SKRIPSI

IMPLEMENTASI SOFT VOTING CLASSIFIER UNTUK DIAGNOSIS PNEUMONIA

Disusun dan diajukan oleh:

ARIES WAHYU SYAPUTRA
D121 17 1528

PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2023

LEMBAR PENGESAHAN SKRIPSI

IMPLEMENTASI SOFT VOTING CLASSIFIER UNTUK DIAGNOSIS PNEUMONIA

Disusun dan diajukan oleh

ARIES WAHYU SYAPUTRA

D121171528

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 27 November 2023 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping,
Dr. Amil Ahmad Ilham, S.T., MIT.	A. Ais Prayogi, ST., M.Eng
Nip. 197310101998021001	Nip. 198305102014041001

Ketua Program Studi,

Prof. Dr. Ir. Indrabayu, S.T., M.T., M.Bus.Sys., IPM, ASEAN. Eng.
Nip. 19750716 200212 1 004

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama	:	Aries Wahyu Syaputra
NIM	:	D121171528
Program Studi	:	Teknik Informatika
Jenjang	:	S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Implementasi Soft Voting Classifier untuk Diagnosis Pneumonia

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

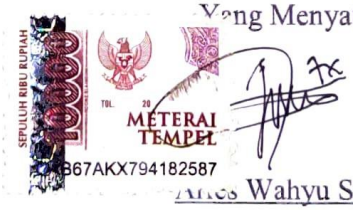
Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 4 Agustus 2023

Yang Menyatakan



Aries Wahyu Syaputra

ABSTRAK

ARIES WAHYU SYAPUTRA. *Implementasi Soft Voting Classifier untuk Diagnosis Pneumonia* (dibimbing oleh Dr. Amil Ahmad Ilham, S.T., M.IT., dan A. Ais Prayogi Alimuddin, S.T., M.Eng.)

Pneumonia adalah infeksi akut pada jaringan paru-paru (alveolus) yang dapat disebabkan oleh berbagai mikroorganisme seperti virus, jamur dan bakteri. Hal ini dapat menyebabkan penyakit ringan hingga yang mengancam jiwa pada orang-orang dari segala usia, namun pneumonia ini menjadi penyebab kematian menular terbesar pada anak-anak di seluruh dunia. Berdasarkan data yang diperoleh dari Rumah Sakit Ibnu Sina YW-UMI, terdapat beberapa variabel yang berhubungan dengan pneumonia sehingga dilakukan analisis untuk mengklasifikasikan penyakit tersebut. Sehingga, penelitian ini membangun tiga model klasifikasi *machine learning*, yaitu *K-Nearest Neighbor*, *Support Vector Machine*, dan *Decision Tree*. Evaluasi model akan dilakukan untuk membandingkan hasil klasifikasi dari komposisi data latih dan data uji yang sama. Model *Ensemble Soft Voting* dibangun untuk mengkombinasikan hasil prediksi ketiga model klasifikasi yang bertujuan untuk mencapai kinerja yang lebih baik dan optimal dibandingkan kinerja pada satu model klasifikasi saja. *Soft Voting* memiliki kinerja dengan memasukkan prediksi dari beberapa pengklasifikasi berdasarkan rata-rata probabilitas dari *output* kelas hasil prediksi beberapa algoritma yang digunakan. Berdasarkan hasil pengujian performa, model *Soft Voting* memiliki performa yang lebih unggul dibandingkan ketiga model lainnya dengan akurasi sebesar 98,3%, kemudian diikuti oleh model *Decision Tree* dengan akurasi 97,1%, *Support Vector Machine* dengan akurasi 94,3%, dan *K-Nearest Neighbor* dengan akurasi 92,6%. Adapun dalam hasil pengujian waktu komputasi, model *Decision Tree* yang menunjukkan waktu komputasi tercepat, yaitu 0,0046 detik, selanjutnya ada *K-Nearest Neighbor* dengan 0,0121 detik, kemudian *Support Vector Machine* dengan 7,1949 detik, dan *Soft Voting* dengan 7,2568 detik. Dapat dilihat pada konteks performa *Soft Voting* yang lebih optimal dan dalam konteks kecepatan memprediksi *Decision Tree* yang lebih unggul. Pada penelitian ini dengan kasus diagnosis penyakit pneumonia, umumnya akurasi dari diagnosis adalah aspek yang sangat penting dalam konteks medis, karena kesalahan dalam prediksi dapat berakibat fatal. Oleh karena itu, diputuskan menggunakan model *Soft Voting* yang lebih memiliki tingkat akurasi tertinggi untuk mengklasifikasikan penyakit pneumonia.

Kata Kunci: Pneumonia, *K-Nearest Neighbor*, *Support Vector Machine*, *Decision Tree*, *Ensemble Soft Voting*

ABSTRACT

ARIES WAHYU SYAPUTRA. *Implementation of Soft Voting Classifier for Pneumonia Diagnosis* (supervised by Dr. Amil Ahmad Ilham, S.T., M.IT., and A. Ais Prayogi Alimuddin, S.T., M.Eng.)

Pneumonia is an acute infection of the lung tissue (alveolus) that can be caused by various microorganisms such as viruses, fungi and bacteria. It can cause mild to life-threatening illness in people of all ages, but it is the largest infectious cause of death in children worldwide. Based on the data obtained from Ibnu Sina Hospital YW-UMI, there are several variables associated with pneumonia so an analysis was conducted to classify the disease. Thus, this study builds three machine learning classification models, namely K-Nearest Neighbor, Support Vector Machine, and Decision Tree. Model evaluation will be conducted to compare classification results from the same composition of training data and test data. The Soft Voting Ensemble model is built to combine the prediction results of the three classification models which aims to achieve better and optimal performance than the performance of one classification model alone. Soft Voting has a performance by including predictions from several classifiers based on the average probability of the class *output* of the prediction results of several algorithms used. Based on the performance test results, the Soft Voting model has superior performance compared to the other three models with an accuracy of 98.3%, followed by the Decision Tree model with 97.1% accuracy, Support Vector Machine with 94.3% accuracy, and K-Nearest Neighbor with 92.6% accuracy. As for the computation time test results, the Decision Tree model shows the fastest computation time, which is 0.0046 seconds, followed by K-Nearest Neighbor with 0.0121 seconds, then Support Vector Machine with 7.1949 seconds, and Soft Voting with 7.2568 seconds. It can be seen in the context of Soft Voting's more optimal performance and in the context of Decision Tree's superior prediction speed. In this study with the case of pneumonia diagnosis, generally the accuracy of diagnosis is a very important aspect in the medical context, as errors in prediction can be fatal. Therefore, it was decided to use the Soft Voting model which has the highest accuracy rate for classifying pneumonia diseases.

Keywords: Pneumonia, *K-Nearest Neighbor, Support Vector Machine, Decision Tree, Ensemble Soft Voting*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
K-NN	K-*Nearest Neighbor*
SVM	*Support Vector Machine*
PSO	*Particle Swarm Optimization*
TP	*True Positive,* jumlah data positif yang terklasifikasi dengan benar oleh sistem untuk kelas ke-i
TN	*True Negative*, jumlah data negatif yang terklasifikasi dengan benar oleh sistem untuk kelas ke-i
FN	*False Negative*, jumlah data negatif namun terklasifikasi salah oleh sistem untuk kelas ke-i
FP	*False Positive*, jumlah data positif namun terklasifikasi salah oleh sistem.
b	Bias
$x = x_1, x_2, \ldots . x_D)^T$	Variabel *input*
$w = (w_0, w_1, \ldots . w_D)^T$	Parameter bobot
$\phi(x)$	Fungsi transformasi fitur
(x_i, x_j)	Vektor *input*
γ (gamma)	Konstanta untuk percepatan fungsi
r	Koefisien
d	Derajat *polynomial*
c (*complexity*)	Konstanta untuk jarak margin
w_j	Bobot yang dapat diberikan pada pengklasifikasi *j*
S	Himpunan kasus
n	Jumlah partisi n
pi	Proporsi Si terhadap S
A	Fitur
$\|Si\|$	Proporsi jumlah kasus dalam S
$\|S\|$	Porsi Si terhadap S

DAFTAR LAMPIRAN

KATA PENGANTAR

Assalamu'alaikum Warohmatullahi Wabarokatuh

Alhamdulillah, segala puji dan syukur penulis panjatkan kehadirat Allah Subhanahu wa ta'ala, Tuhan yang Maha Esa atas limpahan rahmat dan hidayah-Nya, sehingga dapat menyelesaikan tugas akhir dengan judul "Implementasi *Soft Voting Classifier* untuk Diagnosis Pneumonia" sebagai salah satu syarat dalam menyelesaikan jenjang Strata-1 di Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin.

Dalam penyusunan penelitian ini disajikan hasil penelitian terkait dengan judul yang telah diangkat dan telah melalui proses pencarian dari berbagai sumber baik jurnal penelitian, prosiding pada seminar-seminar nasional/internasional, buku maupun dari situs-situs di internet.

Penulis menyadari bahwa dalam proses pengerjaan tugas akhir ini tidak akan lepas dari doa, bantuan, bimbingan serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Kedua orang tua penulis, Bapak H. Muhammad Sila T. dan Ibu Hj. Matahari yang tidak pernah lelah dalam mendidik, mendoakan, memberikan semangat, serta bantuan dalam berbagai bentuk kepada penulis.
2. Bapak Dr. Amil Ahmad Ilham, S.T., M.IT., selaku pembimbing I dan Bapak A. Ais Prayogi Alimuddin, S.T., M.Eng., selaku pembimbing II sekaligus sebagai Dosen Pembimbing Akademik, yang selalu menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa untuk mengarahkan penulis dalam proses penyelesaian tugas akhir ini.
3. Bapak Prof. Dr. Indrabayu, S.T., M.T., M.Bus.sys., selaku Ketua Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin atas segala bimbingan, motivasi dan dukungan selama masa perkuliahan.
4. Segenap Dosen dan Staf Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah banyak memberikan ilmu dan pengalaman, serta bantuan kepada penulis selama menuntut ilmu di kampus tercinta ini.
5. Bapak dr. Okto Mara Fandi Harahap, M.Ked(Paru), Sp.P., dan Bapak dr. Ida Bagus Putu Ekaruna, Sp.P selaku dokter paru yang senantiasa memberikan masukan terkait pneumonia.
6. Bapak Dr. Muhammad Arafah, S.Kom., M.T., yang telah membantu penulis dalam pencarian informasi mengenai rumah sakit.
7. Bapak Budi, Ibu Luly, Kak Ima, dan seluruh staf di Rumah Sakit Ibnu Sina YW-UMI yang telah memberikan izin penelitian, dan membantu penulis selama penelitian, serta Muhammad Andar Sugianto dan Prasetya Abdi Putra yang telah membantu penulis dalam pengumpulan data.
8. Muhammad Zulfahmi Sadrah, S.T., sebagai teman diskusi yang tidak pernah bosan untuk dijadikan tempat konsultasi mengenai *source code* yang digunakan dalam penyelesaian tugas akhir ini dan juga semasa perkuliahan.

9. Orang orang yang pernah spesial sebagai moodbooster dalam pengerjaan skripsi ini sehingga dapat diselesaikan dengan lancar. Terima kasih telah menjadi teman cerita, teman ngonten dan teman hidup.
10. Kak Fadel Ramadhan, Amiruddin, Wahyu Faisal, Fauzan Anwar, Muhammad Bishram, Ahmad Reza, Ikhwan Ramadhan, Muhammad Fadhil, Fitriani Nasir, serta seluruh anggota Recognizer yang telah membantu penulis sejak awal perkuliahan dan selalu menjadi *supportive system* secara tidak langsung dalam pengerjaan skripsi.
11. Serta teman-teman ataupun pihak-pihak lain yang tidak dapat disebutkan dan tanpa sadar telah menjadi inspirasi dan membantu penulis dalam menyelesaikan tugas akhir.

Akhir kata, penulis berharap semoga segala bantuan dalam bentuk apapun mendapatkan berkah dari yang Maha Kuasa dan berkenan membalas segala kebaikan dari semua pihak yang telah banyak membantu. Semoga Tugas Akhir ini dapat memberikan manfaat bagi pengembangan ilmu. Aamiin.

Assalamu'alaikum Warohmatullahi Wabarokatuh

Makassar, Januari 2023

Penulis

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pneumonia adalah bentuk infeksi pernapasan akut yang menyerang paru- paru. Paru-paru terdiri dari kantung kecil yang disebut alveolus, yang terisi udara saat orang sehat bernapas. Ketika seseorang menderita pneumonia, alveolus dipenuhi dengan nanah dan cairan, yang membuat pernapasan terasa sakit dan membatasi oksigenasi. Infeksi ini biasanya ditularkan melalui kontak langsung dengan orang yang terinfeksi. Virus dan bakteri dapat menyebar melalui tetesan udara dari batuk atau bersin. Selain itu, pneumonia dapat menyebar melalui darah, terutama selama dan sesaat setelah lahir (WHO, 2019).

Berdasarkan data dari *World Health Organization* tahun 2019, pneumonia merupakan penyebab utama kematian menular pada anak-anak di seluruh dunia. Pneumonia membunuh 740.180 anak di bawah usia lima tahun pada tahun 2019, terhitung 14% dari semua kematian anak usia di bawah lima tahun. Diperkirakan 19.000 anak meninggal akibat pneumonia pada tahun 2018. Estimasi global menunjukkan bahwa setiap satu jam ada 71 anak di Indonesia yang tertular pneumonia (UNICEF, 2019).

Gejala pneumonia memiliki kemiripan dengan gejala batuk biasa, sehingga masyarakat tidak terlalu percaya dan khawatir dengan gejala yang biasanya muncul. Hal ini disebabkan kurangnya pengetahuan masyarakat tentang pneumonia (Josefa *et al*, 2019). Umumnya pasien yang datang ke dokter sudah mengalami gejala pneumonia berat, kondisi ini bisa membahayakan kesehatan dan keselamatan penderita pneumonia, sehingga penanganannya juga menjadi tidak maksimal (Gustri *et al*, 2018).

Dengan demikian, maka pemanfaatan *Data Mining* dapat menjadi solusi yang dapat membantu proses diagnosis pneumonia berdasarkan gejala, pemeriksaan fisik serta pemeriksaan penunjang. Salah satu teknik *Data Mining* dalam diagnosis penyakit adalah teknik klasifikasi yang dimana merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui (Rahmawati *et al*, 2019).

Penelitian yang memprediksi pneumonia pada daerah Bangladesh dengan pendekatan *machine learning* menggunakan enam algoritma yang berbeda. Dan tiga di antaranya adalah K-NN, *Support Vector Machine,* dan *Decision Tree*. Hasil penelitian ini menunjukkan nilai akurasi yang didapatkan dari 30% data testing pada algoritma K-NN sebesar 90%, *Support Vector Machine* 85% dan *Decision Tree* sebesar 91%. Dari keenam

hasil algoritma yang digunakan, algoritma *Decision Tree* yang menghasilkan akurasi terbaik (Hasan *et al*, 2021).

Adapun juga penelitian serupa yang mendeteksi pneumonia covid dan pneumonia umum menggunakan komparasi algoritma klasifikasi *machine learning*. Dan tiga di antaranya merupakan *K-Nearest Neighbor, Support Vector Machine*, dan *Decision Tree*. Hasil penelitian menunjukkan nilai akurasi pada *K-Nearest Neighbor* sebesar 88,24%, *Support Vector Machine* sebesar 86,48%, dan *Decision Tree* sebesar 89,82%. Dari ketiga algoritma tersebut Decision Tree menunjukkan nilai akurasi yang terbaik (Chenglong, 2020).

Adapun pada penelitian lainnya dengan menentukan model prediksi *stage* awal pada pasien kanker paru-paru terbaik berdasarkan beberapa algoritma klasifikasi. Adapun tiga di antaranya adalah *K-Nearest Neighbor*, *Support Vector Machine*, dan *Decision Tree*. Hasil penelitian yang didapatkan pada *K-Nearest Neighbor* sebesar 85%, *Support Vector Machine* sebesar 90%, dan *Decision Tree* sebesar 77%. Dari ketiga algoritma tersebut *Support Vector Machine* yang menunjukkan nilai akurasi yang terbaik (Haijing Tang *et al*, 2018).

Dan pada penelitian lainnya yang memprediksi pneumonia pada pasien penderita skizofrenia dengan pendekatan *machine learning* menggunakan tujuh algoritma. Dari ketujuh algoritma tersebut, tiga di antaranya K-NN, *Support Vector Machine,* dan *Decision Tree*. Dari ketujuh hasil algoritma yang digunakan *Decision Tree* menunjukkan akurasi prediksi yang optimal dibanding algoritma lainnya dengan akurasi sebesar 95% (Kuo *et al*, 2019).

Berdasarkan penelitian di atas, penulis akan mengembangkan sistem diagnosis pneumonia dengan menggunakan metode *Soft Voting Ensemble. Soft Voting Ensemble* merupakan kombinasi dari beberapa pengklasifikasi dimana keputusan dibuat berdasarkan keputusan individu yang digabungkan dengan nilai probabilitas untuk menentukan bahwa data termasuk dalam kelas tertentu. Dalam metode *Soft Voting Ensemble*, prediksi dibobot berdasarkan kepentingan pengklasifikasi dan menggabungkannya untuk mendapatkan jumlah probabilitas terbobot. Label target dengan jumlah probabilitas tertimbang terbesar dipilih karena memiliki nilai *voting* terbesar (Sherazi *et al*, 2021).

Dengan adanya sistem ini, diharapkan dapat memberikan hasil performa model terbaik dari klasifikasi penyakit pneumonia dan dapat membantu dalam mendiagnosis pneumonia sehingga dapat memberikan informasi sebagai dasar untuk melakukan tindakan yang diperlukan untuk pengobatan, pengendalian dan pencegahan.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah pada tugas akhir ini adalah sebagai berikut:

a. Bagaimana mengidentifikasi penyakit pneumonia menggunakan beberapa algoritma klasifikasi.
b. Bagaimana menguji performa dari klasifikasi penyakit pneumonia dengan teknik *Soft Voting*.

1.3 Tujuan Penelitian/Perancangan

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

a. Mengidentifikasi penyakit pneumonia menggunakan beberapa algoritma klasifikasi.
b. Memberikan performa terbaik dari klasifikasi penyakit pneumonia dengan teknik *Soft Voting*.

1.4 Manfaat Penelitian/Perancangan

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

a. Memberikan dorongan penggunaan teknologi *data mining* dengan teknik *Soft Voting* untuk mendapatkan informasi dalam diagnosis penyakit pneumonia.
b. Memberikan informasi sebagai dasar pengambilan keputusan atau tindakan untuk pengendalian dan pencegahan pneumonia di masa yang akan datang.

1.5 Ruang Lingkup/Asumsi perancangan

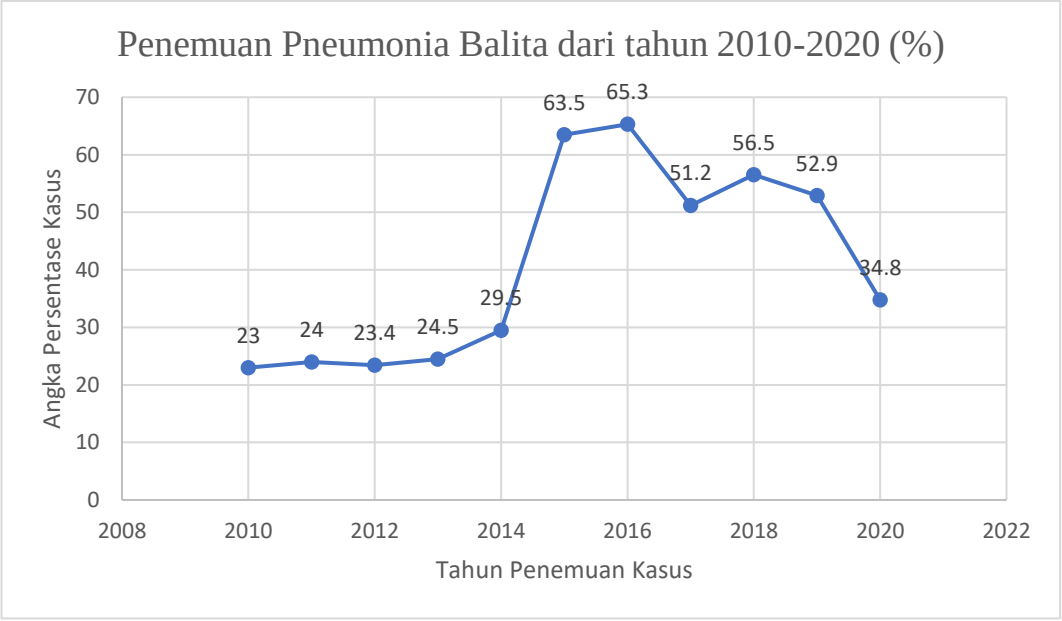
Ruang lingkup dari penelitian ini adalah sebagai berikut:

a. Pengambilan data dilakukan di Rumah Sakit Ibnu Sina YW-UMI, yang berupa data rekam medis pasien pneumonia dari tahun 2018 hingga tahun 2022, yang kemudian dibentuk dalam format *excel* (.csv).
b. Hasil klasifikasi terdiri dari dua kelas, yaitu “Pneumonia Berat” dan “Pneumonia Ringan”.

BAB II
TINJAUAN PUSTAKA

2.1 Pneumonia

Pneumonia adalah infeksi akut pada jaringan paru-paru (alveolus) yang dapat disebabkan oleh berbagai mikroorganisme seperti virus, jamur dan bakteri. Hal ini dapat menyebabkan penyakit ringan hingga yang mengancam jiwa pada orang-orang dari segala usia, namun pneumonia ini menjadi penyebab kematian menular terbesar pada anak-anak di seluruh dunia. Begitupun yang terjadi di Indonesia, pneumonia ini menjadi penyakit penyebab kematian balita terbesar. Sampai saat ini program pengendalian pneumonia lebih diprioritaskan pada pengendalian pneumonia balita. Salah satu upaya untuk mengatasi penyakit ini yaitu dengan meningkatkan penemuan pneumonia pada balita. Berikut cakupan penemuan kasus pneumonia pada balita di Indonesia pada tahun 2010-2020 yang dapat dilihat pada Gambar 1 (Kemenkes RI, 2021).



Gambar 1 Cakupan Penemuan Pneumonia Balita di Indonesia Tahun 2010-2020 (Kemenkes RI, 2021)

Pada Gambar 1 dapat dilihat cakupan penemuan pneumonia pada balita di Indonesia berkisar antara 20% hingga 30% dari tahun 2010 sampai dengan 2014, dan cakupan dari tahun 2015 hingga 2019 terjadi peningkatan dikarenakan adanya perubahan angka perkiraan kasus dari 10% menjadi 35,5%. Namun, pada tahun

2020 terjadi penurunan kembali menjadi 34,8%. Penurunan ini lebih disebabkan oleh dampak pandemi *Covid-19*, dimana adanya stigma pada penderita *Covid-19* sehingga menyebabkan penurunan jumlah kunjungan balita yang mengalami batuk atau sesak napas di puskesmas pada tahun 2020 (Kemenkes RI, 2021).

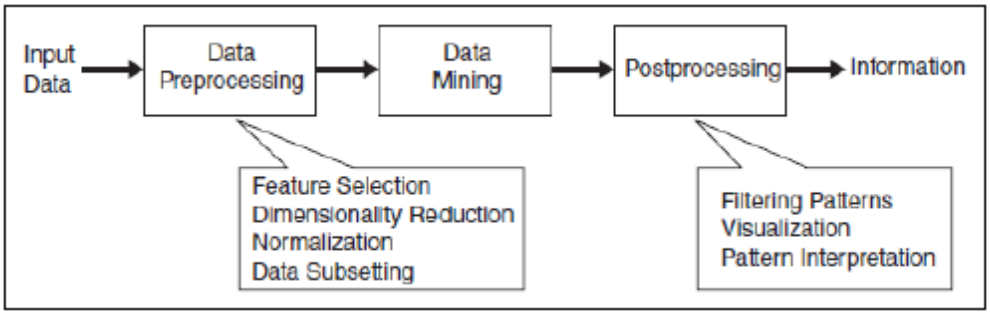
Orang yang berisiko terkena pneumonia tidak hanya pada usia anak, namun juga segala usia termasuk orang dewasa di atas usia 65 tahun dan orang dengan masalah kesehatan yang sudah ada sebelumnya (WHO, 2021).

Tingkat keparahan pneumonia dapat dinilai dari ada tidaknya kriteria yang berdasarkan organisasi *American Thoracic Society/Infectious Diseases Society of America* (ATS/IDSA) yang terdiri dari kriteria minor dan mayor. Adapun kriteria minor terdiri dari frekuensi napas ≥ 30 x/menit, rasio PAO2/FIO2 ≤ 250, *infiltrate multilobar*, penurunan kesadaran/disorientasi, uremia (BUN ≥ 20 mg/dL), leukopenia (< 4000/dL), trombositopenia (< 100.000/dL), hipotermia (< 36°C), dan hipotensi yang memerlukan resusitasi cairan yang agresif. Untuk kriteria mayor terdiri dari perlunya menggunakan ventilator mekanik dan adanya syok septik yang memerlukan vasopressor. Indikasi rawat ICU untuk penderita pneumonia bila didapatkan minimal 3 kriteria minor dan 1 kriteria mayor (Joshua *at al*, 2019).

2.2 *Data Mining*

Data mining adalah proses untuk menemukan informasi yang berguna dalam penyimpanan data yang besar. Teknik *Data Mining* digunakan untuk menjelajahi kumpulan data dalam skala yang besar untuk menemukan pola baru yang sebelumnya tidak diketahui dan berguna (Tan *et al*, 2019).

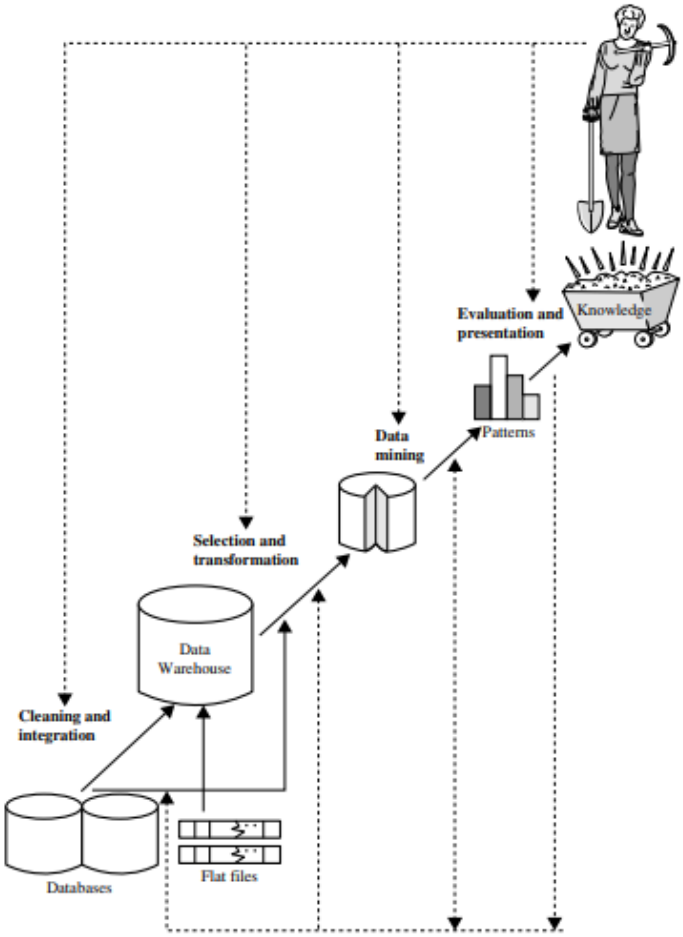
Tan et al menyebutkan bahwa *Data Mining* merupakan bagian penting dalam proses *Knowledge Discovery in Database* (KDD), yang merupakan keseluruhan proses untuk mengubah data mentah menjadi informasi yang berguna. Proses KDD ini terdiri dari beberapa langkah, mulai dari *preprocessing* data hingga *post processing* hasil dari *Data Mining*, seperti yang ditampilkan pada Gambar 2.



Gambar 2 Proses *Knowledge Discovery in Database* (Tan *et al*, 2019)

Pada Gambar 2 dapat dilihat proses dimulai dengan memasukkan data yang disimpan dalam penyimpanan data terpusat. Kemudian, dilakukan *preprocessing* untuk mengubah data mentah menjadi format yang sesuai untuk analisis selanjutnya. Pada proses *preprocessing data* dengan melakukan *feature selection*, *dimensionality reduction*, *normalization*, dan *data subsetting*. Setelah itu dilakukan proses integrasi hasil *Data Mining* ke dalam sistem pendukung keputusan. Integrasi tersebut memerlukan langkah *post processing* untuk memastikan bahwa hanya hasil yang valid dan berguna yang digabungkan ke dalam sistem pendukung keputusan. Contoh *post processing* adalah visualisasi yang memungkinkan analisis untuk mengeksplorasi data dan hasil penambangan data dari berbagai sudut pandang (Tan *et al*, 2019).

Adapun tahapan *Data Mining* lainnya menurut Jiawei Han *et al*, dapat dilihat pada Gambar 3.



Gambar 3 Tahap *Data Mining* (Han *et al*, 2011)

Gambar 3 menampilkan tahapan *Data Mining*. Jiawei Han *et al* menyebutkan proses dari *Data Mining* atau *Knowledge Discovery from Database* (KDD) yang terbagi ke dalam 7 tahap, yaitu sebagai berikut (Han *et al*, 2011).

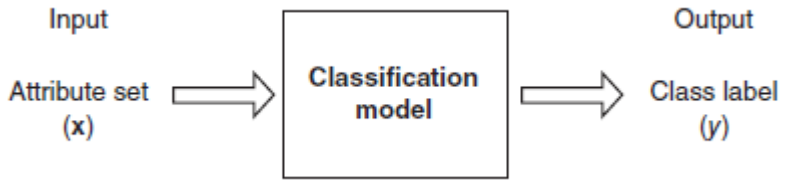
a. *Data cleaning*: tahap untuk membersihkan data yang hilang, *noise* dan tidak konsisten.
b. *Data integration*: tahap dimana beberapa sumber data dapat digabungkan.
c. *Data selection*: tahap untuk data yang relevan dengan analisis diambil dari *database*.
d. *Data transformation*: tahap untuk data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk penambangan dengan melakukan operasi *summary* atau *aggregation*.
e. *Data Mining*: tahap penting dimana metode cerdas diterapkan untuk mengekstrak pola data.

f. *Pattern evaluation*: tahap untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan ukuran keterkaitannya (*distance/interestingness measure*).
g. *Knowledge presentation*: tahap dimana teknik gambaran visualisasi yang digunakan untuk menyajikan pengetahuan kepada pengguna.

2.3 Klasifikasi

Klasifikasi adalah proses menemukan model (fungsi) yang menggambarkan dan membedakan kelas atau konsep data. Model diturunkan berdasarkan analisis *dataset* pelatihan (objek data yang label kelasnya diketahui). Model digunakan untuk memprediksi label kelas untuk objek yang label kelasnya tidak diketahui. Klasifikasi termasuk algoritma *supervised learning* dengan mengelompokkan data ke dalam kelas-kelas (Han *et al*, 2011).

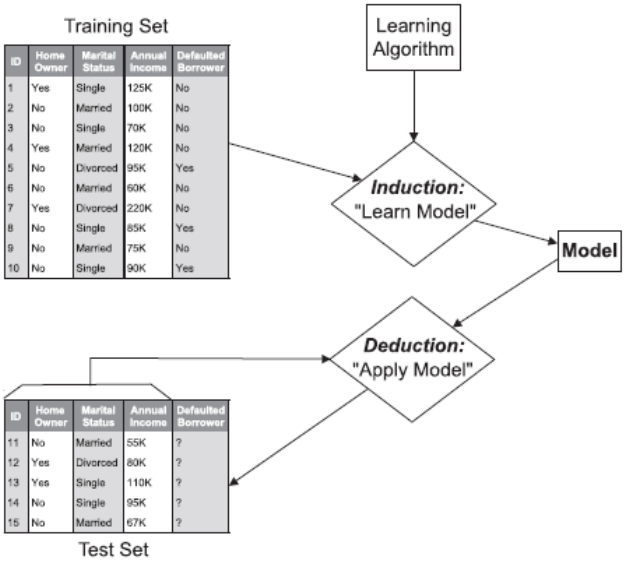
Dalam proses klasifikasi data yang diinputkan berupa data *record* atau data sampel. Pada setiap *record* dikenal dengan *instance* atau contoh yang ditentukan oleh sebuah tuple (x,y). Dimana x adalah himpunan atribut dan y adalah atribut tertentu yang menyatakan sebagai label *class*. Berikut ini mengilustrasikan proses klasifikasi pada Gambar 4 (Tan *et al*, 2019).



Gambar 4 Proses Klasifikasi (Tan *et al*, 2019)

Klasifikasi biasanya digambarkan sebagai pendekatan sistematis untuk membangun model klasifikasi dari sebuah *dataset input*. Model yang dibangun dengan sebuah algoritma pembelajaran haruslah sesuai dengan data *input* dan memprediksi dengan benar label kelas dari *record* yang belum pernah terlihat sebelumnya. Dengan demikian, kunci utama dari algoritma pembelajaran adalah

membangun model dengan kemampuan generalisasi yang baik, yaitu model yang secara akurat memprediksi label kelas dari *record* yang tidak diketahui sebelumnya.



Gambar 5 Pendekatan Umum untuk Pembangunan Model Klasifikasi
(Tan *et al*, 2019)

Gambar 5 menunjukkan pendekatan umum untuk penyelesaian masalah klasifikasi. Pertama, *training data* berisi *record* yang mempunyai label kelas yang diketahui haruslah tersedia. *Training data* digunakan untuk membangun model klasifikasi, yang kemudian diaplikasikan ke *testing data*, yang berisi *record-record* dengan label kelas yang belum diketahui. (Tan *et al*, 2019).

2.4 *K-Nearest Neighbor (*K-NN*)*

K-Nearest Neighbor adalah metode untuk mengklasifikasikan sebuah sampel data berdasarkan kategori mayoritas banyaknya K data pelatihan yang terdekat dengannya (Bramer, 2020).

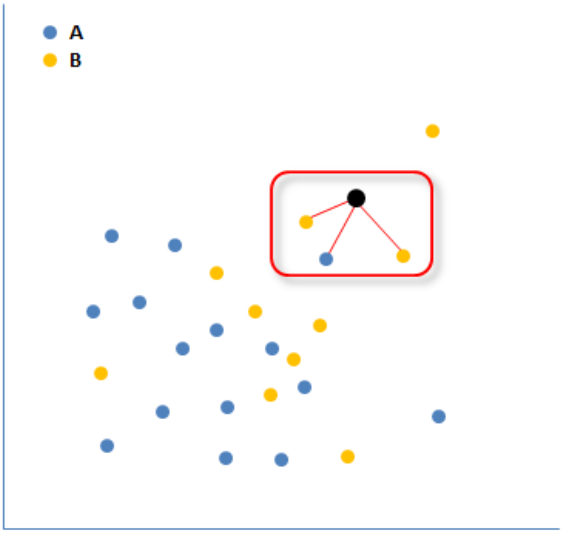
Algoritma K-NN merupakan algoritma *supervised learning* yang dapat digunakan dalam proses klasifikasi dan regresi yang bekerja dengan mengambil sejumlah K data (tetangga) terdekat sebagai acuan untuk menentukan kelas dari data baru. Algoritma ini mengklasifikasikan data berdasarkan *similarity* atau kemiripan ataupun kedekatan dengan data lainnya (Afifah, 2020).

Secara umum, cara kerja algoritma K-NN adalah sebagai berikut:

a. Tentukan jumlah tetangga (K) yang akan digunakan untuk pertimbangan penentuan kelas.

b. Hitung jarak data baru terhadap seluruh data yang ada dalam *dataset*.

c. Ambil sebanyak K data atau tetangga terdekat, kemudian tentukan kelas dari data baru berdasarkan kelas mayoritas dari tetangga terdekat.



Gambar 6 Ilustrasi Algoritma K-NN (Afifah, 2020)

Dari Gambar 6, ada sejumlah titik data yang terbagi menjadi dua kelas yaitu A (biru) dan B (kuning). Misalnya ada data baru (hitam) yang akan kita prediksi kelasnya menggunakan algoritma KNN. Dari contoh di atas, nilai K yang digunakan adalah 3. Setelah dihitung jarak antara titik hitam ke masing-masing titik data lainnya, didapatkan 3 titik terdekat yang terdiri dari 2 titik kuning dan satu titik biru seperti yang diilustrasikan di dalam kotak merah, maka kelas untuk data baru (titik hitam) adalah B (kuning) (Afifah, 2020).

Adapun metode perhitungan jarak pada algoritma K-NN untuk menentukan titik data mana yang paling dekat dengan titik kueri tertentu, jarak antara titik kueri dan titik data lainnya perlu dihitung. Metrik jarak ini membantu membentuk batasan keputusan, yang mengarahkan kueri partisi ke kelas yang berbeda. Untuk menentukan titik serupa terdekat, dapat menggunakan perhitungan jarak seperti persamaan berikut:

a. *Euclidean Distance*

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \qquad (1)$$

b. *Manhattan Distance*

$$d(x,y) = \sum_{i=1}^{m} |(x_i - y_i)| \quad (2)$$

c. *Minkowsky Distance*

$$d(x,y) = \left(\sum_{i=1}^{m} |(x_i - y_i| \, r \right)^{1/r} \quad (3)$$

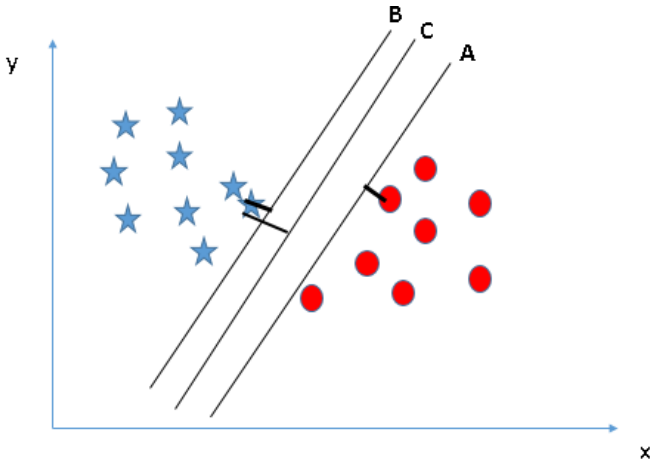
d. *Chebyshev Distance*

$$d(x,y) = \max_{i=1} |x_i - y_i| \quad (4)$$

2.5 *Support Vector Machine (*SVM*)*

Support Vector Machine (SVM) adalah sistem pembelajaran yang menggunakan ruang hipotesis dalam bentuk fungsi linier dalam ruang fitur (*feature space*) berdimensi tinggi dan dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan menerapkan *learning* bias yang berasal dari teori pembelajaran statistik (Krisantus, 2007).

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik tahun 1992 sebagai rangkaian harmonis konsep-konsep utama di dalam bidang *pattern recognition* (Colanus, 2017). SVM bertujuan untuk mencari *hyperplane* yang dapat memisahkan kelas-kelas dengan jarak (*margin/gap*) yang maksimal antara *border-line* (*support vectors*) kelas satu dengan *border-line* kelas lain, seperti pada Gambar 7 (Ray, 2017).



Gambar 7 Visualisasi SVM (Ray, 2017)

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah antara dua *class* pada *input space*. Gambar 7 memperlihatkan beberapa *pattern* yang merupakan anggota dari dua buah *class* : *class* 1 (dinotasikan dengan +1) dan *class 2* (dinotasikan dengan –1). *Pattern* yang tergabung pada *class* 1 disimbolkan dengan bintang, sedangkan *pattern* pada *class* 2 disimbolkan dengan lingkaran. *Hyperplane* pemisah terbaik antara kedua *class* dapat ditemukan dengan mengukur *margin hyperplane* tersebut dan mencari titik maksimalnya. *Margin* adalah jarak antara *hyperplane* dan data terdekat dari masing-masing *class*. Subset data *training set* yang paling dekat ini disebut sebagai *support vector*. Garis yang terletak di tengah kedua garis pada gambar 7 menunjukkan *hyperplane* yang terbaik (Ray, 2017).

Proses pembelajaran SVM adalah untuk menentukan *support vector*, kita hanya cukup mengetahui fungsi *kernel* yang dipakai, dan tidak perlu mengetahui wujud dari fungsi non-linear. Berikut adalah persamaan dari fungsi SVM (Colanus, 2017).

$$f(x) = w^2\phi(x) \; + \; b \tag{5}$$

Dalam SVM terdapat fungsi *kernel* yang terbagi dua yaitu *kernel linier* dan *kernel non-linier*. *Kernel linier* digunakan ketika data yang akan diklasifikasi dapat terpisah dengan sebuah garis (*hyperplane*). Sedangkan *kernel non-linier* digunakan ketika data hanya dapat dipisahkan dengan garis lengkung atau sebuah bidang pada ruang dimensi tinggi. Contoh *kernel non-linier* adalah *Polynomial, Gaussian RBF, Sigmoid, Additive*. Adapun fungsi *kernel* yang umum digunakan adalah sebagai berikut.

a. *Kernel Linier*

$$K(x_i, x_j) = x_i, x_j \tag{6}$$

b. *Kernel Polynomial*

$$K(x_i, x_j) = (\gamma(x_i, x_j) + r)^d \;, \qquad \gamma > 0 \tag{7}$$

c. *Kernel RadialS Basic Function* (RBF)

$$K(x_i, x_j) = \exp\left(\gamma \left\| x_i - x_j \right\|^2\right), \qquad \gamma > 0 \tag{8}$$

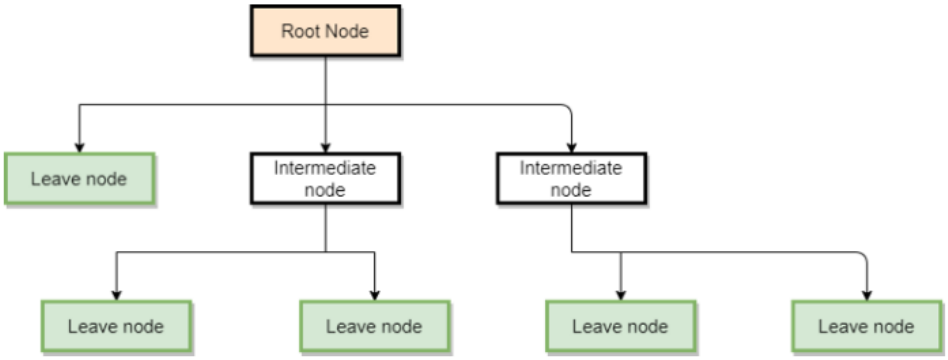
d. *Kernel Sigmoid*

$$K(x_i, x_j) = \tanh(\gamma(x_i, x_j) + r) \ , \qquad \gamma > 0 \tag{9}$$

2.6 *Decision Tree*

Decision Tree adalah metode pohon keputusan di mana setiap *node* mewakili fitur (atribut), setiap tautan (cabang) mewakili keputusan (aturan), dan setiap daun mewakili hasil (nilai kategorikal atau kontinu) (Sanjeevi, 2017).

Decision Tree ini dibangun menggunakan dua jenis elemen, yaitu *node* dan cabang. Di setiap *node*, salah satu fitur data akan dievaluasi untuk membagi pengamatan dalam proses pelatihan atau untuk membuat titik data tertentu mengikuti jalur tertentu saat membuat prediksi. Metode ini dibangun dengan mengevaluasi fitur yang berbeda secara rekursif dan menggunakan fitur yang terbaik untuk membagi data pada setiap *node*. Gambar 8 struktur umum *Decision Tree* (Thorn, 2020).



Gambar 8 Struktur Umum *Decision Tree* (Thorn, 2020)

Pada gambar 8, kita dapat mengamati tiga jenis *node*:

a. *Root node* adalah *node* yang memulai grafik. Dalam *Decision Tree*, jenis *node* ini mengevaluasi variabel yang paling baik membagi data.
b. *Intermediate node* adalah *node* dimana variabel dievaluasi tetapi bukan *node* akhir tempat prediksi dibuat.
c. *Leave node* adalah *node* akhir dari *Decision Tree*, tempat prediksi kategori atau nilai numerik dibuat.

2.7 *Particle Swarm Optimization*

Particle Swarm Optimization (PSO) adalah algoritma metaheuristik yang sering digunakan dalam menyelesaikan masalah optimasi. Algoritma ini mengambil inspirasi dari perilaku kolektif yang dapat diamati dalam kelompok-kelompok alami, seperti kawanan burung atau ikan (Shami *et al*, 2022). PSO terdiri dari tiga komponen penting, yaitu: partikel, komponen kognitif dan komponen sosial, dan kecepatan partikel. Setiap partikel bertindak sebagai representasi dari sebuah solusi. Pembelajaran partikel terbagi menjadi dua faktor, yaitu pengalaman partikel (*cognitive learning*) dan kombinasi pembelajaran dari seluruh kelompok (*social learning*) (Cholissodin dan Riyandani, 2016).

Dalam algoritma PSO, proses pencarian solusi dilakukan oleh suatu populasi yang terdiri dari beberapa partikel. Populasi dibentuk secara acak dengan nilai minimum dan maksimum yang telah ditentukan sebagai batas bawah dan batas atas. Setiap partikel mencari solusi dengan mengeksplorasi ruang pencarian dalam menyesuaikan posisi terbaik pribadi mereka (*local best*), serta penyesuaian terhadap posisi partikel terbaik dari seluruh populasi partikel (*global best*) selama proses pencarian. Sejumlah iterasi dijalankan untuk mencari posisi terbaik dari setiap partikel sampai posisi tersebut menjadi relatif stabil atau telah mencapai batas iterasi yang telah ditentukan. Pada setiap iterasi, performa setiap solusi (posisi partikel) akan dievaluasi dengan memasukkan solusi tersebut ke dalam fungsi kecocokan (*fitness function*) (Sasongko, 2016).

2.8 Ensemble Learning

Ensemble Learning adalah metode yang bertujuan untuk menggabungkan keputusan dari beberapa algoritma pembelajaran untuk meningkatkan hasil akurasi (terutama algoritma pembelajaran yang lemah) dan membuat model yang lebih baik untuk melakukan prediksi. Penggunaan metode *ensemble* lebih baik jika dibandingkan dengan hanya menggunakan satu algoritma pembelajaran (Onan, *et al*., 2016).

Adapun salah satu jenis dari teknik *ensemble* adalah *Voting Ensemble*. *Voting Ensemble* adalah salah satu teknik *ensemble learning* dalam *machine learning*, dimana beberapa model yang berbeda dijalankan pada *dataset* yang sama dan hasil

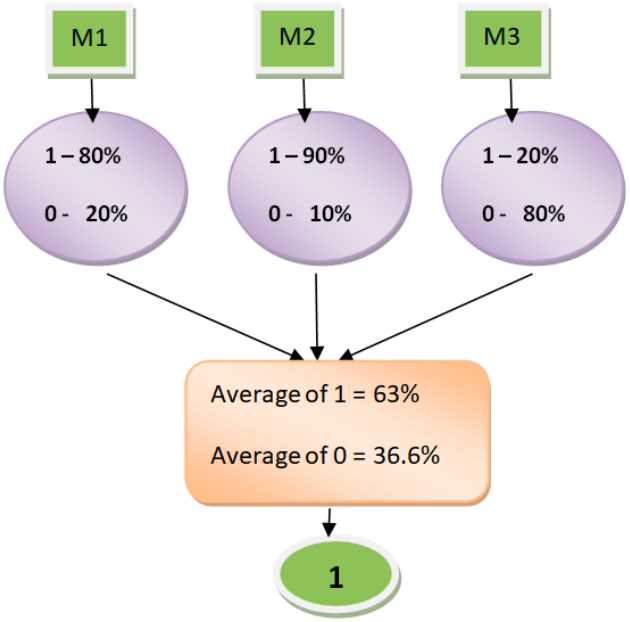
prediksi dari masing-masing model digabungkan untuk menghasilkan prediksi akhir yang lebih akurat dan konsisten. Ada dua jenis *Voting Ensemble* yang cukup populer, yaitu *Hard Voting* dan *Soft Voting* (Peppes *et al*., 2021). *Hard Voting* bekerja sangat sederhana dengan menggabungkan prediksi setiap model klasifikasi dan mengambil kelas prediksi dengan suara terbanyak (Géron, 2019). Sedangkan *Soft Voting* adalah metode yang lebih kompleks yang memperhitungkan probabilitas setiap prediksi oleh masing-masing model klasifikasi dasar (Peppes *et al*., 2021).

Voting Ensemble sangat berguna dalam situasi dimana terdapat beberapa model yang berbeda dengan hasil prediksi yang berbeda pula. Hal ini dapat mengurangi risiko terhadap satu model yang membuat prediksi yang tidak akurat dikarenakan ada model lain yang dapat membuat prediksi yang akurat (Aashish, 2021). Dalam kasus ini, menggunakan *ensemble learning* dapat meningkatkan akurasi dan keandalan dari model dengan memperhitungkan probabilitas setiap prediksi dari beberapa model klasifikasi yang digunakan.

2.9 *Soft Voting Classifier*

Soft Voting adalah metode pemungutan suara dalam pemilihan model atau algoritma dimana setiap model memberikan "suara" berdasarkan probabilitas prediksi kelas yang dihasilkan untuk setiap contoh pada *dataset*. Dalam *Soft Voting*, model dengan probabilitas prediksi rata-rata tertinggi dianggap sebagai pemenang dan dipilih untuk melakukan prediksi pada contoh baru. *Soft Voting* memungkinkan model yang lebih unggul dalam hal prediksi probabilitas untuk memiliki pengaruh yang lebih besar pada hasil akhir.

Soft Voting Classifier merupakan sebuah pendekatan *ensemble learning* yang digunakan untuk menggabungkan hasil prediksi dari beberapa model pembelajaran mesin yang berbeda dalam proses klasifikasi. Dalam metode klasifikasi ini, nilai rata-rata probabilitas dari setiap kelas yang diberikan oleh setiap model akan digunakan untuk menentukan kelas dengan rata-rata probabilitas prediksi tertinggi (Oliveira *et al, 2022*).



Gambar 9 *Ensemble Soft Voting* (Kumar, 2020)

Dalam konsep *Soft Voting Classifier*, setiap model menghasilkan nilai probabilitas atau skor prediksi pada setiap kelas, dan juga setiap model akan diberikan sebuah bobot. Bobot yang diberikan pada setiap pengklasifikasi dapat diterapkan dengan tepat berdasarkan persamaan 10 (Kumar, 2020).

$$\hat{y} = arg\ max_i \sum_{j=1}^{m} w_j p_{ij} \tag{10}$$

Nilai probabilitas atau skor prediksi ini akan dijumlahkan untuk setiap kelas dan diambil nilai rata-ratanya, kemudian dikalikan dengan bobot yang dihasilkan untuk mendapatkan nilai akhir dari probabilitas atau skor untuk setiap kelas. Selanjutnya, kelas dengan nilai probabilitas atau skor paling tinggi akan diambil sebagai *output* prediksi akhir.

Kelebihan dari penggunaan *Soft Voting Classifier* adalah kemampuannya dalam menghasilkan prediksi yang lebih stabil dan akurat dibandingkan dengan hanya menggunakan satu model saja. Hal ini dikarenakan *Soft Voting Classifier* memanfaatkan kemampuan dari setiap model dalam memprediksi dan mampu menyeimbangkan kekurangan yang dimiliki oleh setiap model.

Soft Voting lebih efektif daripada *Hard Voting* karena memperhitungkan probabilitas prediksi kelas yang dihasilkan oleh setiap model dan memperhitungkan bobot dari setiap model. Hal ini memungkinkan model yang lebih baik dalam hal prediksi probabilitas untuk memiliki pengaruh yang lebih besar pada hasil akhir dan dapat menghasilkan kinerja yang lebih baik dalam banyak kasus.

2.10 *Confusion Matrix*

Confusion Matrix merupakan tabel pencatat hasil kerja klasifikasi. *Confusion Matrix* berisi informasi tentang klasifikasi yang dapat diprediksi dengan sistem klasifikasi. Kinerja sistem umumnya dievaluasi menggunakan data dalam bentuk matriks (Santra & Christy, 2012). Melalui *Confusion Matrix* keakuratan, tingkat kesalahan, ketepatan dan nilai penarikan dapat diketahui. Kuantitas *Confusion Matrix* dapat diringkas menjadi 2 nilai, yaitu akurasi dan laju *error*. Dengan mengetahui jumlah data yang diklasifikasikan dengan benar, dapat diketahui akurasi hasil prediksi dan dengan mengetahui jumlah data yang diklasifikasikan secara salah, dapat diketahui laju *error* dari prediksi yang dilakukan.

Pada pengukuran kinerja menggunakan *Confusion Matrix*, terdapat 4 istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. Nilai *True Negative (TN)* merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan *False Positive (FP)* merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, *True Positive (TP)* merupakan data positif yang terdeteksi benar. *False Negative (FN)* merupakan kebalikan dari *True Positive*, sehingga data positif, namun terdeteksi sebagai data negatif (Saifullah, 2019).

Berdasarkan nilai *True Negative (TN)*, *False Positive (FP)*, *False Negative (FN)*, dan *True Positive (TP)* dapat diperoleh nilai akurasi, presisi, *error* dan *recall*. Namun untuk penelitian ini hanya akan membahas tentang akurasi dan *error* pada sistem. Akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasi data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan persamaan 11 (Saifullah, 2019).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} * 100\% \quad (11)$$

Adapun, *Precision* adalah perbandingan nilai antara *True Positive* (TP) dengan banyaknya data yang **diprediksi positif**. Dalam kata lain, *precision* memberi kita informasi tentang seberapa akurat model kita dalam memprediksi kelas positif, dituliskan dengan persamaan 12.

$$Precision = \frac{TP}{TP + FP} * 100\% \quad (12)$$

Selanjutnya, *Recall* adalah perbandingan nilai antara *True Positive (TP)* dengan banyaknya data yang **sebenarnya positif**. Dalam kata lain, *recall* memberi kita informasi tentang seberapa akurat model kita dalam menemukan kelas positif, dituliskan dengan persamaan 13.

$$Recall = \frac{TP}{TP + FN} * 100\% \quad (13)$$

Diketahui, jika nilai salah satu dari *Precision* atau *Recall* tinggi, maka nilai salah satu yang lain cenderung lebih rendah. Sebagai contoh jika nilai *Precision* sangat tinggi atau mendekati skor 1.0, maka nilai dari persamaan *Recall* akan cenderung lebih rendah. Jika hanya menggunakan 2 persamaan ini, dapat terjadi bias pada kondisi dunia nyata. Oleh sebab itu tidak boleh hanya menggunakan skor ini untuk melakukan pemilihan model. Untuk mengatasi hal ini, diperlukan persamaan *F1-Score* yang dituliskan sebagai persamaan 14.

$$F1\ Score = 2x\ \frac{Precision\ x\ Recall}{Precision +\ Recall} * 100\% \quad (14)$$

Nilai yang didapatkan kemudian menggunakan rumus *F1-Score* adalah *harmonic mean* dari *precision dan recall*. *F1-Score* adalah perbandingan antara nilai rata-rata *precision* dan nilai *recall* yang berguna untuk menyeimbangkan *precision dan recall*. Nilai terbaik dari *F1-Score* adalah 1.0 dan nilai terburuk adalah 0. Secara representasi, jika *F1-Score* punya skor yang baik mengindikasikan bahwa model klasifikasi memiliki *precision* dan *recall* yang baik pula (Saifullah, 2019).

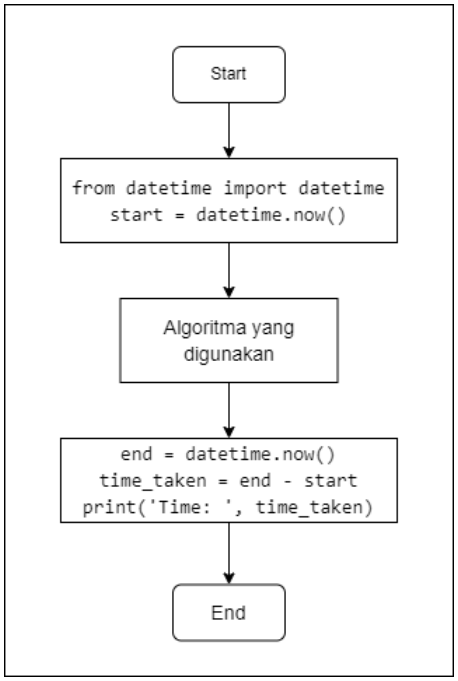
2.11 Waktu Komputasi

Waktu komputasi pada *Data Mining* merujuk pada waktu yang dibutuhkan untuk mengeksekusi tugas-tugas pemrosesan data, seperti pengolahan data, pelatihan model, evaluasi model, dan sebagainya. Waktu komputasi dapat bervariasi tergantung pada kompleksitas tugas, ukuran kumpulan data, sumber daya perangkat keras dan perangkat lunak yang tersedia, serta teknik pemrosesan data yang digunakan. Waktu komputasi digunakan untuk menghitung berapa lama sebuah perintah atau algoritma diselesaikan oleh komputer (Dedi Gunawan, 2016).

Penggunaan algoritma yang kompleks dan data yang sangat besar dapat memperpanjang waktu komputasi dan membutuhkan sumber daya perangkat keras yang lebih tinggi. Selain itu, waktu komputasi dapat dipengaruhi oleh teknik pemrosesan data yang digunakan, seperti pemrosesan paralel dan distribusi, serta penggunaan perangkat keras yang dioptimalkan untuk pemrosesan data tertentu, seperti unit pemrosesan grafis (GPU) dan teknologi *grid computing*.

Oleh karena itu, waktu komputasi merupakan faktor penting yang perlu dipertimbangkan dalam merancang dan melaksanakan tugas-tugas *Data Mining*. Peningkatan kinerja komputasi dapat dicapai melalui pengoptimalan algoritma dan penggunaan sumber daya perangkat keras dan perangkat lunak yang optimal untuk tugas-tugas *Data Mining* yang spesifik.

Pada penelitian ini waktu komputasi digunakan sebagai perbandingan seberapa efisien algoritma dalam memprediksi pneumonia. Gambar 10 menunjukkan *flowchart* dari waktu komputasi.



Gambar 10 *Flowchart* Waktu Komputasi

Pada gambar 10 menunjukkan alur proses waktu komputasi, dimana ketika program pertama kali dijalankan perintah yang dilakukan adalah memasukkan waktu mulai kedalam variabel *start. Start* berfungsi untuk memulai perhitungan waktu komputasi dari algoritma yang akan digunakan. Selanjutnya, program mulai melakukan komputasi terhadap algoritma yang telah ditetapkan. Adapun proses komputasi setiap algoritma dilakukan secara bertahap. Pada penelitian ini ada empat algoritma yang digunakan, antara lain *K-Nearest Neighbor*, *Support Vector Machine, Decision Tree* dan *Soft Voting*. Setelah program selesai melakukan komputasi terhadap algoritma yang telah ditetapkan, maka program akan mencatat waktu selesai yang disimpan pada variabel *end,* yang selanjutnya akan digunakan untuk mengukur selisih waktu dari proses komputasi algoritma.

Waktu komputasi menjadi penting, terutama dalam aplikasi *real-time* seperti deteksi penipuan, sistem rekomendasi, atau aplikasi lainnya yang memerlukan respons cepat. Dalam kasus seperti ini, efisiensi waktu dari algoritma menjadi faktor kunci dalam memilih algoritma yang tepat. Oleh karena itu, penting untuk mempertimbangkan keseimbangan antara waktu komputasi dan akurasi model saat memilih algoritma untuk aplikasi klasifikasi data tertentu.