

**ANALISIS PERBANDINGAN ALGORITMA NAÏVE BAYES DAN
SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASI
KELANGSUNGAN PEMBERIAN ASI EKSLUSIF
DI INDONESIA**

*COMPARATIVE ANALYSIS OF THE NAÏVE BAYES AND SUPPORT
MACHINE LEARNING ALGORITHM IN CLASSIFYING THE CONTINUITY
OF EXCLUSIVE BREASTFEEDING IN INDONESIA*

MUH ZARKAWI YAHYA



**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023**

**ANALISIS PERBANDINGAN ALGORITMA NAÏVE BAYES DAN
SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASI
KELANGSUNGAN PEMBERIAN ASI EKSLUSIF
DI INDONESIA**

Tesis

sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

Muh Zarkawi Yahya

H062211004

Kepada

**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2023**

TESIS

ANALISIS PERBANDINGAN ALGORITMA NAÏVE BAYES DAN SUPPORT VECTOR MACHINE DALAM MENGLASIFIKASI KELANGSUNGAN PEMBERIAN ASI EKSKLUSIF DI INDONESIA

MUH ZARKAWI YAHYA

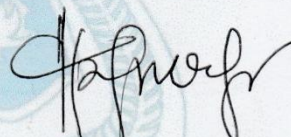
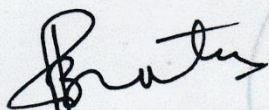
H062211004

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin pada tanggal 18 Agustus 2023 dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping



Sri Astuti Thamrin, S.Si., M.Stat., Ph.D.
NIP. 19740713 199903 2 001

Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Ketua Program Studi
Magister Statistika,

Dekan Fakultas Matematika dan
Ilmu Pengetahuan Alam
Universitas Hasanuddin



Dr. Dr. Georgina M. Tinungki, M.Si.
NIP. 19620926 198702 2 001



Dr. Eng. Amiruddin, S.Si., M.Si.
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis berjudul “Analisis Perbandingan Algoritma Naïve Bayes dan Support Vector Machine dalam Mengklasifikasi Kelangsungan Pemberian ASI Eksklusif di Indonesia” adalah benar karya saya dengan arahan dari komisi pembimbing (Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. dan Dr. Erna Tri Herdiani, S.Si., M.Si.). Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari isi tesis ini telah dipublikasikan di Jurnal (*International Journal of Research Publications*, Volume 130, Halaman 25–32, dan DOI: <https://doi.org/10.47119/IJRP1001301820235301>) sebagai artikel dengan judul “*Application of Naïve Bayes and Support Vector Machine Algorithms to Classify Exclusive Breastfeeding in Indonesia*”.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 18 Agustus 2023



Muh Zarkawi Yahya
NIM. H062211004

UCAPAN TERIMA KASIH

Puji dan syukur ke hadirat Allah karena atas pertolongan-Nyalah sehingga tesis ini dapat diselesaikan sebagai salah satu syarat untuk menyelesaikan studi pada program Magister di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin. Tidak lupa pula penulis mengirimkan shalawat serta salam kepada bimbingan kita, Rasulullah Muhammad ﷺ yang telah memberikan banyak pedoman kehidupan bagi manusia.

Dalam menyelesaikan tesis ini, penulis tentunya tidak luput dari berbagai kekeliruan dan hambatan. Namun, pada akhirnya tesis ini dapat diselesaikan berkat adanya kontribusi dari berbagai pihak yang selalu memberikan motivasi, bantuan, dan bimbingan bagi penulis. Oleh karena itu, disamping rasa syukur ke hadirat Allah, penulis juga mengucapkan terima kasih dan rasa hormat kepada kedua orang tua penulis, Ayahanda Dr. H. Muhammad Yahya, M.Ag dan Ibunda Hj. Kasturi, S.St yang telah menjadi orang terhebat dalam membesarkan dan mendidik penulis. Hanya doalah balasan terbaik yang dapat penulis sampaikan. Selanjutnya, ucapan terima kasih juga penulis sampaikan kepada:

1. Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc. selaku rektor Universitas Hasanudin.
2. Bapak Dr. Eng. Amiruddin, S.Si., M.Si. selaku dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.
3. Ibu Dr. Anna Islamiyati, S.Si., M.Si. selaku ketua Departemen Statistika sekaligus dosen pada program studi magister Statistika. Terima kasih atas perhatian yang diberikan kepada kami mahasiswa pascasarjana, juga atas saran untuk perbaikan tesis penulis.
4. Ibu Dr. Dr. Georgina Maria Tinungki, M.Si. selaku ketua program studi Magister Statistika sekaligus selaku tim penguji tesis penulis. Terima kasih atas perhatian yang diberikan kepada kami mahasiswa pascasarjana, juga atas saran untuk perbaikan tesis penulis.
5. Ibu Sri Astuti Thamrin, S.Si., M.Stat., Ph.D. selaku pembimbing utama penulis. Terima kasih atas segala kebaikan dan perhatian yang ditujukan kepada penulis, juga kesediaan dan kesabaran meluangkan waktu, tenaga dan pikiran untuk membimbing penulis sehingga tesis ini dapat diselesaikan.

6. Ibu Dr. Erna Tri Herdiani, S.Si., M.Si selaku pembimbing pertama penulis. Terima kasih atas kesediaan dan kesabaran meluangkan waktu untuk membimbing penulis, juga arahan untuk perbaikan tesis penulis.
7. Ibu Dr. Nurtiti Sunusi, S.Si., M.Si selaku tim penguji tesis penulis. Terima kasih atas saran dan arahan yang diberikan kepada penulis dalam menyelesaikan tesis ini.
8. Staff di Departemen Statistika FMIPA Unhas. Terima kasih atas bantuan terkait dengan persuratan yang diberikan kepada kami mahasiswa pascasarjana.
9. Rand Corporation yang menyediakan data secara *online* sehingga memudahkan penulis untuk menyelesaikan tesis ini.
10. Isra' Khusunul Khatimah Salahuddin istriku tersayang, yang selalu memberi semangat dan dukungan serta doa dalam menyelesaikan tesis ini.
11. Teman-teman mahasiswa program Magister Statistika angkatan empat: Adik Nurul, saudari Widia, Kak Muammar, dan Kakak Nunu. Sebuah kehormatan bisa diberikan kesempatan untuk belajar bersama.
12. Teman-teman mahasiswa program Magister Statistika lainnya: Misbah, Ardi, Amel, Arya, Mila, Samsir, Qardawi, Andis, Kak Gio, Kak Imma, Kak Tisen, Kak Gery, Kak Nunu, Kak Mumun, Kak Nurul, Kak Mita, Kak Alfi, kak muhsinil, dan lainnya yang penulis tidak bisa sebutkan satu persatu.

Kepada seluruh pihak yang tidak sempat disebutkan atas keikutsertaan membantu menyelesaikan tesis ini, penulis ucapkan terima kasih. Semoga bantuan yang diberikan kepada penulis mendapatkan balasan dari Allah berupa kebaikan. Semoga tesis ini dapat bermanfaat bagi pembaca.

Penulis,

Muh Zarkawi Yahya

ABSTRAK

MUH ZARKAWI YAHYA. **Analisis Perbandingan Algoritma Naïve Bayes dan Support Vector Machine dalam Mengklasifikasi Kelangsungan Pemberian ASI Eksklusif di Indonesia** (dibimbing oleh Sri Astuti Thamrin dan Erna Tri Herdiani).

Badan kesehatan *World Health Organization* (WHO) dan *United Nations International Children's Emergency Fund* (UNICEF) merekomendasikan inisiasi menyusukan dini dalam waktu 1 jam dari kelahiran bayi. Pemerintah Indonesia mendukung program WHO dan UNICEF dengan menyediakan program pemberian ASI Eksklusif pada tahun 2012. Air Susu Ibu (ASI) eksklusif diinisiasi sejak 6 bulan pertama kehidupan dilanjutkan pengenalan nutrisi yang memadai dan aman komplementer (padat) dengan terus menyusui sampai 2 tahun atau lebih. Faktor-faktor yang mempengaruhi keberhasilan program eksklusif terdiri dari 9 faktor. Oleh karena itu, akan disediakan factor apa yang berpengaruh terhadap keberhasilan pemberian ASI eksklusif. Metode yang digunakan adalah algoritma Naïve Bayes dan Support Vector Machine (SVM) dengan menggunakan JASP. Hasilnya menunjukkan bahwa SVM memiliki nilai akurasi yaitu 87.6%, yang lebih baik dibandingkan dengan metode Naïve Bayes yaitu 86.2%, pada data ASI eksklusif di Indonesia.

Kata Kunci: Naïve Bayes, Support Vector Machine, ASI Eksklusif

ABSTRACT

MUH ZARKAWI YAHYA. **Application of Naïve Bayes and Support Vector Machine Algorithms to Classify Exclusive Breastfeeding in Indonesia** (supervised by Sri Astuti Thamrin and Erna Tri Herdiani).

The World Health Organization (WHO) and the United Nations International Children's Emergency Fund (UNICEF) both advocate for initiation of breastfeeding within one hour of a baby's birth. The Indonesian government supported the WHO and UNICEF programs by implementing exclusive breastfeeding programs in 2012. Exclusive breastfeeding is recommended for the first 6 months of a child's life, followed by the introduction of appropriate and safe complementary nutrition while continuing breastfeeding for up to 2 years or more. There are 9 factors that can influence the success of the exclusive breastfeeding program, and these factors will be taken into consideration. To analyze the impact of these factors, the Naive Bayes algorithm and Support Vector Machine (SVM) using JASP were employed. The results indicate that SVM achieved a higher accuracy rate of 87.6% compared to the Naive Bayes method, which yielded an accuracy rate of 86.2%, when applied to exclusive breastfeeding data in Indonesia.

Keywords: Naïve Bayes, Support Vector Machine, Exclusive Breastfeeding.

DAFTAR ISI

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA.....	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK.....	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Defenisi ASI Eksklusfi.....	5
2.2 Manfaat Pemberian ASI	7
2.3 Klasifikasi Pemberian ASI Eksklusif	8
2.4 Naïve Bayes.....	9
2.5 Support Vector Machine (SVM)	11
2.6 Preprocessing	14
2.7 Split Data	14
2.8 Confusion Matrix.....	15
BAB III METODE PENELITIAN.....	16
3.1 Data dan Sumber Data.....	16
3.2 Variabel Penelitian	16
3.3 Metode Analisis	17
3.4 Diagram alir penelitian.....	18
BAB IV HASIL DAN PEMBAHASAN.....	19
4.1 Deskripsi Data.....	19
4.2 Preprocessing	19
4.3 Split Data	20
4.4 Klasifikasi.....	20

4.4.1 Klasifikasi Support Vector Machine	20
4.3.1 Klasifikasi Metode Naïve Bayes.....	27
4.5 Pembahasan	33
4.5.1 Hasil Pengujian Klasifikasi dengan Support Vector Machine	33
4.5.2 Hasil Pengujian Klasifikasi dengan Naïve Bayes	33
4.5.3 Hasil Perbandingan SVM dan Naïve Bayes.....	34
BAB V KESIMPULAN DAN SARAN.....	36
5.1 Kesimpulan.....	36
5.2 Saran.....	36
DAFTAR PUSTAKA.....	37
LAMPIRAN	40

DAFTAR TABEL

Tabel 4. 1 Data Sebelum Melakukan Preprocessing	19
Tabel 4. 2 Data Setelah melakukan preprocessing	20
Tabel 4. 3 Skenario Split Data	20
Tabel 4. 4 Skenario Pengujian Pada Support Vector Machine	21
Tabel 4. 5 Confusion Matrix SVM Pengujian 1	22
Tabel 4. 6 Accuracy, Precision, Recall pada SVM Pengujian 1	22
Tabel 4. 7 Confusion Matrix SVM Pengujian 2	22
Tabel 4. 8 Accuracy, Precision, Recall pada SVM Pengujian 2	22
Tabel 4. 9 Confusion Matrix SVM Pengujian 3	23
Tabel 4. 10 Accuracy, Precision, Recall pada SVM Pengujian 3	23
Tabel 4. 11 Confusion Matrix SVM Pengujian 4	23
Tabel 4. 12 Accuracy, Precision, Recall pada SVM Pengujian 4	24
Tabel 4. 13 Confusion Matrix SVM Pengujian 5	24
Tabel 4. 14 Accuracy, Precision, Recall pada SVM Pengujian 5	24
Tabel 4. 15 Confusion Matrix SVM Pengujian 6	25
Tabel 4. 16 Accuracy, Precision, Recall pada SVM Pengujian 6	25
Tabel 4. 17 Confusion Matrix SVM Pengujian 7	25
Tabel 4. 18 Accuracy, Precision, Recall pada SVM Pengujian 7	25
Tabel 4. 19 Confusion Matrix SVM Pengujian 8	26
Tabel 4. 20 Accuracy, Precision, Recall pada SVM Pengujian 8	26
Tabel 4. 21 Nilai Rata-Rata Accuracy, Precision, dan Recall pada SVM	26
Tabel 4. 22 Skenario Pengujian Pada Naive Bayes	27
Tabel 4. 23 Confusion Matrix Naive Bayes Pengujian 1	28
Tabel 4. 24 Accuracy, Precision, Recall pada SVM Pengujian 1	28
Tabel 4. 25 Confusion Matrix Naive Bayes Pengujian 2	28
Tabel 4. 26 Accuracy, Precision, Recall pada Naive Bayes Pengujian 2	29
Tabel 4. 27 Confusion Matrix Naive Bayes Pengujian 3	29
Tabel 4. 28 Accuracy, Precision, Recall pada Naive Bayes Pengujian 3	29
Tabel 4. 29 Confusion Matrix Naive Bayes Pengujian 4	29
Tabel 4. 30 Accuracy, Precision, Recall pada Naive Bayes Pengujian 4	30
Tabel 4. 31 Confusion Matrix Naive Bayes Pengujian 5	30
Tabel 4. 32 Accuracy, Precision, Recall pada Naive Bayes Pengujian 5	30
Tabel 4. 33 Confusion Matrix Naive Bayes Pengujian 6	31
Tabel 4. 34 Accuracy, Precision, Recall pada Naive Bayes Pengujian 6	31
Tabel 4. 35 Confusion Matrix Naive Bayes Pengujian 7	31
Tabel 4. 36 Accuracy, Precision, Recall pada Naive Bayes Pengujian 7	31
Tabel 4. 37 Confusion Matrix Naive Bayes Pengujian 8	32
Tabel 4. 38 Accuracy, Precision, Recall pada Naive Bayes Pengujian 8	32
Tabel 4. 39 Nilai Rata- rata Accuracy, Precision, Recall pada Naive Bayes	32
Tabel 4. 40 Hasil Perbandingan Pengujian SVM dan Naive Bayes	34

DAFTAR GAMBAR

Gambar 2. 1 Cakupan Pemberian ASI Eksklusif Pada Bayi 0-6 Bulan Menurut Provinsi Tahun 2021	6
Gambar 2. 2 Klasifikasi Bayes	11
Gambar 2. 3 Hyperplane.....	14
Gambar 3. 1 Diagram Alir Penelitian.....	18
Gambar 4. 1 Grafik Perbandingan Akurasi, Presisi, Recall.....	35

DAFTAR LAMPIRAN

Lampiran 1 Head (100) Data Mentah Indonesia Family Life Survey Gelombang 5 (IFLS-5) di Indonesia	41
Lampiran 2 Head (100) Data Hasil Preprocessing Indonesia Family Life Survey Gelombang 5 (IFLS-5) di Indonesia	44

BAB I

PENDAHULUAN

1.1 Latar Belakang

ASI mengandung nutrisi alamiah terbaik bagi bayi karena mengandung kebutuhan energi dan zat yang dibutuhkan selama enam bulan pertama kehidupan bayi (Hasniati et al., 2015). Badan kesehatan World Health Organization (WHO) dan United Nations International Children's Emergency Fund (UNICEF) merekomendasikan inisiasi menyusui dini dalam waktu 1 jam dari lahir ASI eksklusif selama 6 bulan pertama kehidupan dan pengenalan nutrisi yang memadai dan aman komplementer (padat) makanan pada 6 bulan bersama dengan terus menyusui sampai 2 tahun atau lebih. Namun, banyak bayi dan anak-anak tidak menerima makan optimal, di mana hanya sekitar 36% dari bayi usia 0 sampai 6 bulan di seluruh dunia yang diberikan ASI eksklusif selama periode tahun 2007 sampai dengan tahun 2014 (WHO, 2017).

WHO (2017) melaporkan bahwa secara global rata-rata angka pemberian ASI eksklusif di dunia pada tahun 2017 hanya sebesar 38%, WHO menargetkan pada tahun 2025 angka pemberian ASI eksklusif pada usia 6 bulan pertama kelahiran meningkat setidaknya 50%. Kementerian kesehatan Indonesia tahun 2017 melaporkan perempuan di Indonesia 96% menyusui anak mereka namun hanya 42% yang memberikan ASI eksklusif selama 6 bulan (RISKESDAS, 2018).

ASI eksklusif di Indonesia yang diharapkan belum mencapai target yaitu sebesar 80%. (RISKESDAS, 2018) melaporkan di Indonesia proporsi pemberian ASI pada bayi dan anak usia 0 sampai 5 bulan sebesar 37,3%. Upaya pemerintah untuk melindungi, mendukung dan mempromosikan pemberian ASI eksklusif maka PP Nomor 33 tahun 2012 tentang pemberian ASI eksklusif 6 bulan, peraturan ini melaksanakan ketentuan pasal 129 ayat (2) UU Nomor 36 Tahun 2009 tentang kesehatan.

Permasalahan pokok terkait masih rendahnya penggunaan ASI eksklusif di Indonesia adalah faktor kurangnya pengetahuan ibu hamil, sosial budaya, keluarga dan masyarakat akan pentingnya ASI eksklusif, serta instansi kesehatan yang belum sepenuhnya mendukung peningkatan pemberian ASI eksklusif (Prabhakara, 2010).

Penerapan data mining sangat di perlukan dalam bidang kesehatan sebagai inovasi dalam mengambil keputusan yang akurat sehingga bermanfaat untuk pengembangan informasi diberikan menjadi lebih akurat. (Wilson & Haridas, 2018) Data mining adalah teknik yang digunakan untuk membangun machine learning. Machine learning adalah teknik kecerdasan buatan modern yang belajar membangun model dengan menggunakan data empiris. (Wilson & Haridas, 2018)

Untuk mengatasi masalah di atas, maka diperlukan suatu sistem dalam pengolahan data seperti klasifikasi, prediktif, dan klustering yang masuk didalam Supervised Learning (sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada. (Huddleston & Brown, 2018)

Klasifikasi adalah proses untuk menemukan model atau fungsi yang dapat menggambarkan dan membedakan kelas data atau konsep, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas yang belum diketahui dari suatu objek pengamatan (J. Han & Pei, 2012). Klasifikasi digunakan untuk mengelompokkan data ke dalam beberapa kategori agar lebih mudah untuk diolah dan dianalisis (Wibawa et al., 2018). Metode klasifikasi pada machine learning yang sering digunakan yaitu Regression Trees (CART), Random Forest, Naïve Bayes, Support Vector Machines (SVM), KNN dan Neural Networks (Sihombing & Yuliati, 2021).

Salah satu permasalahan dalam klasifikasi data adalah banyaknya data yang tidak seimbang antara kelas yang berbeda, dan ketika kondisi ketidakseimbangan ekstrim, masalah ini disebut rare event atau imbalanced data (Maalouf & Siddiqi, 2014). Meskipun banyak metode klasifikasi yang baik digunakan untuk klasifikasi data, namun sayangnya belum tentu metode tersebut tepat digunakan pada kondisi imbalanced data, karena metode tersebut didasarkan pada asumsi bahwa banyaknya data terdistribusi secara merata antara kelas yang berbeda.

(King & Zeng, 2017) menyatakan bahwa ketika metode klasifikasi digunakan pada kasus imbalanced data, maka pengklasifikasian cenderung menihilkan peluang dari kelas minoritas karena nilai prediksi cenderung pada kategori yang mayoritas, sehingga tingkat ketepatan klasifikasi yang dihasilkan menjadi kurang baik. Hal ini terjadi terutama untuk set data besar.

Penjelasan ini juga diperkuat dengan adanya beberapa penelitian yang telah menerapkan pendekatan machine learning antara lain Harahap et al., (2021) pada penelitiannya yaitu klasifikasi diagnosa penyakit jantung menggunakan algoritma Random Forest memperoleh hasil akurasi dengan menggunakan Bahasa pemrograman Python adalah sebesar 85,3%. Penelitian lain juga dilakan Hermawan et al., (2022), dalam penelitiannya menggunakan metode algoritma CART (Classification And Regression Tree) untuk memprediksi awal penyakit stroke berdasarkan rekam medis, dengan memperoleh hasil akurasi terbesar sebanyak 74,73%. Kemudian penelitian yang dilakukan oleh (Riany & Testiana, 2022) Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes memperoleh hasil dalam mengklasifikasi data penyakit stroke menghasilkan nilai akurasi sebesar 92,48%. (Wijaya, 2022) dalam mengklasifikasi penderita diabetes melitus menggunakan metode Support Vector Machine (SVM) menghasilkan nilai akurasi sebesar 96%. (Sumarlinda & Lestari, 2022) dalam penelitiannya Aplikasi K-Nearest Neighbor (KNN) untuk klasifikasi penyakit kardiovaskuler menghasilkan performansi dengan akurasi sebesar 75,75%, nilai presisinya adalah 76,78%, sedangkan recall menghasilkan 77,14%.

Berdasarkan hasil penelitian terdahulu, metode naïve bayes dan SVM memiliki nilai akurasi yang tinggi dibandingkan klasifikasi lainnya. Oleh karena itu, pada thesis ini kami membahas perbandingan metode naïve bayes dan SVM. Adapun metode tersebut digunakan dalam mengklasifikasi kelangsungan pemberian ASI eksklusif di Indonesia.

1.2 Rumusan Masalah

Dari latar belakang masalah yang telah dijabarkan di atas, maka didapatkan rumusan masalah yaitu “Bagaimana perbandingan antara algoritma Naïve Bayes dan SVM dalam mengklasifikasi kelangsungan pemberian ASI eksklusif di Indonesia?”

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah “Membandingkan algoritma Naïve Bayes dan SVM dalam mengklasifikasi kelangsungan pemberian ASI eksklusif di Indonesia”

1.4 Batasan Masalah

Dalam penelitian ini permasalahan dibatasi pada perbandingan hasil keakuratan klasifikasi menggunakan metode *Naïve Bayes* dan *Support Vector*

Machine. Data yang digunakan dalam penelitian ini adalah data hasil *Indonesia Family Life Survey* Gelombang 5 (IFLS-5) yang diperoleh dari <https://www.rand.org>.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

1.5.1 Manfaat Teoritis

Hasil penelitian ini diharapkan menambah Khasanah keilmuan mengenai Klasifikasi Kelangsungan Pemberian ASI Eksklusif di Indonesia dan memudahkan bagi peneliti lain dalam mempelajari *Naïve Bayes* dan SVM untuk bidang Statistik.

1.5.2 Manfaat Praktis

- a. Bagi tenaga kesehatan, dapat menggunakan metode *Naïve Bayes* dan SVM dalam mengklasifikasi upaya untuk meningkatkan cakupan pemberian ASI eksklusif di Indonesia
- b. Bagi peneliti selanjutnya dapat dijadikan sebagai informasi awal dalam melakukan penelitian selanjutnya.

1.5.3 Manfaat kebijakan

Hasil penelitian ini diharapkan dapat dijadikan tolak ukur tentang faktor-faktor dalam pemberian ASI eksklusif saat ini sehingga dapat diambil langkah kebijakan strategis agar dapat meningkatkan pengetahuan ibu dalam pemberian ASI eksklusif.

BAB II

TINJAUAN PUSTAKA

Pada bagian ini, dibahas tentang konsep teoritis yang digunakan pada bagian hasil dan pembahasan.

2.1 Defenisi ASI Eksklusfi

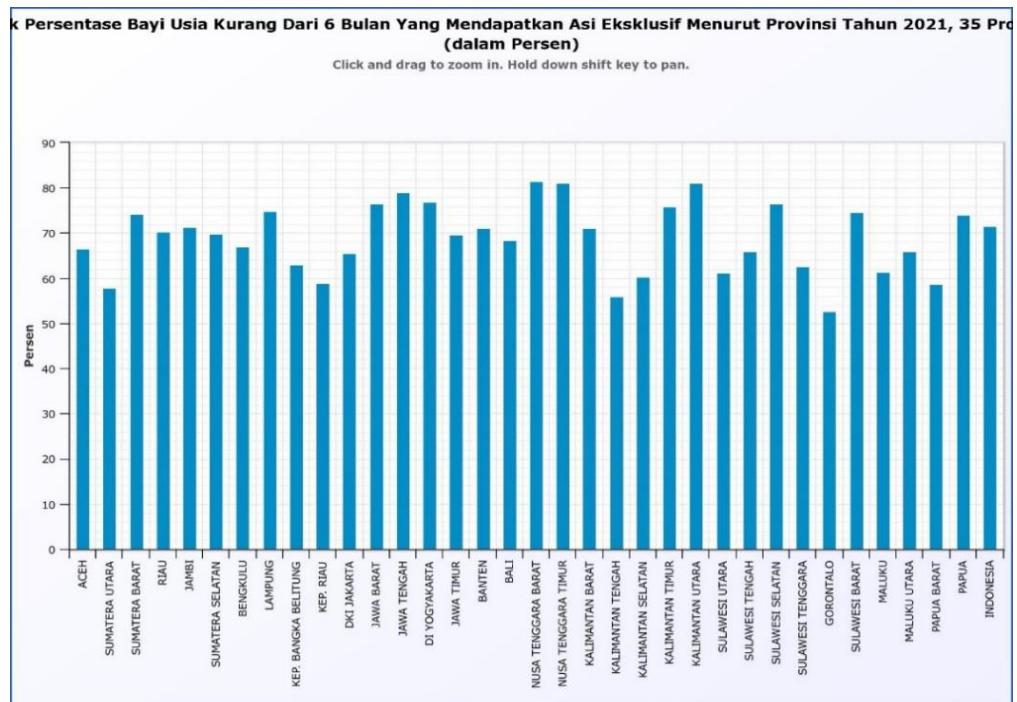
ASI Eksklusif adalah tidak memberi bayi makanan atau minuman lain, termasuk air putih, selain menyusui selama 6 bulan sejak dilahirkan (Infodatin, 2014). Cara pemberian makanan pada bayi yang baik dan benar adalah menyusui bayi secara eksklusif sejak lahir sampai dengan umur 6 bulan dan meneruskan menyusui anak sampai umur 24 bulan. Mulai umur 6 bulan, bayi mendapat makanan pendamping ASI yang bergizi sesuai dengan kebutuhan tumbuh kembangnya (Infodatin, 2014).

ASI merupakan makanan terbaik untuk bayi yang mengandung sel darah putih, protein dan zat kekebalan yang cocok untuk bayi. ASI membantu pertumbuhan dan perkembangan anak secara optimal serta melindungi terhadap penyakit. Bayi diperkenankan untuk diberi makanan tambahan pendamping ASI setelah berusia 6 bulan serta ASI tetap diberikan sampai usia 2 tahun (Sri Purwanti, 2004).

Dari penjelasan di atas sangat berkaitan dengan Peraturan Pemerintah Republik Indonesia Nomor 33 Tahun 2012 tentang pemberian ASI eksklusif Pasal 6 berbunyi “Setiap Ibu yang melahirkan harus memberikan ASI eksklusif kepada bayi yang dilahirkannya”.

Keputusan Menteri Kesehatan (2004) tentang Pemberian ASI secara eksklusif di Indonesia:

- a. Menetapkan ASI eksklusif di Indonesia selama 6 bulan dan dianjurkan dilanjutkan sampai dengan anak berusia 2 tahun atau lebih dengan pemberian makanan tambahan yang sesuai.
- b. Tenaga kesehatan agar menginformasikan kepada semua ibu yang melahirkan untuk memberikan ASI eksklusif dengan mengacu pada 10 langkah keberhasilan menyusui.



Gambar 2. 1 Cakupan Pemberian ASI Eksklusif Pada Bayi 0-6 Bulan Menurut Provinsi Tahun 2021

ASI adalah sumber asupan nutrisi yang penting untuk bayi. ASI diberikan kepada bayi selama 2 tahun. Sedangkan ASI eksklusif diberikan kepada bayi selama 6 bulan dengan tanpa menambahkan makanan atau minuman lainnya. Organisasi Kesehatan Dunia (WHO) seperti yang dilansir Pusat Data dan Informasi Kementerian Kesehatan Indonesia menyebutkan ASI eksklusif yakni Inisiasi menyusui dini (IMD) pada satu jam pertama setelah lahir (Infodatin, 2014).

Hampir seluruh dunia mengakui bahwa ASI merupakan makanan yang paling tepat bagi bayi, tidak ada satu susu formula buatan manusia mana pun yang sanggup menggantikan semua keunggulan ASI. Bahkan Organisasi Kesehatan Sedunia (WHO) dan UNICEF merekomendasikan ASI Eksklusif merupakan gizi terbaik dan sumber makanan utama yang paling sempurna bagi bayi usia 0-6 bulan.

Berbagai penelitian telah membuktikan bahwa menyusui secara eksklusif selama 6 bulan terbukti memberikan risiko yang lebih kecil terhadap berbagai penyakit infeksi seperti: diare, infeksi saluran napas, infeksi telinga, pneumonia, infeksi saluran kemih dan beberapa penyakit lainnya (IDAI, 2013). Pemenuhan asupan nutrisi yang dibutuhkan oleh bayi pada 6 bulan pertama dapat meningkatkan kekebalan tubuh. Pemberian ASI atau menyusui menurut WHO (2017) adalah:

1. Pemberian ASI eksklusif atau menyusui eksklusif adalah hanya memberikan ASI pada bayi dan tidak memberi bayi makanan atau minuman yang lainnya, termasuk air putih, kecuali obat-obatan dan vitamin atau mineral tetes, serta ASI perah yang sampai bayi berusia 6 bulan.
2. Pemberian ASI eksklusif atau menyusui predominant adalah menyusui bayi, tapi pernah memberikan sedikit air atau minuman berbasis air, misalnya teh (biasanya makanan/minuman prelakteal sebelum ASI keluar).
3. Pemberian ASI eksklusif atau menyusui parsial adalah menyusui bayi serta memberikan makanan buatan selain ASI baik, susu formula, bubur atau makanan lainnya, (baik diberikan secara kontinu maupun diberikan sebagai makanan prealektal).

Bayi sehat pada umumnya tidak memerlukan makanan tambahan sampai Usia 6 bulan. Pada keadaan-keadaan khusus dibenarkan untuk mulai memberikan makanan pada setelah bayi berumur 4 bulan tetapi belum mencapai 6 bulan. Misalnya karena terjadi peningkatan berat badan bayi yang kurang dari standar atau didapatkan tanda-tanda lain yang menunjukkan bahwa pemberian ASI eksklusif tidak berjalan dengan baik.

Selain itu, terlepas dari rekomendasi baru UNICEF, masih ada pihak yang tetap mengusulkan pemberian makanan padat mulai ada usia 4 bulan sesuai dengan isi Deklarasi Innocent 1990, yaitu “hanya diberi ASI sampai bayi berusia 4-6 bulan” (Purwanti, 2004). Namun pengetahuan terakhir tentang efek pemberian makanan padat yang terlalu dini telah cukup menunjang pembaharuan definisi ASI eksklusif menjadi “ASI saja sampai usia sekitar 6 bulan.

2.2 Manfaat Pemberian ASI

1. Membantu ikatan batin antara ibu dengan bayi.

Bayi yang sering berada dalam dekapan ibu karena menyusui akan merasakan kasih sayang ibunya. Ia juga akan merasa aman dan tenteram, terutama karena masih mendengar detak jantung sang ibu yang telah dikenalnya sejak dalam kandungan.

2. Membantu menunda kehamilan baru jika menyusui dilakukan rutin.

Cara ini mengandalkan pemberian ASI pada masa menyusui bayi (pasca persalinan). Selama ibu memberi ASI eksklusif dan belum haid, 98% tidak akan hamil pada 6 bulan pertama setelah melahirkan, dan 96% tidak akan hamil sampai bayi berusia 12 bulan.

3. Melindungi kesehatan ibu.

Menyusui dapat mengurangi risiko pendarahan setelah melahirkan, karena pada saat menyusui kadar Oksitoksin yang berguna juga untuk penutupan pembuluh darah sehingga pendarahan lebih cepat berhenti. Selain itu dapat mengurangi anemia, mengecilkan rahim, lebih cepat langsing, dan mengurangi risiko menderita kanker payudara & indung telur.

4. Biayanya lebih rendah daripada pemberian asupan buatan, apalagi susu formula.

Dengan memberi ASI eksklusif, berarti tidak ada pengeluaran untuk membeli susu formula selama 6 bulan, bahkan sampai 2 tahun. Selain itu karena bayi akan lebih jarang sakit, maka pengeluaran untuk ke dokter atau ke rumah sakit juga akan berkurang.

5. Meningkatkan kecerdasan anak.

Dengan memberikan ASI eksklusif selama 6 bulan, akan menjamin tercapainya pengembangan potensi kecerdasan anak secara optimal. Hal ini karena selain sebagai nutria yang ideal, dengan komposisi yang tepat, serta disesuaikan dengan kebutuhan bayi. ASI juga mengandung nutria khusus yang diperlukan otak.

6. Meningkatkan daya tahan tubuh bayi.

Bayi yang baru lahir secara alamiah mendapat imunoglobulin (zat kekebalan tubuh) dari ibunya melalui ari-ari. Namun kadar zat ini akan cepat sekali menurun segera setelah bayi lahir. Ketika zat kekebalan menurun dan tubuh bayi belum mampu memproduksi banyak zat kekebalan, maka ASI adalah cairan hidup yang mengandung zat kekebalan yang akan melindungi bayi dari berbagai penyakit (Prasetyo, 2012).

2.3 Klasifikasi Pemberian ASI Eksklusif

1. Menyusui Exclusive

Menyusui exclusive tidak memberi bayi makanan atau minuman lain, termasuk air putih, selain menyusui kecuali obat-obatan dan vitamin atau mineral tetes, ASI perah juga diperbolehkan. Menurut RISKESDAS (2018) menyusui eksklusif adalah komposit dari pertanyaan bayi masih disusui, sejak lahir tidak pernah mendapatkan makanan dan minuman selain ASI, selama 24 jam terakhir bayi hanya disusui.

2. Menyusui Predominan

Menyusui bayi tetapi pernah memberikan sedikit air atau minuman berbasis air, misalnya teh, sebagai makanan/minuman prelakteal sebelum ASI keluar. Pada RISKESDAS (2018) menyusui predominan adalah bayi masih disusui, selama 24 jam terakhir bayi hanya disusui, sejak lahir tidak pernah mendapatkan makanan atau minuman kecuali minuman berbasis air, yaitu air putih atau air teh.

3. Menyusui Parsial

Menyusui bayi serta diberikan makanan buatan selain ASI, baik susu formula, bubur atau makanan lainnya sebelum bayi berumur enam bulan, baik diberikan secara kontinu maupun diberikan sebagai makanan prelakteal (RISKESDAS, 2018). Menyusui parsial adalah bayi masih disusui, pernah diberikan makanan prelakteal selain makanan atau minuman berbasis air seperti susu formula, biskuit, bubur, nasi lembek, pisang atau makanan yang lain (Infodatin, 2014).

2.4 Naïve Bayes

Metode Naive Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dan efisien untuk pembelajaran mesin dan *Data Mining* (Muin et al., 2016). *Naïve Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada theorem Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *Theorem Bayes*. Ciri utama dari *Naïve Bayes* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi / kejadian (Widianto, 2019).

Naive Bayes bekerja sangat baik dibanding dengan *model classifier* lainnya. Hal ini dibuktikan oleh Xhemali, Hinde Stone dalam jurnalnya "*Naive Bayes vs Decision Trees vs Neural Networks in the Classification of Training Web Pages*" mengatakan bahwa "*Naive Bayes* memiliki tingkat akurasi yang lebih baik dibanding *model classifier* lainnya" (Widianto, 2019).

Pedekatan dari theorem Naïve Bayes menurut (Kusumadewi, 2003) adalah sebagai berikut:

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)}$$

Dimana:

x : Data dengan kelas (label) yang belum diketahui

c : Hipotesis data merupakan suatu class spesifik

$P(c|x)$: Probabilitas hipotesis c berdasarkan kondisi x (*posterior probabilitas*)

$P(c)$: Probabilitas hipotesis c (*prior probabilitas*)

$P(x|c)$: Probabilitas hipotesis x berdasarkan kondisi pada hipotesis c

$P(x)$: Probabilitas hipotesis dari x

Pendekatan Naïve Bayes memiliki asumsi sederhana yaitu asumsinya semua atribut bersifat saling bebas/independent/tidak saling mempengaruhi:

$$\begin{aligned} P(\mathbf{x}|c_i) &= P(x_1, x_2, \dots, x_d|c_i) \\ &= P(x_1|c_i) P(x_2|c_i) \dots P(x_d|c_i) \text{ Saling Bebas} \\ &= \prod_{j=1}^d P(x_j|c_i) \quad (1) \end{aligned}$$

Atribut Numerik

Untuk atribut numerik kami membuat asumsi default bahwa masing-masing adalah normal didistribusikan untuk setiap kelas c_i . Misalkan μ_{ij} dan σ_{ij}^2 menunjukkan rata-rata dan varians untuk atribut X_j untuk kelas c_i . Kemungkinan untuk kelas c_i , pada dimensi X_j , diberikan sebagai

$$P(x_j|c_i) \propto f(x_j|\mu_{ij}, \sigma_{ij}^2) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \exp\left\{-\frac{(x_j - \mu_{ij})^2}{\sigma_{ij}^2}\right\}$$

Klasifikasi Naïve Bayes menggunakan sampel rata-rata $\hat{\mu}_i = (\hat{\mu}_{i1}, \dots, \hat{\mu}_{id})^T$ dan sampel covarian matrix diagonal $\hat{\Sigma}_i = \text{diag}(\sigma_{ij}^2, \dots, \sigma_{id}^2)$ untuk setiap kelas c_i . Algoritma menunjukkan psoude-code untuk klasifikasi naïve bayes. Diberi input dataset \mathbf{D} , metode ini mengestimasi probabilitas prior dan rata-rata untuk setiap kelas.

Atribut Kategori

Asumsi independensi mengarah pada penyederhanaan ukuran dari probability mass function, dalam persamaan (4) dapat ditulis sebagai berikut:

$$P(\mathbf{x}|c_i) = \prod_{j=1}^d P(x_j|c_i) = \prod_{j=1}^d f(\mathbf{X}_j = \mathbf{e}_{jrj}|c_i) \quad (2)$$

Dimana $f(\mathbf{X}_j = \mathbf{e}_{jrj}|c_i)$ adalah probability mass function untuk \mathbf{X}_j , yang dapat di estimasi dari \mathbf{D}_i sebagai berikut:

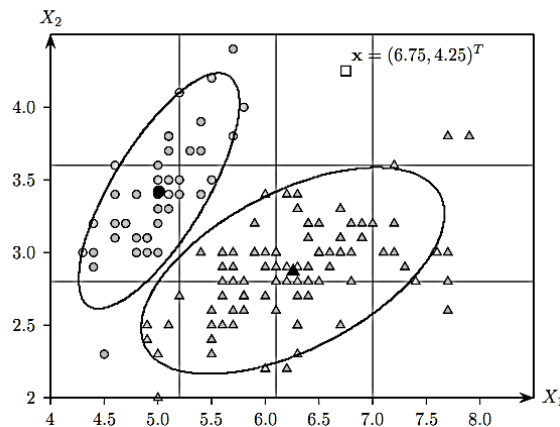
$$\hat{f}(\mathbf{v}_j|c_i) = \frac{n_i(\mathbf{v}_j)}{n_i}$$

di mana $n_i(\mathbf{v}_j)$ adalah frekuensi yang diamati dari nilai $\mathbf{v}_j = \mathbf{e}_j r_j$ sesuai dengan r_j nilai kategori a_{jr_j} untuk atribut X_j , pada kelas c_i . Seperti dalam kasus Bayes, jika count adalah nol, kita dapat menggunakan metode pseudo-count untuk mendapatkan probabilitas Prior. Mengestimasi dengan pseudo-count diberikan adalah sebagai berikut:

$$\hat{f}(\mathbf{v}_j|c_i) = \frac{n_i(\mathbf{v}_j) + 1}{n_i + m_j}$$

Dimana $m_j = |\text{dom}(X_j)|$. (3)

Contoh grafik klasifikasi dengan metode Bayes dapat dilihat pada Gambar 2.



Gambar 2. 2 Klasifikasi Bayes
 Sumber: *Data Mining and Analysis: Fundamental Concepts and Algorithms*

2.5 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi (Santosa, 2007). Teknik SVM digunakan untuk mendapatkan fungsi pemisah (*hyperplane*) yang optimal untuk memisahkan observasi yang memiliki nilai variabel target yang berbeda. Metode *Support Vector Machine* memiliki beberapa keuntungan yaitu:

1. Generalisasi

Generalisasi didefinisikan sebagai kemampuan suatu metode untuk mengklasifikasi suatu *pattern* atau pola, yang tidak termasuk data yang digunakan dalam fase pembelajaran metode itu.

2. Curse of dimensionality

Curse of dimensionality didefinisikan sebagai masalah yang dihadapi suatu metode *pattern recognition* dalam mengestimasi parameter dikarenakan jumlah sampel data yang relatif lebih sedikit dibandingkan dengan dimensional ruang vektor tersebut.

3. Feasibility

SVM dapat diimplementasikan relatif lebih mudah, karena proses penentuan *support vector* dapat dirumuskan dalam *Quadratic Programming* (QP) problem (Nugroho et al., 2003).

SVM memiliki prinsip dasar linier classifier yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada ruang berdimensi tinggi, akan dicari hyperplane (hyperplane) yang dapat memaksimalkan jarak (margin) antara kelas data (Santosa, 2007).

Menurut Santosa, (2007) hyperplane klasifikasi linier SVM dinotasikan:

$$f(x) = \mathbf{w}^T \mathbf{x} + b$$

Sehingga menurut Vapnik dan Cortes (Cortes & Vapnik, 1995) diperoleh persamaan

$$[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \geq 1 \text{ untuk } y_i = +1$$

$$[(\mathbf{w}^T \cdot \mathbf{x}_i) + b] \leq -1 \text{ untuk } y_i = -1$$

dengan \mathbf{x}_i = himpunan data training, $i = 1, 2, \dots, n$ dan y_i = label kelas dari \mathbf{x}_i Untuk mendapatkan hyperplane terbaik adalah dengan mencari hyperplane yang terletak di tengah-tengah antara dua bidang pembatas kelas dan untuk mendapatkan hyperplane terbaik itu, sama dengan memaksimalkan margin atau jarak antara dua set objek dari kelas yang berbeda.

Margin dapat dihitung dengan $\frac{2}{\|\mathbf{w}\|}$

Mencari hyperplane terbaik dapat digunakan metode Quadratic Programming (QP) Problem yaitu meminimalkan

$$\frac{1}{2} \mathbf{w}^T \mathbf{w}$$

Dengan syarat $y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) \geq 1, i = 1, 2, 3 \dots n$

Solusi untuk mengoptimasi oleh Vapnik (Cortes & Vapnik, 1995) diselesaikan dengan menggunakan fungsi Lagrange sebagai berikut

$$L(\mathbf{w}, b, a) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n a_i \{y_i [(\mathbf{w}^T \cdot \mathbf{x}_i) + b] - 1\} \quad (1)$$

dengan α_i = pengganda fungsi Lagrange dan $i = 1, 2, \dots, n$

Nilai optimal dapat dihitung dengan memaksimalkan L terhadap α_i , dan meminimalkan L terhadap \mathbf{w} dan b. Hal ini seperti kasus dual problem

$$\max_a W(a) = \max_a \left(\min_{\mathbf{w}, b} L(\mathbf{w}, b, a) \right) \quad (\text{Gunn, 1998})$$

Nilai minimum dari fungsi lagrange tersebut diberikan oleh

$$\begin{aligned} \frac{\partial L}{\partial b} = 0 &\Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \\ \frac{\partial L}{\partial \mathbf{w}} = 0 &\Rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i \mathbf{x}_i y_i \end{aligned} \quad (2)$$

Untuk menyederhanakannya persamaan (1) harus ditransformasikan ke dalam fungsi Lagrange Multiplier itu sendiri, sehingga menurut Santosa (Santosa, 2007) persamaan (1) menjadi

$$L(\mathbf{w}, b, a) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^n \alpha_i y_i (\mathbf{w}^T \cdot \mathbf{x}_i) - b \sum_{i=1}^n \alpha_i y_i + \sum_{i=1}^n \alpha_i \quad (3)$$

Berdasarkan persamaan (2), maka persamaan (3) oleh (Hastie et al., 2001) menjadi sebagai berikut

$$L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

dan diperoleh dual problem

$$\max_a L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

Dengan batasan, $\alpha_i \geq 0$, $i = 1, 2, \dots, n$ dan $\sum_{i=1}^n \alpha_i y_i = 0$

Data training dengan $\alpha_i \geq 0$, terletak pada hyperplane disebut support vector. Data training yang tidak terletak pada hyperplane tersebut mempunyai $\alpha_i = 0$. Setelah solusi permasalahan quadratic programming ditemukan (nilai α_i), maka kelas dari data yang akan diprediksi atau data testing dapat ditentukan berdasarkan nilai fungsi berikut.

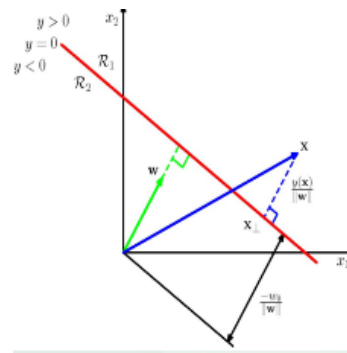
$$f(x_t) = \sum_{s=1}^{ns} \alpha_s y_s \mathbf{x}_s \cdot \mathbf{x}_t + b$$

Keterangan

\mathbf{x}_t = data yang akan diprediksi kelasnya (data testing)

\mathbf{x}_s = data support vector, $s = 1, 2, \dots, ns$

ns = banyak data support vector



Gambar 2. 3 Hyperplane

2.6 Preprocessing

Preprocessing data adalah tahapan dari data mining yaitu suatu proses atau tahapan yang dilakukan untuk mengolah data mentah menjadi data yang berkualitas atau masukan yang baik untuk dilanjutkan ke proses selanjutnya (Kumar & Chadha, 2012). Tahap Preprocessing data ini adalah tahap yang sangat krusial atau sangat penting dan harus dilakukan dengan teliti karena proses data mining membutuhkan data yang konsisten dalam penulisan, benar dalam formatnya, tidak ada data yang kosong, duplikasi data, dan lain-lain (Kumar & Chadha, 2012).

Preprocessing digunakan untuk mempersiapkan teks sebelum digunakan dalam pengujian atau pelatihan dengan tujuan untuk mengurangi noise yang ada pada data sehingga dapat meningkatkan kinerja classifier dan mempercepat proses klasifikasi (Sivakumar & Gunasundari, 2017).

2.7 Split Data

Split data atau pemisahan data adalah metode membagi data menjadi dua bagian atau lebih yang membentuk sub himpunan data. Umumnya, split data memisahkan dua bagian, bagian pertama digunakan untuk mengevaluasi atau uji data dan data lainnya digunakan untuk melatih model (Gillis, 2022).

Split data adalah teknik yang membagi data menjadi dua bagian secara acak, sebagian sebagai data training dan sebagian lainnya sebagai data testing, dengan menggunakan Split data akan dilakukan percobaan training berdasarkan split ratio yang telah ditentukan sebelumnya, untuk kemudian sisa dari split ratio data training akan dianggap sebagai data testing. Data training adalah data yang akan dipakai dalam melakukan pembelajaran sedangkan data testing adalah data yang

belum pernah dipakai sebagai pembelajaran dan akan berfungsi sebagai data pengujian kebenaran atau keakurasian hasil pembelajaran(Witten et al., 2017).

2.8 Confusion Matrix

Confusion matrix adalah sebuah metode yang digunakan untuk menganalisa tingkat kesuksesan sebuah model klasifikasi dalam mengenali data dengan kelas-kelas yang berbeda (J. Han & Pei, 2012). Adapun evaluasi pengukuran kinerja dari sistem menggunakan metode perhitungan menggunakan precision, recall, dan accuracy (Challa et al., 2016). Precision adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Sedangkan accuracy adalah tingkat kedekatan antara nilai prediksi dengan nilai actual (Challa et al., 2016).