

**KLASIFIKASI PENYAKIT INFEKSI KULIT PADA KUCING
MENGUNAKAN ALGORITMA RANDOM FOREST**

SKRIPSI

OLEH:

ADITHYA WAHYUDI NASUTION

188160040



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 18/1/24

Access From (repository.uma.ac.id)18/1/24

**KLASIFIKASI PENYAKIT INFEKSI KULIT PADA KUCING
MENGUNAKAN ALGORITMA RANDOM FOREST**

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana di Fakultas Teknik

Universitas Medan Area

OLEH:

ADITHYA WAHYUDI NASUTION

188160040

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 18/1/24

Access From (repository.uma.ac.id)18/1/24

HALAMAN PENGESAHAN

Judul skripsi : Klasifikasi Penyakit Infeksi Kulit Pada Kucing
Menggunakan Algoritma Random Forest

Nama : Adithya Wahyudi Nasution

Npm : 188160040

Fakultas : Teknik



Tanggal Lulus : 13 September 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 sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 13 September 2023



Adithya Wahyudi Nasution
NIM 188160040

**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 : Adithya Wahyudi Nasution

NPM : 188160040

Program Studi : Teknik Informatika

Fakultas : Teknik

Jenis karya : Tugas Akhir/Skripsi/Tesis

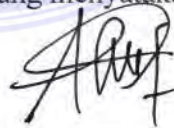
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul : *Klasifikasi Penyakit Infeksi Kulit Pada Kucing Menggunakan Algoritma Random Forest* beserta perangkat 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 tugas akhir/skripsi/tesis 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 : 13 September 2023

Yang menyatakan



(Adithya Wahyudi Nasution)

ABSTRAK

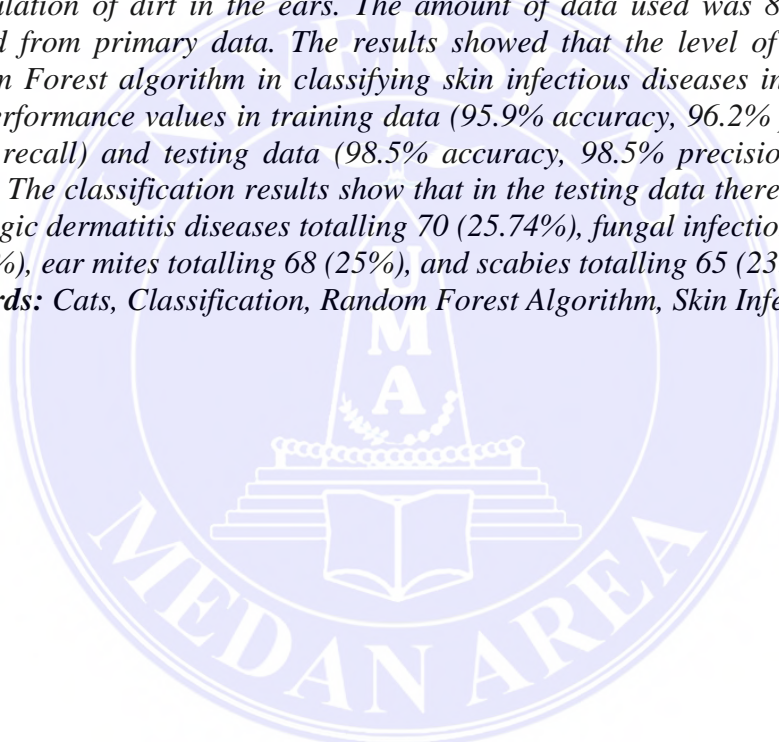
Kucing sebagai hewan peliharaan populer memiliki peran penting dalam kehidupan manusia. Namun, seperti hewan lain, mereka rentan terhadap berbagai penyakit, termasuk infeksi kulit. Infeksi kulit pada kucing dapat disebabkan oleh bakteri, jamur, virus, atau parasit, menyebabkan gejala seperti gatal dan infeksi yang berpotensi fatal. Diagnosa dini dan pengobatan yang tepat sangat penting untuk mencegah dampak buruk. Saat ini, diagnosis sering dilakukan oleh dokter hewan dengan pemeriksaan fisik dan tes laboratorium, yang memakan waktu dan biaya. Solusi lebih efisien diperlukan, seperti penggunaan algoritma pembelajaran mesin. Algoritma Random Forest, yang efektif dalam mengatasi masalah klasifikasi dengan fitur yang kompleks dan tidak linier, menjadi alternatif yang dapat digunakan. Penelitian ini bertujuan untuk mengklasifikasikan penyakit infeksi kulit pada kucing menggunakan algoritma Random Forest. Jenis penyakit yang diklasifikasikan adalah dermatitis alergi, infeksi jamur, kudis dan tungau telinga, dengan menggunakan 8 gejala sebagai atribut, yakni gatal-gatal, kulit kering, kulit kemerahan, bulu rontok, luka, kulit bersisik, kulit berkerak, penumpukan kotoran pada telinga. Jumlah data yang digunakan sebanyak 800 yang bersumber dari data primer. Hasil penelitian menunjukkan tingkat kemampuan algoritma Random Forest dalam melakukan klasifikasi penyakit infeksi kulit pada kucing mencapai nilai performa yang tinggi pada data training (akurasi 95,9%, presisi 96,2%, dan recall 95,9%) serta data testing (akurasi 98,5%, presisi 98,5%, dan recall 98,5,7%). Hasil pengklasifikasian menunjukkan bahwa pada data *testing* terdapat sejumlah penyakit dermatitis alergi berjumlah 70 (25,74%), penyakit infeksi jamur berjumlah 69 (25,37%), penyakit tungau telinga berjumlah 68 (25%), dan penyakit kudis berjumlah 65 (23,90%).

Kata Kunci : Algoritma Random Forest, Infeksi Kulit Pada Kucing, Klasifikasi, Kucing

ABSTRACT

Cats, as popular pets, play an important role in people's lives. However, like other animals, they are susceptible to various diseases, including skin infections. Skin infections in cats can be caused by bacteria, fungi, viruses or parasites, causing symptoms such as itching and potentially fatal infections. Early diagnosis and proper treatment are essential to prevent adverse effects. Currently, diagnosis is often done by veterinarians with physical examinations and laboratory tests, which is time-consuming and costly. More efficient solutions are needed, such as the use of machine learning algorithms. Random Forest algorithm, which is effective in overcoming classification problems with complex and non-linear features, is an alternative that can be used. This study aims to classify skin infection diseases in cats using Random Forest algorithm. The types of diseases classified are allergic dermatitis, fungal infections, scabies and ear mites, using 8 symptoms as attributes, namely itching, dry skin, reddish skin, fur loss, wounds, scaly skin, crusty skin, accumulation of dirt in the ears. The amount of data used was 800 which was sourced from primary data. The results showed that the level of ability of the Random Forest algorithm in classifying skin infectious diseases in cats reached high performance values in training data (95.9% accuracy, 96.2% precision, and 95.9% recall) and testing data (98.5% accuracy, 98.5% precision, and 98.5% recall). The classification results show that in the testing data there are a number of allergic dermatitis diseases totalling 70 (25.74%), fungal infections totalling 69 (25.37%), ear mites totalling 68 (25%), and scabies totalling 65 (23.90%).

Keywords: Cats, Classification, Random Forest Algorithm, Skin Infection in Cats

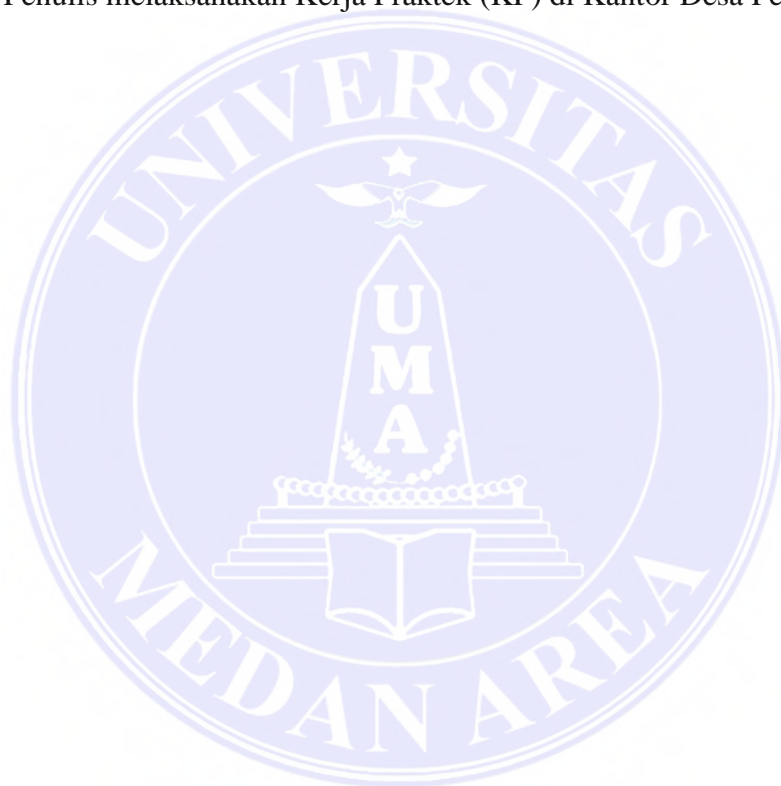


RIWAYAT HIDUP

Penulis dilahirkan di Sibolga. Pada tanggal 27 Agustus 1999 dari ayah Chairuddin Efendi Nasution dan ibu Asrida Hayani Simanjuntak. Penulis merupakan anak pertama dari 4 bersaudara.

Tahun 2018 Penulis lulus dari SMA Negeri 1 Tukka dan pada tahun 2018 terdaftar sebagai mahasiswa Fakultas Teknik Universitas Medan Area.

Penulis melaksanakan Kerja Praktek (KP) di Kantor Desa Pematang Johar.



KATA PENGANTAR

Assalamualaikum Wr. Wb,

Puji syukur ke hadirat Allah SWT, berkat rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi yang berjudul "Klasifikasi Penyakit Infeksi Kulit pada Kucing Menggunakan Algoritma Random Forest". Skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Strata-1 (S.Kom) pada jurusan Teknik Informatika, Fakultas Teknik.

Penulis menyadari bahwa penyusunan skripsi ini tidak lepas dari bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada:

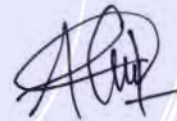
1. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc. selaku Rektor Universitas Medan Area.
2. Bapak Dr. Eng., Supriatno, S.T, M.T.selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom, M,Kom selaku Ketua Prodi Teknik Informatika Universitas Medan Area.
4. Bapak Andre Hasudungan Lubis, S.Ti, M,Sc Dosen Pembimbing I yang telah memberikan bimbingan dan arahan selama penulisan skripsi.
5. Dosen-dosen program studi Teknik Informatika yang mengajar di Fakultas Teknik Universitas Medan Area.
6. Kedua Orang Tua dan keluarga yang selalu menyemangati dan senantiasa mendo'akan penulis tiada henti sehingga dapat menyelesaikan skripsi ini.

7. Seluruh teman-teman informatika angkatan 2018 yang tidak dapat disebutkan satu persatu.
8. Teman-teman kost yang telah menemani selama penulisan skripsi ini.
9. Segala pihak yang tidak dapat disebutkan satu per satu, yang dengan ikhlas memberikan doa dan motivasi, telah memberikan kontribusi yang berarti dalam menyelesaikan skripsi ini. Dukungan dan semangat yang tulus dari mereka telah menjadi faktor penting dalam keberhasilan penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, oleh karena itu saran dan kritik yang membangun sangat penulis harapkan untuk perbaikan di masa yang akan datang.

Wassalamualaikum Wr. Wb

Medan, 13 September 2023

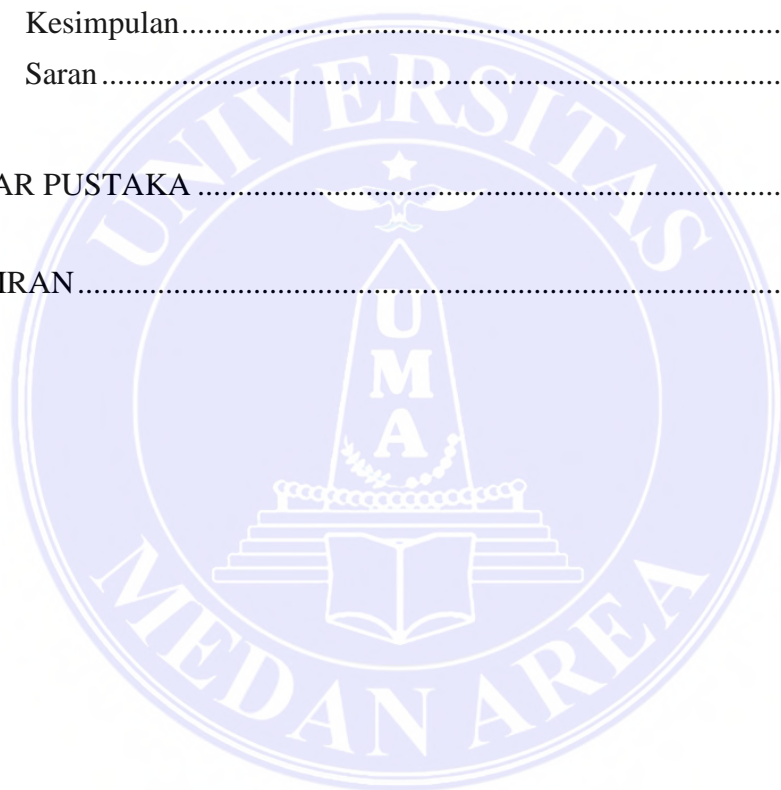


Adithya Wahyudi Nasution
NIM 188160040

DAFTAR ISI

ABSTRAK	iv
ABSTRACT	v
RIWAYAT HIDUP	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Batasan Penelitian	3
1.5 Manfaat	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 <i>Data Mining</i>	6
2.1.1 Tahapan <i>Data Mining</i>	7
2.1.2 Teknik <i>Data Mining</i>	9
2.2 Klasifikasi	10
2.3 Penyakit Infeksi Kulit Pada Kucing	11
2.4 Algoritma <i>Random Forest</i>	16
2.4.1 Langkah-langkah Perhitungan <i>Random Forest</i>	17
2.4.2 Rumus	18
2.5 <i>Confusion Matrix</i>	19
2.6 <i>Google Colab</i>	21
2.7 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	22
2.8 Penelitian Terdahulu	23
BAB III METODOLOGI PENELITIAN	26
3.1 Tahapan Penelitian	26
3.2 Bahan dan Alat Penelitian	28
3.3 Proses Pengumpulan Data	29
3.4 Perhitungan Manual	29

BAB IV HASIL DAN PEMBAHASAN	41
4.1 Hasil.....	41
4.1.1 Pengumpulan Data	41
4.1.2 <i>Data Preprocessing</i>	42
4.1.3 Algoritma Random Forest.....	44
4.1.4 <i>Confusion Matrix</i>	47
4.2 Pembahasan	53
BAB V KESIMPULAN DAN SARAN	56
5.1 Kesimpulan.....	56
5.2 Saran	56
DAFTAR PUSTAKA	58
LAMPIRAN	61



DAFTAR TABEL

Tabel 2. 1 Penyakit.....	15
Tabel 2. 2 Gejala	15
Tabel 2. 3 Confusion Matrix	20
Tabel 2. 4 Penelitian Terdahulu	23
Tabel 3. 1 Data Jenis Penyakit	29
Tabel 3. 2 Data Gejala Penyakit.....	30
Tabel 3. 3 Data Sampel	31
Tabel 4. 1 Hasil Performa Pembagian 60% : 40%	44
Tabel 4. 2 Hasil Performa Pembagian 70% : 30%	45
Tabel 4. 3 Hasil Performa Pembagian 80% : 20%	46
Tabel 4. 4 Perbandingan Penelitian.....	53

DAFTAR GAMBAR

Gambar 3. 1 Tahapan Penelitian	26
Gambar 3. 2 Nilai <i>Entropy</i> dan <i>Gain</i>	32
Gambar 3. 3 Menghitung Kembali Nilai <i>Entropy</i> dan <i>Gain</i> (1.2)	35
Gambar 3. 4 Menghitung Kembali Nilai <i>Entropy</i> dan <i>Gain</i> (1.3)	38
Gambar 3. 5 Pohon Keputusan.....	40
Gambar 4. 1 Dataset Penyakit Kucing	41
Gambar 4. 2 Tampilan Jumlah Kelas	42
Gambar 4. 3 Teknik SMOTE	43
Gambar 4. 4 Pembagian Data.....	43
Gambar 4. 5 Implementasi Algoritma Random Forest	44
Gambar 4. 6 Diagram Batang Ke-1.....	45
Gambar 4. 7 Diagram Batang Ke-2.....	46
Gambar 4. 8 Diagram Batang Ke-3.....	47
Gambar 4. 9 Confusion Matrix Pembagian 60% : 40%	47
Gambar 4. 10 Hasil Klasifikasi Pembagian 60% : 40%	48
Gambar 4. 11 Confusion Matrix Pembagian 70% : 30%	49
Gambar 4. 12 Hasil Klasifikasi Pembagian 70% : 30%	50
Gambar 4. 13 Confusion Matrix Pembagian 80% : 20%	51
Gambar 4. 14 Hasil Klasifikasi Pembagian 80% : 20%	52
Gambar 4. 15 Pohon Keputusan Random Forest	53

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kucing sebagai salah satu hewan peliharaan yang sangat populer, memiliki peran penting dalam hidup manusia. Kucing memiliki keindahan dan kelucuan yang membuat mereka sangat digemari. Pada awalnya, kucing merupakan hewan yang berasal dari alam liar. Kemudian, kucing mengalami proses domestikasi, yaitu pengadopsian dari kehidupan liar menjadi lingkungan kehidupan manusia sehari-hari (Loss dkk., 2022). Namun, seperti hewan lain, kucing juga dapat terkena berbagai jenis penyakit. Salah satu jenis penyakit yang sering dialami oleh kucing adalah infeksi kulit (Gunawan & Fernando, 2021).

Infeksi kulit pada kucing dapat disebabkan oleh berbagai hal, seperti infeksi bakteri, jamur, virus, dan parasit. Penyakit ini dapat menimbulkan gejala seperti ruam kulit, gatal, dan bahkan infeksi yang parah dapat menyebabkan kematian pada hewan tersebut. Oleh karena itu, penting bagi pemilik kucing untuk dapat mengenali dan mengatasi infeksi kulit pada kucing sejak dini (Zen dkk., 2021). Saat ini, diagnosis penyakit pada kucing biasanya dilakukan oleh dokter hewan melalui pemeriksaan fisik dan tes laboratorium. Namun, metode ini membutuhkan waktu yang lama dan biaya yang relatif mahal. Oleh karena itu, dibutuhkan suatu metode yang lebih efisien dan efektif dalam mendeteksi infeksi kulit pada kucing, seperti penggunaan teknologi informasi (Dami, 2022).

Algoritma pembelajaran mesin merupakan salah satu teknologi yang dapat digunakan untuk mengatasi masalah tersebut. Algoritma *Random Forest* adalah

salah satu algoritma pembelajaran mesin yang dapat digunakan untuk melakukan klasifikasi. Algoritma ini memiliki kelebihan dalam mengatasi masalah klasifikasi dengan data yang memiliki banyak fitur dan tidak linier.

Algoritma *Random Forest* telah digunakan untuk mengklasifikasi penyakit kulit pada manusia. Dimana, Khasanah dkk. (2021) telah melakukan penelitian tentang klasifikasi penyakit kanker kulit dengan menggunakan algoritma *Random Forest*. Adapaun penelitian yang dilakukannya menggunakan data citra digital dan menghasilkan klasifikasi dengan akurasi sebesar 84%. Penelitian lainnya yang menggunakan algoritma *Random Forest* dilakukan oleh Muttaqien dkk. (2020) untuk mengklasifikasi penyakit kucing pada umumnya. Penelitian ini menghasilkan akurasi sebesar 96%.

Berdasarkan hal tersebut, penulis memutuskan untuk melakukan penelitian dengan judul "Klasifikasi Penyakit Infeksi Kulit pada Kucing Menggunakan Algoritma *Random Forest*". Tujuan dari penelitian ini adalah untuk menganalisis kemampuan algoritma *Random Forest* dalam melakukan klasifikasi penyakit infeksi kulit pada kucing dan membandingkan hasil klasifikasi dengan metode yang biasa digunakan saat ini.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka yang menjadi pokok permasalahan adalah bagaimana melakukan klasifikasi penyakit infeksi kulit pada kucing menggunakan algoritma *random forest*?

1.3 Tujuan

Adapun Tujuan yang diharapkan adalah:

1. Untuk melakukan klasifikasi penyakit infeksi kulit pada kucing menggunakan algoritma *random forest*.
2. Untuk mengetahui efektivitas dan efisiensi algoritma *Random Forest* dalam melakukan klasifikasi penyakit infeksi kulit pada kucing.

1.4 Batasan Penelitian

Adapun Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Jenis Penyakit kulit kucing yang diklasifikasi berjumlah 4 yaitu Dermatitis alergi, Infeksi jamur, Kudis , Tungau telinga.
2. Jenis gejala yang di klasifikasikan berjumlah 8 yaitu Gatal-gatal, Kulit kering, Kulit kemerahan, Bulu rontok, Luka, Kulit bersisik, Kulit berkerak, Penumpukan kotoran pada telinga.
3. Data yang digunakan pada penelitian ini adalah data primer.
4. Algoritma *Random Forest* diterapkan pada *Google Colab* dengan bahasa pemrograman *Python*.

1.5 Manfaat

Penelitian ini diharapkan memberikan manfaat antara lain:

1. Manfaat secara teoritis

Penelitian ini diharapkan dapat memberikan kontribusi terhadap algoritma yang digunakan, yaitu *Random Forest* dalam mengklasifikasikan penyakit pada hewan khususnya kucing. Sehingga penelitian ini diharapkan untuk membuktikan

kelayakan algoritma *random forest* dalam mengklasifikasi data set penyakit kucing dan yang lainnya.

2. Manfaat secara metodologi

Penelitian ini menerapkan algoritma *Random Forest* untuk mengklasifikasi penyakit pada kucing kedalam sebuah program. Sehingga, penelitian ini diharapkan dapat menjadi rujukan bagi penelitian yang memiliki kesamaan tujuan atau objek yang sama.

3. Manfaat secara praktisi

Penelitian ini juga akan memberikan manfaat bagi bidang akademik. Secara spesifik, penelitian ini akan memberikan solusi alternatif dalam melakukan klasifikasi penyakit infeksi kulit pada kucing. Ini akan membantu meningkatkan hasil penelitian tentang masalah ini dan memberikan informasi penting bagi para ahli dan praktisi dalam bidang kedokteran hewan dan teknologi informasi. Hasil penelitian ini juga dapat menjadi referensi bagi penelitian sejenis yang akan datang, sehingga dapat memperkaya ilmu pengetahuan dan membantu memecahkan masalah yang ada.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi adalah urutan berpikir yang digunakan dalam penulisan dan sangat penting untuk memudahkan pembaca memahami isi laporan. Biasanya, sistematika penulisan dibagi menjadi lima bab dengan penjelasan singkat untuk setiap bab.

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian. Ini membantu pembaca memahami topik penelitian secara umum.

BAB II TINJAUAN PUSTAKA

Bab ini membahas studi literatur dan konsep-konsep teori yang terkait dengan topik penelitian. Ini bertujuan untuk memberikan pemahaman yang lebih mendalam bagi pembaca tentang topik penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian, termasuk desain penelitian, populasi dan sampel penelitian, teknik pengumpulan data, dan analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil dari penelitian yang telah dilakukan serta pembahasan atas hasil tersebut.

BAB V SIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan dari penelitian yang telah dilakukan, serta saran yang diberikan berdasarkan hasil penelitian tersebut.

BAB II

TINJAUAN PUSTAKA

2.1 Data Mining

Data mining adalah proses penggalian informasi yang dilakukan dengan menggunakan teknik-teknik analisis statistik dan komputasi untuk menemukan pola dan hubungan dalam dataset besar. Tujuan utama dari data mining adalah untuk mengidentifikasi pola yang tersembunyi atau tidak terlihat dalam data dan menghasilkan pengetahuan yang dapat membantu dalam pengambilan keputusan. Teknik-teknik *data mining* digunakan untuk memproses data dalam skala besar, mempelajari pola-pola yang muncul dalam data, dan membantu mengambil keputusan berdasarkan hasil analisis data (Ginantra dkk., 2021).

Data mining adalah sebuah metode untuk mengeksplorasi, menganalisis, dan menemukan informasi baru dari data besar dengan tujuan untuk memperoleh pengetahuan yang berguna dan berharga. Dalam *data mining*, analisis dilakukan pada data dengan menggunakan teknik-teknik seperti *clustering*, *association*, *classification*, dan prediksi (Ginantra dkk., 2021). Tujuan dari teknik-teknik tersebut adalah untuk mengelompokkan data berdasarkan pola tertentu, menemukan hubungan antar data, dan memprediksi kejadian di masa depan. Selain itu, *data mining* juga dapat membantu dalam mengoptimalkan pengambilan keputusan dan meningkatkan efisiensi bisnis dengan cara menemukan cara-cara baru untuk meningkatkan kinerja bisnis.

Data mining juga dapat membantu dalam mengeksplorasi data secara visual dan menemukan pola-pola yang tersembunyi di dalam data. Teknik-teknik *data*

mining yang digunakan seperti *clustering*, *association*, dan *classification* dapat membantu dalam mengidentifikasi hubungan antar data dan menemukan pola-pola yang tersembunyi. Pola-pola tersebut dapat digunakan untuk membuat model prediksi atau untuk mengambil keputusan yang lebih baik. Selain itu, *data mining* juga dapat digunakan untuk menemukan anomali atau kejadian yang tidak biasa dalam data, yang dapat membantu dalam mengidentifikasi masalah atau potensi kesalahan dalam proses bisnis. Dengan demikian, data mining dapat membantu dalam meningkatkan kualitas dan efektivitas bisnis dan pengambilan keputusan.

2.1.1 Tahapan Data Mining

Data mining memiliki beberapa tahapan yang harus dilalui sampai mendapatkan *knowledge* yang dapat digunakan untuk berbagai tujuan. Tahapan yang dilakukan dimulai dari pemilihan dan pengumpulan data sampai dengan evaluasi model. Adapun tahapan dalam *data mining* adalah sebagai berikut (Ginantra dkk., 2021):

1. Pemilihan dan pengumpulan data: Tahap awal *data mining* adalah pemilihan dan pengumpulan data yang akan digunakan dalam analisis. Data tersebut dapat diperoleh dari berbagai sumber seperti database, file teks, atau sumber data *online*.
2. *Pre-processing* data: Setelah data terkumpul, tahap selanjutnya adalah melakukan *pre-processing* data. Tujuan dari tahap ini adalah untuk memastikan data yang digunakan sudah bersih dan siap untuk diolah. Beberapa tindakan yang dilakukan pada tahap ini adalah membersihkan data dari *noise* dan *outlier*, melakukan normalisasi data, dan transformasi data.

3. Pemilihan atribut: Setelah data siap, tahap selanjutnya adalah melakukan pemilihan atribut. Pada tahap ini, atribut atau variabel yang tidak relevan atau tidak signifikan dieliminasi dari data. Hal ini dilakukan untuk meningkatkan kualitas data dan efisiensi analisis.
4. Transformasi data: Tahap ini melibatkan penggunaan teknik seperti reduksi dimensi dan *feature extraction* untuk menghasilkan data yang lebih mudah diolah. Tujuannya adalah untuk menghasilkan informasi yang lebih mudah dipahami dan digunakan dalam proses analisis.
5. Pemilihan teknik *data mining*: Tahap ini melibatkan pemilihan teknik atau algoritma *data mining* yang sesuai untuk analisis data. Beberapa teknik *data mining* yang umum digunakan adalah *clustering*, *association rule mining*, *decision tree*, dan *neural network*.
6. Implementasi algoritma: Setelah teknik *data mining* dipilih, tahap selanjutnya adalah mengimplementasikan algoritma pada data. Pada tahap ini, dilakukan proses analisis dan pengolahan data menggunakan algoritma yang telah dipilih.
7. Evaluasi model: Tahap terakhir adalah evaluasi model. Pada tahap ini, model yang telah dibangun diuji dan dinilai berdasarkan kemampuan dan keakuratan dalam memprediksi atau mengklasifikasikan data.

Tahapan di atas merupakan tahapan umum dalam proses *data mining*. Setelah tahapan tersebut selesai, hasil analisis dapat digunakan untuk membangun model prediksi atau untuk mengambil keputusan bisnis yang lebih baik.

2.1.2 Teknik *Data Mining*

Teknik *data mining* adalah metode atau algoritma yang digunakan untuk mengekstraksi informasi atau pola yang berguna dari data yang besar dan kompleks. Beberapa teknik data mining yang umum digunakan adalah sebagai berikut (Ginantra dkk., 2021):

1. *Clustering*: Teknik ini digunakan untuk mengelompokkan objek dalam kumpulan data berdasarkan kesamaan fitur atau karakteristik. Tujuan utama dari *clustering* adalah untuk mengidentifikasi grup atau kelompok yang *homogen* dan *heterogen* dalam data.
2. *Association Rule Mining*: Teknik ini digunakan untuk menemukan hubungan atau korelasi antara dua atau lebih variabel dalam kumpulan data. Teknik ini berguna dalam membuat rekomendasi atau prediksi berdasarkan pola yang ditemukan dalam data.
3. *Decision Tree*: Teknik ini digunakan untuk membangun model prediksi berdasarkan pohon keputusan yang dibangun dari data. Tujuan utama dari *decision tree* adalah untuk memprediksi variabel target berdasarkan variabel *input*.
4. *Neural Network*: Teknik ini digunakan untuk membangun model prediksi berdasarkan proses yang menyerupai fungsi otak manusia. Teknik ini berguna dalam memprediksi variabel target yang kompleks dan sulit dipahami.
5. *Regression*: Teknik ini digunakan untuk membangun model prediksi berdasarkan hubungan *linier* atau *non-linier* antara variabel *input* dan *output*. Tujuan utama dari *regression* adalah untuk memprediksi nilai variabel target berdasarkan nilai variabel *input*.

6. *Text Mining*: Teknik ini digunakan untuk mengekstraksi informasi atau pola dari teks atau dokumen yang besar dan kompleks. Teknik ini berguna dalam menganalisis opini pelanggan, memprediksi tren pasar, atau menemukan topik yang sering dibicarakan dalam media sosial.

7. *Time Series Analysis*: Teknik ini digunakan untuk memprediksi nilai variabel target berdasarkan tren atau pola yang ditemukan dalam data waktu. Teknik ini berguna dalam memprediksi nilai penjualan atau produksi dalam periode waktu tertentu.

Teknik data mining yang digunakan tergantung pada jenis data dan tujuan analisis. Pemilihan teknik yang tepat dapat meningkatkan keakuratan dan efisiensi analisis data.

2.2 Klasifikasi

Klasifikasi adalah proses pengelompokan objek, item, atau data ke dalam beberapa kelompok atau kategori berdasarkan karakteristik yang dicari. Tujuannya adalah untuk mengelompokkan data menjadi bagian-bagian yang lebih mudah dipahami dan dianalisis. Klasifikasi digunakan dalam banyak bidang, seperti ilmu data, teknologi informasi, dan biologi, untuk menentukan kelompok yang sesuai dan memberikan wawasan baru tentang data tersebut (Ginantra dkk., 2021).

Klasifikasi adalah teknik pengolahan data yang banyak digunakan dalam ilmu data, khususnya pada data mining dan machine learning. Teknik ini dilakukan dengan cara membangun model atau algoritma yang dapat memprediksi kategori atau kelas suatu objek berdasarkan karakteristik atau atribut yang dimilikinya. Dalam klasifikasi, data dibagi menjadi dua kelompok utama: data pelatihan dan

data pengujian. Data pelatihan digunakan untuk membangun model atau algoritma, sedangkan data pengujian digunakan untuk menguji keakuratan model atau algoritma yang telah dibuat(Wibawa dkk., 2018).

Ada beberapa jenis algoritma klasifikasi yang umum digunakan, seperti *Decision Tree*, *Random Forest*, *Naive Bayes*, *Support Vector Machine (SVM)*, dan *k-Nearest Neighbor (k-NN)*. Masing-masing algoritma memiliki karakteristik dan kelebihan yang berbeda-beda tergantung pada jenis data yang dianalisis dan tujuan analisis yang ingin dicapai. Dalam memilih algoritma klasifikasi yang tepat, perlu diperhatikan faktor-faktor seperti keakuratan, kecepatan, kompleksitas, dan kemampuan untuk menangani data yang besar dan kompleks(Wibawa dkk., 2018).

Dalam aplikasinya, klasifikasi digunakan dalam banyak bidang, seperti dalam analisis risiko kredit, pengenalan suara, deteksi spam, identifikasi penyakit, dan pengenalan citra. Klasifikasi juga sangat berguna dalam pengambilan keputusan dan prediksi, terutama dalam dunia bisnis dan industri. Dengan memanfaatkan algoritma klasifikasi, organisasi dapat mengoptimalkan pengambilan keputusan, meningkatkan efisiensi, dan mengurangi biaya operasional.

2.3 Penyakit Infeksi Kulit Pada Kucing

Infeksi kulit pada kucing merupakan salah satu masalah kesehatan yang sering terjadi. Penyebab infeksi kulit pada kucing antara lain adalah infeksi jamur, bakteri, virus, dan parasit. Beberapa gejala dari infeksi kulit pada kucing antara lain adalah gatal, kehilangan bulu, dan perubahan warna kulit (Muthiah & Razak, 2023).

Penyakit Infeksi Kulit pada Kucing adalah kondisi medis yang umum terjadi pada kucing. Infeksi kulit terjadi ketika bakteri, jamur, atau parasit menyebar dan berkembang biak pada kulit. Ini dapat menyebabkan berbagai gejala, termasuk ruam kulit, gatal-gatal, kelembapan, kerontokan bulu, dan kemerahan. Ada banyak jenis penyakit kulit yang bisa menyerang kucing, seperti dermatitis alergi, infeksi jamur, kutu, dan infeksi bakteri (Dwiramadhan dkk., 2022).

Dermatitis alergi adalah salah satu jenis penyakit kulit yang paling umum pada kucing. Ini terjadi ketika kucing terkena alergi terhadap bahan-bahan tertentu, seperti serbuk sari, makanan, atau kutu. Infeksi jamur juga dapat mempengaruhi kulit kucing dan menyebabkan rambut rontok, kulit kering, dan bercak-bercak yang berair. Kutu juga dapat menyebabkan gatal-gatal dan iritasi kulit, sementara infeksi bakteri dapat menyebabkan luka, bercak-bercak merah, dan nanah (Adriana, 2021).

Infeksi jamur pada kucing, juga dikenal sebagai dermatofitosis, adalah infeksi kulit dan rambut kucing yang disebabkan oleh jamur. Infeksi ini sangat menular dan sering terjadi pada kucing yang hidup dalam kondisi tidak higienis. Gejala umum dari infeksi jamur pada kucing adalah gatal, rambut rontok, dan kulit berkerak. Infeksi ini dapat diobati dengan antijamur topikal atau oral dan perlu diobati dengan hati-hati untuk mencegah penularan ke kucing lain atau ke manusia (Paryuni, 2022).

Kudis pada kucing, atau scabies, adalah infeksi kulit yang disebabkan oleh tungau *Sarcoptes scabiei*. Tungau ini berkembang biak di dalam kulit dan menyebabkan rasa gatal yang hebat pada kucing. Kudis dapat menyebar melalui kontak langsung dengan hewan yang terinfeksi atau benda-benda yang terkontaminasi dengan tungau. Jika tidak diobati, kudis pada kucing dapat

menyebabkan kerusakan kulit, infeksi sekunder, dan hilangnya bulu pada area yang terinfeksi (Susanto dkk., 2020).

Tungau telinga pada kucing adalah kondisi infeksi telinga yang disebabkan oleh tungau kecil yang dikenal sebagai *Notoedres cati*. Tungau ini dapat menyebar dengan cepat dan menyebabkan gatal yang hebat pada telinga kucing, serta dapat menyebar ke kepala dan leher kucing. Kucing yang terinfeksi tungau telinga akan sering menggaruk atau menggosokkan telinganya pada permukaan kasar dan berusaha membersihkan telinga dengan menggelengkan kepala. Pengobatan yang tepat harus dilakukan untuk menghilangkan tungau dan mencegah infeksi berulang serta kerusakan pada telinga kucing (Yudhana dkk., 2021).

Penanganan Infeksi Kulit pada Kucing tergantung pada penyebab dasarnya. Untuk beberapa kasus, seperti alergi atau infeksi bakteri, mungkin diperlukan pengobatan dengan antibiotik atau antijamur. Terapi topikal seperti sampo, salep, atau semprotan juga dapat membantu mengurangi gatal-gatal dan iritasi. Selain pengobatan medis, menjaga kebersihan kucing dan lingkungannya dapat membantu mengurangi risiko terjadinya infeksi kulit. Rutin memeriksa kucing dan memberinya makanan yang sehat serta berolahraga juga dapat membantu meningkatkan kesehatan kulit kucing secara keseluruhan.

Dalam mengidentifikasi penyakit infeksi kulit pada kucing, terdapat beberapa gejala yang dapat dijadikan sebagai acuan jenis penyakit mana yang diderita oleh kucing. Gejala-gejala tersebut adalah sebagai berikut (Romadiana dkk., 2021).

- Gatal-gatal: Kucing yang terkena penyakit infeksi kulit sering kali merasakan gatal-gatal pada kulitnya. Kucing akan terus-menerus menjilat atau menggaruk area yang terinfeksi untuk meredakan rasa gatal tersebut.
- Kulit kering: Kulit kucing yang terinfeksi dapat menjadi kering dan terkelupas. Hal ini disebabkan oleh hilangnya kelembaban pada kulit dan kerusakan pada lapisan pelindung kulit.
- Kulit kemerahan: Kulit kucing yang terinfeksi dapat menjadi kemerahan atau meradang. Hal ini disebabkan oleh peradangan pada kulit yang disebabkan oleh infeksi.
- Bulu rontok: Kucing yang terkena infeksi kulit sering mengalami kebotakan pada area yang terinfeksi. Hal ini dapat disebabkan oleh kerusakan pada folikel rambut atau penggarukan yang terus menerus.
- Luka: Infeksi kulit dapat menyebabkan luka pada kulit kucing. Luka ini dapat terjadi karena gigitan serangga atau kerusakan pada kulit akibat menggaruk yang terus menerus.
- Kulit bersisik: Infeksi kulit pada kucing juga dapat menyebabkan kulit bersisik. Hal ini disebabkan oleh kelebihan pertumbuhan sel kulit atau pengelupasan kulit yang abnormal.
- Kulit berkerak: Kulit kucing yang terinfeksi juga dapat berkerak atau terdapat endapan kering pada kulit. Hal ini disebabkan oleh kelebihan pertumbuhan sel kulit dan kotoran yang menumpuk pada kulit.
- Penumpukan kotoran pada telinga: Infeksi kulit juga dapat menyebabkan penumpukan kotoran pada telinga kucing. Hal ini disebabkan oleh infeksi jamur atau bakteri pada telinga yang menyebabkan produksi kotoran yang berlebihan.

Gejala-gejala ini digunakan untuk mengklasifikasikan penyakit kulit pada kucing menggunakan algoritma klasifikasi *Random Forest*. Setiap gejala memberikan kontribusi pada pembentukan model klasifikasi dan pengambilan keputusan dalam mengklasifikasikan penyakit kulit pada kucing.

Berdasarkan gejala-gejala tersebut, terdapat gejala yang paling dominan untuk dijadikan rujukan dalam menentukan penyakit infeksi kulit pada kucing. Tabel di bawah ini merangkum jenis-jenis penyakit dan gejala-gejala yang dapat dijadikan sebagai atribut dalam mengklasifikasi penyakit infeksi kulit pada kucing yang diambil dari penelitian penelitian sebelumnya (Adriana, 2021; Paryuni, 2022; Susanto dkk., 2020; Yudhana dkk., 2021).

Tabel 2. 1 Penyakit

Kode penyakit	Nama penyakit
P01	Dermatitis alergi
P02	Infeksi jamur
P03	Kudis
P04	Tungau telinga

Tabel 2. 2 Gejala

Kode Gejala	Nama Gejala
G01	Gatal-gatal
G02	Kulit kering
G03	Kulit kemerahan
G04	Bulu Rontok

G05	Luka
G06	Kulit bersisik
G07	Kulit berkerak
G08	Penumpukan kotoran pada telinga

2.4 Algoritma *Random Forest*

Random forest adalah salah satu algoritma yang digunakan dalam *machine learning* untuk melakukan klasifikasi atau regresi pada dataset (Muttaqien dkk., 2020). Algoritma ini menggunakan teknik *ensemble learning*, yaitu menggabungkan beberapa pohon keputusan (*decision tree*) untuk membentuk sebuah model yang lebih akurat. Setiap pohon keputusan pada *random forest* dilatih dengan subset data yang dipilih secara acak dari dataset yang tersedia, sehingga dapat mengurangi *overfitting* dan membuat model yang lebih stabil (Nalatissifa dkk., 2021).

Pada setiap pohon keputusan dalam *random forest*, pemilihan fitur dilakukan secara acak dengan teknik *bagging (bootstrap aggregating)*. Teknik ini mengambil sejumlah sampel dengan pengembalian dari dataset asli untuk dilatih pada setiap pohon keputusan. Setiap sampel memiliki subset fitur yang dipilih secara acak, sehingga masing-masing pohon keputusan pada *random forest* memiliki keunikan dalam fitur-fitur yang digunakan.

Keuntungan dari menggunakan *random forest* adalah model yang dihasilkan dapat memiliki tingkat akurasi yang tinggi dan dapat mengatasi *overfitting*. Selain itu, *random forest* juga dapat digunakan untuk mengevaluasi pentingnya fitur (*feature importance*) pada dataset. Hal ini dapat membantu dalam

pemilihan fitur yang relevan dan meningkatkan kinerja model. Namun, kelemahan dari random forest adalah kompleksitas yang tinggi dan waktu pelatihan yang relatif lambat pada dataset yang sangat besar.

2.4.1 Langkah-langkah Perhitungan *Random Forest*

Berikut ini adalah langkah-langkah atau proses perhitungan *random forest* secara umum:

1. Memilih jumlah pohon keputusan ($n_estimators$) yang akan digunakan dalam *random forest*.
2. Memilih jumlah fitur yang akan diambil secara acak dari dataset untuk melatih setiap pohon keputusan.
3. Untuk setiap pohon keputusan, memilih sampel secara acak dari dataset dengan penggantian dan latih pohon keputusan pada sampel tersebut dengan fitur yang telah dipilih sebelumnya.
4. Setiap pohon keputusan memberikan prediksi pada data uji yang belum pernah dilihat sebelumnya. Untuk klasifikasi, prediksi dilakukan dengan cara voting (suara terbanyak), sedangkan untuk regresi, prediksi dilakukan dengan cara mengambil nilai rata-rata dari semua prediksi pohon.

Dalam membangun setiap pohon keputusan, terdapat dua tahap utama yang dilakukan yaitu pembentukan pohon dan pruning. Pada tahap pembentukan pohon, data dibagi menjadi dua atau lebih *node*, dengan tujuan untuk meningkatkan *homogenitas* data pada setiap *node*. Pembagian data dilakukan berdasarkan nilai dari salah satu atribut. Pemilihan atribut dilakukan dengan menghitung nilai

impurity atau ketidakmurnian data pada setiap atribut dan memilih atribut yang memiliki nilai *impurity* paling rendah.

Setelah pohon keputusan terbentuk, pruning dilakukan untuk mencegah *overfitting*. *Overfitting* terjadi ketika pohon keputusan terlalu kompleks dan hanya cocok dengan data latih, sehingga tidak mampu melakukan prediksi dengan akurat pada data uji. Pruning dilakukan dengan memangkas pohon keputusan sehingga menjadi lebih sederhana dan generalisasi lebih baik. Pemangkasan pohon dapat dilakukan dengan menghapus salah satu *node* atau cabang yang tidak memberikan kontribusi signifikan terhadap performa pohon.

2.4.2 Rumus

$$Entropy = - \sum p(x) \log_2 p(x) \quad (2.1)$$

di mana:

- $p(x)$ adalah probabilitas dari nilai x dalam suatu himpunan data.
- \log_2 adalah logaritma berbasis 2.

Rumus tersebut digunakan dalam pembelajaran mesin untuk mengukur ketidakpastian dalam himpunan data. Semakin tidak pasti atau tidak teratur himpunan data, semakin tinggi nilai entropy. Sebaliknya, semakin pasti atau teratur himpunan data, semakin rendah nilai entropy.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2.2)$$

Keterangan:

- $Gain(S, A)$: *Gain Information* untuk atribut A pada set data S
- $Entropy(S)$: Nilai *Entropy* untuk set data S

- $|S_v|$: Jumlah sampel pada subset S yang memenuhi kriteria atribut $A = v$
- $|S|$: Jumlah total sampel pada set data S
- $Entropy(S_v)$: Nilai Entropy untuk subset S yang memenuhi kriteria atribut

$A = v$

Rumus tersebut dapat digunakan dalam pengambilan keputusan untuk memilih atribut terbaik dalam membuat pohon keputusan pada algoritma *Random Forest*. Semakin besar *Gain Information* suatu atribut, semakin baik pula atribut tersebut dalam membagi set data menjadi subset yang *homogen*.

2.5 Confusion Matrix

Confusion matrix dapat diartikan sebagai tabel yang berisi nilai-nilai klasifikasi hasil prediksi dari suatu model atau algoritma pada data yang telah diberikan label. Nilai-nilai utama dalam *confusion matrix* yaitu *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN). TP menunjukkan jumlah data positif yang terklasifikasi dengan benar, FP menunjukkan jumlah data negatif yang terklasifikasi secara salah, TN menunjukkan jumlah data negatif yang terklasifikasi dengan benar, dan FN menunjukkan jumlah data positif yang terklasifikasi secara salah (Normawati & Prayogi, 2021).

Confusion matrix sangat penting dalam mengevaluasi performa suatu model klasifikasi karena memberikan gambaran detail mengenai kualitas hasil prediksi. Berbagai metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score* dapat dihitung berdasarkan nilai-nilai TP, FP, TN, dan FN pada *confusion matrix*. *Confusion matrix* juga bermanfaat untuk mengevaluasi model klasifikasi dalam kasus ketidakseimbangan kelas pada data.

Pada umumnya, *confusion matrix* digunakan untuk mengevaluasi model klasifikasi *binary*, namun juga dapat diaplikasikan pada model klasifikasi *multi-class* dengan modifikasi. Dalam kasus *multi-class*, *confusion matrix* berisi matriks $n \times n$, di mana n merupakan jumlah kelas yang ada. TP, FP, TN, dan FN dihitung untuk setiap kelas dan metrik evaluasi dihitung berdasarkan kelas tersebut. Oleh karena itu, *confusion matrix* sangat membantu dalam mengevaluasi performa model klasifikasi dengan lebih detail dan akurat pada data *multi-class*.

Tabel berikut merupakan contoh confusion matriks untuk klasifikasi *biner*.

Tabel 2. 3 *Confusion Matrix*

		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan:

TP (*True Positive*) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1

TN (*True Negative*) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0

FP (*False Positive*) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1

FN (*False Negative*) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

Rumus *confusion matrix* untuk menghitung *accuracy*, *precision*, dan *recall* seperti berikut.

$$accuracy = \frac{TP + TN}{Total} \quad (2.3)$$

$$precision = \frac{TP}{TP + FP} \quad (2.4)$$

$$recall = \frac{TP}{TP + FN} \quad (2.5)$$

2.6 Google Colab

Google Colab adalah *platform* pengembangan proyek berbasis *cloud computing* yang dikelola oleh *Google*. *Platform* ini memungkinkan pengguna untuk menulis dan mengeksekusi kode *Python* menggunakan lingkungan *Jupyter Notebook* yang familiar. *Google Colab* menyediakan berbagai fitur seperti kemampuan untuk menyimpan, mengelola, dan berbagi proyek secara online, dengan dukungan dari *Google Drive* dan *GitHub* (Bisong & Bisong, 2019).

Google Colab juga menyediakan lingkungan pengembangan yang gratis dan dapat digunakan secara online, tanpa memerlukan instalasi atau konfigurasi tambahan pada komputer pengguna. Hal ini sangat memudahkan pengguna untuk mengakses dan menggunakan lingkungan pengembangan yang sama, di mana saja dan kapan saja, asalkan terhubung dengan internet.

Selain itu, *Google Colab* juga dilengkapi dengan akses ke berbagai pustaka populer dalam ekosistem *Python*, seperti *TensorFlow*, *PyTorch*, dan *Scikit-Learn*, yang memungkinkan pengguna untuk dengan mudah membangun dan menguji model *machine learning* atau *deep learning*. Dalam hal ini, *Google Colab*

memberikan akses ke sumber daya komputasi yang cukup besar, seperti unit pemrosesan grafis (GPU) dan unit pemrosesan tensor (TPU), yang dapat digunakan untuk mengakselerasi proses pelatihan model.

2.7 Synthetic Minority Over-sampling Technique (SMOTE)

SMOTE, singkatan dari *Synthetic Minority Over-sampling Technique*, adalah sebuah teknik yang digunakan untuk mengatasi ketidakseimbangan jumlah sampel pada kelas minoritas dalam data. Teknik ini bekerja dengan memilih sampel data dari kelas minoritas dan membuat sampel sintetis baru sehingga jumlahnya seimbang dengan jumlah sampel pada kelas mayoritas. Penggunaan metode *SMOTE* dapat menyebabkan *overfitting*, di mana model terlalu memfokuskan diri pada data yang telah digandakan dari kelas minoritas sehingga mengurangi kemampuan generalisasi (Kasanah dkk., 2019).

Proses *SMOTE* dimulai dengan menghitung jarak antara data pada kelas minoritas. Selanjutnya, nilai persentase *SMOTE* ditentukan untuk menentukan seberapa banyak sampel sintetis yang akan dibuat, kemudian menentukan jumlah k terdekat. Terakhir, data sintetis dibuat dengan menggunakan persamaan sebagai berikut:

$$x_{syn} = x_i + (x_{knn} - x_i) \times \delta \quad (2.6)$$

Dengan x_{syn} adalah data sintesis yang akan diciptakan x_i data yang akan direplikasi, x_{knn} data yang memiliki jarak terdekat dari data yang akan direplikasi dan δ nilai random antara 0 dan 1.

2.8 Penelitian Terdahulu

Tabel 2. 4 Penelitian Terdahulu

No	Penulis	Topik	Hasil
1	Muttaqien, Abdul Hafiz; Utami, Alvi Syahrini; Miraswan, Kanda Januar (2020)	Klasifikasi Penyakit Kucing Menggunakan Metode Random Forest	Dalam penelitian ini, kinerja Random Forest digunakan untuk mengklasifikasi penyakit kucing pada umumnya. Penelitian ini menghasilkan akurasi sebesar 96%.
2	Khasanah, Nurul; Rachman Komarudin, Nurul Afni; Maulana, Yana Iqbal; Salim, Agus (2021)	Klasifikasi Kanker Kulit Menggunakan Algoritma Random Forest Skin Cancer Classification Using Random Forest Algorithm	Dalam penelitian ini, klasifikasi penyakit kanker kulit dengan menggunakan algoritma Random Forest. Adapaun penelitian yang dilakukannya menggunakan data citra digital dan menghasilkan klasifikasi dengan akurasi sebesar 84%.
3	Nalatissifa, Hiya; Gata,	Perbandingan Kinerja Algoritma	Pada hasil penelitian, algoritma Random Forest

	Windu; Diantika, Sri; Nisa, Khoirun (2021)	Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja	memperoleh nilai akurasi, presisi, dan recall yang paling tinggi dibandingkan dengan algoritma Naive Bayes dan SVM, yaitu menghasilkan nilai akurasi sebesar 99.38%, presisi 99.42% dan recall 99.39%.
4	Gunawan, Indra; Fernando, Yusra (2021)	Sistem Pakar Diagnosa Penyakit Kulit Pada Kucing Menggunakan Metode Naive Bayes Berbasis Web	Metode yang digunakan dalam penelitian ini adalah Naive Bayes untuk mencari nilai peluang terbesar munculnya penyakit kulit pada kucing. Dari pengujian yang dilakukan pada 15 data rekam medis dokter, didapatkan tingkat akurasi sistem pakar diagnosa penyakit kulit pada kucing sebesar 80%.
5	Meiriyama, Meiriyama;	Penerapan Algoritma Random Forest	Pada penelitian ini, Algoritma Random Forest yang digunakan untuk

	Sudiadi, Sudiadi (2022)	Untuk Klasifikasi Jenis Daun Herbal	klasifikasi daun herbal memiliki akurasi keseluruhan sebesar 85,33%.
--	----------------------------	--	---

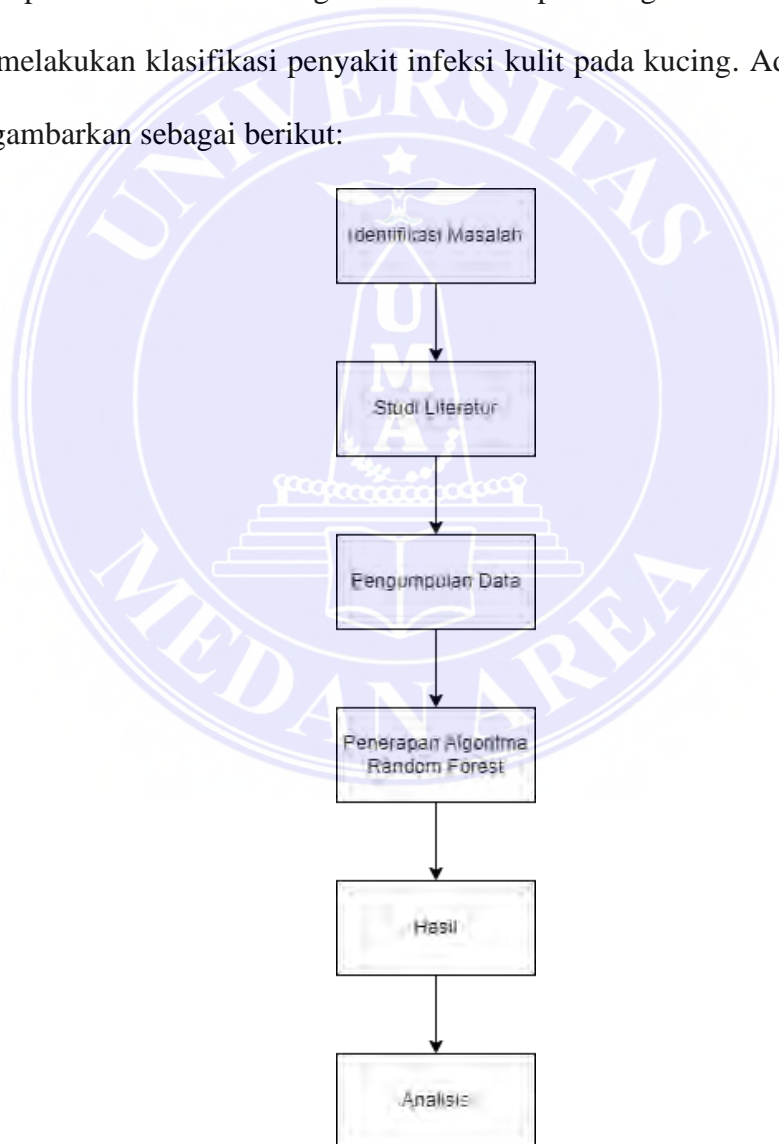


BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Penelitian ini akan melalui beberapa tahapan dalam mencapai tujuan penelitian. Tahapan pertama dimulai dengan identifikasi masalah, dan tahapan terakhir penelitian untuk menganalisis kemampuan algoritma *Random Forest* dalam melakukan klasifikasi penyakit infeksi kulit pada kucing. Adapun diagram alur digambarkan sebagai berikut:



Gambar 3. 1 Tahapan Penelitian

1. Identifikasi Masalah: Penelitian bertujuan untuk menganalisis kemampuan algoritma *Random Forest* dalam melakukan klasifikasi penyakit infeksi kulit pada kucing.
2. Studi Literatur: tahap ini dilakukan dengan membaca dan mempelajari referensi yang relevan dengan topik penelitian. Peneliti melakukan studi literatur untuk memperoleh pemahaman yang cukup mengenai infeksi kulit pada kucing dan algoritma *Random Forest*.
3. Pengumpulan Data: tahap ini dilakukan dengan mengumpulkan data tentang infeksi kulit pada kucing melalui wawancara dan kuesioner yang dilakukan kepada responden yang dipilih berdasarkan kriteria tertentu di lokasi penelitian.
4. Penerapan Algoritma *Random Forest*: Menggunakan algoritma *Random Forest* untuk mengklasifikasikan jenis penyakit infeksi kulit pada kucing. Melakukan proses pelatihan model dengan membagi data menjadi data latih dan data uji. Selanjutnya dilakukan proses tuning parameter untuk meningkatkan akurasi model.
5. Hasil: Menampilkan hasil dari klasifikasi penyakit infeksi kulit pada kucing dengan menggunakan algoritma *Random Forest*. Hasil tersebut akan berupa daftar jenis penyakit infeksi kulit pada kucing beserta akurasi dari model klasifikasi.
6. Analisis: Melakukan analisis terhadap hasil klasifikasi untuk mengetahui tingkat keakuratan dan ketepatan model. Analisis ini dapat memberikan wawasan tentang jenis penyakit infeksi kulit pada kucing yang paling sering terjadi dan faktor apa yang menyebabkannya.

Dengan tahapan penelitian yang terstruktur dan sistematis ini, penulis dapat menghasilkan penelitian yang valid, efektif, dan akurat untuk mencapai tujuan dari skripsi ini.

3.2 Bahan dan Alat Penelitian

Berikut adalah bahan dan alat yang mungkin dibutuhkan untuk melakukan penelitian skripsi ini:

1. Data Penyakit Infeksi Kulit pada Kucing:

Data penyakit infeksi kulit pada kucing merupakan bahan utama yang dibutuhkan untuk melakukan penelitian ini. Data tersebut bisa berupa data medis, gejala, dan faktor risiko yang terkait dengan infeksi kulit pada kucing.

2. *Software* Pemrograman:

Software pemrograman seperti *Python* atau *R* dibutuhkan untuk membuat dan menjalankan algoritma *Random Forest*. Ini juga memungkinkan untuk mengelola dan memproses data yang dibutuhkan untuk penelitian.

3. Perangkat Keras:

Perangkat keras seperti komputer atau laptop dibutuhkan untuk menjalankan *software* pemrograman dan melakukan analisis data.

4. Library dan Package:

Beberapa *library* dan *package* seperti *scikit-learn* atau *caret* diperlukan untuk membuat dan menjalankan algoritma *Random Forest*. Ini juga memungkinkan untuk melakukan validasi dan evaluasi model.

Dengan bahan dan alat di atas, penulis dapat melakukan penelitian yang efektif dan hasil yang akurat dan dapat membantu mencapai tujuan dari skripsi ini.

3.3 Proses Pengumpulan Data

Pengumpulan data merupakan tahap penting dalam sebuah penelitian. Data yang digunakan berasal dari penelitian yang langsung diambil dari sumber aslinya. Pada penelitian ini, hanya menggunakan data primer. Data *primer* diperoleh melalui wawancara dan kuesioner yang dilakukan kepada responden yang dipilih berdasarkan kriteria tertentu dalam pengumpulan data, penting untuk memastikan bahwa data yang digunakan memenuhi kriteria validitas, reliabilitas, dan relevansi untuk memastikan keakuratan hasil penelitian. Dengan menggunakan data *primer*, diharapkan penelitian ini dapat memberikan gambaran yang lebih komprehensif dan akurat terhadap fenomena yang diteliti.

3.4 Perhitungan Manual

Pada tahap ini, dilakukan perhitungan manual *Random Forest* menggunakan microsoft *excel* dengan mengacu pada hasil pengumpulan data sebelumnya. Data yang telah terkumpul kemudian diolah dengan menggunakan metode klasifikasi *Random Forest* untuk mengidentifikasi jenis penyakit infeksi kulit yang dialami oleh kucing.

Tabel 3. 1 Data Jenis Penyakit

Kode penyakit	Nama penyakit
P01	Dermatitis alergi
P02	Infeksi jamur
P03	Kudis
P04	Tungau telinga

Tabel 3. 2 Data Gejala Penyakit

Kode Gejala	Nama Gejala
G01	Gatal-gatal
G02	Kulit kering
G03	Kulit kemerahan
G04	Bulu Rontok
G05	Luka
G06	Kulit bersisik
G07	Kulit berkerak
G08	Penumpukan kotoran pada telinga

Tabel 3. 3 Data Sampel

No	Gejala_1	Gejala_2	Gejala_3	Gejala_4	Penyakit
1	Gatal-Gatal	Kulit Kemerahan	Kulit Bersisik	Kulit Berkerak	Dermatitis Alergi
2	Gatal-Gatal	Kulit Kemerahan	Penumpukan kotoran pada telinga	Bulu Rontok	Tungau Telinga
3	Gatal-Gatal	Luka	Penumpukan kotoran pada telinga	Bulu Rontok	Infeksi Jamur
4	Kulit Kering	Luka	Penumpukan kotoran pada telinga	Kulit Berkerak	Kudis
5	Gatal-Gatal	Kulit Kemerahan	Kulit Bersisik	Bulu Rontok	Dermatitis Alergi
6	Kulit Kering	Luka	Penumpukan kotoran pada telinga	Bulu Rontok	Kudis
7	Gatal-Gatal	Kulit Kemerahan	Kulit Bersisik	Kulit Berkerak	Infeksi Jamur
8	Gatal-Gatal	Kulit Kemerahan	Kulit Bersisik	Bulu Rontok	Tungau Telinga
9	Gatal-Gatal	Kulit Kemerahan	Kulit Bersisik	Bulu Rontok	Kudis
10	Kulit Kering	Luka	Penumpukan kotoran pada telinga	Kulit Berkerak	Dermatitis Alergi
11	Gatal-Gatal	Kulit Kemerahan	Kulit Bersisik	Bulu Rontok	Tungau Telinga
12	Kulit Kering	Luka	Kulit Bersisik	Kulit Berkerak	Infeksi Jamur
13	Gatal-Gatal	Luka	Penumpukan kotoran pada telinga	Bulu Rontok	Dermatitis Alergi
14	Kulit Kering	Kulit Kemerahan	Kulit Bersisik	Bulu Rontok	Infeksi Jamur

Tabel 3.3 adalah data sampel yang akan dilakukan perhitungan menggunakan algoritma *Random Forest*, adapun atribut yang digunakan adalah gejala 1, gejala 2, gejala 3, gejala 4, dan penyakit. Tools yang digunakan untuk melakukan perhitungan adalah *microsoft excel*.

	G	H	I	J	K	L	M	N	O	P	Q
		Node		Keterangan	Jumlah (S)	Penyakit (Dermatitis Alergi)	Penyakit (Tungau Telinga)	Penyakit (Infeksi Jamur)	Penyakit (Kudis)	Entropy	Gain
2											
3											
4											
5		i	Total		14	4	3	4	3	0.992614068	
6											
7			Gejala_1								0.384773
8				Gatal-Gatal	9	3	3	2	1	0.945530556	
9											
10				Kulit Kering	5	1	0	2	2	0	
11											
12			Gejala_2								0.44814576
13				Kulit Kemerahan	8	2	3	2	1	0.952819531	
14											
15				Luka	6	2	0	2	2	0	
16											
17			Gejala_3								0.03708237
18				Kulit Bersisik	8	2	2	3	1	0.952819531	
19				Penumpukan kotoran pada telinga	6	2	1	1	2	0.959147917	
20											
21											
22			Gejala_4								0.35781273
23				Kulit Berkerak	5	2	0	2	1	0	
24											
25				Bulu Rontok	9	2	3	2	2	0.987468751	
26											

Gambar 3. 2 Nilai *Entropy* dan *Gain*

Gambar 3.2 merupakan perhitungan untuk mencari jumlah dari masing-masing atribut, nilai *entropy*, dan nilai *gain*. Adapun rumus *excel* digunakan dalam mencari *entropy* (contoh pada atribut gejala 1) ialah sebagai berikut:

$$=IFERROR((- (L8/K8)*LOG((L8/K8),4))+(- (M8/K8)*LOG((M8/K8),4))+(- (N8/K8)*LOG((N8/K8),4))+(- (O8/K8)*LOG((O8/K8),4)),0)$$

Kemudian untuk mencari nilai *gain* (contoh pada atribut gejala 1) yaitu :

$$=P5-(((K10/$K5)*P10)+((K8/$K5)*P8))$$

Perhitungan *entropy* total dilakukan dengan menggunakan Persamaan (2.1). Rumus tersebut digunakan dalam pembelajaran mesin untuk mengukur ketidakpastian dalam himpunan data. Semakin tidak pasti atau tidak teratur himpunan data, semakin tinggi nilai *entropy*. Sebaliknya, semakin pasti atau teratur himpunan data, semakin rendah nilai *entropy*. *Entropy* untuk setiap Gejala berdasarkan Gambar 3.2 yaitu:

Entropy(Gatal – Gatal)

$$= \left(-\frac{3}{9} * \log_2 \left(\frac{3}{9} \right) \right) + \left(-\frac{3}{9} * \log_2 \left(\frac{3}{9} \right) \right) + \left(-\frac{2}{9} * \log_2 \left(\frac{2}{9} \right) \right) + \left(-\frac{1}{9} * \log_2 \left(\frac{1}{9} \right) \right) = 0.945$$

Entropy(Kulit Kering)

$$= \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5} \right) \right) + \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5} \right) \right) + \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5} \right) \right) + \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5} \right) \right) = 0$$

Entropy(Kulit Kemerahan)

$$= \left(-\frac{2}{8} * \log_2 \left(\frac{2}{8} \right) \right) + \left(-\frac{3}{8} * \log_2 \left(\frac{3}{8} \right) \right) + \left(-\frac{2}{8} * \log_2 \left(\frac{2}{8} \right) \right) + \left(-\frac{1}{8} * \log_2 \left(\frac{1}{8} \right) \right) = 0.952$$

Entropy(Luka)

$$= \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6} \right) \right) + \left(-\frac{0}{6} * \log_2 \left(\frac{0}{6} \right) \right) + \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6} \right) \right) + \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6} \right) \right) = 0$$

Entropy(Kulit Bersisik)

$$= \left(-\frac{2}{8} * \log_2 \left(\frac{2}{8}\right)\right) + \left(-\frac{2}{8} * \log_2 \left(\frac{2}{8}\right)\right) + \left(-\frac{3}{8} * \log_2 \left(\frac{3}{8}\right)\right) + \left(-\frac{1}{8} * \log_2 \left(\frac{1}{8}\right)\right) = 0.952$$

Entropy(Penumpukan Kotoran)

$$= \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6}\right)\right) + \left(-\frac{1}{6} * \log_2 \left(\frac{1}{6}\right)\right) + \left(-\frac{1}{6} * \log_2 \left(\frac{1}{6}\right)\right) + \left(-\frac{2}{6} * \log_2 \left(\frac{2}{6}\right)\right) = 0.959$$

Entropy(Kulit Berkerak)

$$= \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{0}{5} * \log_2 \left(\frac{0}{5}\right)\right) + \left(-\frac{2}{5} * \log_2 \left(\frac{2}{5}\right)\right) + \left(-\frac{1}{5} * \log_2 \left(\frac{1}{5}\right)\right) = 0$$

Entropy(Bulu Rontok)

$$= \left(-\frac{2}{9} * \log_2 \left(\frac{2}{9}\right)\right) + \left(-\frac{3}{9} * \log_2 \left(\frac{3}{9}\right)\right) + \left(-\frac{2}{9} * \log_2 \left(\frac{2}{9}\right)\right) + \left(-\frac{2}{9} * \log_2 \left(\frac{2}{9}\right)\right) = 0.987$$

Entropy(Total)

$$= \left(-\frac{4}{14} * \log_2 \left(\frac{4}{14}\right)\right) + \left(-\frac{3}{14} * \log_2 \left(\frac{3}{14}\right)\right) + \left(\frac{4}{14} * \log_2 \left(\frac{4}{14}\right)\right) + \left(\frac{3}{14} * \log_2 \left(\frac{3}{14}\right)\right) = 0.992$$

Perhitungan gain dilakukan dengan menggunakan Persamaan (2.2). Rumus tersebut digunakan dalam pembelajaran mesin untuk mengukur ketidakpastian dalam himpunan data. Semakin tidak pasti atau tidak teratur himpunan data, semakin tinggi nilai gain. Sebaliknya, semakin pasti atau teratur himpunan data, semakin rendah nilai *gain*. *Gain* untuk setiap atribut berdasarkan Gambar 3.2 yaitu:

$$Gain(Gejala_1) = 0.992 - \left(\left(\frac{5}{14} \right) * 0 \right) + \left(\left(\frac{9}{14} \right) * 0.945 \right) = 0.384$$

$$Gain(Gejala_2) = 0.992 - \left(\left(\frac{6}{14} \right) * 0 \right) + \left(\left(\frac{8}{14} \right) * 0.952 \right) = 0.448$$

$$Gain(Gejala_3) = 0.992 - \left(\left(\frac{6}{14} \right) * 0.959 \right) + \left(\left(\frac{8}{14} \right) * 0.952 \right) = 0.037$$

$$Gain(Gejala_4) = 0.992 - \left(\left(\frac{9}{14} \right) * 0.987 \right) + \left(\left(\frac{5}{14} \right) * 0 \right) = 0.357$$

Setelah mendapatkan semua nilai *entropy* dan *gain*, maka di *filter* nilai gain yang paling tinggi diantara semua atributnya seperti Gambar 3.2. proses selanjutnya adalah menghitung kembali data sampelnya yang bernilai tinggi (jika nilai *entropy* 0 maka proses akan berhenti).

	G	H	I	J	K	L	M	N	O	P	Q
30											
31											
32		Node	Keterangan	Jumlah (S)	Penyakit (Dermatitis Alergi)	Penyakit (Tungau Telinga)	Penyakit (Infeksi Jamur)	Kudis	Entropy	Gain	
33											
34		I.II	Total	8	2	3	2	1	0.952819531		
35											
36			Gejala_1								0.146782222
37			Gatal-Gatal	7	2	3	1	1	0.921185497		
38			Kulit Kering	1	0	0	1	0	0		
39											
40			Gejala_3								0.099601753
41			Kulit Bersisik	7	2	2	2	1	0.975106032		
42			Penumpukan kotoran pada telinga	1	0	1	0	0	0		
43											
44											
45			Gejala_4								0.280639062
46			Kulit Berkerak	2	1	0	1	0	0		
47			Bulu Rentok	6	1	3	1	1	0.896240625		
48											

Gambar 3. 3 Menghitung Kembali Nilai *Entropy* dan *Gain* (1.2)

Adapun rumus *excel* digunakan dalam mencari *entropy* (contoh pada atribut gejala 4) ialah sebagai berikut:

$$=IFERROR((-L47/K47)*LOG((L47/K47),4))+(-M47/K47)*LOG((M47/K47),4)+(-N47/K47)*LOG((N47/K47),4)+(-O47/K47)*LOG((O47/K47),4),0)$$

Kemudian untuk mencari nilai *gain* (contoh pada atribut gejala 4) yaitu :

$$=SP$34-(((L46/$K$34)*P46)+((L47/$K$34)*P47))$$

Perhitungan *entropy* untuk setiap Gejala berdasarkan Gambar 3.3 yaitu:

Entropy(Gatal – Gatal)

$$= \left(-\frac{2}{7} * \log_2 \left(\frac{2}{7} \right) \right) + \left(-\frac{3}{7} * \log_2 \left(\frac{3}{7} \right) \right) + \left(-\frac{1}{7} * \log_2 \left(\frac{1}{7} \right) \right) + \left(-\frac{1}{7} * \log_2 \left(\frac{1}{7} \right) \right) = 0.921$$

Entropy(Kulit Kering)

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

Entropy(Kulit Bersisik)

$$= \left(-\frac{2}{7} * \log_2 \left(\frac{2}{7} \right) \right) + \left(-\frac{2}{7} * \log_2 \left(\frac{2}{7} \right) \right) + \left(-\frac{2}{7} * \log_2 \left(\frac{2}{7} \right) \right) + \left(-\frac{1}{7} * \log_2 \left(\frac{1}{7} \right) \right) = 0.975$$

Entropy(Penumpukan Kotoran)

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

Entropy(Kulit Berkerak)

$$= \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) \\ + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) = 0$$

Entropy(Bulu Rontok)

$$= \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) + \left(-\frac{3}{6} * \log_2\left(\frac{3}{6}\right)\right) + \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) \\ + \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) = 0.896$$

Entropy(Total)

$$= \left(-\frac{2}{8} * \log_2\left(\frac{2}{8}\right)\right) + \left(-\frac{3}{8} * \log_2\left(\frac{3}{8}\right)\right) + \left(-\frac{2}{8} * \log_2\left(\frac{2}{8}\right)\right) \\ + \left(-\frac{1}{8} * \log_2\left(\frac{1}{8}\right)\right) = 0.952$$

Perhitungan *Gain* untuk setiap atribut berdasarkan Gambar 3.3 yaitu:

$$Gain(Gejala_1) = 0.952 - \left(\left(\frac{7}{8}\right) * 0.921\right) + \left(\left(\frac{1}{8}\right) * 0\right) = 0.146$$

$$Gain(Gejala_3) = 0.952 - \left(\left(\frac{7}{8}\right) * 0.975\right) + \left(\left(\frac{1}{8}\right) * 0\right) = 0.099$$

$$Gain(Gejala_4) = 0.952 - \left(\left(\frac{2}{8}\right) * 0\right) + \left(\left(\frac{6}{8}\right) * 0.896\right) = 0.280$$

Setelah mendapatkan semua nilai *entropy* dan *gain*, maka di *filter* nilai *gain* yang paling tinggi diantara semua atributnya seperti Gambar 3.3. proses selanjutnya adalah menghitung kembali data sampelnya yang bernilai tinggi (jika nilai *entropy* 0 maka proses akan berhenti).

Node	Keterangan	Jumlah (S)	Penyakit (Dermatitis Alergi)	Penyakit (Tungau Telinga)	Penyakit (Infeksi jamur)	Kudis	Entropy	Gain
lIII	Total	6	1	3	1	1	0.896240625	
Gejala 1								0.896240625
	Gatal-Gatal	5	1	3	0	1	0	
	Kulit Kering	1	0	0	1	0	0	
Gejala 3								0.095437252
	Kulit Bersisik	5	1	2	1	1	0.960964047	
	Penumpukan kotoran pada telinga	1	0	1	0	0	0	

Gambar 3. 4 Menghitung Kembali Nilai *Entropy* dan *Gain* (1.3)

Adapun rumus *excel* digunakan dalam mencari *entropy* (contoh pada atribut gejala 3) ialah sebagai berikut:

$$=IFERROR((- (L64/K64)*LOG((L64/K64),4))+(- (M64/K64)*LOG((M64/K64),4))+(- (N64/K64)*LOG((N64/K64),4))+(- (O64/K64)*LOG((O64/K64),4)),0)$$

Kemudian untuk mencari nilai *gain* (contoh pada atribut gejala 3) yaitu :

$$=P57-(((L64/$K57)*P64)+((L65/$K57)*P65))$$

Setelah mendapatkan semua nilai *entropy* dan *gain*, maka di filter nilai *gain* yang paling tinggi diantara semua atributnya seperti Gambar 3.4. (jika nilai *entropy* 0 maka proses akan berhenti).

Perhitungan *entropy* untuk setiap Gejala berdasarkan Gambar 3.4 yaitu:

$$Entropy(Gatal - Gatal) = \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{3}{5} * \log_2\left(\frac{3}{5}\right)\right) + \left(-\frac{0}{5} * \log_2\left(\frac{0}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) = 0$$

Entropy(Kulit Kering)

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

Entropy(Kulit Bersisik)

$$= \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{2}{5} * \log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) \\ + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) = 0.960$$

Entropy(Penumpukan Kotoran)

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

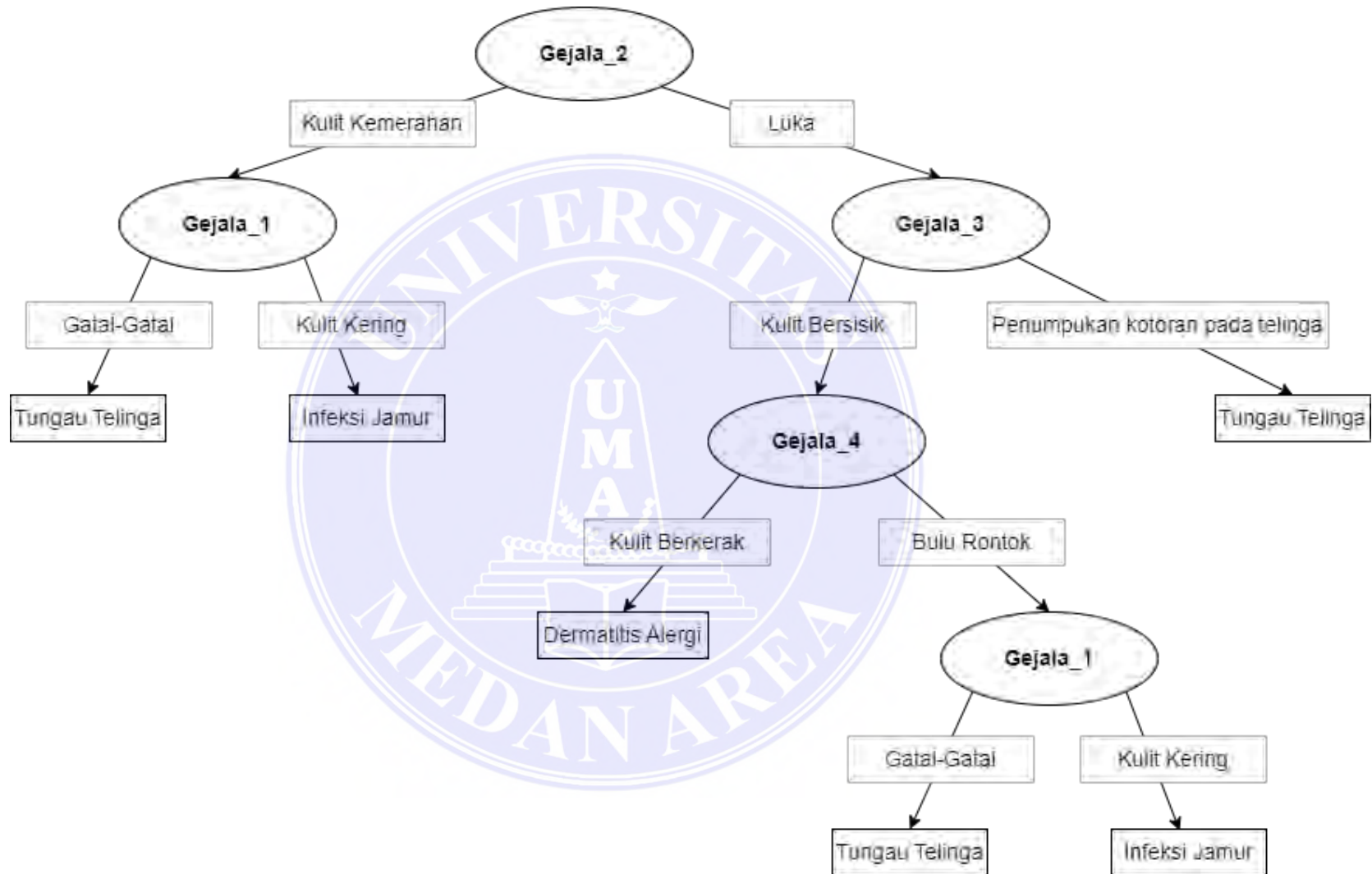
Entropy(Total)

$$= \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) + \left(-\frac{3}{6} * \log_2\left(\frac{3}{6}\right)\right) + \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) \\ + \left(-\frac{1}{6} * \log_2\left(\frac{1}{6}\right)\right) = 0.896$$

Perhitungan *Gain* untuk setiap atribut berdasarkan Gambar 3.4 yaitu:

$$Gain(Gejala_1) = 0.952 - \left(\left(\frac{5}{6}\right) * 0\right) + \left(\left(\frac{1}{6}\right) * 0\right) = 0.896$$

$$Gain(Gejala_3) = 0.952 - \left(\left(\frac{5}{6}\right) * 0.960\right) + \left(\left(\frac{1}{6}\right) * 0\right) = 0.095$$



Gambar 3. 5 Pohon Keputusan

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil yang sudah dikerjakan dapat disimpulkan bahwa :

1. Tingkat kemampuan algoritma *Random Forest* dalam melakukan klasifikasi penyakit infeksi kulit pada kucing mendapatkan nilai performa yang tinggi pada *training* yaitu akurasi 95,9%, presisi 96,2%, dan *recall* 95,9%. Kemudian pada *testing* nilai akurasinya 98,5%, presisi 98,5%, dan *recall* 98,5% untuk proporsi data 80:20.
2. Hasil klasifikasi yang telah dilakukan mendapatkan 4 yang salah saat diprediksi yaitu data sebenarnya penyakit kudis tetapi saat diprediksi mendapatkan hasil penyakit dermatitis alergi dengan berjumlah 3 dan data sebenarnya penyakit dermatitis alergi tetapi saat diprediksi mendapatkan penyakit infeksi jamur dengan sebanyak 1.

5.2 Saran

Adapun saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut.

1. Menambah jumlah data yang digunakan dapat berpengaruh terhadap pembelajaran model *machine learning* dan performanya akan semakin akurat.
2. Menerapkan algoritma klasifikasi lain seperti *Support Vector Machine*, *Multilayer Perceptron*, dan *Decision Tree*.

3. Penelitian ini menggunakan sebanyak 8 gejala penyakit sebagai atribut untuk pengklasifikasiannya. Penelitian kedepan diharapkan menggunakan gejala-gejala yang lain untuk dijadikan sebagai atribut lainnya.



DAFTAR PUSTAKA

- Adriana, L. (2021). *Sistem Pakar Diagnosis Penyakit Kulit Pada Kucing Berbais WEB Dengan Metode Forward Chaining*. Prodi Teknik Informatika.
- Bisong, E., & Bisong, E. (2019). Google colaboratory. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 59–64.
- Dami, J. C. (2022). *Diagnosis Ankilostomiasis Pada Kucing (felis catus) Berdasarkan Gejala Klinis Dan Indentifikasi Molekuler*. Universitas Gadjah Mada.
- Dwiramadhan, F., Wahyuddin, M. I., & Hidayatullah, D. (2022). Sistem Pakar Diagnosa Penyakit Kulit Kucing Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(3), 429–437.
- Ginantra, N. L. W. S. R., ARifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y., Effendy, F., Iskandar, A., Hazriani, H., Sari, I. Y., & others. (2021). *Data mining dan penerapan algoritma*. Yayasan Kita Menulis.
- Gunawan, I., & Fernando, Y. (2021). Sistem Pakar Diagnosa Penyakit Kulit Pada Kucing Menggunakan Metode Naive Bayes Berbasis Web. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, 2(2), 239–247.
- Kasanah, A. N., Muladi, M., Pujiyanto, U. (2019). Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 3(2), 196–201.
- Khasanah, N., Rachman Komarudin, N. A., Maulana, Y. I., & Salim, A. (2021).

Klasifikasi Kanker Kulit Menggunakan Algoritma Random Forest Skin Cancer Classification Using Random Forest Algorithm.

Loss, S. R., Boughton, B., Cady, S. M., Londe, D. W., McKinney, C., O'Connell, T. J., Riggs, G. J., & Robertson, E. P. (2022). Review and synthesis of the global literature on domestic cat impacts on wildlife. *Journal of Animal Ecology*, 91(7), 1361–1372.

Meiriyama, M., & Sudiadi, S. (2022). Penerapan Algoritma Random Forest Untuk Klasifikasi Jenis Daun Herbal. *Jurnal Teknologi Sistem Informasi*, 3(1), 131–138.

Muthiah, S. F., & Razak, R. (2023). *Pengetahuan Dan Praktik Personal Hygiene Terhadap Timbulnya Gejala Dermatofitosis Pada Pemilik Kucing*. Sriwijaya University.

Muttaqien, A. H., Utami, A. S., & Miraswan, K. J. (2020). *Klasifikasi Penyakit Kucing Menggunakan Metode Random Forest*. Sriwijaya University.

Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. *Jurnal Informatika Universitas Pamulang*, 5(4), 578–584.

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.

PARYUNI, A. D. (2022). *Dermatofitosis pada Kucing: Profil Kortisol, Sitokin IL-2, IL-10, dan IL-12 dalam Hubungannya dengan Kesembuhan Selama Proses Terapi*. Universitas Gadjah Mada.

- Romadiana, P., Septryanti, A., & Sari, L. I. (2021). Implementasi Algoritma Certainty Factor Dalam Sistem Pakar Untuk Mendiagnosa Penyakit Kulit Pada Kucing. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 10(3), 545–556.
- Susanto, H., Kartikaningrum, M., Wahjuni, R. S., Warsito, S. H., & Yuliani, M. G. A. (2020). Kasus scabies (*Sarcoptes scabiei*) pada kucing di klinik Intimedipet Surabaya. *Jurnal Biosains Pascasarjana*, 22(1), 37–45.
- Wibawa, A. P., Guntur, M., Purnama, A., Akbar, M. F., & Dwiyanto, F. A. (2018). Metode-metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).
- Yudhana, A., Praja, R. N., Pratiwi, A., & Islamiyah, N. (2021). Diagnosa dan observasi terapi infestasi ektoparasit notoedres cati penyebab penyakit scabiosis pada kucing peliharaan. *Media Kedokteran Hewan*, 32(2), 70.
- Zen, L. E., Nurcahyo, G. W., & Yuhandri, Y. (2021). Metode Forward Chaining dalam Menganalisis Penyakit Kucing Akibat Infeksi Virus. *Jurnal Sistem Informasi Dan Teknologi*, 251–256.

LAMPIRAN

- Source code

```
1. import pandas as pd
2.
3. penyakit = pd.read_excel('/content/800 data kucing srvet.xlsx')
4.
5.
6. penyakit['Penyakit'].value_counts()
7.
8.
9. import matplotlib.pyplot as plt
10.
11. # Membuat jumlah visualisasi dari masing-masing klasifikasi
12.
13. data = penyakit['Penyakit'].value_counts()
14. data.plot(kind='pie',autopct='% .2f%%')
15. plt.show()
16.
17.
18. # Menentukan x & y
19. feature_cols = ['G01', 'G02', 'G03', 'G04', 'G05', 'G06', 'G07', 'G08']
20.
21. X = penyakit[feature_cols]
22. y = penyakit['Penyakit']
23.
24. from sklearn.model_selection import train_test_split
25.
26. # split
27. X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
    random_state=2) # agar splitting tidak berubah, diberi random_state
28.
29. from imblearn.over_sampling import SMOTE
30.
31. smote = SMOTE()
32. X_train, y_train = smote.fit_resample(X_train, y_train)
33. X_test, y_test = smote.fit_resample(X_test, y_test)
34.
35. X_train = pd.DataFrame(X_train, columns=X.columns)
36. X_train
37.
38. X_test = pd.DataFrame(X_test, columns=X.columns)
39. X_test
40.
```

```


41. y_train.value_counts()
42.
43. y_test.value_counts()
44.
45.
46. from sklearn.ensemble import RandomForestClassifier
47.
48. # Ranfor Training
49. ranfor = RandomForestClassifier(n_estimators=100, max_depth=5,
    min_samples_split=10, min_samples_leaf=5,
    criterion="entropy",random_state=2)
50. ranfor.fit(X_train, y_train)
51.
52. y_train_pred = ranfor.predict(X_train)
53. y_test_pred = ranfor.predict(X_test)
54.
55.
56. from sklearn.metrics import accuracy_score
57. from sklearn.metrics import precision_score
58. from sklearn.metrics import recall_score
59. from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
60.
61. print('Performa Pada Training\n')
62. print('Accuracy Training :',accuracy_score(y_train, y_train_pred))
63. print('Precision Training :',precision_score(y_train, y_train_pred,
    average='macro'))
64. print('Recall Training :',recall_score(y_train, y_train_pred,
    average='macro'))
65.
66. print('Performa Pada Testing\n')
67. print('Accuracy Testing :',accuracy_score(y_test, y_test_pred))
68. print('Precision Testing :',precision_score(y_test, y_test_pred,
    average='macro'))
69. print('Recall Testing :',recall_score(y_test, y_test_pred, average='macro'))
70. print('n')
71. print('Confusion Matrix :\n')
72. cm = confusion_matrix(y_test, y_test_pred, labels=ranfor.classes_)
73. disp = ConfusionMatrixDisplay(confusion_matrix=cm,
    display_labels=ranfor.classes_)
74. disp.plot()
75.
76. plt.show()
77.
78.
79. y = pd.DataFrame(y_test_pred)
80. y.value_counts()

```

```
81.  
82.  
83. data = y.value_counts()  
84. data.plot(kind='pie',autopct='% .2f%%')  
85. plt.show()  
86.  
87.  
88. from sklearn.tree import export_graphviz  
89. import graphviz  
90.  
91. dot_data = export_graphviz(ranfor.estimators_[0], out_file=None)  
92. graph = graphviz.Source(dot_data)  
93. graph.render("tree")
```



- Turnitin

 **Similarity Report ID:** oid:29477:47139349

PAPER NAME: **Adithya Wahyudi Nasution 188160040.pdf**
AUTHOR: **Adithya Wahyudi Nasution**

WORD COUNT: **8915 Words**
CHARACTER COUNT: **52975 Characters**

PAGE COUNT: **61 Pages**
FILE SIZE: **700.3KB**

SUBMISSION DATE: **Nov 22, 2023 5:18 AM GMT+7**
REPORT DATE: **Nov 22, 2023 5:19 AM GMT+7**

● **19% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.


- 16% Internet database
- 7% Publications database
- Crossref database
- Crossref Posted Content database
- 13% Submitted Works database

● **Excluded from Similarity Report**

- Small Matches (Less than 10 words)

Summary

- Sk Pembimbing

**UNIVERSITAS MEDAN AREA**
FAKULTAS TEKNIK

Kampus I : Jalan Kolim Nomor 1 Medan Estate/Jalan PGSI Nomor 1 (061) 7366878, 7360168, 7364348, 7366781, Fax.(061) 7366998 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Ganyu Nomor 70 A, (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanama@uma.ac.id

Nomor : 124/FT.6/01.10/II/2023 16 Februari 2023
Lamp : -
Hal : Pembimbing Tugas Akhir

Yth. Pembimbing Tugas Akhir
Andre Hasudungan Lubis, S. Ti, MSc
di
Tempat

Dengan hormat, sehubungan telah dipenuhinya persyaratan untuk memperoleh Tugas Akhir dari mahasiswa atas :

N a m a : Adithya Wahyudi Nasution
N P M : 188160040
Jurusan : Teknik Informatika

Maka dengan hormat kami mengharapkan kesediaan saudara :


Andre Hasudungan Lubis, S. Ti, MSc (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :


"Klasifikasi Penyakit Infeksi Kulit pada Kucing Menggunakan Algoritma *Random Forest*".

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.


Dekan,
Dr. Rahmad Syah, S. Kom, M.Kom

- Pengantar Riset

 **UNIVERSITAS MEDAN AREA**
FAKULTAS TEKNIK

Kampus I : Jalan Kadam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎(061) 7366478, 7360168, 7364346, 7360781, Fax: (061) 7366998 Medan 20223
Kampus II : Jalan Selayubi Nomor 79 / Jalan Sei Serayu Nomor 70 A. ☎ (061) 8226602, Fax: (061) 8226331 Medan 25122
Website: www.fabrik.uma.ac.id. E-mail: univ_medanama@uma.ac.id

Nomor : 395 /FT.6/01.10/V/2023 29 Mei 2023
Lamp : -
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Pimpinan SR Vet Clinic & Petshop
Masjid Al-Falah, Jalan Denai Samping No.296
Di
Medan

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



NO	NAMA	NPM	PRODI
1	Adithya Wahyudi Nasution	188160040	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 pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :


Klasifikasi Penyakit Infeksi Kulit pada Kucing Menggunakan Algoritma *Random Forest*

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.


Dekan,

Dr. Rahmad Syah, S. Kom, M. Kom

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

- Selesai Riset

 **SR VET CLINIC & PET SHOP**
Jl. Denai No.296 C Kel. Tegal Sari Mandala II, Kec. Medan Denai, Kota
Medan, Sumatera Utara 20371, Phone: 0852-6226-2299

Medan, 03 Juli 2023

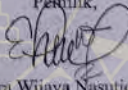
Nomor :
Lamp : -
Hal : Surat Selesai Riset / Penelitian Tugas Akhir
Yth,
Dekan Fakultas Teknik
Universitas Medan Area
Di
Tempat

Dengan hormat,
Bersama ini kami sampaikan bahwa mahasiswa yang tersebut dibawah ini:

No	NAMA	NPM	PROGRAM STUDI
1	Adithya Wahyudi Nasution	188160040	Teknik Informatika

Adalah benar telah menyelesaikan penelitian di SR VET CLINIC & PESHOP sebagai syarat dalam menyelesaikan studinya. Terhitung mulai tanggal 29 Mei 2023 sampai dengan 03 Juli 2023.

Dengan surat ini kami sampaikan untuk dapat diketahui dan digunakan seperlunya.

Pemilik,

Chandra Wijaya Nasution, S.E