

BAB II

TINJAUAN PUSTAKA DAN TEORI

2.1. Tinjauan Pustaka

Metode *Linear Discriminant Analysis* (LDA) digunakan untuk ekstraksi fitur kedalam kelas-kelas yang berbeda. Metode *Linear Discriminant Analysis* (LDA) mampu memisahkan antarkelas menjadi lebih terpisah dengan memaksimalkan nilai *between-class scatter* dan meminimalkan *within-class scatter*. Pada ekstraksi ciri menggunakan *Linear Discriminant Analysis* (LDA) dataset lokasinya tetap namun kelas yang dibentuk menjadi lebih terpisah sehingga dapat menyebabkan jarak antar kelas menjadi lebih besar, sedangkan jarak antar data pelatihan dalam satu kelas akan menjadi lebih kecil (Widyati dkk, 2021).

Pada penelitian sebelumnya, metode *Linear Discriminant Analysis* telah digunakan untuk melakukan Prediksi Kelulusan Mahasiswa. Pada penelitian tersebut, metode *Linear Discriminant Analysis* digunakan untuk penentuan skor prediksi (Sulistio, 2017). Dari hasil penelitian tersebut diperoleh hasil prediksi status kelulusan mahasiswa yang cukup akurat yaitu lebih dari 90% untuk evaluasi terhadap kondisi sebenarnya, maupun rata-rata dari tiap iterasi *k-fold cross validation*. Kelebihan metode ini adalah memiliki nilai akurasi yang tinggi untuk memisahkan sekelompok siswa ke kelas masing-masing (lulus tepat waktu atau terlambat), sehingga metode ini juga dapat digunakan untuk menentukan prediksi penyakit jantung yang diderita oleh seseorang. Dengan angka akurasi yang tinggi diharapkan metode *Linear Discriminant Analysis* dapat memprediksi penyakit jantung dengan akurat.

Pada penelitian lainnya metode *Linear Discriminant Analysis* diimplementasikan Untuk Klasifikasi Pengambilan Mata Kuliah Pilihan (Anggrestianingsih dkk, 2019). Penelitian tersebut melakukan pengujian dan analisis menggunakan metode LDA, dengan menggunakan 30 data latih yang terdiri dari 15 data untuk kelas “ya” dan 15 data untuk kelas “tidak” dengan

menggunakan 5 data uji. Dari skenario pengujian yang ada, tingkat akurasi dengan menggunakan metode LDA tergantung pada kelas data uji, dengan hasil nilai terbesar yaitu 70%.

Penelitian sebelumnya yang berkaitan dengan proses Keakuratan metode *Linear Discriminant Analysis* juga diperkuat oleh penelitian terdahulu yaitu mendeteksi Jenis Autis pada Anak Usia Dini (Budiman dan Afirianto, 2017). Pada penelitiannya mengembangkan sistem untuk pendeteksi Jenis dari penyakit Autis pada anak usia dini, menggunakan data pelatihan sebesar 75% dari jumlah data yang digunakan dan sisanya 25% sebagai data percobaan untuk menguji model sistem yang telah dibentuk dari penggunaan metode *Linear Discriminant Analysis*. Hasil dari penelitiannya menghasilkan nilai akurasi sebesar 88%. Nilai akurasi dapat berubah ketika data pelatihan ditambahkan lagi jumlah datanya.

Penelitian lain yang menggunakan metode LDA yaitu penelitian yang dilakukan oleh Hana Fida (Hana, 2020), penelitian ini membandingkan algoritma *Neural Network* dengan *Linear Diskriminant Analisis* untuk klasifikasi penyakit diabetes. Data yang digunakan terdiri dari 16 variabel dataset gejala dan 1 variabel *class* pengklasifikasi, dengan jumlah data sebanyak 520. Dari data tersebut maka data dibagi menjadi dua dengan presentase 70% data pelatihan dan 30% data pengujian. Setelah dilakukan proses pelatihan dan pengujian, dilakukan klasifikasi menggunakan algoritma *Neural Network* dengan *Linear Diskriminant Analisis* dan dihitung akurasinya menggunakan *confusion matrix* untuk menghasilkan *performance vector*. Penelitian menggunakan algoritma *Linear Diskriminant Analisis* menunjukkan akurasi cukup besar yaitu 90,19 % dengan *precision* sebesar 82.61% dan *recall* sebesar 95,00% dengan akurasi sebesar 95.19 % dengan *precision* sebesar 92,68% dan *recall* sebesar 95,00% dengan algoritma *Neural Network*. Dari hasil penelitian tersebut membuktikan bahwa algoritma *Neural Network* menghasilkan akurasi yang lebih besar untuk klasifikasi penyakit diabetes.

Selain penelitian menggunakan metode *Linear Discriminant Analysis*, banyak penelitian lain yang berkaitan dengan metode prediksi *Quadratic Discriminant Analysis*, pada penelitian tersebut dilakukan perbandingan *Quadratic*

Discriminant Analysis dan *Support Vector Machine* untuk Klasifikasi Tutupan Lahan di DKI Jakarta (Dimas dkk, 2020). Pada penelitian ini data yang digunakan berupa data citra, tujuan dari penelitiannya yaitu untuk membangun sistem yang bertujuan itu pemantauan berbagai penggunaan lahan sesuai dengan rencana dari tata ruang wilayah DKI Jakarta. Hasil penelitian menunjukkan bahwa pada kedua metode, varians dari semua kriteria di atas semakin kecil seiring dengan bertambahnya jumlah data latih. QDA dan SVM memiliki kinerja serupa berdasarkan akurasi dan spesifisitas keseluruhan. Namun, SVM lebih baik daripada QDA pada sensitivitas, jadi dapat disimpulkan masih ada beberapa kekurangan yang dirasakan peneliti pada penelitian yang dilakukan.

2.2. Dasar Teori

2.2.1. Penyakit Jantung

Penyakit Jantung Koroner (PJK) adalah salah satu akibat utama *arteriosklerosis* (pengerasan pembuluh darah nadi) yang dikenal sebagai *arterosklerosis*. Pada keadaan ini, arteri menyempit karena timbunan lemak dalam dinding pembuluh darah. Hal ini bisa membuat aliran oksigen ke jantung terhambat dan menyebabkan *angina* atau rasa nyeri dan tidak nyaman di bagian dada. Apabila tidak diatasi, kondisi ini bisa menyebabkan penyakit jantung koroner. Penyakit Jantung Koroner (PJK) merupakan salah satu bentuk penyakit kardiovaskular yang menjadi penyebab kematian nomor satu di dunia. Penyakit jantung koroner adalah suatu penyakit degeneratif yang berkaitan dengan gaya hidup, dan sosial ekonomi masyarakat. Penyakit ini merupakan problem kesehatan utama yang sering terjadi dan dialami di berbagai Negara maju di dunia (Mori, 2020).

Menurut data dari *Nurses Health Study* dari negara Amerika Serikat, gaya hidup yang sehat dapat mengurangi berbagai risiko penyakit yang menyerang manusia termasuk dari penyakit jantung dengan kisaran yaitu sebanyak 80%. Hidup sehat tersebut termasuk dalam beberapa kategori yang kerap kali manusia lakukan yaitu diantaranya manusia yang menjaga berat badan, tidak merokok, membatasi konsumsi alkohol, berolahraga minimal 30 menit setiap hari, serta mengonsumsi makanan rendah lemak dan tinggi serat memiliki risiko rendah menderita berbagai

penyakit kardiovaskuler. Dari pola hidup sehat yang terus dan konsisten dilakukan menjadi point utama dalam mengurangi terserangnya penyakit jantung (Abufaraj dkk., 2019).

2.2.2. Gejala Penyakit Jantung

Gejala dari penyakit ataupun dari serangan jantung memiliki beberapa bervariasi gejala, tidak semua orang yang mengalami serangan jantung memiliki gejala yang sama atau memiliki tingkat keparahan yang sama. Beberapa orang merasakan nyeri ringan, sementara lainnya merasakan rasa sakit yang amat sangat. Beberapa orang bahkan tidak menunjukkan gejala, sedangkan penderita lain mungkin mengalami tanda pertama berupa serangan jantung secara mendadak. Semakin banyak tanda dan gejala yang dirasakan, semakin besar kemungkinan mengalami serangan jantung (Jabri dkk., 2021).

Berikut ini merupakan beberapa penjelasan terkait gejala yang biasanya menyertai dari penderita penyakit jantung koroner diantaranya adalah gejala nyeri dada, sesak nafas dan jantung berdebar. Gejala nyeri dada biasanya dirasakan oleh sekitar sepertiga penderita penyakit jantung koroner. Nyeri dirasakan dibagian tengah dan menyebar ke leher, lengan, dan dagu. Perasaan nyeri sering disertai rasa seperti diremas atau dicengkeram, dan hal ini disebabkan karena jantung kekurangan darah dan oksigen.

Selain nyeri dada, gejala lain yang muncul pada penderita penyakit jantung yaitu sesak nafas. Gejala sesak nafas berhubungan dengan kesulitan bernafas yang disadari dan dirasakan perlu usaha tambahan untuk mengatasi kekurangan udara. Bila jantung tidak dapat memompa sebagaimana mestinya, cairan cenderung dapat berkumpul di jaringan dan paru, sehingga menyebabkan kesulitan bernafas waktu berbaring.

Keluhan lainnya, yaitu debaran jantung tidak seperti biasanya atau pada normalnya. Debaran jantung lebih keras dari pada biasa atau irama jantung yang tidak teratur (aritmia). Terkadang rasa berdebar-debar juga diikuti dengan keluhan lain seperti keringat dingin, sakit dada, dan sesak nafas (Jabri dkk., 2021).

2.2.3. Faktor Risiko Penyakit Jantung

Banyak faktor yang mempengaruhi munculnya risiko penyakit jantung koroner yaitu dari faktor risiko yang tidak dapat dimodifikasi dan dapat dimodifikasi. Faktor risiko penyakit jantung koroner yang tidak dapat dimodifikasi (tidak dapat diubah) adalah riwayat keluarga, umur dan jenis kelamin.

Adapun beberapa faktor dari risiko penyakit jantung koroner yang dapat dimodifikasi (dapat diubah) diantaranya (Jabri dkk., 2021):

- 1) Hipertensi,
- 2) Merokok,
- 3) Diabetes melitus,
- 4) Dislipidemia (metabolisme lemak yang abnormal),
- 5) Kurang aktifitas fisik (kurang gerak),
- 6) Obesitas,
- 7) Pola makan, dan
- 8) Stress.

2.2.4. Tingkat Risiko Penyakit Jantung

Salah satu klasifikasi yang sering digunakan yaitu klasifikasi berdasarkan abnormalitas struktural jantung yang disusun oleh *American Heart Association/American College of Cardiology (AHA/ACC)* atau berdasarkan gejala berkaitan dengan kapasitas fungsional yang disusun oleh *New York Heart Association (NYHA)*.

Sekolah Pascasarjana

Tabel 2.1 Tabel Klasifikasi Risiko Penyakit Jantung

Klasifikasi menurut ACC/AHA	Klasifikasi menurut NYHA27
<p>Stadium A Memiliki risiko tinggi untuk berkembang menjadi gagal jantung. Tidak terdapat gangguan struktural atau fungsional jantung.</p>	<p>Kelas I Pasien dengan penyakit jantung tetapi tidak ada pembatasan aktivitas fisik. Aktivitas fisik biasa tidak menyebabkan kelelahan berlebihan, palpitasi, dispnea atau nyeri angina.</p>
<p>Stadium B Telah terbentuk penyakit struktural jantung yang berhubungan dengan perkembangan gagal jantung, tidak terdapat tanda dan gejala.</p>	<p>Kelas II Pasien dengan penyakit jantung dengan sedikit pembatasan aktivitas fisik. Merasa nyaman saat istirahat. Hasil aktivitas normal menyebabkan fisik kelelahan, palpitasi, dispnea atau nyeri angina.</p>
<p>Stadium C Gagal jantung yang simptomatis berhubungan dengan penyakit struktural jantung yang mendasari</p>	<p>Kelas III Pasien dengan penyakit jantung yang terdapat pembatasan aktivitas fisik. Merasa nyaman saat istirahat. Aktifitas fisik ringan menyebabkan kelelahan, palpitasi, dispnea atau nyeri angina.</p>
<p>Stadium D Penyakit struktural jantung yang lanjut serta gejala gagal jantung yang sangat bermakna saat istirahat walaupun telah mendapat terapi.</p>	<p>Kelas IV Pasien dengan penyakit jantung yang mengakibatkan ketidakmampuan untuk melakukan aktivitas fisik apapun tanpa ketidaknyamanan. Gejala gagal jantung dapat muncul bahkan pada saat istirahat. Keluhan meningkat saat melakukan aktifitas.</p>

2.2.5. Machine Learning

Machine learning atau pembelajaran mesin merupakan pendekatan dalam AI (*Artificial Intelligence*) yang banyak digunakan untuk menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Sesuai namanya, *machine learning* mencoba menirukan bagaimana proses manusia atau makhluk cerdas belajar dan menggeneralisasi. Ciri khas dari *machine learning* adalah adanya proses pelatihan, pembelajaran atau pelatihan. Oleh karena itu, *machine learning* membutuhkan data untuk dipelajari yang disebut sebagai data pelatihan. Klasifikasi adalah metode dalam *machine learning* yang digunakan oleh mesin untuk memilah atau mengklasifikasikan objek berdasarkan

ciri tertentu sebagaimana manusia mencoba membedakan benda satu dengan yang lain. Sedangkan prediksi atau regresi digunakan oleh mesin untuk menerka keluaran dari suatu data masukan berdasarkan data yang sudah dipelajari dalam pelatihan. Algoritma yang digunakan pada *machine learning* tidak didefinisikan secara eksplisit (Roihan dkk., 2020).

Machine learning merupakan sub dari bidang keilmuan kecerdasan buatan (*Artificial intelligence*) yang banyak untuk diamati dan diteliti yang dapat berguna untuk memecahkan berbagai macam permasalahan kehidupan di banyak bidang. Bidang tersebut seperti bidang kesehatan, keuangan, kemasyarakatan maupun bidang lainnya. Ulasan dari berbagai macam bidang dengan *machine learning* dapat disajikan dalam berbagai macam algoritma yang dapat diimplementasikan. Algoritma tersebut dibagi menjadi tiga kategori dalam *machine learning* diantaranya *supervised learning*, *unsupervised learning* dan *reinforcement learning* (Praiss dkk., 2020).

2.2.6. Klasifikasi Statistik

Klasifikasi statistik atau sering juga disebut sebagai klasifikasi adalah sebuah cara atau metode untuk mengidentifikasi kemana sebuah objek data dikelompokkan. Sebagai contoh adalah klasifikasi *email* ke dalam kategori *spam* atau bukan *spam*, seseorang yang termasuk ke dalam kategori terdeteksi memiliki penyakit jantung atau tidak, dan sebagainya. Klasifikasi juga bisa digunakan untuk membangun sebuah model prediktif untuk memprediksi objek baru. Klasifikasi statistik mempelajari fungsi antara karakteristik sesuatu (misalnya variabel independent) dan keanggotannya (misalnya variabel *output*) melalui suatu proses pembelajaran 'terawasi' (*supervised learning process*) pada kedua jenis variabel (*input* dan *output*) disajikan ke algoritma (Verma dkk., 2021).

Dalam klasifikasi statistik, biasanya tiap observasi dari individual atau objek baru akan dianalisa ke dalam suatu set variabel yang bisa dijelaskan (*explanatory variabel*) atau fitur. Nilai dari fitur ini bisa berupa kategori (seperti golongan darah A, B, AB, atau O), urutan/ordinal (seperti besar, kecil, atau sedang) atau berupa angka (bilangan bulat atau real). Ada juga *classifier* yang bekerja

dengan melakukan perbandingan similaritas observasi lalu dan sekarang (Verma dkk., 2021).

Metodologi dua langkah yang paling umum dari prediksi jenis *classification* adalah model pelatihan (*training*) dan model pengujian (*testing*). Pada model pelatihan sekumpulan data masukan termasuk berbagai label kelas yang aktual digunakan. Setelah suatu ‘model’ dilatih, model tersebut di tes terhadap sampel data yang tersisa untuk penilaian akurasi dan pada akhirnya diimplementasikan untuk penggunaan riil yang digunakan untuk memprediksi kelas – kelas dari data baru (label kelas tidak diketahui) (Caulkins dkk., 2006).

Dalam berbagai masalah klasifikasi yang dijadikan sumber utama untuk memprediksi akurasi adalah *confusion matrix* (matrik klasifikasi). Tabel 2.2 di bawah ini adalah *confusion matrix* pada klasifikasi dengan dua kelas.

Tabel 2.2 Matriks konfusi dua kelas

Prediksi kelas	Kebenaran kelas	
	Positif	Negatif
Positif	Kebenaran positif (TP)	Negatif palsu (FP)
Negatif	Positif palsu (FN)	Kebenaran negatif (TN)

Confusion matrix tersebut terdiri dari *true positive* (TP), *false negatif* (FN), *false positive* (FP) dan *true negatif* (TN). Pada TP merupakan data positif yang terdeteksi benar. FN merupakan kebalikan dari *true positive*, sehingga data positif, namun terdeteksi sebagai data negatif. Untuk FP sendiri, merupakan data negatif namun terdeteksi sebagai data positif, sedangkan TN merupakan data negatif yang terdeteksi dengan benar.

Maka dapat dihitung nilai accuracy, precision, recall dan F-1 score:

- 1) Accuracy menggambarkan seberapa akurat model dalam mengklasifikasikan dengan benar.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (2.1)$$

- 2) Precision menggambarkan akurasi antara data yang diminta dengan hasil prediksi yang diberikan oleh model.

$$Precision = \frac{(TP)}{(TP + FP)} \quad (2.2)$$

- 3) Recall atau sensitivity: menggambarkan keberhasilan model dalam menemukan kembali sebuah informasi.

$$Recall = \frac{(TP)}{(TP + FN)} \quad (2.3)$$

- 4) F-1 Score menggambarkan perbandingan rata-rata precision dan recall yang dibobotkan. *Accuracy* tepatnya digunakan sebagai acuan performansi algoritma jika dataset memiliki jumlah data *False Negatif* dan *False Positif* yang sangat mendekati (*symmetric*). Namun jika jumlahnya tidak mendekati, maka sebaiknya menggunakan *F1 Score* sebagai acuan.

$$F - 1 \text{ SCORE} = \frac{(2 * Recall * Precision)}{(Recall + Precision)} \quad (2.4)$$

Data yang angka-angka pada diagonal dari kiri ke kanan bawa adalah hasil prediksi yang benar, dan angka-angka di luar diagonal adalah hasil prediksi yang salah. Salah satu cara untuk melakukan klasifikasi adalah dengan menggunakan algoritma klasifikasi.

2.2.7. Algoritma klasifikasi

Algoritma klasifikasi merupakan algoritma yang banyak dipakai untuk keperluan klasifikasi pada bidang ilmu *machine learning*. Algoritma klasifikasi biasanya digunakan sebagai metode perhitungan untuk melakukan prediksi data baru dengan karakteristik data lama. Algoritma klasifikasi juga biasanya merujuk kepada rumus atau fungsi matematika yang juga mempunyai kemampuan prediksi dengan mengklasifikasi data baru ke dalam kelas tertentu. Beberapa contoh

Algoritma klasifikasi adalah *linear classifier*, *support vector machine*, *quadratic classifier*, *kernel estimation*, *boosting*, *decision tree*, *neural networks*, dan *learning vector quantization*. Semua algoritma klasifikasi mempunyai keuntungan dan kerugian masing–masing, tergantung dari macam-macam faktor seperti tipe data set, jumlah kelas dan lain–lain (Dutta dkk., 2020).

2.2.8. Klasifikasi Linear

Pada klasifikasi statistik, target atau objektif dari klasifikasi adalah menentukan atau mengidentifikasi ke mana sebuah objek dimasukkan dalam beberapa grup atau kelas. Hal ini dilakukan dengan menggunakan kombinasi linear dari beberapa karakteristik atau fitur sebuah objek yang diolah menjadi fungsi prediksi linear untuk setiap kelas yang ada. Setelah itu objek baru akan ditentukan untuk masuk ke kelas tertentu berdasarkan hasil skor dengan cara memasukkan karakteristik objek baru ke dalam fungsi *predictor*. Nilai fungsi yang didapat dari proses klasifikasi ini akan menentukan kelas dari objek setelah diinterpretasi. Salah satu keunggulan dari klasifikasi linear adalah kecepatannya dalam melakukan klasifikasi terutama untuk objek yang hanya mempunyai dua kelas. Klasifikasi linear bisa dicari dengan menggunakan model generatif dan model diskriminatif, yang membedakan model generatif dan diskriminatif adalah prosedur pelatihan yang digunakan untuk mencari fungsi diskriminan dan bagaimana hasil skor dari fungsi diskriminan tersebut diinterpretasikan. Contoh algoritma yang digunakan dalam model diskriminatif adalah metode *linear discriminant analysis* (Trajdos dan Burduk, 2021).

2.2.9. Linear Discriminant Analysis (LDA)

Linear discriminant analysis, disingkat LDA adalah generalisasi diskriminan linear fisher, yaitu sebuah metode yang digunakan dalam ilmu statistika, pengenalan pola dan pembelajaran mesin untuk mencari kombinasi linear fitur yang menjadi ciri atau yang memisahkan dua atau beberapa objek atau peristiwa. Kombinasi yang diperoleh dapat dijadikan pengklasifikasi linear atau biasanya digunakan untuk proses reduksi dimensionalitas sebelum pengklasifikasian. Tujuan dari *Linear Discriminant Analysis* (LDA) adalah mengklasifikasikan objek ke dalam beberapa kelas berdasarkan ciri yang

menggambarkan objek tersebut, di dalam *Linear discriminant analysis* objek mempunyai dua variabel yaitu variabel kelas/terikat (*dependent variabel*) dan variabel atribut/bebas (*independent variabel*), variabel terikat mempunyai ikatan dengan variabel bebas yang menggambarkan variabel tersebut (Wei dan Tan, 2021).

Variabel bebas nantinya akan digunakan untuk menentukan kombinasi linear dari objek tersebut. LDA bekerja dengan menggunakan analisa matriks penyebaran yang bertujuan menemukan proyeksi optimal sehingga dapat memproyeksikan data input pada ruang dengan dimensi yang lebih kecil pada semua pola (*pattern*) dapat dipisahkan semaksimal mungkin. Dalam LDA variabel *dependent* merupakan kelas dari objek yang biasanya mempunyai nilai nominal/nama dari kelas dan variabel *independent* merupakan fitur yang menggambarkan objek tersebut biasanya bernilai skalar. Sebelum melakukan prediksi, LDA membutuhkan fase pelatihan untuk menentukan fungsi diskriminan. Fase pelatihan ini membutuhkan objek yang telah terklasifikasi beserta sejumlah variabel ciri/independen variabel. Menurut (Wei dan Tan 2021) Langkah-langkah dalam menghitung tingkat pelatihan dalam LDA dimulai dengan mengelompokkan pelatihan data ke dalam matriks sejumlah kelasnya dinotasikan dengan X_i , dengan i merupakan jumlah kelas. Setelah mengelompokkan pelatihan data dalam matriks, berikutnya perlu menghitung rata-rata matriks dari tiap kelas (μ_i). Kemudian menghitung nilai rata-rata global dari keseluruhan data (μ) matriks, setelah didapatkan nilai rata-rata global berikutnya perlu menghitung data terkoreksi rerata (*mean corrected data*/ X_i^0) dengan mengurangkan setiap nilai pada X_i dengan nilai rata-rata global (μ). Langkah berikutnya adalah menghitung *covariance* matriks (C_i) dari masing-masing X_i dengan rumus:

$$C_i = \frac{(X_i^0)^T X_i^0}{n_i} \quad (2.5)$$

Dengan,

n_i = banyak baris pada grup X_i

Setelah didapatkan kovarians matriks maka dapat dihitung nilai varians global matriks (C) dengan rumus :

$$C = \frac{1}{N} \sum_{i=1}^g n_i C_i \quad (2.6)$$

Dengan,

N = jumlah baris dari keseluruhan data.

Setelah itu hitung invers dari kovarians global matriks $C(C^{-1})$. Setelah didapatkan matriks inversnya, hitung peluang dari tiap kelas (P_i). Dengan begitu bisa menghitung fungsi diskriminan (f_i), dengan rumus :

$$f_i = \mu C^{-1} X_k^T - \frac{1}{2} \mu C^{-1} \mu_i^T + \ln(p_i) \quad (2.7)$$

Dengan,

f_i = fungsi discriminant

μ = rerata global

μ_i = rerata ciri dari grup i

Setelah setiap kelas diketahui, akurasi fungsi diskriminan (f_i) bisa dilakukan dengan melakukan pemetaan ulang tiap grup dalam pelatihan data dengan fungsi diskriminan (f_i). Objek dari data pelatihan akan masuk ke dalam kelas tergantung dari nilai maksimum. Setelah rumus diskriminan di ketahui dan di uji akurasinya, maka tingkat pelatihan bisa dinyatakan selesai, dan fungsi dikriminan ini sudah bisa digunakan untuk mengklasifikasikan objek baru ke dalam kelas yang ditentukan.

Dengan menghitung X_k^T dari objek baru dan kemudian memasukkannya ke dalam fungsi diskriminan masing–masing kelas. Objek baru akan di kelompokkan ke dalam kelas yang diwakili oleh fungsi diskriminan yang mempunyai nilai terbesar.

2.2.10. Correlation Coefficient

Correlation coefficient adalah nilai yang digunakan untuk mengukur korelasi antara dua pasang set data. Nilai dari korelasi ini bisa dikategorikan menjadi korelasi positif, korelasi negatif dan tanpa korelasi (Wei dan Tan, 2021).

Korelasi positif terjadi jika satu variabel naik, maka variabel lainnya juga akan ikut naik, dan begitu pun sebaliknya. Korelasi negatif terjadi jika satu variabel naik, maka variabel lainnya turun, dan begitupun sebaliknya. Sedangkan keadaan tidak adanya korelasi terjadi jika kenaikan variabel tidak mempengaruhi variabel lainnya sama sekali.

Adapun untuk menghitung koefisien korelasi dapat digunakan persamaan berikut: (2.8)

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}}$$

Dengan,

r = nilai koefisien korelasi

\bar{x} = nilai rata-rata x

\bar{y} = nilai rata-rata y

Sekolah Pascasarjana