

BAB II

TINJAUAN PUSTAKA

2.1 Tinjauan Pustaka

Dalam penelitian ini yang berjudul Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Dengan Metode Support Vector Machine disimpulkan bahwa hasil akurasi SVM sebesar 90% untuk confusion matrix. Beberapa penelitian yang pernah dilakukan digunakan sebagai acuan dalam proses penelitian ini.

Penelitian yang dilakukan oleh Nurlaelatul Maulidah, dkk tentang Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes. Berdasarkan penelitian ini disimpulkan bahwa hasil akurasi SVM lebih tinggi dengan tingkat akurasi sebesar 78.04% dan Naive Bayes dengan tingkat akurasi 76.98%. (Maulidah et al., 2021).

Selanjutnya penelitian oleh Lutfi Budi Ilmawan dan Muhammad Aliyazid Mude tentang Perbandingan Metode Klasifikasi Support Vector Machine dan Naive Bayes untuk Analisis Sentimen pada Ulasan Tekstual di Google Play Store. Berdasarkan penelitian tersebut disimpulkan bahwa hasil akurasi dari metode SVM adalah sebesar 81.46% dan metode Naive Bayes sebesar 75.41%. (Ilmawan & Mude, 2020)

Aris Budianto dkk, pernah menuliskan tentang Perbandingan K-Nearest Neighbor (Knn) Dan Support Vector Machine (Svm) Dalam Pengenalan Karakter Plat Kendaraan Bermotor. Pada penelitian ini didapatkan hasil akurasi pengujian pengenalan plat kendaraan dengan metode Support Vector Machine dengan akurasi 95%. Sedangkan menggunakan metode KNN mendapatkan akurasi pengujian 80%. (Budianto et al., 2019)

Penerapan Algoritma Support Vector Machine (Svm) Untuk Klasifikasi Pencemaran Nama Baik Di Media Sosial Twitter yang dilakukan oleh Fatwa Abdusyukur dengan data dari twitter. Dari hasil modelling atau pelatihan dan evaluation atau pengujian pada model SVM dengan menggunakan 6000 data tweet, diperoleh hasil akurasi tertinggi sebesar 87.7%. (Abdusyukur, 2023).

Penelitian yang dilakukan oleh Dimsyiar dkk tentang Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes didapatkan hasil akurasi sebesar 85,25% dengan dataset yang digunakan berjumlah 304 dan data yang digunakan sebagai data latih berjumlah 242 data. (Dimsyiar M Al Hafiz et al., 2021).

Selanjutnya penelitian yang dilakukan oleh Ahmad Yogiarto dkk tentang Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung, didapatkan tingkat akurasi sebesar 64,03%, nilai presesinya sebesar 64.58%, dan recall yang didapatkan sebesar 75.15% dengan 303 data sample dengan 14 atribut. (Yogiarto et al., 2024)

Penelitian dengan judul serupa juga pernah dilakukan oleh Afifa dengan judul Implementasi Metode Support Vektor Machine (Svm) Dalam Prediksi

Penyakit Jantung Dengan Menggunakan Python 3 didapatkan hasil akurasi sebesar 71% dengan dataset yang digunakan berjumlah 7000 data dan 11 atribut. (Afifa, 2024)

Adapun detail dari tinjauan pustaka ditunjukkan pada tabel 2.1

Tabel 2.1 Review Penelitian Terkait

No	Judul	Objek	Penelitian, Publikasi, Tahun	Metode	Tools	Hasil
1	PREDIKSI PENYAKIT DIABETES MELITUS MENGGUNAKAN METODE SUPPORT VECTOR MACHINE DAN NAIVE BAYES.	data sekunder database kesehatan Dataset Diabetes yang diambil dari dataset Kaggle .	Nurlaelatul Maulidah, Riki Supriyadi, Dwi Yuni Utami, Fuad Nur Hasan, Ahmad Fauzi, Ade Christian. Indonesian Journal on Software Engineering 2021	Support Vector Machine dan Naive Bayes	-	Nilai akurasi untuk model algoritma Support Vector Machine sebesar 78,04 persen dan nilai akurasi algoritma Naive Bayes sebesar 76,98 persen.

2	PERBANDINGAN METODE KLASIFIKASI SUPPORT VECTOR MACHINE DAN NAIVE BAYES UNTUK ANALISIS SENTIMEN PADA ULASAN TEKSTUAL DI GOOGLE PLAY STORE	Data training yang akan diolah untuk menghasilkan model klasifikasi berasal dari corpus	Lutfi Budi Ilmawana dan Muhammad Aliyazid Mudeb. ILKOM Jurnal Ilmiah 2020	Support Vector Machine dan Naive Bayes	python	Hasil akurasi dari Support Vector Machine Classifier memiliki nilai akurasi sebesar 81,46% dan Naïve Bayes classifier sebesar 75,41%.
3	PERBANDINGAN K-NEAREST NEIGHBOR	Tanda Nomor Kendaraan	Aris Budiantol Dwi, Maryono, Rosihan Ariyuana. Jurnal Ilmiah	K-nearest Neighbor Dan Support Vector Machine	-	Presentase akurasi pengenalan dengan metode SVM adalah 95%. Sedangkan Pengenalan dengan menggunakan metode

	(KNN) DAN SUPPORT VECTOR MACHINE (SVM) DALAM PENGENALAN KARAKTER PLAT KENDARAAN BERMOTOR	Bermotor atau TNKB	Pendidikan Teknik Kejuruan 2018			KNN adalah 80%.
4	PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI	data tweet dengan rentang waktu dari bulan oktober 2022 sampai dengan februari 2023	Fatwa Abdusyukur. KOMPUTA : Jurnal Ilmiah Komputer dan Informatika 2023	Support Vector Machine	-	Berdasarkan hasil pelatihan dan pengujian yang sudah dilakukan, algoritma support vector machine (SVM) memiliki performa yang baik dengan tingkat akurasi tertinggi sebesar 87,7%.

	PENCEMARAN NAMA BAIK DI MEDIA SOSIAL TWITTER					
5	SISTEM PREDIKSI PENYAKIT JANTUNG MENGUNAKAN METODE NAIVE BAYES	menggunakan dataset yang disediakan oleh UCI Machine learning.	Dimsyar M Al Hafiz, Khoirul Amaly, Javen Jonathan, M. Teranggono Rachmatullah, Rosidi. Jurnal Rekayasa Elektro Sriwijaya 2021	Naive Bayes	python	Berdasarkan eksperimen data uji berjumlah 61 data, 52 diantaranya terprediksi benar dengan nilai keakuratan sebesar 85,25%
6	IMPLEMENTASI METODE K- NEAREST NEIGHBORS (KNN) UNTUK KLASIFIKASI	Dataset penyakit jantung dari database Cleveland dan UCI Machine LearningRepository	Ahmad Yogiarto, Ahmad Homaidi, Zachol Fatah. G- Tech:Jurnal Teknologi Terapan 2024	K-Nearest Neighbors	RapidMiner	Hasil pada metode KNN dengan parameter K=5 ini nilai Akurasi yang didapatkan sebesar 64,03%, nilai presesinya sebesar 64.58%, dan recall yang didapatkan sebesar 75.15%.

	PENYAKIT JANTUNG					
7	IMPLEMENTASI METODE SUPPORT VEKTOR MACHINE (SVM) DALAM PREDIKSI PENYAKIT JANTUNG DENGAN MENGGUNAKAN PYTHON 3	Dataset yang diperoleh dari website Kaggle.com yaitu Cardiovascular Disease dataset	Afifa (2024)	Support Vector Machine	Python3	Dalam penelitian ini didapatkan akurasi sebesar 71.52% dengan menggunakan kernel RBF non linear. Didapatkan juga hasil precision sebesar 81%, recall 67%, dan f1-Score 73%.
8	IMPLEMENTASI DATA MINING UNTUK	Dataset diperoleh dari public repository Kaggle	Muhammad Hilmy Khairy (2025)	Support Vector Machine	Python	Berdasarkan pengujian dalam menggunakan confusion matrix didapatkan hasil akurasi

MEMPREDIKSI PENYAKIT JANTUNG DENGAN METODE SUPPORT VECTOR MACHINE (SVM)						sebesar 90% , hasil precision sebesar 91%, recall 90%, dan F1 score 90%.
--	--	--	--	--	--	---

2.2 Dasar Teori

2.2.1 Penyakit jantung

Penyakit jantung merujuk pada berbagai kondisi yang memengaruhi jantung yang bertanggung jawab memompa darah ke seluruh tubuh. Potensi penyakit jantung melibatkan berbagai faktor dan penyebab yang dapat meningkatkan kemungkinan seseorang mengalami masalah jantung. beberapa faktor utama penyebab penyakit jantung seperti Tekanan darah tinggi, Kadar kolesterol, Hr max, dan nyeri dada (Santoso et al., 2024). Penyakit Jantung disebabkan oleh penumpukan plak di dinding arteri. Hal tersebut berhubungan dengan faktor risiko, seperti usia, jenis kelamin, merokok, dislipidemia, dan tekanan darah tinggi (redi ruldani et al., 2015).

2.2.2 Data mining

Data mining adalah proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau mencari dari data yang terdapat pada basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar (Putro et al., 2020).

Dalam data mining terdapat 6 fungsi-fungsi yaitu : (Siti Masripah, 2015)

1. Fungsi deskripsi (description)

Fungsi deskripsi merupakan langkah menggambarkan sekumpulan

data secara ringkas. Banyak cara yang dapat digunakan untuk memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya. Contoh dari penggambaran fungsi deksripsi yaitu Deskripsi Grafis, Deskripsi Lokasi, dan Deskripsi Keragaman.

2. Fungsi estimasi (estimation)

Fungsi estimasi diidentifikasi untuk membuat perkiraan suatu hal yang sudah memiliki data. Fungsi estimasi terdiri dari dua cara yaitu Estimasi Titik dan Estimasi Selang Kepercayaan.

3. Fungsi prediksi (prediction)

Fungsi prediksi adalah memperkirakan hasil dari hal yang belum diketahui, dan digunakan untuk mendapatkan hal baru yang akan muncul selanjutnya.

4. Fungsi klasifikasi (classification)

Fungsi klasifikasi merupakan proses menggolongkan suatu data. Contoh algoritma klasifikasi : Algoritma Mean Vector, Algoritma K-nearest Neighbor, Algoritma ID3, Algoritma C4.5, dan Algoritma C5.0

5. Fungsi pengelompokan (cluster)

Fungsi pengelompokan merupakan sebuah proses pengelompokan data. Data yang dikelompokkan disebut objek atau catatan yang memiliki kemiripan atribut kemudian dikelompokkan pada kelompok yang berbeda. Contoh algoritma yang digunakan seperti : Algoritma Hirarchical Clustering, Algoritma Partitional Clustering, Algoritma

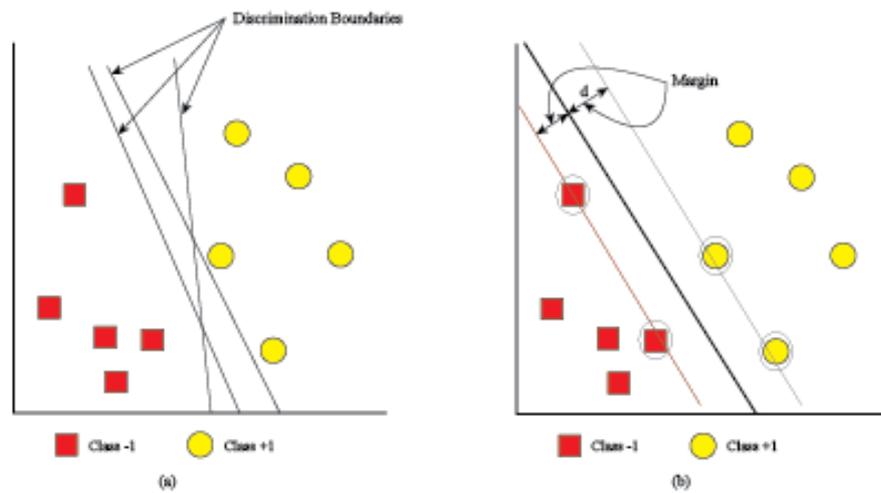
Single Linkage, Algoritma Complete Linkage, Algoritma Average Linkage, Algoritma K- Means

6. Fungsi asosiasi (association)

Fungsi asosiasi difungsikan untuk menemukan aturan asosiasi (association rule) yang berguna dalam mengidentifikasi item-item yang menjadi objek. Algoritma yang digunakan seperti algoritma Generalized Association Rules, Quantitative Association Rule, Asynchronous Parallel Mining.

2.2.3 Support vector machine

Vapnik pada tahun 1992 memperkenalkan Support Vector Machine (SVM) sebagai teknik klasifikasi yang efisien untuk masalah nonlinier. Konsep SVM secara sederhana dapat dijelaskan sebagai upaya untuk menemukan hyperplane optimal yang bertindak sebagai pemisah antara dua kelas dalam ruang input. SVM berupaya mencari fungsi pemisah (hyperplane) dengan memaksimalkan jarak antar kelas. Dengan cara ini, SVM dapat menjamin kemampuan generalisasi yang tinggi untuk data yang akan datang (Salsabila Dwi Fitri et al., 2024).



Gambar 2.1 Hyperplane Terbaik Yang Memisahkan Kedua Class Negatif Dan Positif (Salsabila Dwi Fitri et al., 2024)

Gambar 2.1 menggambarkan dua kelas data yang terpisah. Garis merah tebal menunjukkan hyperplane yang memisahkan kedua titik tersebut, sehingga titik data pada salah satu sisinya diberi label kelas negatif -1 dan label kelas positif +1. Titik data (vektor) terdekat dengan hyperplane disebut vektor pendukung, garis merah kecil yang memotong hyperplane memiliki pengaruh terbesar. Jarak antara hyperplane dan support vector disebut margin. Model linier secara umum yang dipakai dalam SVM untuk menghasilkan hyperplane adalah sebagai berikut (Fitriyah et al., 2020) :

$$y = w \cdot x + b \tag{1}$$

Dengan

x = vector baris berdimensi k (banyaknya fitur)

y = nilai target dari himpunan data x

w = vektor baris yang menjadi parameter bobot

b = bias atau error

Dua pemisah sejajar dengan hyperplane yang diperoleh dari data terdekat pada masing - masing kelas yang diformulasikan sebagai berikut:

$$(w \cdot x) + b = 1, \text{ untuk kelas positif} \quad (2)$$

$$(w \cdot x) + b = -1, \text{ untuk kelas negatif} \quad (3)$$

Pada permasalahan SVM, margin dimaksimalkan, sehingga :

$$\text{Margin} = \frac{1}{|w|} \quad (4)$$

Untuk mendapatkan hyperplane terbaik dapat digunakan metode Quadratic Programming (QP) yaitu dengan cara meminimalkan

$$\min \tau(w) = \frac{1}{2} |w|^2 \quad (5)$$

dengan syarat

$$y(w \cdot x + b) \geq 1 \quad (6)$$

Beberapa kernel yang umum dipakai pada SVM adalah:

1. Polynomial. Kernel trick polynomial cocok digunakan untuk menyelesaikan masalah klasifikasi, dimana dataset pelatihan sudah normal.
2. Radial Basis Function (RBF) atau Gaussian. Kernel trick radial basis function atau gaussian merupakan kernel yang paling banyak digunakan

untuk menyelesaikan masalah klasifikasi untuk dataset yang tidak terpisah secara linear, dikarenakan akurasi pelatihan dan akurasi prediksi yang sangat baik pada kernel ini.

2.2.4 Confusion matrix

Confusion Matrix digunakan sebagai pengukur kinerja setelah mengolah data mining dengan model klasifikasi. Pada dasarnya, Confusion Matrix digunakan untuk memberikan informasi perbandingan dari hasil klasifikasi yang dilakukan oleh algoritma yang digunakan dengan hasil klasifikasi sebenarnya (Romadloni et al., 2022). Disajikan pada Tabel 2.2 Confusion Matrix.

Tabel 2.2 Confusin Matrix

PREDICTED	ACTUAL	
	TRUE	FALSE
TRUE	TP (True Positive)	FP (False Positive)
FALSE	TN (True Negative)	FN (False Negative)

Pada tabel di atas, terdapat empat istilah yang merepresentasikan hasil proses klasifikasi pada confusion matrix, yaitu:

1. True Positive (TP), data positif yang diprediksi benar.
2. True Negative (TN), data negatif yang diprediksi benar.
3. False Positive (FP), error tipe 1, data negatif namun diprediksi sebagai data positif.
4. False Negative (FN), error tipe 2, data positif namun diprediksi sebagai data negatif.

Pada prakteknya confusion matrix digunakan untuk menghitung nilai akurasi, presisi, dan recall.

1. Nilai akurasi, menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar.

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (7)$$

2. Presisi, menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model.

$$Presisi = \frac{TP}{TP + FP} \quad (8)$$

3. Recall atau sensitivity, menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

4. F1 Score, merupakan perbandingan rata-rata presisi dan recall yang dibobotkan

$$F1 \text{ Score} = 2 * (\text{Recall} * \text{Presisi}) / (\text{Recall} + \text{Presisi}) \quad (10)$$

2.2.5 Python

Python adalah scripting language yang berorientasi objek. Bahasa pemrograman ini dapat digunakan untuk pengembangan perangkat lunak dan bisa dijalankan melalui berbagai sistem operasi (Manalu & Gunadi, 2022). Sebagai bahasa pemrograman tingkat tinggi, Python menjadi pilihan banyak pengembang dalam menciptakan berbagai jenis aplikasi. Bahasa ini terkenal mudah dipelajari,

terutama oleh pemula, berkat sintaksnya yang sederhana dan mudah dipahami. Selain itu, Python menyediakan beragam pustaka dan kerangka kerja yang memudahkan proses pengembangan aplikasi di berbagai bidang (Dayu Renita et al., 2023).