

BAB II

LANDASAN TEORI

2.1 Pertemanan

Kata pertemanan diambil dari akar katanya “teman” yaitu yang bermakna orang dekat yang akan diundang/orang yang bersama-sama bekerja (berbuat, berjalan)/ lawan (bercakap-cakap) (KBBI, 2017). Pertemanan terdiri dari 1 suku kata, kata tersebut mempunyai 1 kata terkait. Sedangkan kata pertemanan sendiri telah mengalami penambahan berupa imbuhan pe-an, yang memiliki arti sesuatu yang lebih pada sebuah aktivitas dari kata dasarnya.

Pertemanan adalah suatu bentuk kedekatan hubungan yang meliputi kesenangan, penerimaan, kepercayaan, penghargaan, bantuan yang saling menguntungkan, saling mempercayai, pengertian, dan spontanitas (Santrock, 2002). Pertemanan didefinisikan sebagai keintiman, kebersamaan, dan hubungan diadik yang secara konseptual dan metodologis berbeda dari penerimaan teman sebaya (Franco & Levitt, 1998). Pertemanan merupakan tahapan perkembangan yang penting selama rentang kehidupan (Hartup & Stevens, 1999), khususnya pada masa remaja (Demir & Urberg, 2004).

2.2 Jejaring Sosial (*Social Network*)

Social Networking merupakan sebuah bentuk layanan *internet* yang ditujukan sebagai komunitas *online* bagi orang yang memiliki kesamaan aktivitas, ketertarikan pada bidang tertentu, atau kesamaan latar belakang tertentu. *Social networking* lazim disebut sebagai jaringan pertemanan. Yang berbasis internet, Layanan *social network* biasanya berbasis *web*, dilengkapi dengan beragam fitur bagi penggunaannya agar dapat saling berkomunikasi dan berinteraksi.

Ide Jejaring Sosial muncul untuk menghubungkan orang-orang dari seluruh belahan dunia. Kehadiran jejaring sosial diawali dengan munculnya “Sixdegrees.com” pada tahun 1997 sebagai situs jejaring sosial pertama di dunia.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Tahun 1999 dan 2000 muncul situs jejaring sosial bernama Lunarstorm, Live Journal, dan Cyword sengan sistem informasi searah.

Tahun 2011 muncul lagi jejaring sosial baru bernama Google plus yang dibuat oleh perusahaan internset Google. Jejaring sosial ini sepertinya sengaja dibuat untuk menyaingi situs milik Mark Zukerberg (Facebook). Dengan tampilan dan fitur yang sangat mirip dengan Facebook, Google plus disebut-sebut akan menjadi jejaring sosial populer.

2.3 Facebook

Pada tahun 2006 kehadiran Facebook yang didirikan oleh Mark Zukerberg menggeser posisi jejaring sosial pada saat itu. Facebook yang telah diluncurkan pada tahun 2004 telah memiliki 750 juta pengguna.

Facebook adalah sebuah situs web yang dimiliki Mark Zukerberg, penggunaan Facebook merupakan suatu hal yang sudah lumrah dikalangan masyarakat Indonesia saat ini, bahkan hampir semua masyarakat telah mendaftarkan dan memiliki akun Facebook, pada sebahagian lainnya bahkan memiliki lebih dari satu akun.

Facebook merupakan media jejaring sosial yang memungkinkan pengguna untuk menambah teman, mengirim pesan, dan memperbaharui profil pribadi yang dapat dilihat oleh semua orang (Haase dan Young, 2010). Penggunaan utama Facebook mengarah pada agaimana menjaga hubungan dengan orang-orang di sekitar atau mungkin dengan orang yang sudah lama tidak berjumpa sehingga tetap menjalin komunikasi tanpa batas ruang, jarak dan waktu. Pengelola Facebook pun juga aktif berinovasi dalam hal pengembangan Facebook seperti kemasam tampilan agar pengguna mudah dalam menggunakan, pengaturan privasi dan sebagainya. Facebook juga menyediakan fitur *group* yang dapat dijadikan pengguna untuk bergabung dengan orang yang memiliki tujuan tertentu berdasarkan karakteristik seperti tempat kerja, sekolah, perguruan tinggi atau karakteristik lainnya (Rizal dalam Mulkan et.al, 2011).

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

2.3.1 Data profil pengguna Facebook

Berikut ini akan dijelaskan beberapa data yang ada pada data profil untuk pengguna jejaring sosial Facebook.

- 1) Nama Akun Pengguna (*Username*)
Hal ini dilakukan agar dapat membedakan antara akun yang satu dengan akun yang lain, karena berisi data yang unik dan pasti berbeda pada setiap akun penggunanya.
- 2) Nama Akun
Pada tiap-tiap pengguna memiliki data nama, sebagai panggilan atau fungsi penanda dan pencarian agar lebih mudah ditemukan.
- 3) Tanggal lahir
Berisi tentang data kelahiran, hal ini biasanya untuk melihat rentang usia dari pengguna, data kelahiran juga bisa disembunyikan pada pengaturan privasi penggunanya.
- 4) Jenis kelamin
Ini berisi tentang jenis kelamin penggunanya.
- 5) Alamat.
Berisi tentang alamat tinggal dari penggunanya
- 6) Sekolah
Berisi tentang alamat sekolah dan nama sekolah yang pernah atau yang sedang ditempuh.
- 7) Kesukaan
Berisi data seputar hal-hal yang disukai, termasuk masalah halaman dan publik figur.
- 8) Buku
Berisi tentang informasi dan buku kesukaan.
- 9) Status Hubungan
Berisi data hubungan terhadap akun pengguna lainnya, biasanya dalam hal keluarga, pernikahan dan sebagainya.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

10) Persinggahan

Berisi data tempat tinggal sementara atau tempat yang pernah dikunjungi.

2.4 Rekomendasi dan Sistem Rekomendasi

Menurut kamus besar bahasa Indonesia (KBBI) rekomendasi adalah sebuah saran atau sebuah pengarahan dalam memilih sesuatu berdasarkan aspek tertentu. Rekomendasi juga memungkinkan untuk mengambil sebuah keputusan berdasarkan pertimbangan yang sangat kompleks, beberapa rekomendasi bahkan memiliki parameter pertimbangan yang sangat tinggi, sehingga dalam melakukan rekomendasi akan menemui hasil yang memuaskan.

Sistem rekomendasi adalah suatu program yang melakukan prediksi sesuatu item, seperti rekomendasi film, musik, buku, berita, termasuk pula rekomendasi subjek berupa orang dan lain sebagainya. Sistem ini berjalan dengan mengumpulkan data dari user secara langsung maupun tidak.

Pengumpulan data secara langsung dapat dilakukan sebagai berikut :

1. Meminta *user* untuk melakukan *rating* pada sebuah *item*.
2. Meminta *user* untuk melakukan ranking pada *item* favorit setidaknya memilih satu *item* favorit.
3. Memberikan beberapa pilihan *item* pada *user* dan memintanya memilih yang terbaik.
4. Meminta user untuk mendaftar *item* yang paling disukai atau *item* yang tidak disukainya.

Pengumpulan data dengan tidak langsung berhubungan dengan seorang *user*, dilakukan dengan cara seperti berikut:

1. Mengamati *item* yang dilihat oleh seorang user pada sebuah akun *web e-commerce*.
2. Mengumpulkan data transaksi pada sebuah toko *online*.

Data hasil pengumpulan, kemudian dilakukan perhitungan dengan Metode tertentu yang kemudian hasil tersebut dikembalikan lagi kepada user sebagai sebuah rekomendasi item dengan parameter dari user tersebut. Sistem rekomendasi juga merupakan salah satu alternatif sebagai mesin pencari suatu *item* yang dicari oleh *user*.

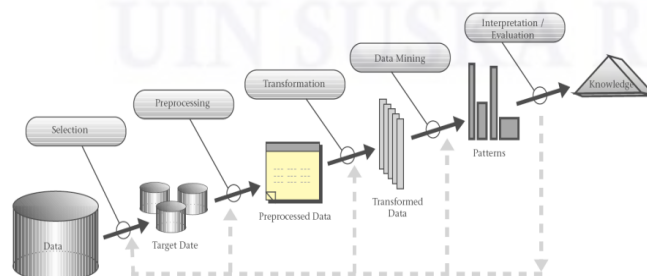
2.5 Data Mining

Data *mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Data mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Turban, dkk. 2005).

2.5.1 Hubungan Data Mining dan Knowledge Data Discovery

Data mining bukanlah teknologi utuh yang berdiri sendiri., namun istilah *data mining* lebih dikenal dibandingkan dengan *knowledge data discovery* (KDD), terutama para pelaku bisnis dan pelajar. Pada aplikasinya, sebenarnya *data mining* merupakan bagian dari proses *KDD*. Sebagai komponen dalam *KDD* adalah *data mining* terutama berkaitan dengan ekstraksi dan penghitungan pola-pola dari data yang dianalisa.

Proses KDD secara garis besar dapat di jelaskan sebagai berikut :



Gambar 2.1 Hubungan KDD dengan proses data mining (Fayyad, 1996)

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- 1) Pemahaman terhadap domain dari aplikasi, relevansinya terhadap pengetahuan yang ada dan *goal* dari *end-user*.
- 2) Menciptakan himpunan data target: pemilihan himpunan data, atau memfokuskan pada subset variable atau sampel data, dimana penemuan akan dilakukan.
- 3) Pemrosesan pendahuluan dan pembersihan data: operasi dasar seperti penghapusan *noise* dilakukan.
- 4) Proyeksi dan pengurangan data: pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai.
- 5) Pemilihan tugas *data mining*: pemilihan *goal* dari proses KDD misalnya klasifikasi, regresi, *clustering*, dan lain-lain.
- 6) Pemilihan algoritma *data mining* untuk pencarian.
- 7) *Data mining* pencarian pola-pola yang diinginkan di himpunan representasi. Penterjemahan pola-pola yang dihasilkan dari *data mining* langkah ini, kemungkinan dapat kembali langkah awal untuk iterasi lebih lanjut.
- 8) Konsolidasi pengetahuan yang ditemukan: pendokumentasian hasil, pencarian penyelesaian apabila ada konflik dengan pengetahuan yang telah dipercaya sebelumnya.

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak di duga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data (Larose, 2005).

Data mining dapat di bagi menjadi beberapa tahapan keempat dalam proses data KDD, untuk setiap data sehingga menghasilkan output yang baik. Adapun tahapan tersebut bersifat interaktif dimana pemakai terlibat langsung atau dengan perantara basis pengetahuan. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. *Data selection*

Pemilihan seleksi data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Dan data hasil seleksi yang akan digunakan untuk proses data *mining* disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing*

Yang menjadi fokus dalam data *mining* adalah proses *cleaning* sebelum dilakukannya data mining. Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data. Tahapan ini *cleaning* dilakukan dengan tujuan untuk mendapatkan data yang konsisten.

3. *Transformation*

Transformasi data menggunakan reduksi dimensional atau metode transformasi. Pada tahap ini data input yang digunakan terlebih dahulu dinormalisasikan agar data disesuaikan dengan kebutuhan algoritma agar sebaran datanya tidak terlalu jauh.

4. Normalisasi Data

Untuk mencari normalisasi data dapat menggunakan rumus berikut ini :

$$v^i = \frac{v - \min_a}{\max_a - \min_a} (new_max_a - new_min_a) + new_min_a \dots\dots\dots (2. 1)$$

Keterangan :

- v^i : Data baru setelah normalisasi
- v : Data sebelum normalisasi
- new_max_a : Batas nilai max baru adalah 1
- new_min_a : Batas nilai min baru adalah 0
- max_a : Nilai maximum pada kolom
- min_a : Nilai minimum pada kolom

5. *Data Mining* dan teks *mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu, teknik atau metode dalam data *mining* sangatlah bervariasi.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Text Mining (Penambangan teks) adalah proses ekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, dll. Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan. Penambangan teks dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan penambangan data. Bentuk proses yang umum dilakukan oleh penambangan teks diantaranya adalah perangkuman otomatis, kategorisasi dokumen, penggugusan teks (Turban, 2011, dikutip oleh www.wikipedia.org, 2013)

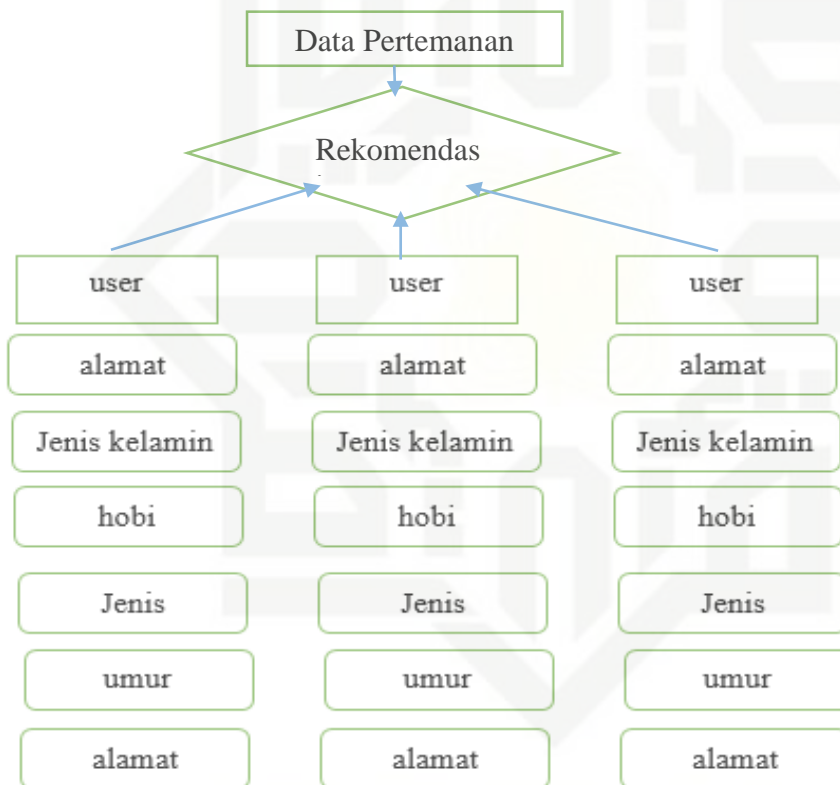
Pada dasarnya *text mining* merupakan bidang interdisiplin yang mengacu pada perolehan informasi (*information retrieval*), *data mining*, pembelajaran mesin (*mechine learning*), statistik, dan komputasi linguistik (Jiawei dkk, 2012 dikutip oleh Joni Iskandar, 2015). Tujuan utama dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan pada *text mining* adalah teks yang tidak terstruktur atau minimal semi terstruktur (Eldira, 2010). Aplikasi yang paling umum dilakukan *text mining* saat ini misalnya penyaringan *spam*, analisis sentimen, mengukur preferensi pelanggan, meringkas dokumen, pengelompokan topik penelitian dan banyak lainnya.

6. *Interpretation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

2.6 Klasifikasi Modified K-Nearest Neighbor

Klasifikasi adalah proses untuk menemukan sebuah model berdasarkan kelas-kelas yang digunakan sebagai pembeda antara kelas satu dengan kelas yang lain. *Decision tree* adalah penerapan metode klasifikasi yang paling populer, dengan metode ini sebuah *item* dapat dikelompokkan dan dimodelkan pada sebuah pohon keputusan, sehingga dapat dengan mudah dimengerti. Contoh *decision tree* dapat dilihat pada gambar berikut yang menerangkan bagaimana user biasa dipilih oleh sebagai saran rekomendasi.



Gambar 2.2 Skema Rekomendasi (Titania Dwi Andini, 2013)

Modified K-Nearest Neighbor adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan Metode pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan *learning* biasa yang berasal dari teori pembelajaran *statistic*.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Menurut Kusriani dan Luthfi (2009 : 93), K- Nearest Neighbor merupakan Metode pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama yaitu berdasarkan pencocokan bobot dari sejumlah atribut yang ada dan memiliki kriteria kesamaan (*similarity*). Tujuan dari Metode ini untuk mengklasifikasikan objek baru berdasarkan atribut dan *training sample*. *Classifier* tidak menggunakan model apapun untuk dicocokkan tetapi berdasarkan pada memori.

Ide utama dari metode ini adalah hal pertama yang dilakukan adalah perhitungan validitas untuk semua data yang terdapat pada data latih. Selanjutnya, dilakukan perhitungan *Weight Voting* pada semua data uji menggunakan validitas data (Parvin,dkk : 2010). Berikut tahapan dalam metode Modified K-Nearest Neighbor :

- a. Menentukan nilai k dan untuk nilai k yang baik ialah nilai bilangan ganjil agar mudah melakukan *voting*
- b. Hitung nilai validitas
- c. Hitung *weight voting*
- d. Menentukan kelas mayoritas dari k buah data latih dengan *weight voting* tertinggi
- e. Menghasilkan model klasifikasi untuk menentukan kelas pada data uji. Output yang di dapat adalah hasil klasifikasi berupa rekomendasi pertemanan.

2.6.1 Euclidean Distance

Jarak *Euclidean* paling sering digunakan untuk menghitung jarak. Jarak *euclidean* berfungsi untuk menguji ukuran yang dapat digunakan sebagai kedekatan jarak antara dua objek. Berikut persamaan yang di gunakan untuk memperoleh nilai jarak *Euclidean* (Parvin et al,2010) :

$$d = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \dots\dots\dots (2.2)$$

Hak Cipta Dilindungi Undang-Undang
 1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Keterangan :

- d : Jarak dari data p ke q
- pi : elemen ke-i dari data p
- qi : elemen ke-i dari data q
- n : jumlah elemen dari data p dan data q

Kasus khusus dimana klasifikasi diprediksikan berdasarkan data pembelajaran yang paling dekat (dengan kata lain, k = 1) disebut Metode nearest neighbor (Kusrini dan Luthfi E.T, 2009)

Nearest Neighbor didefinisikan berdasar persamaan sebagai berikut :

$$Similarity(problem, case) = \frac{S1*W1+S2*W2+\dots+Sn*Wn}{W1+W2+\dots+Wn} \dots\dots\dots (2.3)$$

Keterangan :

- S : similarity (nilai kemiripan)
- W : weight (bobot yang diberikan)

Prinsip kerja K-Nearest Neighbor(KNN) adalah mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan. perhitungan untuk mencari jarak dengan (d) adalah jarak dan (p) adalah dimensi data :

$$d_i = \sqrt{\sum_{l=1}^p (x_{2l} - x_{1l})^2} \dots\dots\dots (2.4)$$

Dengan :

- x1 : sampel data
- x2 : data uji
- i : variabel data
- d : jarak
- p : dimensi data

Untuk perhitungan kemiripan kasus dapat menggunakan rumus sebagai berikut.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$\text{Similarity (T, S)} = \frac{\sum_{i=1}^n f(T_i, S_i)}{w_i} \dots\dots\dots(2.5)$$

Keterangan :

- p : kasus baru
- q : kasus yang ada dalam penyimpanan
- n : jumlah atribut dalam tiap kasus
- i : atribut individu antara 1 sampai dengan n
- f : fungsi similarity atribut i antara kasus p dan q
- w : bobot yang diberikan pada atribut ke-i

Ide utama dari metode ini adalah hal pertama yang dilakukan adalah perhitungan validitas untuk semua data yang terdapat pada data latih. Selanjutnya, dilakukan perhitungan *Weight Voting* pada semua data uji menggunakan validitas data (Parvin,dkk : 2010). Berikut tahapan dalam metode Modified K-Nearest Neighbor :

1. Menentukan nilai k
 2. Hitung nilai validitas
 3. Hitung *weight voting*
 4. Menentukan kelas mayoritas dari k buah data latih dengan *weight voting* tertinggi
 5. Menghasilkan model klasifikasi untuk menentukan kelas pada data uji.
- Output yang didapat adalah hasil klasifikasi dari data rekomendasi teman.

2.6.2 Validitas Data Training

Dalam Metode Modified K-Nearest Neighbor setiap data latih akan melalui tahap validasi terlebih dahulu. Validasi setiap titik dihitung sesuai dengan tetangganya dan dilakukan hanya satu kali. Nilai validasi dari data latih inilah nantinya yang digunakan untuk melakukan pembobotan pada data latih yang akan digunakan untuk menentukan kelas suatu data *testing*. Untuk memvalidasi sebuah

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

data latih, harus ditentukan terlebih dahulu parameter K yang akan digunakan. Parameter K ini merupakan sebuah nilai yang merepresentasikan jumlah tetangga yang digunakan untuk melakukan proses validasi. Diantara K tetangga terdekat dari data latih x, nilai validasi $validity(x)$ setiap data latih dilakukan dengan menghitung jumlah data yang memiliki label yang sama dengan label x. Berikut ialah rumus untuk menghitung nilai validasi setiap data latih:

$$validity(x) = \frac{1}{k} \sum_{i=1}^k S(lbl(x), lbl(Ni(y))) \dots\dots\dots(2.6)$$

- Keterangan :
- Validity(x) : validitas data ke-x
 - K : jumlah data tetangga terdekat
 - Lbl(x) : kelas dari data latih ke-x
 - Lbl(y) : kelas dari data latih terdekat dari x
 - S : fungsi similaritas data

Fungsi S memperhitungkan kesamaan antara satu data latih dan data latih lainnya. Berikut ialah definisi dari fungsi S:

$$S(a,b) = \begin{cases} 1 & a=b \\ 0 & a \neq b \end{cases} \dots\dots\dots(2.7)$$

- Keterangan :
- a : kelas a pada data latih
 - b : kelas lain selain a pada data latih

2.6.3 Weight voting

Dalam metode MKNN, pertama *weight* masing-masing tetangga di hitung dengan menggunakan $1/(de+0,5)$. Kemudian validitas dari setiap data pada data latih dikalikan dengan *weight* berdasarkan pada jarak *Euclidean*.

Weight voting berpengaruh terhadap data yang mempunyai nilai validitas lebih tinggi dan paling dekat dengan data. Adapun persamaan *weight voting* (Parvin et al, 2010) adalah sebagai berikut:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

$$W(i) = \text{validity}(i) \times \frac{1}{de+0,5} \dots\dots\dots(2.8)$$

Keterangan :

W(i) : Perhitungan weight voting

Validity(i) : Nilai Validitas

de : Jarak Euclidean

2.6.4 Evaluasi

Evaluasi merupakan proses untuk mengetahui tingkat keberhasilan klasifikasi dari sistem yang dibuat dengan menentukan nilai akurasi. *Confusion matrix* adalah sebuah metode yang dibuat untuk menganalisa tingkat kesuksesan dari suatu model klasifikasi dalam mengenali data kelas-kelas yang berbeda (M. K. Jiawei Han, 2012). *Confusion matrix* dilakukan untuk mengetahui kemampuan dari sebuah model klasifikasi dengan pengukuran tingkat akurasi seberapa baik *classifier* dalam melakukan klasifikasi. Akurasi dari model klasifikasi yang di buat dapat di hitung dengan persamaan berikut:

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \dots\dots\dots(2.9)$$

Keterangan :

TP : Jumlah data yang diklasifikasikan dengan benar

TN : Jumlah data yang diklasifikasikan dengan benar

FN : Jumlah data yang diklasifikasikan dengan salah

FP : Jumlah data yang diklasifikasikan dengan benar

P : Jumlah dari TP dan FN

N : Jumlah dari FP dan TN