BAB II

TINJAUAN PUSTAKA

II.1. Penelitian Terkait

Pada penelitian berjudul “Implementation of the K-Nearest Neighbor (KNN) Method for Analysis FLIP Financial Technology Application User Satisfaction Sentiment”,penelitian dilakukan oleh (Rahayu et al. 2022), menggunakan metode yaitu Studi Literatur, Crawling, Translate, Preprocessing, Labeling, Split Data, Supervised Learning, dan Evaluasi. Data yang digunakan dalam penelitian ini merupakan data ulasan atau review yang diberikan oleh pengguna Flip melalui situs Google Play Store. Hasil pengujian dan pembahasan yang dilakukan, maka dapat diperoleh pengguna terhadap aplikasi Flip menunjukkan hasil yang positif seperti rating yang diterima di Google Play Store. Selain itu, algoritma K-Nearest Neighbor dengan pembobotan TF-IDF layak digunakan untuk proses analisis sentimen pengguna terhadap aplikasi Flip yang ditunjukkan oleh tingkat akurasi klasifikasi sebesar 76,68%.

Pada penelitian berjudul “Confirming Customer Satisfaction With Tones of Speech”, penelitian dilakukan oleh (Ko et al. 2022), menggunakan metode yaitu desain penelitian, protokol pengumpulan data, mfcc processing, LSTM model construction, SVM model Construction, dan Auto-encoder Neural Networks. Data yang digunakan dalam penelitian ini merupakan data MFCC diekstraksi dari data suara sebagai fitur; karena ukuran dataset terbatas, Auto Encoder digunakan untuk lebih mengurangi fitur suara; model

berdasarkan Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNNs) dan Support Vector Machine (SVM) dibangun untuk memprediksi kepuasan. Hasil penelitian ini mengembangkan model diskriminatif untuk mendeteksi tingkat kepuasan pelanggan dengan menganalisis informasi implisit yang terkandung dalam nada suara pelanggan. Metodologi studi melibatkan tiga langkah: ekstraksi fitur, pengurangan fitur, dan konstruksi model. Fitur yang diekstraksi melalui teknik berbasis MFCC digabungkan dengan fitur prosodik yaitu voice pitch dan voice intensity, untuk mendapatkan vektor fitur 45 dimensi. AE digunakan untuk mengurangi fitur untuk konstruksi model. Vektor fitur yang direkonstruksi kemudian diumpankan ke SVM dan LSTM-RNN untuk menyusun model diskriminatif.

Pada penelitian berjudul "Analysis of Visitor Satisfaction Levels Using the K-Nearest Neighbor Method" penelitian dilakukan oleh (Violita et al. 2023), menggunakan metode penelitian yaitu Metode K-Nearest Neighbor. Data yang digunakan dalam penelitian ini ialah 45 data pengunjung. Hasil Penelitian ini dilakukan dengan proses klasifikasi menggunakan metode K-Nearest Neighbor yang akan diproses dalam data mining. Setelah dilakukan klasifikasi dengan metode K-Nearest Neighbor, didapatkan hasil bahwa terdapat 41 pengunjung yang puas di Suzuya Mall Rantauprapat dan hasil yang didapatkan dari klasifikasi dengan metode K-Nearest Neighbor adalah 41 pengunjung. Hasil tersebut menunjukkan bahwa banyak pengunjung yang puas dengan Suzuya Mall Rantauprapat. Artinya Suzuya Mall Rantauprapat dapat memberikan kepuasan kepada pengunjung, mulai dari lokasi yang

mudah dijangkau, sarana dan prasarana yang lengkap, tempat parkir yang luas, pelayanan staf yang ramah serta tempat yang bersih dan bersih. Parameter ini memiliki hasil positif dari pengunjung.

Pada penelitian berjudul "Comparison Of Prediction Analysis Of Gofood Service Users Using The Knn & Naive Bayes Algorithm With Rapidminer Software", penelitian dilakukan oleh (Yuliarina and Hendry 2022), menggunakan metode penelitian yaitu algoritma KNN dan Naive Bayes. Data yang digunakan yaitu data kepuasan pelanggan yang akan digunakan untuk melakukan prediksi kepuasan pengguna layanan GoFood. Hasil penelitian pengujian yang dilakukan, bahwa algoritma KNN merupakan algoritma yang terbaik dibandingkan dengan Naive Bayes, walaupun hasil Accuracy yang diperoleh dari kedua algoritma tersebut keduanya memiliki hasil dengan nilai yang baik. Namun KNN merupakan algoritma terbaik untuk digunakan dalam prediksi tingkat kepuasan pengguna layanan GoFood pada mahasiswa, dengan tingkat Accuracy sebesar 98,80% dan Recall 100% sedangkan Naive Bayes memperoleh nilai Accuracy sebesar 94,10% dan Recall sebesar 94,43%.

Pada penelitian berjudul "An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students During the COVID-19", penelitian dilakukan oleh (Abdelkader et al. 2022), menggunakan metode penelitian yaitu 11 algoritma FS berbasis pembungkus, K-NN dan SVM. Data yang digunakan yaitu kumpulkan melalui kuesioner yang dirancang khusus untuk menentukan bagaimana siswa

dipengaruhi oleh OL. Hasil penelitian menunjukkan bahwa akurasi keseluruhan berdasarkan fitur yang dipilih telah ditingkatkan masing-masing sebesar 2% dan 8% untuk k-NN dan SVM, dibandingkan dengan menggunakan semua fitur; akurasi rata-rata keseluruhan untuk k-NN dan SVM mencapai 100% dengan algoritma FS. Algoritma SFO dengan k-NN dan SVM memiliki performa terbaik dalam hal kemampuan eksplorasi dan eksploitasi (fitness). Itu hanya menentukan empat fitur. Kami menyimpulkan bahwa empat fitur (bukan 20 fitur) memengaruhi SSL dan cukup untuk memprediksi SSL dengan OL dengan akurasi prediksi 100%. Fitur minimal namun krusial yang dipilih adalah: "Ceramah disajikan dengan gaya yang menarik", "Metode pengajaran dalam kursus ini mendorong saya untuk berpartisipasi aktif selama kelas", "Kuis telah disiapkan dalam derajat", dan"Siswa dilatih tentang cara menyelesaikan ujian online dengan merancang kuis eksperimental". Ini dapat membantu PT untuk memprediksi SSL pada tahap awal dan menyajikan diagnosis dan terapi untuk menghindari hambatan dalam proses pendidikan dan mencapai hasil yang paling signifikan selama krisis akut seperti COVID-19. Di masa mendatang, karena model Random Forest (RF) dapat dengan sempurna menyesuaikan hubungan input-output dengan kompleksitas tinggi yang tidak terbatas, model ini dapat dicoba dengan 11 metode FS pada dataset real-time yang diusulkan.

Pada penelitian berjudul “Predicting Airline Customer Satisfaction using k-nn Ensemble Regression Models”, penelitian dilakukan oleh (García et al. 2019), menggunakan metode penelitian yaitu Algoritma KNN dan

Model Ensemble BAGGING. Data yang digunakan yaitu menggunakan database nyata dari 129.890 sampel dari perusahaan penerbangan, untuk memverifikasi manfaat model ansambel untuk memprediksi kepuasan pelanggan. Hasil penelitian menampilkan rata-rata RMSE dan MAE selama lima putaran. Untuk setiap model prediksi, telah memplot hasil untuk pelajar dasar (3-nn) dan juga hasil model regresi ansambel 3- nn saat memvariasikan jumlah subsampel bootstrap dari D. Perhatikan bahwa garis sejajar dengan sumbu X sesuai dengan kasus base learner, yang menunjukkan bahwa hasil yang dicapai dengan belajar langsung dari training set D, model regresi individu mencapai nilai error tertinggi (y 0,7985, ÿ 0,6365), sedangkan regresi ansambel model muncul sebagai model pembelajaran dengan kesalahan keseluruhan terendah (y 0,5664 ketika model angka atau regresi ditetapkan ke 30). Ini adalah perbedaan antara kedua model y 5% saat RMSE digunakan dan ÿ 11% dalam kasus MAE. Di sisi lain, saat memvariasikan jumlah bootstrap dari baik tingkat RMSE dan MAE menurun seiring bertambahnya jumlah subsampel bootstrap, meskipun konfigurasi pengklasifikasi dasar tidak bervariasi di sepanjang subsampel bootstrap yang berbeda. Dari hasil yang dilaporkan di sini, tampak bahwa masalah prediksi kepuasan pelanggan dapat ditangani dengan lebih baik menggunakan pendekatan ansambel daripada model tunggal. Demikan pula ketika jumlah model dasar bertambah, dan kesalahan berkurang.

Pada penelitian berjudul “The K-NN Method Was Used To Assess Student Satisfaction With The Services Provided By Employees Of Research

And Service Institutions" peneliti dilakukan oleh (Wathani et al. 2021), menggunakan metode penelitian yaitu metode algoritma K-Means. Data yang digunakan yaitu data yang diperoleh dari data sensus penduduk yang dilakukan oleh Badan Pusat Statistik (BPS). Hasil penelitian yaitu perhitungan klasifikasi kepuasan mahasiswa terhadap pegawai LPPM adalah klasifikasi kepuasan mahasiswa terhadap pegawai LPPM dengan menggunakan Algoritma K-NN (Nearest Neighbor) dapat diterapkan. Dari data yang telah diperoleh dapat ditentukan hasil prediksi kepuasan mahasiswa untuk data selanjutnya adalah "PUAS" atau mahasiswa puas terhadap pelayanan yang diberikan pegawai LPPM. Dari perhitungan yang telah dilakukan, KNN lebih cocok dengan jumlah data yang banyak atau memiliki jumlah data yang banyak sehingga dapat memperoleh hasil yang lebih akurat.

Pada penelitian berjudul "Implementasi Algoritma Decision Tree dan Naive Bayes Untuk Klasifikasi Sentimen Terhadap Kepuasan Pelanggan Starbucks" peneliti dilakukan oleh (Asih et al. 2023), menggunakan metode penelitian yaitu algoritma Decision tree & Naive Bayes. Data yang digunakan yaitu data uji yang diperoleh dari tweet. Hasil penelitian ini dimana dari proses klasifikasi sentimen terhadap kepuasan pelanggan starbucks diperoleh kategori netral, dapat dilihat dari ulasan menggunakan kata kunci "starbuck OR starbucks OR #starbucks" didapatkan hasil yakni komentar positif sebanyak 476 tweet dengan jumlah pesentase sebesar 19,2%, komentar neutral sebanyak 1743 tweet dengan jumlah pesentase sebesar 70,3 % dan komentar negatif sebanyak 258 tweet dengan jumlah pesentase sebesar

10,4%, sehingga dapat ditarik kesimpulan berdasarkan perhitungan polarity tersebut ulasan komentar terhadap stabuck memiliki kategori puas. Pada penelitian ini dapat ditarik kesimpulan bahwa kinerja algoritma Decision Tree lebih baik dari pada algoritma Naive Bayes, dapat dilihat dari penjelasan berikut. Algoritma Decision Tree pada hasil akurasi nilai sebesar 83%. Sedangkan algoritma Naive Bayes pada hasil akurasi nilai sebesar 74%.

Pada penelitian berjudul "Perbandingan Akurasi Algoritma Naive Bayes, K-NN Dan SVM Dalam Memprediksi Penerimaan Pegawai", peneliti dilakukan oleh (Sinaga, Hayadi, and Situmorang 2022), menggunakan metode Naive Bayes, K-NN Dan SVM. Data yang digunakan yaitu melalui data jumlah pelamar. Hasil penelitian ini setelah menggunakan metode Naïve Bayes, K Nearest Neighbor (KNN) dan Supervise Vector Machine (SVM) untuk Memprediksi Kelulusan Pelamar di Politeknik Bisnis Indonesa didapatkan hasil hasil kinerja metode SVM lebih baik dari metode Naïve Bayes dan K-NN. Dengan 33 data uji yang digunakan diperoleh SVM memiliki nilai akurasi 84.9%, presisi 85.1% sedangkan K-NN memiliki nilai akurasi 81.8% , presisi 84.1% dan Naïve Bayes memiliki nilai akurasi 78.8% dan presisi 80.1%.

Pada penelitian berjudul "Classification Of Visitor Satisfaction at The Museum Using The Naïve Bayes Algorithm" , peneliti dilakukan oleh (Ayunda Dwi Agustin, Tacbir Hendro Pudjiantoro 2021), menggunakan metode Naive Bayes. Hasil penelitian dilakukan mengenai penerapan metode naive bayes untuk mengklasifikasi kepuasan pengunjung

terhadap pelayanan di PP-IPTEK dapat disimpulkan bahwa, metode naive bayes memanfaatkan data training untuk menghasilkan probabilitas seetiap kriteria untuk class yang berbeda, sehingga nilai tersebut dapat dioptimalkan untuk pengklasifikasian pengunjung daru dengan data testing yang sudah ada. Didapatkan tingkat akurasi tinggi yaitu sebesar 91,00%, dan mendapatkan tingkat kepuasan pengunjung terhadap pelayanan di PP-IPTEK sebesar 85,00%. Data training sangat berpengaruh pada hasil pengujian, karena data training dijadikan dasar penentuan kepuasan pelanggan. Maka dari itu untuk penelitian selanjutnya diharapkan data training atau atribut yang digunakan dalam penentuan kepuasan pelanggan dengan menggunakan metode naïve bayes jumlahnya lebih banyak karena dapat berpengaruh terhadap hasil akhir.

Adapun rekapitulasi penelitian terkait dapat dilihat pada **Tabel II.1** sebagai berikut :

Tabel II.1 Rekapitulasi Penelitian Terkait

No	Judul Penelitian & Peneliti	Metode	Data	Hasil Penelitian
1	Implementation of the K-Nearest Neighbor (KNN) Method for Analysis FLIP Financial Technology Application User Satisfaction Sentiment. Peneliti Oleh : Sri Rahayu, Yumarlin MZ, Jemmy Edwin Bororing, dan Rahmat Hadiyat	Studi Literatur, Crawling, Translate,Preprocessing,Labeling, SplitData,Supervised Learning, dan Evaluasi. (Metode Algoritma KNN)	Data Flip melalui situs Google Play Store	Hasil yang positif seperti rating yang diterima di Google Play Store. Selain itu, algoritma K-Nearest Neighbor dengan pembobotan TF-IDF layak digunakan untuk proses analisis sentimen pengguna terhadap aplikasi Flip yang ditunjukkan oleh tingkat akurasi klasifikasi sebesar 76,68%.
2	Confirming Customer Satisfaction With Tones of Speech. Peneliti Oleh : Yen Huei Ko, Ping-Yu Hsu, Yu-Chin Liu, And Po-Chiao Yang.	Desain Penelitian, Protokol Pengumpulan Data, MFCC Processing, LSTM Model Construction, SVM Model Construction, dan Auto-encoder Neural Networks	Data MFCC (Mel Frequency Cepstral Coefficients) diekstraksi dari data suara sebagai fitur.	Tingkat kepuasan dibagi menjadi tiga kategori: puas, netral, dan tidak puas. Model SVM yang dikembangkan dikombinasikan dengan AE mencapai akurasi tertinggi sebesar 79,31% dalam mengidentifikasi tingkat kepuasan peserta; itu mengungguli model LSTM-RNN dengan AE, yang mencapai akurasi 75,86%.

3	Analysis of Visitor Satisfaction Levels Using the K-Nearest Neighbor Method Peneliti Oleh : Putri Violita, Gomal Juni Yanris, Mila Nirmala Sari Hasibuan, dan Universitas Labuhanbatu	Metode K-Nearest Neighbor.	Data kuesioner 41 pengunjung	Klasifikasi dengan metode K-Nearest Neighbor adalah 41 pengunjung. Hasil tersebut menunjukkan bahwa banyak pengunjung yang puas dengan Suzuya Mall Rantauprapat. Artinya Suzuya Mall Rantau prapat dapat memberikan kepuasan kepada pengunjung, mulai dari lokasi yang mudah dijangkau, sarana dan prasarana yang lengkap, tempat parkir yang luas, pelayanan staf yang ramah serta tempat yang bersih dan bersih.
4	Comparison Of Prediction Analysis Of Gofood Service Users Using The Knn & Naive Bayes Algorithm With Rapidminer Software Peneliti Oleh : Agista Nindy Yuliarina, dan Hendry	Metode algoritma KNN dan Naive Bayes.	Data kepuasan pelanggan pengguna layanan GoFood.	Hasil KNN merupakan algoritma terbaik untuk digunakan dalam prediksi tingkat kepuasan pengguna layanan GoFood pada mahasiswa, dengan tingkat Accuracy sebesar 98,80% dan Recall 100% sedangkan Naive Bayes memperoleh nilai Accuracy sebesar 94,10% dan Recall sebesar 94,43%

5	An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students During the COVID-19 Peneliti Oleh : Hanan E. Abdelkader, Ahmed G., Amra. Abohany, And Shaymaa E. Sorour	11 algoritma FS berbasis pembungkus, K-NN dan SVM.	Data kuesioner ulasan siswa tentang kursus OL	Hasil penelitian menunjukkan bahwa akurasi keseluruhan berdasarkan fitur yang dipilih telah ditingkatkan masing-masing sebesar 2% dan 8% untuk k-NN dan SVM, dibandingkan dengan menggunakan semua fitur; akurasi rata-rata keseluruhan untuk k-NN dan SVM mencapai 100% dengan algoritma FS. Algoritma SFO dengan k-NN dan SVM memiliki performa terbaik dalam hal kemampuan eksplorasi dan eksploitasi (fitness).
6	Predicting Airline Customer Satisfaction using k-nn Ensemble Regression Models Peneliti Oleh : V. García1, R. Florencia-Juárez, J. P. Sánchez-Solís, G. Rivera-Zarate, R. Contreras-Masse	Algoritma KNN dan Model Ensemble BAGGING.	Data database nyata dari 129.890 sampel dari perusahaan penerbangan.	Hasil Model regresi individu mencapai nilai error tertinggi (y 0,7985, 0,6365), sedangkan regresi ansambel model muncul sebagai model pembelajaran dengan kesalahan keseluruhan terendah (y 0,5664 ketika model angka atau regresi ditetapkan ke 30).

				Perbedaan antara kedua model y 5% saat RMSE digunakan dan y 11% dalam kasus MAE. Dari hasil tampak bahwa masalah prediksi kepuasan pelanggan dapat ditangani dengan lebih baik menggunakan pendekatan ansambel daripada model tunggal. Demikan pula ketika jumlah model dasar bertambah, dan kesalahan berkurang.
7	The K-NN method was used to assess student satisfaction with the services provided by employees of research and service institutions Peneliti oleh : Muhammad Rais Wathani1, Dian Agustini, Muthia Farida, Mokhamad Ramdhani Raharjo	Algoritma K-Means	Data Sensus penduduk yang dilakukan oleh Badan Pusat Statistik (BPS).	Hasil penelitian Dari hasil perhitungan diketahui bahwa penentuan kelompok data uji berdasarkan label mayoritas dari k tetangga terdekat. Karena nilai k = 10 maka 10 diambil jarak terkecil. KNN lebih cocok dengan jumlah data yang banyak atau memiliki jumlah data yang banyak sehingga dapat memperoleh hasil yang lebih akurat.

8	Implementasi Algoritma Decision Tree dan Naive Bayes Untuk Klasifikasi Sentimen Terhadap Kepuasan Pelanggan Starbucks Peneliti Oleh : Tiyas Asih Qurnia Putri, Agung Triayudi, dan Rima Tamara Aldisa	Algoritma Decision Tree dan Naive Bayes	Data uji yang diperoleh dari tweet dengan kata kunci "Starbucks"	Hasil Algoritma Decision Tree pada hasil akurasi nilai sebesar 83%. Sedangkan algoritma Naive Bayes pada hasil akurasi nilai sebesar 74%.
9	Perbandingan Akurasi Algoritma Naive Bayes, K-NN Dan SVM Dalam Memprediksi Penerimaan Pegawai Peneliti Oleh : Novendra Adisaputra Sinaga, B. Herawan Hayadi , Zakarias Situmorang	Algoritma Naive Bayes, K-NN Dan SVM.	Data melalui data jumlah pelamar	Hasil penelitian dengan 33 data uji yang digunakan diperoleh SVM memiliki nilai akurasi 84.9%, presisi 85.1% sedangkan K-NN memiliki nilai akurasi 81.8% , presisi 84.1% dan Naïve Bayes memiliki nilai akurasi 78.8% dan presisi 80.1%.
10	Classification Of Visitor Satisfaction at The Museum Using The Naive Bayes Algorithm Peneliti Oleh : Ayunda Dwi Agustin, Tacbir Hendro Pudjiantoro, Puspita Nurul Sabrina	Algoritma Naive Bayes	Data yang digunakan pada penelitian yaitu data kuisioner dari tahun 2019.	Hasil didapatkan tingkat akurasi tinggi yaitu sebesar 91,00%, dan mendapatkan tingkat kepuasan pengunjung terhadap pelayanan di PP-IPTEK sebesar 85,00%.

II.2. Data Mining

Data mining adalah disiplin ilmu yang mempelajari cara mengekstraksi pengetahuan atau menemukan pola dari data. Data itu sendiri adalah fakta yang direkam yang tidak bermakna, dan pengetahuan adalah pola, aturan, atau model yang muncul dari data. Oleh karena itu, data mining disebut Knowledge Discovery in Database (KDD).

Analisis tinjauan koleksi, atau data mining, adalah analisis yang meninjau kumpulan data untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara yang berbeda dari sebelumnya, sehingga pemilik data dapat memahami dan menggunakannya.

Mengetahui jenis aplikasi data, memilih, membersihkan, mengintegrasikan, mengurangi, dan mengubah data adalah beberapa proses yang digunakan dalam proses data mining, yang juga interaktif dan iteratif. Selain itu, data mining juga menggunakan pengetahuan tentang proses yang dihasilkannya untuk memilih metode untuk menginterpretasikan hasilnya.(Fadli et al. 2022)

Menurut Larose dalam (Pahlevi et al., 2018) dalam bukunya yang berjudul “Discovering Knowledge in Data: An Introduction to Data Mining”, data mining dibagi menjadi beberapa kelompok berdasarkan tugas/pekerjaan yang dapat dilakukan, seperti sebagai:

1. Deskripsi

   Terkadang peneliti dan analis hanya ingin mencoba mencari cara untuk mendeskripsikan pola dan tren yang terkandung dalam data.

Deskripsi pola tren sering memberikan penjelasan yang mungkin untuk suatu pola atau tren.

2. Estimasi

   Estimasi hampir sama dengan klasifikasi, hanya saja variabel target yang diestimasi lebih numerik daripada kategorikal. Model dibangun dengan menggunakan data baris lengkap (record) yang memberikan nilai variabel target sebagai nilai prediktif. Selanjutnya pada ulasan selanjutnya dibuat estimasi nilai variabel target berdasarkan nilai variabel prediksi.

3. Prediksi

   Prediksi hampir sama dengan klasifikasi dan estimasi, hanya saja pada prediksi nilai hasil akan berada di masa yang akan datang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi juga dapat digunakan (dalam kondisi yang sesuai) untuk prediksi.

4. Klasifikasi

   Dalam klasifikasi, ada variabel kategori target. Misalnya, klasifikasi pendapatan dapat dipisahkan menjadi tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

5. Clustering

   Clustering adalah pengelompokan catatan, pengamatan, atau perhatian dan membentuk kelas objek yang memiliki kesamaan. Cluster adalah kumpulan record yang mirip satu sama lain dan memiliki record yang berbeda di cluster lain.

6. Asosiasi

Tugas asosiasi dalam data mining adalah menemukan atribut yang muncul pada satu waktu. Salah satu implementasi asosiasi tersebut adalah market basket analysis atau algoritma apriori (Pahlevi and Amrin 2020).

II.3. Algoritma K-Nearest Neighbors (KNN)

K-Nearest Neighbor (K-NN) adalah pembelajaran berbasis instance metode pembelajaran yang diawasi atau algoritma yang menangguhkan semua perhitungan hingga belajar setelah klasifikasi. sebagian besar sampel milik tertentu kategori, maka sampel juga termasuk kategori ini. Umumnya, metode voting dapat digunakan dalam tugas klasifikasi,yaitu, label kategori yang paling banyak muncul dalam k sampel adalah dipilih sebagai hasil prediksi. Metode rata-rata bisa digunakan dalam tugas regresi, yaitu rata-rata dari nilai riil label keluaran dari k sampel digunakan sebagai hasil prediksi.(Zhang et al. 2020)

Algoritma K-Nearest Neighbor (KNN) adalah salah satu metode klasifikasi dalam data mining, dimana KNN mengklasifikasikan sekumpulan data berdasarkan data pembelajaran yang telah diklasifikasikan atau diberi label. KNN termasuk dalam kelompok supervied learning, yaitu hasil dari instance query yang baru diklasifikasi berdasarkan kedekatan mayoritas dengan kategori yang ada di KNN(Dewi and Dwidasmara 2020).

Cover dan Hart memperkenalkan K Nearest Neighbor pada tahun 1968. K-Nearest Neighbor adalah metode klasifikasi pemalas karena

algoritma ini menyimpan semua nilai data pelatihan dan menunda proses pembentukan model klasifikasi hingga data uji diberikan untuk prediksi. Gagasan dalam metode k Nearest Neighbor adalah untuk mengidentifikasi sampel k dalam set pelatihan yang variabel independen x mirip dengan u, dan menggunakan sampel k ini untuk mengklasifikasikan sampel baru ini ke dalam kelas, v. F adalah fungsi yang halus, sebuah ide yang masuk akal adalah mencari sampel dalam data pelatihan kami yang berada di dekatnya (dalam hal variabel independen) dan kemudian menghitung v dari nilai y untuk sampel ini. Jarak atau ukuran ketidaksamaan dapat dihitung antara sampel dengan mengukur jarak menggunakan jarak Euclidean.(Putra, M. H. Pardede, and Syahputra 2022)

Menurut Kusrini dan Luthfi Nearest Neighbor adalah suatu pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada.(Ma'ruf 2021)

K-Nearest neighbor adalah pendekatan untuk mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada. Rumus untuk menghitung kedekatan antara dua kasus tersebut adalah:

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

Keterangan :

T : Kasus baru

S : Kasus yang ada salam penympanan

n : Jumlah atribut dalam masing-masing kasus

i : atribut individu antara 1 sampai n

f : Fungsi similarity atribut antara kasus T dan S

w : Bobot yang diberikan pada atribut ke-i

Kedekatan biasanya berada pada nilai antara 0 s/d 1. Nilai 0 artinya kedua kasus mutlak tidak mirip, sebaliknya untuk nilai 1 kasus mirip dengan mutlak.(Linof, Gordon S & Berry 2011)

Adapun langkah-langkah klasifikasi algoritma KNN adalah sebagai berikut:

1. Tentukan parameter nilai k = banyaknya jumlah tetangga terdekat
2. Hitung jarak antara data baru dengan semua data training
3. Urutkan jarak dan tetapkan tetangga terdekat berdasarkan jarak minimum ke-k
4. Periksa kelas dari tetangga terdekat
5. Gunakan mayoritas sederhana dari kelas tetangga terdekat sebagai nilai prediksi data baru (Putra, M. H. Pardede, and Syahputra 2022).

II.4. Algoritma Naive Bayes

Kemudahan membangun dan menafsirkan datanya dan kinerjanya yang luar biasa, Naive Bayes banyak digunakan untuk menyelesaikan klasifikasi masalah dalam aplikasi dunia nyata. Teorema Bayes dengan asumsi independensi antara prediktor adalah dasar pembelajaran algoritma

Naive Bayes. Ini menunjukkan bahwa fitur-fitur kelas tidak bergantung pada fitur lainnya. Baik variabel kategoris maupun kontinyu dapat diklasifikasikan dengan metode Naive Bayes. (Religia and Maulana 2021).

Naive Bayes adalah salah satu algoritma pembelajaran induktif yang paling efisien untuk pengolahan data dan pembelajaran mesin. Walaupun menggunakan asumsi keidependenan atribut (tidak ada kaitan antar atribut), kinerja naive bayes tetap kompetitif dalam proses klasifikasi. Namun, meskipun asumsi ini jarang terjadi pada data, kinerja pengklasifikasian naive bayes tetap cukup tinggi, seperti yang dibuktikan oleh banyak penelitian empiris.

Teorema Bayes berasal dari klasifikasi Naive Bayes, yang diciptakan oleh ilmuwan Inggris Thomas Bayes, yang bertujuan untuk memprediksi peluang di masa depan berdasarkan pengalaman sebelumnya. Metode "naive", di mana kondisi antar atribut saling bebas dianggap, menggabungkan teorema ini. Pada sebuah kumpulan data, setiap baris atau dokumen I dianggap sebagai vector dari nilai-nilai atribut <x1,x2,...,x3>, di mana tiap nilai menjadi peninjauan atribut Xi (iЄ[1,n])), dan setiap baris memiliki label kelas ci (c1,c2,...,ck) sebagai nilai variabel kelas C. Oleh karena itu, untuk melakukan klasifikasi, nilai probabilitas p(C=ci|X=xj) dapat dihitung. Dikarenakan pada Naive Bayes diasumsikan setiap atribut saling bebas, maka persamaan yang didapat adalah sebagai berikut :

- Peluang bersyarat atribut Xi dengan nilai xi diberikan kelas c ditunjukkan oleh peluang p(C=ci|X=xj). Dalam Naïve Bayes, kelas C bertipe kualitatif, sedangkan atribut Xi dapat bertipe kuantitatif atau kualitatif.

- Jika atribut Xi bertipe kuantitatif, persamaan peluang p(X=xi|C=cj) akan sangat kecil. Akibatnya, persamaan ini tidak dapat diandalkan untuk masalah atribut bertipe kuantitatif. Jadi, distribusi normal (Gaussian) adalah salah satu dari banyak metode yang dapat digunakan untuk menangani atribut kuantitatif :

$$\hat{f} = N(X_i; \mu_c, \sigma_c) = \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{(X_i-\mu_c)^2}{2\sigma_c^2}},$$

Ataupun kernel density estimation (KDE) :

$$\hat{f} = \frac{1}{n_c} \sum_j N(X_i; \mu_{ij}, \sigma_c),$$

Diskritisasi adalah metode distribusi tambahan yang menangani atribut kuantitatif (numerik). Diskritisasi sendiri adalah proses yang terjadi selama proses persiapan data atau preprocessing data, di mana atribut numerik X diubah menjadi atribut nominal X. Ketika atribut numerik didiskritisasi, hasil klasifikasi Naive Bayes akan lebih baik daripada ketika atribut numerik diasumsikan dengan pendekatan distribusi seperti yang disebutkan di atas.(Syarli and Muin 2019)

II.5. Kepuasan Wali Murid

Kepuasan pelanggan merupakan faktor penting yang tergantung pada kualitas, produktivitas, dan karakteristik layanan yang ditawarkan oleh

penyedia layanan tertentu. Kepuasan pelanggan yang lebih tinggi dapat meningkatkan kemauan pelanggan untuk membeli produk.(Dampage et al. 2021)

Salah satu kunci keberhasilan bisnis dalam bidang apa pun adalah kepuasan pelanggan, jika perusahaan dapat memenuhi kebutuhan pelanggannya dengan layanan yang tersedia, pelanggan akan secara otomatis menjadi lebih loyal terhadap barang dan jasa yang ditawarkan oleh perusahaan, dan perusahaan dapat meningkatkan keuntungan mereka. Kepuasan pelanggan, atau kepuasan pelanggan, bergantung pada bagaimana produk bekerja jika dibandingkan dengan harapan pelanggan. Pelanggan akan kecewa jika kinerja produk tidak memenuhi harapan. Jika kinerja produk memenuhi harapan, konsumen puas.(Ma'ruf 2021)

Kepuasan pelanggan terdapat juga di instansi lembaga pendidikan yaitu di sekolah yang dimana salah satunya wali murid sebagai konsumen nya. Wali murid menjadi pelanggan eksternal utama yang harus dipuaskan dalam penyelenggaraan layanan pendidikan. Pada dasarnya wali murid sebagai pelanggan eksternal bisa mendapatkan kepuasan dengan kapasitas lembaga itu sendiri dan terutama ketika anak-anaknya sebagai pelanggan internal juga dapat dipuaskan.

Wali Murid sebagai pelanggan sekolah (perlu diyakini bahwa sekolah yang akan dipilih adalah sekolah yang memberikan layanan yang relevansesuai dengan kebutuhan dan tuntutan zaman. Namun saat ini banyak lembaga pendidikan saling berlomba untuk memperoleh

siswa dengan menunjukkan eksistensi sekolah lewat label internasional, akreditasi, prestasi yang diraih, tingkat kelulusan yang tinggi, fasilitas sarana dan prasarana yang memadai dan mendukung, program unggulan, dan layanan pendidikan lainnya yang membedakannya dari sekolah lainnya. Akan tetapi disisi lain siswa dan orang tua memiliki kebebasan untuk memilih sekolah yang menurut penilaian mereka memiliki kualitas yang baik (SA'IDU 2016)

Menurut Zeithaml, Bitner, & Gremler (2009) ada lima dimensi yang dapat digunakan dalam menentukan kualitas layanan, yaitu:

1. Tangible (berwujud) yaitu berupa penampilan fasilitas fisik, peralatan, staff, dan material yang dipasang. Menggambarkan wujud secara fisik dan layanan yang akan diterima oleh pelanggan. Contohnya seperti keadaan gedung, fasilitas restoran, desain restoran, dan kerapian penampilan staff
2. Empathy (empati) yaitu kepedulian dan perhatian secara pribadi yang diberikan kepada pelanggan. Layanan yang diberikan oleh para staff harus dapat menunjukkan kepeduliannya kepada pelanggan.
3. Reliability (keandalan) yaitu kemampuan untuk memberikan jasa yang dijanjikan dengan handal dan akurat. Jika dilihat dalam bidang usaha jasa restoran maka sebuah layanan yang handal adalah ketika seorang staff mampu memberikan pelayanan sesuai yang dijanjikan dan membantu penyelesaian masalah yang

dihadapi pelanggan dengan cepat.

4. Responsiveness (cepat tanggap) yaitu kemauan untuk membantu pelanggan dan memberikan jasa dengan cepat. Jika dilihat lebih mendalam pada layanan yang cepat tanggap di sebuah restoran, bisa dilihat dari kemampuan staff yang cepat memberikan pelayanan serta menangani keluhan pada pelanggan.
5. Assurance (jaminan) yaitu pengetahuan, sopan santun, dan kemampuan staff untuk menimbulkan keyakinan dan kepercayaan. Dalam sebuah jasa restoran kepastian menjadi hal yang penting untuk dapat diberikan kepada pelanggannya, seperti jaminan keamanan dan keselamatan dalam bertransaksi dan kerahasiaan pelanggan yang terjamin (Hartono 2013)

II.6. Metode Algoritma K-NN Dan Naive Bayes Untuk Mengukur Tingkat Kepuasan Wali Murid

Metode Algoritma K-NN : Metode yang digunakan untuk melakukan klasifikasi berdasarkan data yang paling mirip atau "dekat" dengan data yang ingin diprediksi. K dalam K-NN adalah jumlah tetangga terdekat yang akan digunakan untuk memprediksi kategori atau label baru untuk data, dan metode ini didasarkan pada gagasan bahwa data dengan karakteristik yang sama cenderung memiliki kategori atau label yang sama.

Langkah-langkah dalam menggunakan metode K-NN adalah sebagai berikut:

- Menghitung dan menentukan nilai K yang akan digunakan.

- Menghitung jarak antara data baru dan data yang sudah ada dalam dataset. Untuk menghitung jarak ini, dapat menggunakan metrik yang serupa dengan jarak geometris salag satu contoh ialah Euclidean.
- Jarak yang telah dihitung, Anda dapat memilih K sebagai tetangga terdekat..
- Untuk memprediksi label data baru, gunakan sebagian besar label tetangga terdekat.

Metode K-NN dapat diterapkan pada data kepuasan wali murid dengan mengukur tingkat kesamaan antara atribut atau fitur pada data tersebut. Metode ini akan mencari tetangga terdekat dari data baru dan mengukur tingkat kepuasan dengan menggunakan sebagian besar label tetangga tersebut.

Metode Naive Bayes didasarkan pada teorema Bayes, yang menyatakan bahwa setiap fitur atau atribut dalam data tidak bergantung satu sama lain. Metode ini menghitung probabilitas kondisional dari masing-masing atribut dan kemudian menggabungkannya untuk memprediksi label atau kategori baru untuk data.

Proses menggunakan metode Naive Bayes adalah sebagai berikut :

- Menghitung kemungkinan prior untuk setiap label atau kategori dalam dataset.
- Berdasarkan label atau kategori, menghitung probabilitas kondisional untuk masing-masing atribut.

- Mengalikan probabilitas untuk masing-masing atribut dengan probabilitas kondisional.
- Sebagai prediksi untuk data baru, pilih label atau kategori dengan probabilitas tertinggi.

Untuk data kepuasan wali murid, metode Naive Bayes dapat diterapkan. Ini dilakukan dengan menghitung probabilitas kondisional dari setiap atribut kepuasan berdasarkan label kepuasan, seperti puas atau tidak puas. Kemudian, metode Naive Bayes akan menggabungkan probabilitas atribut untuk mengukur label kepuasan wali murid.

Kepuasan adalah respons emosional yang ditunjukkan oleh pelanggan setelah membeli produk atau layanan. Kepuasan berasal dari evaluasi pengalaman mengonsumsi produk atau layanan dan perbandingan kinerja aktual terhadap harapan. Evaluasi ini mencakup bahwa produk yang dipilih sekurang-kurangnya memenuhi atau melampaui harapan pelanggan, sedangkan ketidakpuasan terjadi ketika hasil (autcome) tidak memenuhi harapan.

Kepuasan wali murid merujuk pada seberapa puas orang tua atau wali murid dengan pendidikan yang diberikan oleh institusi pendidikan.Kepuasan wali murid dapat mencakup berbagai hal, seperti kualitas pendidikan, ketersediaan fasilitas, dan komunikasi sekolah-orang tua. Untuk meningkatkan kepuasan wali murid, penting untuk memahami dan memenuhi harapan mereka. Kepuasan wali murid yang tinggi dapat berdampak positif pada keberlanjutan dan reputasi institusi pendidikan, serta

mempromosikan kolaborasi yang kuat antara sekolah dan orang tua untuk mendukung pendidikan anak-anak mereka.

Algoritme KNN (K-Nearest Neighbors) dan Naive Bayes adalah algoritma yang termasuk dalam kategori supervised learning (pembelajaran dengan pengawasan), di mana setiap data latih memiliki label atau kelas yang diketahui sebelumnya. Algoritme ini digunakan untuk mengukur tingkat kepuasan wali murid. Untuk mengukur tingkat kepuasan wali murid menggunakan algoritma KNN, berikut adalah langkah-langkah umum :

- Pra-pemrosesan data:
  - Kumpulkan data tentang kepuasan wali murid. Data ini dapat terdiri dari fitur atau atribut yang relevan, seperti partisipasi dalam kegiatan sekolah, hubungan dengan guru, program studi, fasilitas, dll., dan label atau kategori kepuasan, seperti apakah mereka puas atau tidak puas.
  - Melakukan pra-pemrosesan data, termasuk membersihkan data, menangani nilai yang hilang, dan mengkodekan variabel jika diperlukan.
- Bagi dataset:
  - Dataset dibagi menjadi dua subset : subset pelatihan (set pelatihan) dan subset pengujian (set pengujian). Sebagian besar data biasanya digunakan untuk pelatihan, sementara sebagian kecil digunakan untuk pengujian

- Metode K-NN:
  - Tentukan berapa banyak tetangga terdekat (K) yang akan digunakan untuk K-NN.
  - Hitung perbedaan antara data baru dari subset pelatihan dan data baru dari subset pengujian. Untuk mengukur jarak, gunakan metrik jarak seperti jarak Euclidean.
  - Berdasarkan jarak yang dihitung sebelumnya, pilih K sebagai tetangga terdekat.
- Metode Naive Bayes:
  - Tentukan kemungkinan prioritas masing-masing label kepuasan dalam subset pelatihan.
  - Berdasarkan label kepuasan pada subset pelatihan, hitung probabilitas kondisional untuk masing-masing atribut kepuasan.
  - Mengalikan probabilitas untuk masing-masing atribut dengan probabilitas kondisional.