

BAB 2

LANDASAN TEORI

2.1 Diabetes

Diabetes Melitus merupakan penyakit gangguan metabolik akibat pankreas tidak memproduksi cukup insulin atau tubuh tidak dapat menggunakan insulin yang diproduksi secara efektif. Insulin adalah hormon yang mengatur keseimbangan kadar gula darah. Akibatnya terjadi peningkatan konsentrasi glukosa di dalam darah (hiperglikemia) [1]. Diabetes dapat dibagi menjadi beberapa tipe yang berbeda. Namun, ada dua tipe utama, diabetes tipe 1 dan diabetes tipe 2, menurut etiopatologi kelainan tersebut. Diabetes tipe 2 tampaknya merupakan bentuk diabetes yang paling umum (90 % dari semua pasien diabetes), terutama ditandai dengan resistensi insulin. Penyebab utama diabetes tipe 2 termasuk gaya hidup, aktivitas fisik, kebiasaan diet dan faktor keturunan [2], sedangkan penyebab diabetes tipe 1 adalah ketidakmampuan pankreas untuk memproduksi cukup insulin, sehingga glukosa di dalam darah tidak dapat masuk ke dalam sel. Gangguan pada pankreas ini diduga karena proses autoimun, yaitu ketika sistem kekebalan tubuh seseorang menyerang sel-sel tubuh yang sehat. Pada diabetes tipe 1, sistem kekebalan tubuh tersebut menyerang dan merusak sel beta pada pankreas, sehingga tidak dapat memproduksi cukup insulin. Beberapa hal yang bisa menjadi penyebab kerusakan sel beta pankreas, antara lain infeksi virus (enterovirus, virus Epstein-Barr, virus rubella, rotavirus, serta virus gondongan), konsumsi obat-obatan tertentu (pyrinuron dan streptomisin), serta pengaruh gluten [3].

2.2 Naïve Bayes

Naïve Bayes merupakan salah satu algoritma dalam *machine learning* yang menerapkan teori Bayes dalam klasifikasi [4] yang ditemukan oleh ilmuwan dari Inggris bernama Thomas Bayes, dimana memprediksi peluang di masa depan

berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Teorema tersebut dikombinasikan dengan naïve dimana diasumsikan kondisi antar atribut saling tidak terkait atau atribut yang independen sehingga diasumsikan bahwa ada atau tidak adanya ciri khas dari kelas tertentu tidak ada hubungannya dengan ciri dari kelas lainnya. Persamaan dari teorema bayes adalah seperti pada rumus (2.1)

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (2.1)$$

$P(A|B)$ = probabilitas bersyarat dari peristiwa A yang terjadi mengingat B benar

$P(B|A)$ = probabilitas bersyarat dari peristiwa B yang terjadi mengingat A benar

$P(A), P(B)$ = probabilitas independen satu sama lain.

Untuk menjelaskan tentang teorema Naïve Bayes perlu diketahui terlebih dahulu bahwa proses pengklasifikasian memerlukan sejumlah pejumlah untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut karena itu teorema bayes pada rumus (2.2) disesuaikan menjadi sebagai berikut [5].

$$P(C|F_1 \dots F_n) = \frac{p(C)p(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (2.2)$$

Penjelasan :

$p(F_1 \dots F_n|C)$ adalah peluang kemunculan karakteristik-karakteristik sampel pada kelas C atau disebut juga dengan likelihood

$P(C|F_1 \dots F_n)$ adalah peluang masuknya sampel karakteristik tertentu dalam kelas C atau disebut juga dengan posterior probability

$P(F_1 \dots F_n)$ adalah peluang kemunculan karakteristik-karakteristik sampel secara global atau disebut juga dengan evidence

$p(C)$ adalah peluang munculnya kelas C disebut juga dengan Class Prior Probability

Maka rumus di atas menjelaskan bahwa posterior adalah class prior probability dikali dengan likelihood dibagi dengan evidence oleh karena itu rumus diatas bisa ditulis secara sederhana menjadi sebagai berikut [5].

$$Posterior = \frac{Prior \times likelihood}{evidence} \quad (2.3)$$

2.3 Gaussian Naïve Bayes

Ketika berhadapan dengan data kontinu, asumsi yang khas adalah bahwa nilai kontinu yang berkorelasi dengan setiap kelas didistribusikan sesuai dengan distribusi Gaussian. Data pelatihan dibagi berdasarkan kelas dan mean dan standar deviasi dari setiap kelas dihitung. Oleh karena itu untuk memperkirakan probabilitas kumpulan data kontinu, persamaan berikut dapat digunakan[17].

$$P(x_i|c_j) = \frac{1}{\sqrt{2\pi\sigma_{i,j}^2}} * e^{-\frac{1}{2}\left(\frac{x_i-\mu_{i,j}}{\sigma_{i,j}}\right)^2} \quad (2.4)$$

Dimana :

x = variable

c = class

$\mu_{i,j}$ = mean

$\sigma_{i,j}$ = standard deviation

2.4 Confusion Matrix

Confusion Matrix merupakan sebuah tabel yang terdiri atas banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi, tujuannya adalah digunakan untuk menentukan kinerja suatu model klasifikasi. Tabel *Confusion Matrix* dapat membantu dalam memperoleh nilai dari perhitungan tersebut karena berisikan data prediksi yang positif dan negatif yang dihasilkan oleh sistem dan data actual yang positif dan negatif di dunia nyata . Berikut contoh bentuk table *Confusion Matrix* [6].

Tabel 2. 1 Confusion Matrix

		Aktual	
		Positif	Negatif
Sistem	Positif	TP	FP
	Negatif	FN	TN

Penjelasan:

1. True Positives (TP) = kasus dimana hasil sistem memprediksi pasien terkena diabetes dan pasien benar-benar memiliki penyakit diabetes
2. True Negatives (TN) = Kasus dimana hasil sistem memprediksi pasien tidak diabetes dan faktanya pasien memang tidak diabetes
3. False Positives (FP) = Kasus dimana hasil sistem memprediksi pasien terkena diabetes tetapi sebenarnya pasien tidak terkena diabetes (biasa disebut juga dengan error tipe 1)
4. False Negatives (FN) = Kasus dimana hasil sistem memprediksi pasien tidak terkena diabetes padahal faktanya pasien terkena diabetes (biasa disebut juga dengan error tipe 2

2.5 Performance Evaluation

- Akurasi

Perhitungan akurasi dilakukan dengan cara membagi jumlah data yang diklasifikasikan secara benar dengan total sample data testing yang diuji [7].

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.5)$$

- *Precision*

Menghitung nilai precision dengan cara membagi jumlah data benar yang bernilai positif (TP) dibagi dengan jumlah data benar yang bernilai positif (TP) dan data salah yang bernilai positif (FN) [7].

$$Precision = \frac{TP}{TP+FP} \quad (2.6)$$

- *Recall*

Recall dihitung dengan cara membagi data benar yang bernilai positif (TP) dengan hasil penjumlahan dari data benar yang bernilai positif (TP) dan data salah yang bernilai negatif (FN) [7].

$$Recall = \frac{TP}{TP+FN} \quad (2.7)$$

- F1-Score

Nilai F1-Score didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall* kemudian dikalikan dua [7].

$$F1 - Score = 2 * \frac{precision*recall}{precision+recall} \quad (2.8)$$

