnanya lebih terang dan ukurannya lebih lebar. Warna daun teh tua berwarna hijau tua, dan permukaan daunnya lebih licin dibandingkan daun teh muda. tanaman teh merupakan tanaman secara fisik terbilangan sangat lemah, tanaman teh sangat membutuhkan perawatan yang sangat insentif karna sangat mudah di serang penyakit dan hama (Ibrahim, Lestary, & Hanafi, 2022).

Teh adalah salah satu minuman yang mempunyai banyak kepentingan karena di dalam banyak kandungan antioksidannya yang sangat tinggi, biasanya teh hijau disajikan pada minuman teh hijau, teh putih, teh hitam, sebanyak 78% produksi teh hitam sangat di minati oleh Negara-negara barat, 20% nya produksi teh hijau diminati oleh Negara-negara asia, dan 2% nya teh oolong yang diolah secara (fermentasi parsial) banyak diminati Negara-negara Tiongkok Selatan. Teh ini juga mengandung flavonoidnya yang cukup tinggi dan sanagat bagus untuk kesehatan karena bisa mengurangi resiko penyakit kardiovaskular kanker dan kesehatan mulut.

Karena sangat meningkatnya minat masyarakat mengkonsumsi teh maka dimasukkannya teh kelompok minuman dengan sifat fungsional (Anjarsari, 2022).

2.1.1. Penyakit Tanaman Teh

Penyakit tanaman teh yang terjadi di

a) *Blister Blight* (cacar daun)

Pada daun teh yang terserang tampak ada bintik selaput kecil yang pertama berukuran kecil samapai mebesar mencapai 10-15 mm, bercak melebar, daun berlobang, dibagian bawah daun tampak ada selaput putih.



Gambar 2. 1 Cacar Daun

b) Busuk Daun (*Phytophthora infestans*)

Pada daun tampak bercak-bercak coklat, tunas mengering, daun terlepas dari tangkai mati stek.



Gambar 2. 2 Busuk Daun

c. Mati ujung (*die back*)

Gejala pada daun dimulai bercak kecil berwarna coklat, kemudian melebar, pusat bercak keabu-abuan dengan tepinya berwarna coklat.



Gambar 2. 3 Mati Ujung

2.2. Data mining

Data mining adalah proses pengelompokan dan pengolahan data yang berfungsi untuk menentukan informasi penting pada data. Data mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database* dengan cara menggunakan teknik statistic, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang di dalam *database*. Data mining merupakan bagian dari tahapan proses *Knowledge Discovery in Database*(KDD). Kita dapat melakukan pengklasifikasian, memprediksi, memprekirakan dan mendapatkan informasi lain yang bermanfaat dari kumpulan jumlah data besar (Utomo & Mesran, 2020).

2.3. Klasifikasi

Klasifikasi berasal dari kata latin “*classis*” klasifikasi yang artinya adalah pengelompokan sesuatu dengan proses membedakan atau mendistribusikan jenis dalam pengempokan yang berbeda, secara umum dapat dikatakan bahwa batasan klasifikasi usaha menata alam pengetahuan ke dalam tata urtan sistematis.

Tujuan klasifikasi adalah entitas benda yang sama atau hampir sama, tujuan memisahkan entitas benda yang berbeda dengan cara memberikan nomor klasifikasi. Dengan kata lain tujuan klasifikasi adalah untuk mempermudah dalam menyajikan bahan perpustakaan, sehingga akan memudahkan dalam pencariannya oleh pemustaka. Pemberian nomor klasifikasi ini supaya nantinya buku-buku dengan subjek yang sama akan tergabung dalam satu jajaran di rak. Nomor klasifikasi ini dicetak pada label yang kemudian ditempel pada punggung bahan perpustakaan. Dengan demikian, buku tentang pendidikan akan menyatu dalam tata urutan yang sistematis (Subroto, 2019)

2.4. Pohon Keputusan

Pohon keputusan merupakan metode yang sangat populer digunakan, pohon keputusan bisa mengubah data fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan yang mudah di pahami dengan bahasa alami dan bisa di terapkan dengan bahasa Basis Data seperti *Structured Query Language*, menjadi *record* tertentu

Pohon keputusan memiliki node yang mempresentasikan atribut yang telah di uji, merupakan suatu pembagian hasil uji serta node daun (*leaf*), level node teratas

sebuah keputusan adalah node akar (*root*) memiliki peran penting dalam sebuah atribut dan kelas tertentu (Elisa, 2017)

2.5. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang sudah banyak digunakan untuk klasifikasi data memiliki atribut-atribut numerik, kemampuan dari hasil klasifikasi yang dapat digunakan memprediksi nilai atribut bertipe diskret dari *record* yang baru, algoritma C4.5 adalah sebuah pengembangan dari metode algoritma ID3, pengembangan dari metode ID3 ini adalah bisa mengatasi missing data, kontinu dan pruning (Elisa, 2017)

Secara garis besar algoritma C4.5 untuk membangun pohon keputusan sebagai berikut :

1. Pilih atribut sebagai akar,
2. Buang cabang untuk tiap-tiap nilai,
3. Ulangi proses untuk setiap cabang sampai semua kasus memiliki kelas yang sama.

Untuk memilih atribut akar, didasarkan pada nilai gain dari atribut-atribut tertinggi yang ada, Untuk menghitung gain digunakan rumus seperti yang tertera dalam Persamaan (2.1) berikut :

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{A_i}{S} * \text{Entropy}(s_i) \quad (2.1)$$

Keterangan:

S: himpunan kasus

A: atribut

N: jumlah partisi atribut

$A \setminus S_i$: jumlah kasus pada partisi ke-i

$|S_r|$: jumlah kasus dalam S

Sementara itu, penghitung nilai entropi dapat dilihat dari persamaan berikut :

$$\text{Entropy}(S) = \sum_{f=1}^n -p_i * \log_2 p_i$$

Keterangan:

S: himpunan kasus

A: atribut

n: jumlah partisi S

p_i : proporsi dari S_i terhadap S

Sebagai contoh, untuk membentuk pohon keputusan dapat dilakukan dengan dengan menentukan atribut sebagai akar lalu mencari nilai entropi dan again (Asroni, Masajeng Respati, & Riyadi, 2018).

2.6. Confusion Matrix

Confusion matrix dapat mengukur kinerja metode klasifikasi dan memperoleh informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang sebenarnya. Ada 4 istilah representasi hasil proses klasifikasi: *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), *False Negative* (FN), Nilai *True Negative* (TN) sebagai jumlah data negative yang terdeteksi dengan benar, sedangkan merupakan *False Positive* merupakan data negative namun terdeteksi sebagai data positif. Dibawah ini tabel *Confusion Matrix* (Karsito & Susanti, 2019).

Tabel 2. 1 Confusion Matrix

Nilai prediksi	Nilai Aktual	
	TP	TN
	FP	TN

Rumus untuk perhitungan *confusion matrix* untuk menghitung *accuracy*,

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (2.3)$$

2.7. Python

Python dibuat pada akhir 1980-an oleh Guido van Rossum di Centrum Wiskunde & informatica (CWI) di Belanda, *python* merupakan bahasa pemrograman serba guna yang bisa dijalankan pada hampir semua arsitektur sistem dan bisa digunakan untuk berbagai aplikasi dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif, *Python* bisa juga dibilang bahasa pemrograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat *source code* mudah dibaca Pemformatannya tidak berantakan secara visual, dan sering kali menggunakan kata kunci bahasa Inggris di mana bahasa lain menggunakan tanda baca (Ryan, Cooper, & Tauer, 2013).

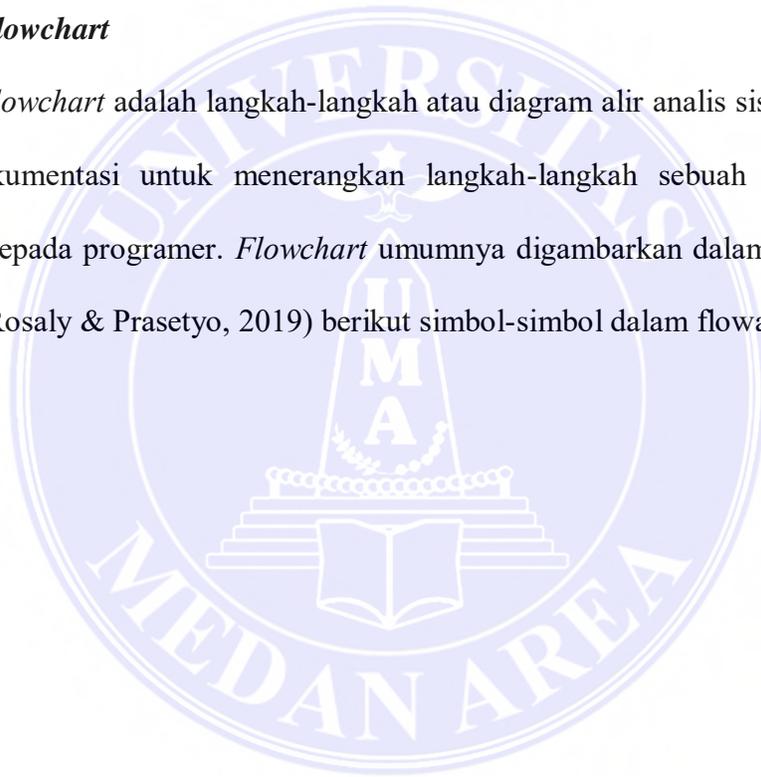
2.8. Google Colab

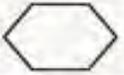
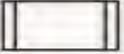
Google Colab merupakan produk dari *Google Research*, *Google Colab* sering digunakan untuk menulis dan mengeksekusi kode *python* melalui browser, dan sangat cocok juga untuk pengembangan *machine learning*, *Deep learning* dan data analysis. Kelebihan dari *Google Colab* yakni tanpa perlu konfigurasi, akses gratis ke General Processing Utlis dan mudah untuk di bagikan, *Google Colab Notebook* programmer

bisa mengkombinasikan kode yang bisa dieksekusi dan rich text dalam satu dokumen tunggal bersama dengan *image*, *HTML*, *LaTeX*. *Google Colab* dapat berkolaborasi dengan pengguna lainnya melalui berbagai *coding* secara *online*, *Colab* merupakan layanan *notebook Jupyter* yang dihosting dan dapat digunakan tanpa penyiapan, serta menyediakan akses gratis ke resource komputasi termasuk GPU (Soen, Marlina, & Renny, 2022).

2.9. *Flowchart*

Flowchart adalah langkah-langkah atau diagram alir analisis sistem dan sebagai bukti dokumentasi untuk menerangkan langkah-langkah sebuah sistem yang dibangun kepada programmer. *Flowchart* umumnya digambarkan dalam bentuk simbol-simbol (Rosaly & Prasetyo, 2019) berikut simbol-simbol dalam flowchart



	Flow Direction symbol Yaitu simbol yang digunakan untuk menghubungkan antara simbol yang satu dengan simbol yang lain. Simbol ini disebut juga connecting line.		Simbol Manual Input Simbol untuk pemasukan data secara manual on-line keyboard
	Terminator Symbol Yaitu simbol untuk permulaan (start) atau akhir (stop) dari suatu kegiatan		Simbol Preparation Simbol untuk mempersiapkan penyimpanan yang akan digunakan sebagai tempat pengolahan di dalam storage.
	Connector Symbol Yaitu simbol untuk keluar - masuk atau penyambungan proses dalam lembar / halaman yang sama.		Simbol Predefine Proses Simbol untuk pelaksanaan suatu bagian (sub-program)/prosedure
	Connector Symbol Yaitu simbol untuk keluar - masuk atau penyambungan proses pada lembar / halaman yang berbeda.		Simbol Display Simbol yang menyatakan peralatan output yang digunakan yaitu layar, plotter, printer dan sebagainya.
	Processing Symbol Simbol yang menunjukkan pengolahan yang dilakukan oleh komputer		Simbol disk and On-line Storage Simbol yang menyatakan input yang berasal dari disk atau disimpan ke disk.
	Simbol Manual Operation Simbol yang menunjukkan pengolahan yang tidak dilakukan oleh computer		Simbol magnetik tape Unit Simbol yang menyatakan input berasal dari pita magnetik atau output disimpan ke pita magnetik.
	Simbol Decision Simbol pemilihan proses berdasarkan kondisi yang ada.		Simbol Punch Card Simbol yang menyatakan bahwa input berasal dari kartu atau output ditulis ke kartu
	Simbol Input-Output Simbol yang menyatakan proses input dan output tanpa tergantung dengan jenis peralatannya.		Simbol Dokumen Simbol yang menyatakan input berasal dari dokumen dalam bentuk kertas atau output dicetak ke kertas.

Gambar 2. 4 Flowchart

2.10. Penelitian Terdahulu

Penelitian ini menggunakan beberapa penelitian terdahulu yang terkait dengan klasifikasi dan Algoritma C4.5 terdapat beberapa penelitian yang melakukan klasifikasi penyakit tanaman teh namun menggunakan algoritma yang lain. Terdapat pula beberapa penelitian menggunakan algoritma C4.5 untuk mengklasifikasi objek lainnya. Tabel dibawah ini merangkum beberapa penelitian terdahulu yang terkait dengan penelitian ini.

Tabel 2. 2 Penelitian Terdahulu

No	Nama Penulis (Tahun)	Judul	Algoritma	Hasil Penelitian
1	(Mardi, 2018)	Data Mining Rekam Medis Untuk Menentukan Penyakit Terbanyak menggunakan DECISION TREE C4.5	DECISION TREE C4.5	Hasil penelitian yang di peroleh dengan menggunakan algoritma C4.5 sangat mudah untuk digunakan dalam memperoleh pohon keputusan
2	(Suryadi, 2019)	Sistem Pakar Pendiaknosaan Hama Dan Penyakit Pada Tanaman Teh Menggunakan Metode Case- Based Reasoning Berbasis	K-NEAREST NEIGHBOUR RETRIEVAL (K-NN)	Perancangan sistem pakar dapat mempermudah dalam melakukan pendiagonasaan terhadap hama dan penyakit tanaman teh,dari penelurusan gejala-gejala yang dirasakan dengan

		Android		hasil diagnose dari sebuah hama atau penyakit tanaman the berdasarkan fakta yang terjadi sebelumnya.
3	(Sartika & Yupianti, 2020)	Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5 RSUD Hasanuddin Damrah manna	C4.5	Hasil penelitian ini menjelaskan bahwa setelah adanya sisitem ini pihak rumah sakit telah terbantu dalam pengklasifikasian penyakit yang terjadi akibat oleh kelenjar tiroid
4	(Fazri Ernawati Erlansari, 2019)	Sistem Pakar Penyakit Dan Hama Pada Tanaman Teh Menggunakan Certainty	Certainty Factor	Hasil peneletian ini menggunakan certainty factor 90% sesusi dengan pengujian yang ada dan keakuratannya

		Factor Berbasis Android		
--	--	-------------------------------	--	--

Berikut adalah perbandingan penelitian terdahulu dengan penelitian yang dibuat oleh penulis:

1. Penelitian yang dibuat (Mardi, 2018) menggunakan *DECISION TREE C4.5* Untuk Menentukan Penyakit Terbanyak perbedaan dengan penelitian ini yaitu peneliti mengklasifikasi penyakit tanaman teh
2. Penelitian yang di buat (Suryadi, 2019) menggunakan *NEAREST NEIGHBOUR RETRIEVAL (K-NN)* untuk Sistem Pakar Pendiaknosaan hama dan Penyakit pada tanaman teh, perbedaan yaitu penelitian menggunakan Algoritma C4.5 mengklasifikasi penyakit tanaman teh
3. Selanjutnya penelitian yang dibuat (Sartika & Yupianti, 2020) dengan menerapkan Algoritma C4.5 untuk klasifikasi penyakit tiroid, perbedaan yang dilakukan dengan penelitian ini adalah klasifikasi penyakit tanaman teh
4. Penelitian yang dibuat (Fazri Ernawati Erlansari, 2019) menggunakan Algoritma *Certainty Factor* untuk sistem pakar hama dan penyakit pada tanaman teh, perbedaan yaitu peneliti menggunakan klasifikasi dan Algoritma C4.5.

Dari penjelasan tertera diatas, dapat disimpulkan bahwa penelitian dengan judul klasifikasi penyakit tanaman teh menggunakan algoritma C4.5 layak untuk diteliti.



BAB III METODOLOGI PENELITIAN

Metodologi penelitian digunakan sebagai pedoman dalam pelaksanaan penelitian agar hasil yang dicapai tidak menyimpang dari tujuan yang telah dilakukan sebelumnya sebelumnya. Metode penelitian mencakup prosedur dan teknik penelitian. Metode penelitian merupakan langkah penting untuk memecahkan masalah-masalah penelitian serta mengembangkan bidang keilmuan yang di geluti.

3.1. Pengumpulan Data

Pada sub bab ini akan membahas mengenai jenis data dan sumber data.

3.1.1. Jenis Data Penelitian

Adapun jenis data yang digunakan penulis dalam penelitian ini menggunakan data primer. Data primer yang didapatkan berasal dari PTPN IV Sidamanik. Dimana dengan mengumpulkan semua data yang dibutuhkan. Berikut ini adalah tabel data yang digunakan pada penelitian ini:

Tabel 3. 1 Data Jenis Penyakit

NO	Kode Penyakit	Nama Penyakit
1	0	Cacar daun
2	1	Busuk daun
3	2	Mati ujung

Tabel 3. 2 Data Jenis Penyakit

No	Kode Ciri-Ciri Penyakit	Ciri-Ciri Penyakit	Keterangan
1	C1	Timbul bercak cokelat pada daun induk	Jika “ya” mengalami = 1, jika “tidak” = 0
2	C2	Daun berlobang	Jika “ya” mengalami = 1, jika “tidak” = 0
3	C3	Tunas mengering	Jika “ya” mengalami = 1, jika “tidak” = 0
4	C4	Bercak melebar	Jika “ya” mengalami = 1, jika “tidak” = 0
5	C5	Daun terlepas dari tangkai	Jika “ya” mengalami = 1, jika “tidak” = 0
6	C6	Mati stek	Jika “ya” mengalami = 1, jika “tidak” = 0
7	C7	Daun berbintik selaput kecil tembus cahaya	Jika “ya” mengalami = 1, jika “tidak” = 0
8	C8	Bercak kecil berwarna cokelat, kemudin melabar	Jika “ya” mengalami = 1, jika “tidak” = 0
9	C9	Pusat bercak keabu-abuan	Jika “ya” mengalami = 1, jika “tidak” = 0
10	C10	Tepi daun berwarna cokelat	Jika “ya” mengalami = 1, jika “tidak” = 0

Tabel 3. 3 Data Tanaman

No	Blok	C1	C2	C3	C4	C5	C6	C7	PENYAKIT
1	B1	0	1	0	1	0	0	1	Busuk daun
2	B2	1	0	1	0	1	1	0	Cacar daun
3	B3	0	1	1	1	0	0	1	Busuk daun
4	B4	1	0	1	1	0	0	1	Mati ujung
5	B5	0	1	1	0	1	0	1	Cacar daun
6	B6	0	1	0	1	0	1	1	Busuk daun
7	B7	0	1	0	1	0	0	0	Busuk daun
8	B8	0	1	0	1	0	0	0	Busuk daun
9	B9	0	1	0	0	0	0	0	Mati ujung
10	B10	1	0	1	1	1	1	0	Mati ujung

3.1.2. Metode Pengumpulan Data

Berikut pengumpulan data yang akan dilakukan dalam pembuatan skripsi:

1. Tahap Wawancara

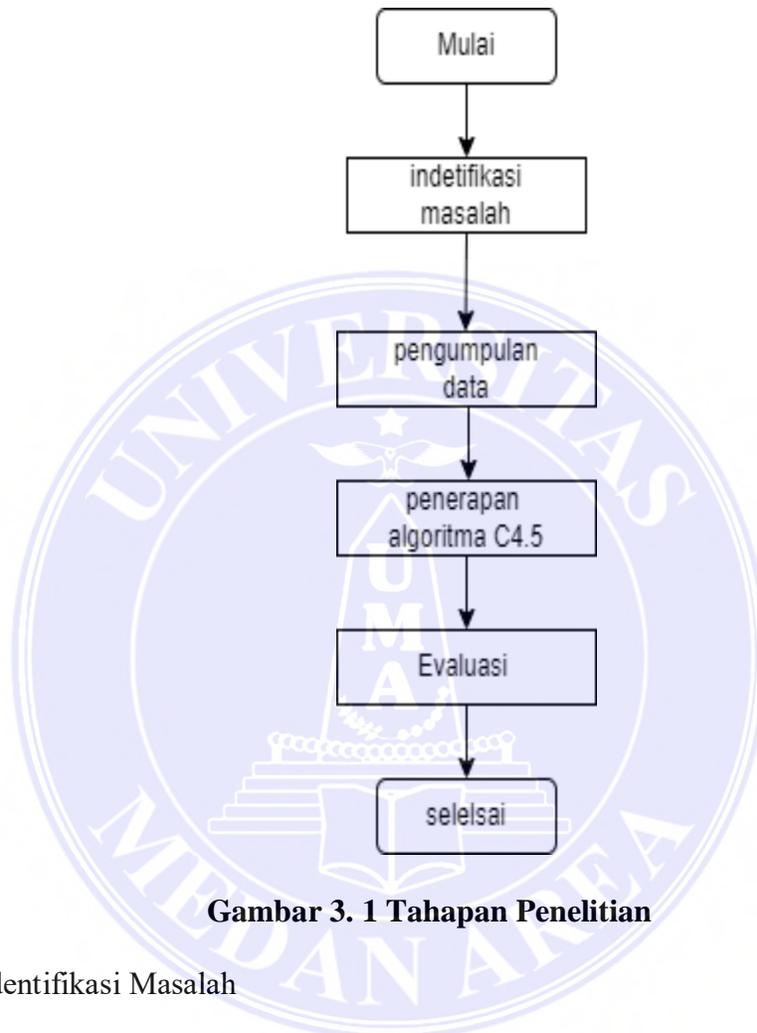
Pengambilan data primer pada pembuatan skripsi ini menggunakan teknik wawancara terhadap asisten lapangan. Pengumpulan data diambil secara langsung dari tempat penelitian.

2. Studi Pustaka

Penulis mencari berbagai sumber referensi ke pustaka sebagai acuan penelitian yang mendukung pada permasalahan kasus laporan yang dibuat

3.2. Tahapan Penelitian

Berikut adalah diagram alir penelitian



Gambar 3. 1 Tahapan Penelitian

1. Identifikasi Masalah

Adapun masalah pada penelitian adalah bagaimana cara menerapkan algoritma C4.5 untuk mengklasifikasi penyakit tanaman teh

2. Studi Pustaka

Pada bagian pencarian referensi dilakukan untuk mendukung judul penelitian judul penelitian, baik itu berupa buku maupun artikel jurnal. Pencarian tersebut dilakukan untuk menyelesaikan suatu masalah pada penelitian.

3. Pengumpulan Data

Pengumpulan data yang dilakukan peneliti dengan melakukan observasi, wawancara dengan asisten lapangan, peneliti akan memperoleh data ini dari PTPN IV SIDAMANIK dengan cara.

4. Penerapan Algoritma C4.5

Setelah semua data yang diperlukan lengkap, maka langkah selanjutnya, maka akan dilakukan transformasi atau mengubah nilai atribut ke dalam bentuk data yang sesuai agar dapat dikelola menggunakan Algoritma C4.5 sehingga dapat data set untuk melakukan langkah selanjutnya.

5. Evaluasi

Dalam penelitian ini dilakukan tahap evaluasi supaya dihasilkan sebuah sistem pengklasifikasi penyakit pada tanaman teh yang efektif dan akurat . dengan cara menerapkan metode *confusion matrix*.

3.3. Kebutuhan Perangkat

Pada penelitian yang akan dilakukan terdapat beberapa kebutuhan yang akan digunakan untuk melakukan penelitian, kebutuhan perangkat pada penelitian ini penulis menggunakan *hardware* dapat dilihat sebagai berikut :

1. *Processor Intel® Celeron® Prosesor N4000*
2. RAM 4 GB
3. SSD 500 GB

Software yang digunakan adalah sebagai berikut:

1. Sistem Operasi *Windows 10 Home single*
2. *Google Colab*

3. Python

3.4. Proses Penerapan Perhitungan Manual C4.5

Pada perhitungan manual berikut disini peneliti hanya membuat contoh perhitungan manual dengan identifikasi penyakit cacar daun dan daun busuk, data yang digunakan hanya menggunakan dummy data.

Jika terdapat (*true*) =1, jika tidak (*false*) =

Selanjutnya di bagi ke dalam 2 kelas yaitu kelas cacar daun dan busuk daun.

Tabel 3. 4 Contoh Kasus

No	Blok	C1	C2	C3	C4	C5	C6	C7	Penyakit
1	B1	0	1	0	1	0	0	1	Busuk Daun
2	B2	1	0	1	0	1	1	0	Cacar Daun
3	B3	0	1	1	1	0	0	1	Busuk Daun
4	B4	1	0	1	1	0	0	1	Mati Ujung
5	B5	0	1	1	0	1	0	1	Cacar Daun
6	B6	0	1	0	1	0	1	1	Cacar Daun
7	B7	0	1	0	1	0	0	0	Busuk Daun
8	B8	0	1	0	1	0	0	0	Busuk Daun
9	B9	0	1	0	0	0	0	0	Mati Ujung
10	B10	1	0	1	1	1	1	0	Mati Ujung

3.5. Pilih Atribut Akar

Pada penentuan sebuah atribut akar, akan menghitung nilai *gain* dan memilih nilai *gain* yang tertinggi, untuk menghitung nilai *gain* dan nilai *entropy* dengan menggunakan rumus ruzukan no 2.2. Berikut contoh perhitungan.

1. *Entropy* Total

Dimana diketahui jumlah data adalah 10 record dengan kelas cacar daun 6 dan kelas daun busuk 4

$$Entropy\ total = \left(-\frac{6}{10} * \log_2\left(\frac{6}{10}\right)\right) + \left(-\frac{4}{10} * \log_2\left(\frac{4}{10}\right)\right) = 0,970950594$$

2. Perhitungan *entropy* setiap atribut

- a) Yang mana jumlah data C1. False 7, 5 record dengan kelas cacar daun dan kelas busuk daun 2. Jumlah C1 True adalah 3 dengan kelas cacar daun 1 dan busuk daun 2.

$$Entropy\ false = \left(-\frac{5}{7} * \log_2\left(\frac{6}{7}\right)\right) + \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right)$$

$$= 0,863120569$$

$$Entropy\ true = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right)$$

$$= 0,918295834$$

$$Gain = 0,970950594 \left(\left(\frac{7}{10} * 0,863120569\right) + \left(\frac{3}{10} * 0,918295834\right) \right)$$

$$= 0,642254947$$

- b) C2, yang mana jumlah data false 3 dengan kelas cacar daun 1 dan kelas busuk daun 2, dan jumlah data true 7 dengan kelas cacar daun 5 dan kelas busuk daun 2.

$$Entropy\ false = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right)$$

$$= 0,918295834$$

$$\begin{aligned} \text{Entropy true} &= \left(-\frac{5}{7} * \log_2\left(\frac{5}{7}\right)\right) + \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right) \\ &= 0,863120569 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,970950594 \left(\left(\frac{3}{10} * 0,918295834\right) + \left(\frac{7}{10} * 0,8632120569\right) \right) \\ &= 1,299646242 \end{aligned}$$

C3, yang mana jumlah data false 5, dengan kelas cacar daun 1 dan kelas busuk daun 4, dan jumlah data true 5 dengan kelas cacar daun 2 dan kelas busuk daun 3.

$$\begin{aligned} \text{Entropy false} &= \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) \\ &= 0,721928095 \end{aligned}$$

$$\begin{aligned} \text{Entropy true} &= \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{2}{5} * \log_2\left(\frac{3}{5}\right)\right) \\ &= 0,970950594 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,970950594 \left(\left(\frac{6}{10} * 0,918295834\right) + \left(\frac{4}{10} * 0,8632120569\right) \right) \\ &= 1,095461844 \end{aligned}$$

c) C4, yang mana jumlah data false 3 dengan kelas cacar daun 1 dan kelas busuk daun 2, dan jumlah data true 7 dengan kelas cacar daun 5 dan kelas busuk daun 2.

$$\begin{aligned} \text{Entropy false} &= \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) \\ &= 0,918295834 \end{aligned}$$

$$\begin{aligned} \text{Entropy True} &= \left(-\frac{5}{7} * \log_2\left(\frac{5}{7}\right)\right) + \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right) \\ &= 0,863120569 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,970950594 \left(\left(\frac{3}{10} * 0,918295834\right) + \left(\frac{7}{10} * 0,8632120569\right) \right) \\ &= 1,115987075 \end{aligned}$$

- d) C5 yang mana jumlah data false 6 dengan kelas cacar daun 5 dan kelas busuk daun 1 dan jumlah data true 4 dengan kelas cacar daun 1 dan kelas busuk daun 3.

$$\begin{aligned} \text{Entropy false} &= \left(-\frac{5}{6} * \log_2\left(\frac{5}{6}\right)\right) + \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) \\ &= 0,650022422 \end{aligned}$$

$$\begin{aligned} \text{Entropy True} &= \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) \\ &= 0,811278124 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,970950594 \left(\left(\frac{6}{10} * 0,650022422\right) + \left(\frac{4}{10} * 0,811278124\right) \right) \\ &= 0,905448391 \end{aligned}$$

- e) C6, yang mana jumlah data false 6 dengan kelas cacar daun 5 dan kelas busuk daun 1, dan jumlah true 4 dengan kelas cacar daun 1 dan kelas daun busuk 3.

$$\begin{aligned} \text{Entropy False} &= \left(-\frac{5}{6} * \log_2\left(\frac{5}{6}\right)\right) + \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) \\ &= 0,650022422 \end{aligned}$$

$$\text{Entropy True} = \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right)$$

$$= 0,811278124$$

$$\begin{aligned} \text{Gain} &= 0,970950594 \left(\left(\frac{6}{10} * 0,650022422 \right) + \left(\frac{4}{10} * 0,811278124 \right) \right) \\ &= 0,905448391 \end{aligned}$$

- f) C7 yang mana jumlah data false 3 dengan kelas cacar daun 1 dan kelas busuk daun 2, dan jumlah data true 7 dengan kelas cacar daun 5 dan kelas busuk daun 2.

$$\begin{aligned} \text{Entropy False} &= \left(-\frac{1}{3} * \log_2 \left(\frac{1}{3} \right) \right) + \left(-\frac{2}{3} * \log_2 \left(\frac{2}{3} \right) \right) \\ &= 0,91829583 \end{aligned}$$

$$\begin{aligned} \text{Entropy True} &= \left(-\frac{5}{7} * \log_2 \left(\frac{5}{7} \right) \right) + \left(-\frac{2}{7} * \log_2 \left(\frac{2}{7} \right) \right) \\ &= 0,86312057 \end{aligned}$$

$$\begin{aligned} \text{Gain} &= 0,970950594 \left(\left(\frac{3}{10} * 0,91829583 \right) + \left(\frac{7}{10} * 0,86312057 \right) \right) \\ &= 1,299646242 \end{aligned}$$

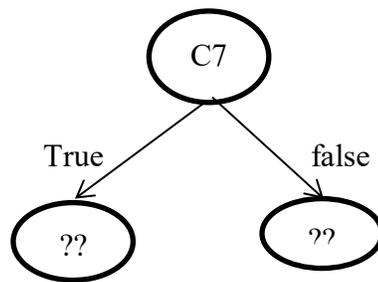
Hasil perhitungan diatas dapat disimpulkan dalam tabel berikut :

Tabel 3. 5 Perhitungan Node 1

	Jumlah Data	Cacar Daun	Busuk Daun	Entropy	Gain
Total	10	6	4	0,970950594	
C1					
FALSE	7	5	2	0,863120569	0,642254947
TRUE	3	1	2	0,918295834	
C2					

FALSE	3	1	2	0,918295834	0,091277446
TRUE	7	5	2	0,863120569	
C3					
FALSE	5	1	4	0,721928095	1,095461844
TRUE	5	2	3	0,970950594	
C4					
FALSE	3	1	2	0,918295834	1,115987075
TRUE	7	5	2	0,863120569	
C5					
FALSE	6	5	1	0,650022422	0,905448391
TRUE	4	1	3	0,811278124	
C6	6	5	1	0,650022422	0,905448391
FALSE	4	1	3	0,811278124	
TRUE					
C7					
FALSE	3	1	2	0,918295834	1,299646242
TRUE	7	5	2	0,863120569	

Berdasarkan hasil perhitungan yang dibuat pada tabe 3.5 nilai *Gain* yang tertinggi terdapat pada atribut C7. Yang mana atribut tersebut akan menjadi akar simpul.



Gambar 3. 2 Pohon Keputusan

Berikut untuk node 1.1 maka kita akan menghitung kembali nilai *entropy* dari true dan false C7, setelah melakukan perhitungan *entropy*, selanjutnya hitung kembali nilai *gain*.

1) *Entropy* dan *gain* true C7

Untuk melakukan perhitungan selanjutnya menghitung nilai *entropy* dan *gain* C7 true

$$\begin{aligned} \text{Entropy total C7 true} &= \left(-\frac{5}{7} * \log_2\left(\frac{5}{7}\right)\right) + \left(-\frac{2}{7} * \log_2\left(\frac{2}{7}\right)\right) \\ &= 0,86312 \end{aligned}$$

$$\begin{aligned} \text{a) Entropy C1 false} &= \left(-\frac{4}{6} * \log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} * \log_2\left(\frac{2}{6}\right)\right) \\ &= 0,9183 \end{aligned}$$

$$\begin{aligned} \text{Entropy c1 true} &= \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) \\ &= 0 \end{aligned}$$

$$\text{Gain} == 0,86312 - \left(\left(\frac{6}{7} * 0,9183\right) + \left(\frac{1}{7} * 0\right)\right) = 0,07601$$

$$\begin{aligned} \text{b) Entropy C2 false} &= \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{1}{1}\right)\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy C2 true} &= \left(-\frac{4}{6} * \log_2\left(\frac{4}{6}\right)\right) + \left(-\frac{2}{6} * \log_2\left(\frac{2}{6}\right)\right) \\ &= 0,9183 \end{aligned}$$

$$\text{Gain} = 0,86312 - \left(\left(\frac{1}{7} * 0\right) + \left(\frac{6}{7} * 0,918295834\right)\right) = 0,07601$$

$$\begin{aligned} \text{c) Entropy C3 false} &= \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) \\ &= 0,81128 \end{aligned}$$

$$\begin{aligned} \text{Entropy C3 true} &= \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) \\ &= 0,9183 \end{aligned}$$

$$\text{Gain} = 0,86312 - \left(\left(\frac{4}{7} * 0,81128\right) + \left(\frac{3}{7} * 0,9183\right)\right) = 0,79309$$

$$\text{d) Entropy C4 false} = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$\begin{aligned} \text{Entropy C4 true} &= \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) \\ &= 0,72193 \end{aligned}$$

$$\text{Gain} = 0,86312 - \left(\left(\frac{2}{7} * 1\right) + \left(\frac{5}{7} * 0,72193\right)\right) = 0,80735$$

$$\text{e) Entropy C5 false} = \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right)$$

$$= 0,72193$$

$$Entropy\ C5\ true = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right)$$

$$= 1$$

$$Gain = 0,86312 - \left(\left(\frac{5}{7} * 0,72193\right) + \left(\frac{2}{7} * 1\right)\right) = 0,63317$$

f) $Entropy\ C6\ false = \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right)$

$$= 0,72193$$

$$Entropy\ C4\ true = \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right)$$

$$= 1$$

$$Gain = 0,86312 - \left(\left(\frac{5}{7} * 0,72193\right) + \left(\frac{2}{7} * 1\right)\right) = 0,63317$$

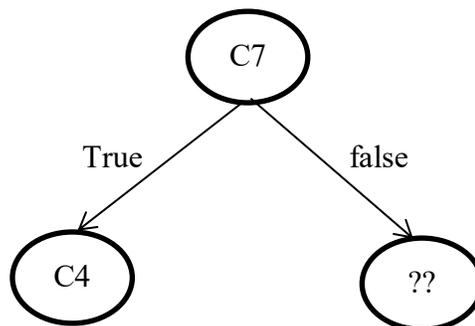
Dengan jumlah data = 7 dengan kelas cacar daun 5 dan busuk daun 2, maka hasil dari perhitungan itu dibuat dalam sebuah tabel.

Tabel 3. 6 Tabel Perhitungan

	Jumlah Data	Cacar Daun	Busuk Daun	Entropy	Gain
Total	7	5	2	0,86312	
C1					
FALSE	6	4	2	0,9183	0,07601
TRUE	1	1	0	0	
C2					

FALSE	1	1	0	0	0,07601
TRUE	6	4	2	0,9183	
C3					
FALSE	4	3	1	0,81128	0,79309
TRUE	3	2	1	0,9183	
C4					
FALSE	2	1	1	1	0,80735
TRUE	5	4	1	0,72193	
C5					
FALSE	5	4	1	0,72193	0,63317
TRUE	2	1	1	1	
C6	5	4	1	0,72193	0,63317
FALSE	2	1	1	1	
TRUE					

Dari tabel 3.6 dilihat bahwa nilai *gain* tertinggi ialah C4, yang berarti C4 ialah akar selanjutnya



Gambar 3. 3 Pohon Keputusan

2) *Entropy* dan *gain* C7 False

$$\begin{aligned} \text{Entropy total C7 false} &= \left(-\frac{2}{3} * \log_2 \left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} * \log_2 \left(\frac{1}{3}\right)\right) \\ &= 0,9183 \end{aligned}$$

$$a) \text{ Entropy C1 false} = \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1}\right)\right) = 0$$

$$\begin{aligned} \text{Entropy C1 true} &= \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2}\right)\right) \\ &= 0 \end{aligned}$$

$$\text{Gain} = 0,9183 - \left(\left(\frac{1}{3} * 0\right) + \left(\frac{2}{3} * 0\right)\right) = 0,9183$$

$$b) \text{ Entropy C2 false} = \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2}\right)\right)$$

$$= 0$$

$$\begin{aligned} \text{Entropy C2 true} &= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1}\right)\right) \\ &= 0 \end{aligned}$$

$$\text{Gain} = 0,9183 - \left(\left(\frac{1}{3} * 0\right) + \left(\frac{2}{3} * 0\right)\right) = 0,9183$$

$$c) \text{ Entropy C3 false} = \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1}\right)\right)$$

$$= 0$$

$$\text{Entropy C3 true} = \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2}\right)\right)$$

$$= 0$$

$$Gain = 0,9183 - \left(\left(\frac{1}{3} * 0 \right) + \left(\frac{2}{3} * 0 \right) \right) = 0,9183$$

$$\begin{aligned} \text{d) Entropy C4 false} &= \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right) + \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1} \right) \right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy C4 true} &= \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2} \right) \right) + \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2} \right) \right) \\ &= 0 \end{aligned}$$

$$Gain = 0,9183 - \left(\left(\frac{1}{3} * 0 \right) + \left(\frac{2}{3} * 1 \right) \right) = 1,585$$

$$\begin{aligned} \text{e) Entropy C5 false} &= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1} \right) \right) + \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right) \\ &= 0 \end{aligned}$$

$$\text{Entropy C5 true} = \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2} \right) \right) + \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2} \right) \right)$$

$$Gain = 0,9183 - \left(\left(\frac{1}{3} * 0 \right) + \left(\frac{2}{3} * 1 \right) \right) = 0,9183$$

$$\text{f) Entropy C6 false} = \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right) + \left(-\frac{2}{1} * \log_2 \left(\frac{2}{1} \right) \right)$$

$$\text{C5 true} = \left(-\frac{0}{2} * \log_2 \left(\frac{0}{2} \right) \right) + \left(-\frac{2}{2} * \log_2 \left(\frac{2}{2} \right) \right)$$

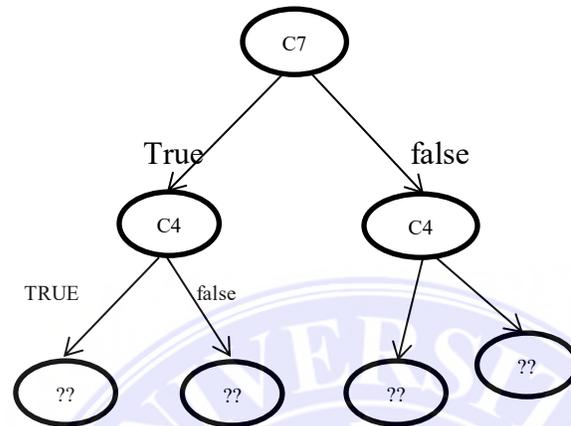
$$Gain = 0,9183 - \left(\left(\frac{1}{3} * 0 \right) + \left(\frac{2}{3} * 1 \right) \right) = 0,9183$$

Dengan jumlah data = 3 dengan kelas cacar daun 2 dan busuk daun 1, maka hasil dari perhitungan itu dibuat dalam sebuah tabel

Tabel 3. 7 Tabel Perhitungan

	Jumlah Data	Cacar Daun	Busuk Daun	Entropy	Gain
Total	3	2	1	0,9183	
C1					
FALSE	1	1	0	0	0,9183
TRUE	2	0	2	0	
C2					
FALSE	2	0	2	0	0,9183
TRUE	1	1	0	0	
C3					
FALSE	1	1	0	0	0,9183
TRUE	2	2	0	0	
C4					
FALSE	1	0	1	0	1,58496
TRUE	2	1	1	1	
C5					
FALSE	1	1	0	0	0,9183
TRUE	2	0	2	0	
C6	1	1	0	0	0,9183
FALSE	2	0	2	0	
TRUE					

Dari tabel 3.7 dilihat bahwa nilai *gain* tertinggi ialah C4, yang berarti C4 ialah akar selanjutnya



Gambar 3. 4 Pohon keputusan

Selanjutnya menghitung *entropy* dan *gain* C4 dari turunan C7 true

3) *Entropy* dan *gain* C4 true

Dengan jumlah data 5, dan kelas cacar daun 4, dan busuk daun 1.

$$\begin{aligned} \text{Entropy total C4 true} &= \left(-\frac{4}{5} * \log_2 \left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) \\ &= 0,72193 \end{aligned}$$

$$\begin{aligned} \text{a) Entropy C1 false} &= \left(-\frac{3}{4} * \log_2 \left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4}\right)\right) \\ &= 0,81128 \end{aligned}$$

$$\begin{aligned} \text{Entropy C1 true} &= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2 \left(\frac{1}{1}\right)\right) \\ &= 0 \end{aligned}$$

$$\text{Gain} = 0,72193 - \left(\left(\frac{4}{5} * 0,81128\right) + \left(\frac{1}{5} * \right)\right) = 0,0729056$$

$$\text{b) Entropy C2 false} = \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$\begin{aligned} \text{Entropy C2 true} &= \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) \\ &= 0,81128 \end{aligned}$$

$$\text{Gain} = 0,72193 - \left(\left(\frac{4}{5} * 0,81128\right) + \left(\frac{1}{5} * 0,81128\right)\right) = 0,0729056$$

$$\begin{aligned} \text{c) Entropy C3 false} &= \left(-\frac{3}{3} * \log_2\left(\frac{3}{3}\right)\right) + \left(-\frac{0}{3} * \log_2\left(\frac{0}{3}\right)\right) \\ &= 0 \end{aligned}$$

$$\begin{aligned} \text{Entropy C3 true} &= \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) \\ &= 1 \end{aligned}$$

$$\text{Gain} = 0,72193 - \left(\left(\frac{3}{5} * 0\right) + \left(\frac{2}{5} * 1\right)\right) = 0,32192809$$

$$\begin{aligned} \text{d) Entropy C5 false} &= \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) \\ &= 0,81128 \end{aligned}$$

$$\begin{aligned} \text{Entropy C5 true} &= \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) \\ &= 0 \end{aligned}$$

$$\text{Gain} = 0,72193 - \left(\left(\frac{4}{5} * 0,81128\right) + \left(\frac{1}{5} * 0,81128\right)\right) = 0,0729056$$

$$\begin{aligned} \text{e) Entropy C6 false} &= \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) + \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) \\ &= 0 \end{aligned}$$

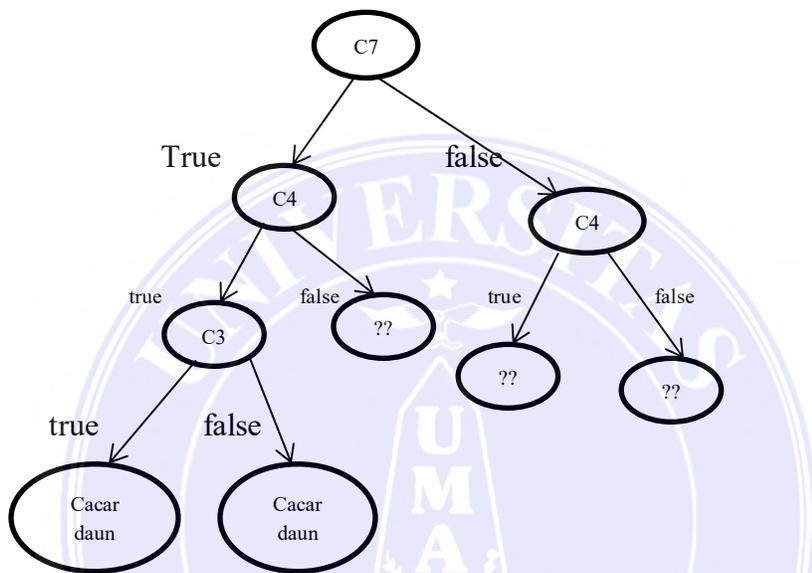
$$Entropy\ C6\ true = \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right)$$

$$Gain = 0,72193 - \left(\left(\frac{4}{5} * 0,81128\right) + \left(\frac{1}{5} * 0,81128\right)\right) = 0,0729056$$

Tabel 3. 8 Perhitungan

	Jumlah Data	Cacar Daun	Busuk Daun	Entropy	Gain
Total	5	4	1	0,72193	
C1					0,0729056
FALSE	4	3	1	0,81128	
TRUE	1	1	0	0	
C2					0,0729056
FALSE	1	1	0	0	
TRUE	4	3	1	0,81128	
C3					0,32192809
FALSE	3	3	0	0	
TRUE	2	1	1	1	
C5					0,0729056
FALSE	4	3	1	0,81128	
TRUE	1	1	0	0	
C6					0,0729056
FALSE	4	3	1	0,81128	
TRUE	1	1	0	0	

Dari tabel diatas nilai *gain* tertinggi nya adalah C3, yang berarti C3 akar selanjutnya dan *gain* true dari setia ciri-ciri dari tabel diatas 0 maka tidak perlu di hitung lagi dan *entropy* false dari C3 0 tidak perlu di hitung lagi



Gambar 3. 5 Pohon Keputusan

4) *Entropy* dan *gain* C4 false dari turunan turunan akar C7 true

Dengan jumlah data 2 dengan kelas cacar daun 1 dan kelas busuk daun 1

$$\begin{aligned}
 \text{a) } \textit{Entropy} \text{ total C4 false} &= \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) \\
 &= 1
 \end{aligned}$$

$$\begin{aligned}
 \textit{Entropy} \text{ C1 false} &= \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2}\right)\right) \\
 &= 1
 \end{aligned}$$

$$\textit{Entropy} \text{ true C1} = \left(-\frac{4}{5} * \log_2 \left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right)$$

$$Gain = 1 - \left(\left(\frac{2}{2} * 0,81128 \right) + \left(\frac{0}{2} * 0 \right) \right) = 0$$

$$b) \text{ Entropy C2 false} = \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0} \right) \right) + \left(-\frac{0}{0} * \log_2 \left(\frac{0}{0} \right) \right) = 0$$

$$\begin{aligned} \text{Entropy C2 true} &= \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2} \right) \right) + \left(-\frac{1}{2} * \log_2 \left(\frac{1}{2} \right) \right) \\ &= 0 \end{aligned}$$

$$Gain = 1 - \left(\left(\frac{2}{2} * 0,81128 \right) + \left(\frac{0}{2} * 0 \right) \right) = 0$$

$$c) \text{ Entropy C3 false} = \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right) + \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1} \right) \right)$$

$$\begin{aligned} \text{Entropy C2 true} &= \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1} \right) \right) + \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right) \\ &= 0 \end{aligned}$$

$$Gain = 1 - \left(\left(\frac{1}{2} * 0 \right) + \left(\frac{1}{2} * 0 \right) \right) = 1$$

$$d) \text{ Entropy C5 false} = \left(-\frac{1}{1} * \log_2 \left(\frac{0}{1} \right) \right) + \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right)$$

$$= 0$$

$$\text{Entropy C5 true} = \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right) + \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1} \right) \right)$$

$$= 0$$

$$Gain = 1 - \left(\left(\frac{1}{2} * 0 \right) + \left(\frac{1}{2} * 0 \right) \right) = 1$$

$$e) \text{ Entropy C6 false} = \left(-\frac{1}{1} * \log_2 \left(\frac{1}{1} \right) \right) + \left(-\frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \right)$$

$$= 0$$

$$Entropy\ C6\ true = \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right)$$

$$= 0$$

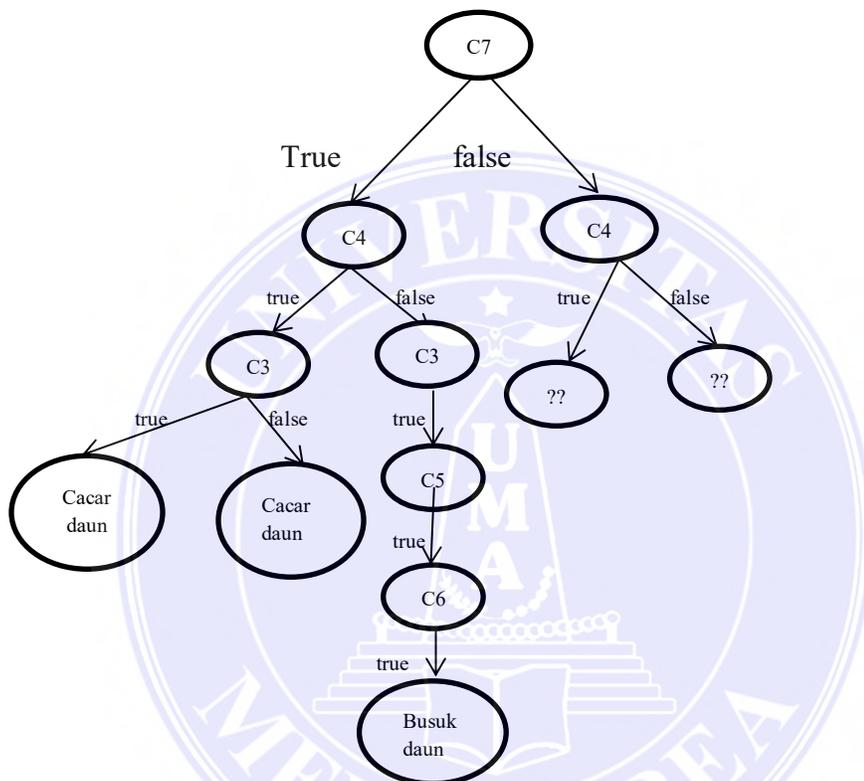
$$Gain = 1 - \left(\left(\frac{1}{2} * 0\right) + \left(\frac{1}{2} * 0\right)\right) = 1$$

Tabel 3. 9 Perhitungan

	Jumlah Data	Cacar Daun	Busuk Daun	Entropy	Gain
Total	2	1	1	1	
C1					
FALSE	2	1	1	1	0
TRUE	0	0	0	0	
C2					
FALSE	0	0	0	0	0
TRUE	2	1	1	1	
C3					
FALSE	1	0	1	0	1
TRUE	1	1	0	0	
C5					
FALSE	1	1	0	0	1
TRUE	1	0	1	0	
					1
C6	1	1	0	0	
FALSE	1	0	1	0	

TRUE					
------	--	--	--	--	--

Dari tabel diatas nilai *gain* tertinggi ada 3 yaitu C3, C5, dan C6 karena memiliki jumlah data yang sama, maka akar selanjutnya adalah C3, C5, dan C6



Gambar 3. 6 Pohon Keputusan

Berdasarkan hasil perhitungan dan pohon keputusan yang telah dibuat, langkah selanjutnya disimpulkan sebuah aturan *rule* sebagai berikut:

If C7= true, and C4=true, and C3=true then cacar daun

If C7=true and C4=true, and C3=false then cacar daun

If C7 true, and C4 false, and C3 true, and C5 true, and C6 tue then busuk daun

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil pembahasan dan uji coba yang telah dilakukan, dapat disimpulkan:

1. Dengan menggunakan metode C4.5 telah berhasil mengklasifikasi penyakit tanaman teh.
2. Pengujian algoritma C4.5 menggunakan *confusion matrix* memperoleh akurasi sebesar 95%

5.2. Saran

Saran untuk pengembangan penelitian ini dimasa yang akan datang sebagai berikut:

1. Diharapkan dapat menambahkan jenis penyakit seperti akar merah anggur, penyakit akar merah bata, penyakit akar hitam, penyakit leher akar, selain cacar daun, busuk daun dan mati ujung.
2. Pada penelitian selanjutnya diharapkan menggunakan data lebih banyak agar tingkat akurasi nya lebih baik, dan bagaimana jika menggunakan cara atau perhitungan yang lain sebagai pembanding atau mendapatkan hasil yang lebih beragam.

DAFTAR PUSTAKA

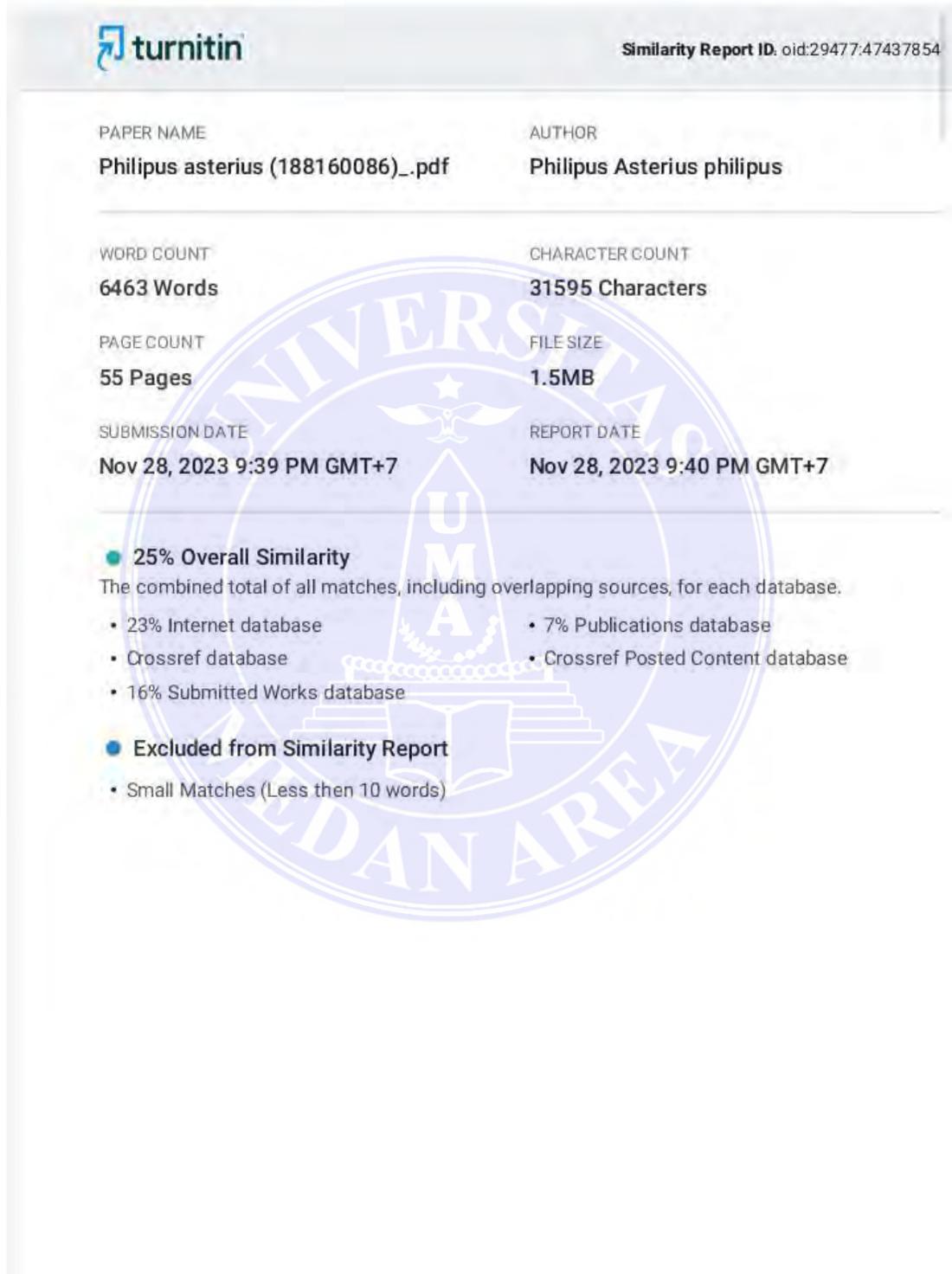
- Anjarsari, I. R. D. (2022). Rekayasa budidaya dan penanganan pascapanen untuk meningkatkan kualitas teh Indonesia sebagai minuman fungsional kaya antioksidan. *Antioksidan dalam Produk Teh*, 21(2), 152–158.
- Asroni, A., Masajeng Respati, B., & Riyadi, S. (2018). Penerapan Algoritma C4.5 untuk Klasifikasi Jenis Pekerjaan Alumni di Universitas Muhammadiyah Yogyakarta. *Semesta Teknika*, 21(2), 158–165.
<https://doi.org/10.18196/st.212222>
- Burhan, A. B. (2018). Pemanfaatan Teknologi Informasi dan Komunikasi untuk Pengembangan Ekonomi Pertanian dan Pengentasan Kemiskinan. *Jurnal Komunikasi Pembangunan*, 16(2), 233–247.
<https://doi.org/10.46937/16201826338>
- Damanik, Y., Okprana, H., & Sormin, R. K. (2022). Sistem Pendukung Keputusan Penentuan Produk Teh Terbaik Menggunakan Metode Moora Pada PTPN IV Sidamanik, 1(1), 24–33.
- Elisa, E. (2017). Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti. *Jurnal Online Informatika*, 2(1), 36.
<https://doi.org/10.15575/join.v2i1.71>
- Fazri Ernawati Erlansari. (2019). Sistem Pakar Penyakit Dan Hama Pada Tanaman Pada Tanaman Teh Menggunakan Certainty Factor Berbasis Android.
- Ibrahim, N. U. R., Lestary, G. A. Y. U., & Hanafi, F. S. (2022). Klasifikasi Tingkat

- Kematangan Pucuk Daun Teh menggunakan Metode Convolutional Neural Network. *Jurnal Energi Elektrik, Teknik Telekomunikasi & Teknik Elektronika.*, 10(1), 162–176.
- Karsito, & Susanti, S. (2019). Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia. *Jurnal Teknologi Pelita Bangsa*, 9, 43–48.
- Ma'rifah, Z. (2020). *Mengenal Teh Hijau*. Alprin. Retrieved from https://books.google.co.id/books?id=km_-DwAAQBAJ
- Mardi, Y. (2018). Data Mining Rekam Medis Untuk Menentukan Penyakit Terbanyak Menggunakan Decision Tree C4.5. *Jurnal Sains Dan Informatika*, 4(1), 40–53. <https://doi.org/10.22216/jsi.v4i1.3077>
- Rosaly, R., & Prasetyo, A. (2019). Pengertian Flowchart Beserta Fungsi dan Simbol-simbol Flowchart yang Paling Umum Digunakan. *Https://Www.Nesabamedia.Com*, 2, 2.
- Ryan, Cooper, & Tauer. (2013). o Title No Title No Title. *Paper Knowledge . Toward a Media History of Documents*, 12–26.
- Sartika, D., & Yupianti, Y. (2020). Klasifikasi Penyakit Tiroid Menggunakan Algoritma C4.5 (Studi Kasus : Rumah Sakit Umum Daerah (RSUD) Hasanuddin Damrah Manna). *Rekayasa*, 13(1), 71–76. <https://doi.org/10.21107/rekayasa.v13i1.5912>
- Soen, G. I. E., Marlina, & Renny. (2022). Implementasi Cloud Computing dengan Google Colaboratory Pada Aplikasi Pengolah Data Zoom Participants. *Journal Informatic Technology And Communication*, 6(1), 24–30.

- Subroto, G. (2019). Klasifikasi bahan pustaka. *Pustakawan Perpustakaan UM*, (Ddc), 1–13.
- Sularno, S., & Anggraini, P. (2017). PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI TINGKAT KEGANASAN HAMA PADA TANAMAN PADI (Studi Kasus : Dinas Pertanian Kabupaten Kerinci). *Jurnal Sains Dan Informatika*, 3(2), 161. <https://doi.org/10.22216/jsi.v3i2.2779>
- Suryadi, A. (2019). Sistem Pakar Pendiaknosaan Hama Dan Penyakit Pada Tanaman Teh Menggunakan Metode Case-Based Reasoning Berbasis Android. *Jursima*, 7(2), 107. <https://doi.org/10.47024/js.v7i2.168>
- Swastina, L. (2018). Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa. *Gema Aktualita*, 2(1), 93–98.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>

LAMPIRAN

Lampiran 1 Hasil Turnitin



The image shows a Turnitin similarity report for a document titled "Philipus asterius (188160086).pdf" by Philipus Asterius philipus. The report includes statistics on word count (6463 words), character count (31595 characters), page count (55 pages), and file size (1.5MB). It also shows the submission date (Nov 28, 2023 9:39 PM GMT+7) and the report date (Nov 28, 2023 9:40 PM GMT+7). The overall similarity is 25%, with a breakdown of matches from various databases: Internet (23%), Publications (7%), Crossref (Crossref database and Crossref Posted Content database), and Submitted Works (16%). Excluded matches include small matches (less than 10 words). A large watermark of Universitas Medan Area is visible in the background of the report.

PAPER NAME	AUTHOR
Philipus asterius (188160086).pdf	Philipus Asterius philipus

WORD COUNT	CHARACTER COUNT
6463 Words	31595 Characters

PAGE COUNT	FILE SIZE
55 Pages	1.5MB

SUBMISSION DATE	REPORT DATE
Nov 28, 2023 9:39 PM GMT+7	Nov 28, 2023 9:40 PM GMT+7

- **25% Overall Similarity**
The combined total of all matches, including overlapping sources, for each database.
 - 23% Internet database
 - 7% Publications database
 - Crossref database
 - Crossref Posted Content database
 - 16% Submitted Works database
- **Excluded from Similarity Report**
 - Small Matches (Less than 10 words)

Lampiran 2 SK Pembimbing



Lampiran 3 SK Penelitian dan Pengambilan Data Tugas Akhir

 **UNIVERSITAS MEDAN AREA**
FAKULTAS TEKNIK
Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax. (061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8225331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

30 Nopember 2022

Nomor : 347 /FT.6/01.10/XI/2022
Lamp : -
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Pimpinan PTPN IV Unit Kebun Sidamanik
Kec. Sidamanik
Di
Simalungun

Dengan hormat,
Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Philipus Asterius	188160086	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana lengkap pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :

Klasifikasi Penyakit pada Tanaman Teh menggunakan Algoritma C4.5 (Studi Kasus : PTPN IV Unit Kebun Sidamanik)

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.


Dr. Rahnada Syah, S. Kom, M. Kom

Tembusan :
1. Ka. BAMAI
2. Mahasiswa
3. File

Lampiran 4 Surat Selesai Riset



UNIT USAHA TEH
PT PERKEBUNAN NUSANTARA IV
SIMALUNGUN-SUMATERA UTARA-INDONESIA

KANTOR UNIT USAHA : BAH BUTONG TELP : (0622) 25617
KANTOR PUSAT : JL. LETJEND SUPRAPTO NO. 2 MEDAN TELP : (061) 45773117

SURAT KETERANGAN
BUT/SK/09/1/2023

Yang bertanda tangan dibawah ini :

Nama : Hwin Dwi Putera
Jabatan : Manajer Unit Usaha Teh
Alamat : PT Perkebunan Nusantara IV Unit Usaha Teh

Menerangkan dengan sebenarnya bahwa Mahasiswa Fakultas Teknik Universitas Medan Area atas nama sbb :

NO	NAMA	Program Study	NIM
1	PHILIPUS ASTERIUS	Informatika	188160086

Telah selesai melakukan Penelitian dan Pengambilan Data Tugas Akhir di PT.Perkebunan Nusantara IV Unit Usaha Teh Sidamanik, terhitung mulai tanggal 29 Desember 2022 s/d 30 Januari 2023 dengan judul Penelitian :

"Klasifikasi Penyakit Tanaman Teh Menggunakan Algoritma C4.5 (Studi Kasus : PTPN IV UNIT KEBUN SIDAMANIK"

Demikian surat keterangan ini diperbuat untuk dapat dipergunakan seperlunya.

Sidamanik, 31 Januari 2023
PT Perkebunan Nusantara IV
Unit Usaha Teh

Hwin Dwi Putera
Manajer

AKHLAK - Amanah, Kompeten, Harmonis, Loyal, Adaptif, Kolaboratif.