

BAB II

KAJIAN PUSTAKA

A. Kajian Teoritis

1. Penjurusan

Penjurusan di SMA, MA, atau SMK merupakan suatu hal yang wajib dan tidak terelakkan dari dunia pendidikan, penjurusan dilakukan untuk mengarahkan siswa agar menekuni karir yang diinginkan dan sesuai dengan kemampuan (Muhammad Daniel, 2022). Penempatan siswa pada dasarnya didasarkan pada tuntutan layanan kesamaan dan perbedaan anak, sehingga melahirkan pemikiran pentingnya manajemen peserta didik (Muhammad Daniel, 2022).

Lembaga pendidikan setiap tahunnya terus berusaha meningkatkan mutu pendidikan khususnya SMK dengan harapan lulusannya dapat memiliki keterampilan dan keahlian lebih dibandingkan sekolah sederajat, hal tersebut dilakukan demi meningkatkan kualitas lulusan sehingga siap memasuki dunia kerja (Muhammad Daniel, 2022). Antusias siswa lulusan SMP maupun MTs untuk masuk SMK cukup besar, tetapi kebanyakan mereka kurang matang untuk memilih jurusan yang ada sesuai kemampuannya, akibatnya cukup banyak siswa baru yang gagal ditengah jalan ketika mereka sudah diterima di SMK (Muhammad Daniel, 2022). Tidak sedikit juga kasus siswa yang merasa dirinya

tidak cocok dengan jurusan yang dipilihnya ketika pembelajaran sudah berlangsung di kelas(Muhammad Daniel, 2022).

2. Artificial Intelligent

Artificial Intelligence atau kecerdasan buatan adalah sistem komputer yang mampu melakukan tugas-tugas yang biasanya membutuhkan kecerdasan manusia. Teknologi ini dapat membuat keputusan dengan cara menganalisis dan menggunakan data yang tersedia di dalam sistem. Proses yang terjadi dalam *Artificial Intelligence* mencakup *learning, reasoning, dan self-correction*. Proses ini mirip dengan manusia yang melakukan analisis sebelum memberikan keputusan(Sobron & Lubis, 2021).

Menurut John Mc Carthy, 1956, Artificial Intelligence adalah untuk mengetahui dan memodelkan proses-proses berpikir manusia dan mendesain mesin agar dapat menirukan perilaku manusia. Cerdas, berarti memiliki pengetahuan ditambah pengalaman, penalaran (bagaimana membuat keputusan dan mengambil tindakan), moral yang baik(Sobron & Lubis, 2021). Keinginan untuk mencapai situasi yang memberikan kemampuan kepada perusahaan agar tetap tumbuh dan berkembang serta mampu menghasilkan keuntungan yang wajar dimotivasi oleh kekhawatiran yang sering menghantui pikiran para pebisnis sehubungan dengan perubahan dinamis pada lingkungan industry yang memasuki revolusi Industri 4.0. Pada era ini permintaan terhadap produk yang sesuai dengan keinginan konsumen dan

dapat di peroleh dengan waktu yang relative singkat menjadi tema utama dalam pengembangan produk. Factor biaya per unit minimum dan mutu produk yang tinggi tidak lagi memiliki kekuatan untuk membangun daya saing. Daya saing ditentukan oleh kinerja dalam pengiriman , kemampuan memenuhi keinginan konsumen dan kualitas keterlibatan dalam penanganan isu-isu lingkungan(Sobron & Lubis, 2021).

3. Machine Learning

Artificial Intelegence memiliki beberapa cabang seperti *Machine learning* dan *Computer vision*. *Machine learning* adalah cabang AI yang paling umum digunakan. Machine learning memungkinkan mesin untuk belajar menggunakan algoritma dari data dan membuat keputusan berdasarkan pola-pola yang ditemukan dalam data tersebut (Ananto et al., 2023).

Machine learning menggunakan model algoritma untuk dapat bekerja dengan baik. Algoritma yang digunakan dalam machine learning terbagi menjadi tiga tipe yaitu *supervised*, *unsupervised*, dan *reinforcement learning*. Setiap tipe dari machine learning memiliki proses yang berbeda beda (Ananto et al., 2023).

4. Data Mining

Data mining adalah proses pengelolaan sebuah data yang besar sehingga dari data- data tersebut dapat memberikan

informasi yang akurat dan dapat mempermudah dalam pemecahan masalah dan pengambilan keputusan (Putry, 2022). Adapun nama-nama lain dari data mining adalah *Pattern Analysis*, *Knowledge Extraction*, *Information Harvesting*, dan sebagainya. Terdapat enam pengelompokan dalam *data mining* yaitu (Putry, 2022)

1. Estimasi

Untuk kelompok estimasi hampir sama dengan klasifikasi, kecuali variabel target dari estimasi lebih mengarah ke numerik daripada ke arah kategori. Estimasi membangun sebuah model dimana model tersebut didapatkan dari record lengkap yang menyediakan nilai yang diambil dari variabel target sebagai nilai prediksi.

2. Deskripsi

Kelompok deskripsi adalah cara untuk menggambarkan pola yang terdapat di dalam data yang dimiliki.

3. Klasifikasi

Klasifikasi adalah suatu proses pembelajaran yang memiliki fungsi untuk menentukan setiap himpunan atribut dari sebuah objek atau dapat dikatakan juga klasifikasi merupakan proses pengelompokan suatu data.

4. Prediksi

Digunakan untuk memperkirakan nilai di masa depan atau masa mendatang dan juga menerka nilai yang belum diketahui.

5. Asosiasi

Asosiasi digunakan untuk menemukan atribut yang muncul dalam satu waktu.

6. Clustering

Clustering adalah pengamatan, pengelompokan record, membentuk kelas objek-objek yang mempunyai kemiripan.

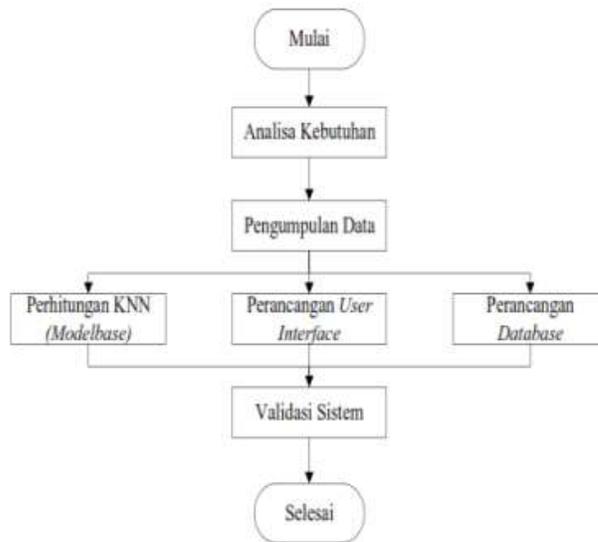
5. Algoritma K Nearest Neighbor (KNN)

Algoritma KNN merupakan sebuah algoritma yang dikenal dengan non numerik dalam data mining, yang bisa digunakan untuk melakukan klasifikasi ataupun regresi. Dalam melakukan klasifikasi menggunakan sebuah algoritma tertentu membutuhkan sebuah dataset yang terdiri dari data training dan data testing (A'yuniyah & Reza, 2023). K-Nearest Neighbor (KNN) termasuk algoritma supervised learning dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Nanti kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi (Qiudandra et al., 2022).

K-Nearest Neighbor adalah suatu pendekatan untuk menghitung kedekatan antara kasus lama, yaitu berdasarkan pada pencocokan bobot dari sejumlah fitur yang ada (Qiudandra et al., 2022). Algoritma ini bekerja dengan berdasarkan pada jarak terpendek dari sample uji ke sample latih untuk menentukan KNNnya. Setelah mengumpulkan KNN, kemudian diambil

mayoritas dari KNN untuk dijadikan prediksi dari sample uji (Argina, 2020).

K-Nearest Neighbor (KNN) termasuk algoritma *supervised learning* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN (Argina, 2020). Nanti kelas yang paling banyak muncul yang akan menjadi kelas hasil klasifikasi. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing (Qiudandra et al., 2022).



Gambar 2. 1 Flowchart KNN(Argina, 2020)

6. Google Collaboratory

Google Collaboratory atau dikenal dengan *Google Colab* adalah sebuah tool yang digunakan untuk tujuan penelitian secara gratis dan menggunakan cloud atau sistem penyimpanan awan. *Google Collaboratory* dibangun dengan menggunakan elemen

dari jupyter serta sebagian besar library tersedia untuk diterapkan ke dalam sebuah penelitian salah satunya adalah data mining. *Google Collaboratory* pada dasarnya mempunyai kesamaan fungsi dengan *Jupyter Notebook*, letak perbedaannya adalah *Google Collaboratory* dapat diakses secara online serta gratis (Setiadi et al., 2023).

7. Website

Website disebut juga site, situs, situs web atau portal. Merupakan kumpulan halaman web yang berhubungan antara satu dengan lainnya, halaman pertama sebuah *website* adalah *home page*, sedangkan halaman demi halamannya secara mandiri disebut *web page*, dengan kata lain *website* adalah situs yang dapat diakses dan dilihat oleh para pengguna *internet* diseluruh dunia (Afdal, 2019). *Website* adalah situs yang dapat diakses dan dilihat oleh para pengguna *Internet* di seluruh dunia (Tangkudung et al., 2019).

7.1 Python

Python adalah bahasa pemrograman interpretatif multiguna dengan filosofi perancangan yang berfokus pada tingkat keterbacaan kode. *Python* diklaim sebagai bahasa yang menggabungkan kapabilitas, kemampuan, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas

pustaka standar yang besar serta komprehensif. *Python* juga didukung oleh komunitas yang besar (Syahrudin & Kurniawan, 2018). *Python* dapat digunakan untuk berbagai keperluan pengembangan perangkat lunak dan dapat berjalan di berbagai platform sistem operasi (Syahrudin & Kurniawan, 2018).

Salah satu keuntungan utama menggunakan Python adalah kemampuan untuk berinteraksi langsung dengan kode, menggunakan terminal atau alat lain seperti Notebook Jupyter, yang akan kita bahas segera. Pembelajaran (Raharjo, 2016).

7.2 Numpy

NumPy adalah pustaka fungsi yang memungkinkan Anda melakukan banyak tugas manipulasi data umum dengan Python. Banyak interaksi antara array NumPy dan Python sangat mirip dengan apa yang akan Anda lakukan dengan variabel Python biasa. Untuk tugas yang lebih lanjut, NumPy juga memiliki fungsi yang bekerja dengan aljabar linier, transformasi Fourier, dan matriks (Putra et al., 2023).

7.3 Pandas

Pandas adalah metode untuk membuat visualisasi data lebih interaktif, mudah dibaca, dan mudah dianalisa. Visualisasi yang dilakukan adalah mengubah data tabel yang kaku menjadi bentuk grafik, diagram, dan sebagainya yang mampu memperlihatkan perubahan dan perbedaan data menjadi lebih jelas. Alhasil, tampilan data hasil analisis biasa menjadi lebih cantik dan berbentuk grafik dengan visual yang baik menggunakan matplotlib dan pandas.(Badri Tamam, 2022)

7.4 Streamlit

Streamlit merupakan framework berbasis Python yang bersifat open source. Framework ini dibuat untuk memudahkan developer dalam membangun program berbasis web di bidang data science dan machine learning yang interaktif. Salah satu kelebihan dari streamlit adalah developer tidak perlu mengatur tampilan website dengan CSS, HTML dan javascript karena framework streamlit telah menyediakannya melalui fungsi-fungsi yang terdapat pada framework tersebut.(Putranto et al., 2023)

7.5 Pickle

Pickle adalah sebuah modul dari bahasa pemrograman python, yang digunakan untuk menyimpan dan membaca data ke dalam/dari sebuah file. File pickle tersebut yang akan di implementasikan kedalam aplikasi. Setelah file pickle dimasukkan kedalam aplikasi. Kemudian file pickle tersebut akan membaca file dari database, yang kemudian mendapatkan output berupa model yang telah dibuat.(Arsa Paskha A et al., 2022)

7.6 Scikit-learn

Scikit – Learn adalah modul python yang mengintegrasikan berbagai algoritma pembelajaran mesin untuk masalah yang diawasi dan tidak diawasi skala menengah. Modul ini sangat efisien untuk data mining dan analisis data. Data mining adalah proses menemukan pola dari data yang tidak diketahui dan dilakukan secara otomatis atau semiotomatis guna mendapatkan pola dan struktur bermakna (Riadi Silitonga, 2019).

8. Flowchart

Flowchart atau sering disebut dengan diagram alir merupakan suatu jenis diagram yang merepresentasikan algoritma atau langkah-langkah instruksi yang berurutan dalam sistem (Rosaly & Prasetyo, 2020). Dengan adanya *flowchart*, setiap urutan

proses dapat digambarkan menjadi lebih jelas. Selain itu, ketika ada penambahan proses baru dapat dilakukan dengan mudah menggunakan (Rosaly & Prasetyo, 2020).

flowchart ini. Pada dasarnya, *flowchart* digambarkan dengan menggunakan simbol-simbol. Setiap simbol mewakili suatu proses tertentu. Sedangkan untuk menghubungkan satu proses ke proses selanjutnya digambarkan dengan menggunakan garis penghubung (Rosaly & Prasetyo, 2020).

Pada dasarnya, dalam merancang *flowchart* tidak ada ketentuan mutlak yang harus dipenuhi. Hal itu dikarenakan *flowchart* dibuat berdasarkan pemikiran untuk menganalisa suatu permasalahan dalam bisnis. Hanya saja, Anda dapat merancang *flowchart* ketika Anda telah mengetahui simbol-simbol standar yang umum digunakan dalam proses pembuatan *flowchart*. Berikut akan dijelaskan mengenai simbol-simbol *flowchart* diantaranya (Rosaly & Prasetyo, 2020):

	Process Simbol yang digunakan untuk menyatakan suatu proses yang dilakukan komputer.
	Terminator Simbol yang menyatakan awal atau akhir dari suatu program.
	Decision Simbol pilihan yang dapat menghasilkan dua kemungkinan jawaban antara ya atau tidak .
	Input/Output Simbol yang digunakan untuk proses input atau output.
	Predefine Proses Simbol untuk melakukan suatu bagian atau prosedur.
	Document Simbol yang menyatakan bahwa input berasal dari dokumen dalam bentuk fisik atau output yang perlu dicetak.

Gambar 2. 2 Simbol FLOWchart

B. Kajian Empiris

1. Penerapan Model Klasifikasi *K-Nearest Neighbor* Dalam Pencarian Keseuaian Pekerjaan

Penerapan Model *Klasifikasi K-Nearest Neighbor* dalam Pencarian Kesesuaian Pekerjaan bertujuan untuk membantu perusahaan atau lembaga dalam menilai kesesuaian pelamar dengan bidang yang dibutuhkan. Teknologi informasi berbasis mobile sangat dibutuhkan untuk mempermudah pelamar dalam mencari lowongan kerja yang sesuai dengan keahlian mereka. Dengan adanya aplikasi mobile yang mengklasifikasikan lowongan kerja dan menyediakan menu pencarian informasi lowongan kerja sesuai dengan skill yang dimiliki, pelamar dapat lebih mudah menemukan pekerjaan yang sesuai. Model *K-Nearest Neighbor (KNN)* diterapkan dalam aplikasi ini untuk memastikan kecocokan antara pelamar dan lowongan kerja. Hasilnya adalah pembuatan sistem pencarian kesesuaian pekerjaan yang lebih efektif dengan menggunakan metode *K-Nearest Neighbor*.

2. Implementasi Pengolahan Citra dan *Klasifikasi K-Nearest Neighbour* Untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi

Implementasi Pengolahan Citra dan *Klasifikasi K-Nearest Neighbor* untuk Membangun Aplikasi Pembeda Daging Sapi dan Babi bertujuan untuk mengatasi masalah kurangnya

ketelitian masyarakat dalam membedakan antara daging sapi dan babi. Permasalahan ini sering dimanfaatkan oleh penjual daging di pasar untuk melakukan pencampuran daging sapi dengan daging babi (oplosan), terutama saat permintaan daging sapi meningkat. Untuk mengatasi masalah ini, dikembangkan sistem pengolahan citra dan klasifikasi *K-Nearest Neighbor* yang dapat membangun aplikasi pembeda daging sapi dan babi. Hasilnya adalah implementasi KNN untuk membedakan daging sapi dan babi, membantu masyarakat dalam membuat pilihan yang lebih tepat saat membeli daging.

3. Penerapan Algoritma *K-Nearest Neighbor* pada Klasifikasi Status Gizi Balita (Studi Kasus Posyandu Desa Aras Kabu)

Penerapan Algoritma *K-Nearest Neighbor* pada Klasifikasi Status Gizi Balita di Posyandu Desa Aras Kabu bertujuan untuk mengatasi masalah sistem informasi yang kurang memadai. Saat ini, pemrosesan data balita, data petugas, data orang tua balita, dan data gizi dilakukan secara manual menggunakan buku besar, yang nantinya dilaporkan kepada kepala desa tanpa bantuan komputer. Metode ini tidak efektif dan akurat. Untuk menyelesaikan permasalahan ini, modernisasi arsip pendataan gizi dilakukan dengan menerapkan algoritma *K-Nearest Neighbor* dalam klasifikasi status gizi balita. Hasil dari penerapan algoritma ini adalah

peningkatan efisiensi dan akurasi dalam menentukan status gizi balita.

4. Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning

Implementasi Algoritma *K-Nearest Neighbor* (K-NN) untuk Analisis Sentimen Publik terhadap Pembelajaran Online bertujuan untuk memahami variasi sentimen publik terkait metode pembelajaran ini, yang menjadi norma baru selama pandemi COVID-19. Sentimen publik terhadap pembelajaran online sangat bervariasi, mulai dari positif hingga negatif, tergantung pada faktor-faktor seperti infrastruktur teknologi, kualitas pengajaran, dan kesiapan siswa serta guru. Untuk mengatasi tantangan ini, data sentimen publik di media sosial dan platform online lainnya yang sangat besar dan tidak terstruktur dianalisis menggunakan algoritma *K-Nearest Neighbor*. Analisis manual tidak efisien dan rentan terhadap bias, sehingga implementasi algoritma ini membantu dalam mengklasifikasikan sentimen publik secara lebih akurat. Hasil dari implementasi algoritma *K-Nearest Neighbor* ini adalah peningkatan akurasi dalam analisis sentimen publik terhadap pembelajaran online.

5. K-Nearest Neighbor Based on Exploratory Data Analysis of Curriculum Models of Chinese Early Childhood Education

K-Nearest Neighbor Berdasarkan Analisis Data Eksplorasi Model Kurikulum Pendidikan Anak Usia Dini di China bertujuan untuk mengatasi kebingungan yang sering muncul di kalangan pendidik dan orang tua dalam memilih model kurikulum yang paling sesuai untuk perkembangan anak. Pendidikan anak usia dini di China memiliki berbagai model kurikulum yang digunakan di berbagai institusi, dan perbedaan ini sering kali membingungkan. Kurangnya alat analisis yang dapat mengkategorikan dan mengklasifikasikan model kurikulum berdasarkan data empiris menambah kesulitan dalam pengambilan keputusan. Untuk mengatasi masalah ini, pendekatan *K-Nearest Neighbor (KNN)* berbasis *Exploratory Data Analysis (EDA)* diterapkan. Hasilnya adalah implementasi klasifikasi model kurikulum di China dengan algoritma *K-Nearest Neighbor*, yang membantu dalam pengambilan keputusan yang lebih informatif dan akurat.

Dari 5 jurnal diatas dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* dapat diterapkan diberbagai permasalahan, dan dapat menyelesaikan masalah dengan efektif. Pada penelitian ini sedikit berbeda dengan penelitian penelitian diatas Algoritma *K-Nearest Neighbor* diterapkan dalam meklasifikasikan pemilihan jurusan di studi kasus SMK PGRI 1 Ngawi.

C. Kerangka Berfikir

Penelitian ini diawali dengan identifikasi masalah. Masalah utama yang dihadapi adalah bagaimana menentukan jurusan yang tepat bagi siswa berdasarkan data yang tersedia. Data ini mencakup berbagai informasi seperti nilai akademik pada saat di bangku Sekolah Menengah Pertama. Langkah pertama dalam penelitian ini adalah mengumpulkan informasi yang relevan dan mengidentifikasi kriteria serta parameter yang akan digunakan dalam analisis menggunakan algoritma K-NN (K-Nearest Neighbor). Kriteria ini akan membantu dalam mengelompokkan siswa ke jurusan yang paling sesuai dengan potensi dan minat mereka.

Selanjutnya, dalam tahap proses, berbagai tools, metode, data, dan populasi akan digunakan untuk mencapai tujuan penelitian. Tools yang akan digunakan termasuk bahasa pemrograman dan framework untuk pengembangan web, yang akan digunakan untuk membangun sistem berbasis web. Algoritma KNN akan diimplementasikan untuk analisis data yang cepat dan akurat. Metode penelitian mencakup penerapan algoritma KNN untuk mengelompokkan data siswa berdasarkan kriteria yang telah ditentukan. Metode ini melibatkan beberapa langkah, termasuk pemilihan parameter K yang optimal, normalisasi data, dan pengujian kinerja model. Data yang digunakan dalam penelitian ini mencakup nilai akademik siswa matapelajaran di bangku SMP. Populasi penelitian terdiri dari siswa-siswa dari sekolah yang akan menggunakan sistem ini untuk menentukan jurusan

mereka. Sampel data diambil dari populasi ini untuk pengembangan dan pengujian sistem.

Pada tahap akhir dari penelitian ini adalah sistem berbasis web yang mampu menganalisis data siswa dan memberikan klasifikasi jurusan menggunakan algoritma KNN. Sistem ini dirancang untuk memiliki berbagai fitur yang memudahkan pengguna dalam menginput data siswa, memproses data menggunakan algoritma KNN, dan menghasilkan klasifikasi jurusan. Antarmuka pengguna yang intuitif akan disediakan untuk berbagai pengguna sistem, termasuk calon siswa, dan administrator, sehingga mereka dapat dengan mudah memasukkan data, melihat hasil analisis, dan mengelola sistem.

Dengan kerangka berpikir ini, penelitian ini tidak hanya bertujuan untuk mengembangkan sistem yang praktis dan efektif untuk mengklasifikasi jurusan siswa, tetapi juga memberikan kontribusi teoritis dalam pengembangan ilmu pengetahuan di bidang penerapan algoritma KNN dalam sistem berbasis web. Penelitian ini diharapkan dapat menjadi referensi bagi peneliti selanjutnya yang ingin mengembangkan atau melakukan penelitian lebih lanjut terkait penggunaan algoritma KNN dalam sistem penentuan jurusan bagi siswa. Selain itu, sistem yang dikembangkan dapat membantu pihak sekolah atau lembaga pendidikan dalam membuat keputusan yang lebih objektif dan berdasarkan data, serta membantu siswa dalam menentukan jurusan yang paling sesuai dengan minat dan kemampuan

mereka. Orang tua dan wali siswa juga dapat menggunakan sistem ini sebagai alat bantu untuk memahami potensi anak mereka dan mendukung pengambilan keputusan yang lebih baik dalam memilih jurusan. Pengguna sistem (administrator atau guru) dapat menghemat waktu dan usaha dalam proses penentuan jurusan, karena sistem ini menyediakan analisis yang cepat dan efisien berdasarkan data yang tersedia. Dengan demikian, penelitian ini memiliki manfaat yang signifikan baik dari segi teoritis maupun praktis.