

PERBANDINGAN KINERJA METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK KLASIFIKASI PEKERJAAN BERDASARKAN NILAI MATA KULIAH (STUDI KASUS: ALUMNI PROGRAM STUDI PENDIDIKAN TEKNOLOGI INFORMASI FAKULTAS ILMU KOMPUTER UNIVERSITAS BRAWIJAYA)

Jonathan Wibisono¹, Aswin Suharsono², Satrio Hadi Wijoyo³

Program Studi Pendidikan Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Brawijaya

Email: ¹joejostill@student.ub.ac.id, ²aswin@ub.ac.id, ³satriohadi@ub.ac.id

Abstrak

Seiring dengan pertumbuhan era digital, kebutuhan akan keterampilan di bidang teknologi menjadi krusial dalam pasar kerja. Perguruan tinggi, termasuk Program Studi Pendidikan Teknologi Informasi Fakultas Ilmu Komputer Universitas Brawijaya, memiliki peran penting dalam mempersiapkan mahasiswa dengan keterampilan yang diperlukan. Tujuan penelitian ini menjelaskan hubungan antara nilai mata kuliah terhadap pekerjaan para alumni menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor. Dalam penelitian ini, metode analisis yang digunakan adalah Naive Bayes dan K-Nearest Neighbor. Pada masa sebelumnya, algoritma klasifikasi seperti Naive Bayes dan K-Nearest Neighbor telah digunakan dalam bidang pendidikan dan penelitian pekerjaan. Penelitian terdahulu menunjukkan hasil yang menjanjikan dalam memprediksi masa tunggu kerja alumni dan tingkat kelulusan mahasiswa. Namun, penelitian ini fokus pada analisis hubungan antara nilai mata kuliah dan pekerjaan yang dijalani oleh alumni Program Studi Pendidikan Teknologi Informasi. Proses dimulai dengan pengumpulan data hingga pengujian. Melalui penggunaan metode Naive Bayes dan K-Nearest Neighbor, hasil penelitian ini mencapai tingkat akurasi sebesar 66,66%. Meskipun nilai akurasi ini dapat memberikan wawasan awal tentang hubungan antara nilai mata kuliah dan pekerjaan alumni, masih diperlukan penelitian lebih lanjut untuk memperdalam pemahaman akan faktor-faktor yang mempengaruhi kesuksesan alumni bagi lulusan program ini.

Kata kunci: Pendidikan Teknologi Informasi, Analisis Nilai Mata Kuliah, Metode Klasifikasi, Naive Bayes, KNearest Neighbor

Abstract

Along with the growth of the digital era, the need for skills in the field of technology has become crucial in the job market. Higher education, including the Information Technology Education Study Program, Faculty of Computer Science, Brawijaya University, has an important role in preparing students with the necessary skills. The aim of this research is to explain the relationship between course grades and alumni work using the Naïve Bayes and K-Nearest Neighbor algorithms. In this research, the analysis methods used are Naive Bayes and K-Nearest Neighbor. In the past, classification algorithms such as Naive Bayes and K-Nearest Neighbor have been used in the fields of education and occupational research. Previous research shows promising results in predicting alumni job waiting periods and student graduation rates. However, this research focuses on analyzing the relationship between course grades and the work undertaken by alumni of the Information Technology Education Study Program. The process starts with data collection to testing. Through the use of the Naive Bayes and K-Nearest Neighbor methods, the results of this research achieved an accuracy level of 66.66%. Although this accuracy value can provide initial insight into the relationship between course grades and alumni employment, further research is still needed to deepen understanding of the factors that influence alumni success for graduates of this program..

Keywords: *Information Technology Education, Course Grade Analysis, Classification Methods, Naive Bayes, K-Nearest Neighbor*

Fakultas Ilmu Komputer

Universitas Brawijaya

1

1. PENDAHULUAN

Masa sekarang ini adalah masa sebagian besar orang telah menggunakan sistem digital pada keseharian mereka (Rahayu, 2019). Pada era digital saat ini, keterampilan dalam bidang digital sangat penting untuk menjadi kompetitif (Hidayah, 2020). Salah satu tempat untuk meningkatkan keterampilan dalam bidang digital adalah perguruan tinggi. Perguruan tinggi merupakan sebuah lembaga yang mengadakan pendidikan tinggi. Perguruan tinggi bisa diambil oleh seseorang, setelah orang itu lulus dari SMA sederajat (Kaligis, et al., 2013). Perguruan tinggi memberikan kesempatan untuk mengembangkan keterampilan dan pengetahuan di berbagai bidang studi (Kaligis, et al., 2013). Keterampilan dan pengetahuan yang diperoleh dari perguruan tinggi sangat diperlukan untuk memasuki pasar kerja dan berpengaruh terhadap kesuksesan mahasiswa. Keterampilan dan pengetahuan tersebut dapat diperoleh melalui mata kuliah yang disediakan oleh perguruan tinggi.

Mahasiswa yang memiliki kemauan atau minat terhadap suatu mata kuliah akan berusaha lebih keras daripada yang tidak berminat (Ekawati, et al., 2021). Usaha yang lebih keras tersebut dapat mempengaruhi prestasi belajar mahasiswa. Prestasi tersebut dapat dilihat dari nilai mata kuliah. Minat juga merupakan salah satu faktor yang mempengaruhi perencanaan karir mahasiswa. Faktor yang mempengaruhi perencanaan karir antara lain kemampuan, minat dan prestasi (Walidain, 2022).

Salah satu lembaga perguruan tinggi yang mengajarkan mata kuliah di bidang digital yaitu Program Studi (Prodi) Pendidikan Teknologi Informasi (PTI) pada Fakultas Ilmu Komputer di Universitas Brawijaya. Program Sarjana Pendidikan Teknologi Informasi (PTI) dibentuk tahun 2014 dan mulai menerima mahasiswa pada tahun 2015 (UB, 2021). Mata kuliah yang dapat dipelajari pada Program Studi Pendidikan Teknologi Informasi meliputi Mata Kuliah

Pengetahuan Konten, Mata Kuliah Pengetahuan Pedagogis, Mata Kuliah Pengetahuan Teknologi, Mata Kuliah Pengetahuan Konten Pedagogis, Mata Kuliah Pengetahuan Konten Teknologi, Mata Kuliah Teknologi Mata Kuliah

Pengetahuan Pedagogi, Mata Kuliah Pengetahuan Konten Pedagogis Teknologi, Mata Kuliah Penelitian dan Kerja Lapangan, dan Mata Kuliah Umum (UB, 2021). Mata kuliah tersebut diajarkan untuk mencapai tujuan dari Program Studi PTI yaitu menghasilkan lulusan yang mampu berperan sebagai edukator di bidang pengembangan teknologi informasi, IT Developer, konten developer dalam bidang edukasi, ilmu data pada bidang edukasi, dan entrepreneur pada bidang teknologi informasi (UB, 2021). Program Studi PTI telah mencetak sebanyak 148 lebih alumni (UB, 2021).

Dalam mempersiapkan mahasiswa dalam perencanaan dunia kerja, perguruan tinggi juga memiliki lembaga tersendiri. Di Universitas Brawijaya terdapat Unit Pengembangan Karir dan Kewirausahaan (UPKK). UPKK Memiliki tujuan Sebagai pusat mempersiapkan mahasiswa dan alumni guna mampu bersaing di dalam dunia kerja serta memiliki jiwa wirausaha dalam menghadapi era globalisasi (Brawijaya, 2015).

UPKK mengadakan pembinaan karir mahasiswa. Kegiatan pembinaan tersebut dikemas dalam bentuk workshop ataupun seminar yang banyak membahas tentang persiapan karir bagi fresh graduate (Brawijaya, 2015). Hasil penelitian tentang hubungan antara nilai mata kuliah dengan pekerjaan dapat dijadikan referensi untuk pembinaan karir mahasiswa.

Beberapa penelitian terdahulu telah membahas implementasi klasifikasi pada bidang pendidikan dan pekerjaan. Penelitian terdahulu yang dilakukan Rachmadiansyah, et al.(2022) tentang prediksi masa tunggu kerja alumni dengan metode Naive Bayes memperlihatkan tingkat keakuratan sebesar 81,82%. Penelitian terdahulu kedua dilakukan Samudra, et al. (2020) memprediksi tingkat lulus mahasiswa dengan

implementasi model K-Nearest Neighbor menunjukkan tingkat akurasi sebesar 76%.

Bersumber dari latar belakang yang telah dideskripsikan, penulis bermaksud untuk melakukan penelitian skripsi dengan topik analisis hubungan antara nilai mata kuliah dan pekerjaan pada alumni Program Studi Pendidikan Teknologi Informasi Fakultas Ilmu Komputer Universitas Brawijaya. Metode yang akan digunakan dalam penelitian ini adalah Naive Bayes dan K-Nearest Neighbor. Tujuan dari penelitian ini adalah menjelaskan hasil analisis hubungan antara nilai mata kuliah terhadap pekerjaan para alumni menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor.

2. LANDASAN KEPUSTAKAAN

2.1 PEKERJAAN ALUMNI

Pekerjaan adalah kegiatan untuk mendapatkan uang setelah menyelesaikannya (Bengkulu, 2016). Pekerjaan yang mengharuskan orang memiliki pendidikan dan keahlian yang khusus disebut profesi (Bengkulu, 2016). Menurut KBBI, Alumni adalah orang menyelesaikan suatu sekolah atau perguruan tinggi. Dapat disimpulkan, pekerjaan alumni adalah kegiatan yang dilakukan oleh orang yang sudah menyelesaikan sekolah atau pendidikan tinggi guna memperoleh imbalan. Pekerjaan alumni dapat dijadikan suatu indikator keberhasilan suatu program studi dalam menghasilkan lulusan yang kompeten dan siap kerja. Pekerjaan alumni dapat diukur berdasarkan jenis pekerjaan, bidang pekerjaan, dan tingkat jabatan.

2.2 NILAI MATA KULIAH

Menurut KBBI, nilai dapat diartikan sebagai angka kepandaian. Mata kuliah adalah kumpulan pelajaran yang ditempuh oleh mahasiswa pada jenjang pendidikan tinggi (UNAIR, 2023). Mata kuliah harus memenuhi suatu tujuan dari visi dan misi suatu program studi (UNAIR, 2023). Nilai mata kuliah dapat dijadikan acuan indikator keberhasilan mahasiswa dalam suatu program studi. Nilai mata kuliah menunjukkan seberapa besar mahasiswa telah menguasai materi dan keterampilan yang diajarkan dalam mata kuliah tersebut. Nilai mata kuliah bisa menjadi variabel dalam memprediksi kesuksesan mahasiswa di masa depan.

2.3 MACHINE LEARNING

Machine Learning adalah sebuah proses pada pemrograman komputer untuk mengoptimalkan kriteria kinerja menggunakan data atau sebuah pengalaman yang telah terjadi pada masa lalu (Alpaydin, 2014). Untuk mengoptimalkan kriteria kinerja, *machine learning* membutuhkan sebuah metode. Metode yang didefinisikan mempunyai parameter tertentu, dan untuk menjalankannya dibutuhkan data latih atau pengalaman masa lalu (Alpaydin, 2014). Metode ini dapat digunakan untuk memprediksi di masa depan atau untuk mengumpulkan informasi dari data.

Machine learning memiliki beberapa macam pendekatan yang umum, yaitu, *supervised learning* (pembelajaran terarah), *unsupervised learning* (Pembelajaran tanpa label), dan *reinforcement learning* (pembelajaran penguatan) (Alpaydin, 2014). *Supervised learning* adalah jenis *machine learning* dimana metode belajar dari data yang sudah memiliki label (Alpaydin, 2014). Contoh penerapan *supervised learning* yaitu memprediksi harga rumah, jumlah kamar, dll. *Unsupervised learning* adalah jenis *machine learning* dari data yang tidak memiliki label (Alpaydin, 2014). Contoh dari *unsupervised learning* adalah mengelompokkan pelanggan dari perilaku mereka.

2.4 PROGRAM STUDI PENDIDIKAN TEKNOLOGI INFORMASI UNIVERSITAS BRAWIJAYA

Program Sarjana Pendidikan Teknologi Informasi (PTI) dibentuk tahun 2014 dan mulai menerima mahasiswa pada tahun 2015 (UB, 2021). Mata kuliah yang dapat dipelajari pada Program Studi Pendidikan Teknologi Informasi meliputi Mata Kuliah Pengetahuan Konten, Mata Kuliah Pengetahuan Pedagogis, Mata Kuliah Pengetahuan Teknologi, Mata Kuliah Pengetahuan Konten Pedagogis, Mata Kuliah Pengetahuan Konten Teknologi, Mata Kuliah Teknologi Mata Kuliah Pengetahuan Pedagogi, Mata Kuliah Pengetahuan Konten Pedagogis Teknologi, Mata Kuliah Penelitian dan Kerja Lapangan, dan Mata Kuliah Umum (UB, 2021). Pada Program Studi Pendidikan Teknologi Informasi terdapat 5 *career opportunity* yaitu edukator di bidang pengembangan teknologi

informasi, *IT Developer*, konten *developer* dalam bidang edukasi, ilmu data pada bidang edukasi, dan *entrepreneur* pada bidang teknologi informasi.

2.5 KLASIFIKASI

Klasifikasi adalah tipe *supervised learning* dalam *machine learning*. Klasifikasi memiliki metode belajar dengan mengidentifikasi data menjadi suatu kategori berdasarkan pola atau karakteristik yang ditemukan dari data yang telah diberi label sebelumnya (Alpaydin, 2014). Tujuan dari klasifikasi adalah menentukan kelas dari data baru berdasarkan data yang telah dipelajari. Contoh penerapan dari klasifikasi adalah menentukan email spam atau bukan, menentukan penyakit pada pasien, dll (Alpaydin, 2014). Ada berbagai jenis algoritma dalam klasifikasi, seperti *Neural Network*, *Naïve Bayes*, *Decision Tree*, *Support Vector Machines (SVM)*, dll (Alpaydin, 2014). Algoritma klasifikasi digunakan untuk memprediksi atau mengambil kesimpulan berdasarkan pola pada data yang memiliki kelas.

2.6 KLASIFIKASI NAÏVE BAYES

Naive Bayes merupakan salah satu metode klasifikasi yang didasarkan pada probabilitas dan Teorema Bayesian yang mengasumsikan bahwa setiap variabel X bersifat bebas (Rahman, et al., 2020). Metode ini sudah banyak digunakan dalam berbagai tugas, antara lain klasifikasi dokumen, klasifikasi teks, dan analisis sentimen. Naive bayes punya beberapa kelebihan yaitu, mudah untuk diimplementasikan, dan punya hasil yang baik saat diterapkan pada banyak kasus (Rahman, et al., 2020). Secara matematis, Naive Bayes memiliki rumus statistik yang dapat dilihat di bawah:

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (1)$$

Keterangan rumus:

X = Data dengan class yang belum diketahui

C_i = Kelas yang tersedia

$P(C_i|X)$ = Probabilitas munculnya kelas C_i berdasarkan kondisi x
 $P(C_i)$ = Probabilitas munculnya C_i

$P(X|C_i)$ = Probabilitas X berdasarkan kondisi C_i

$P(X)$ = Probabilitas dari X

2.7 KLASIFIKASI K-NEAREST NEIGHBOR

K-Nearest Neighbor yaitu suatu metode yang dipakai untuk mengklasifikasikan suatu objek, yang didasarkan pada banyak k data latih yang memiliki jarak terdekat dengan objek tersebut (Samudra, et al., 2020). K adalah jumlah tetangga dengan jarak terdekat yang akan dijadikan penentu kelas suatu objek (Alpaydin, 2014). Algoritma ini dijalankan dengan melakukan pemungutan suara dari K tetangga yang terpilih. Kelas yang jumlah suaranya terbanyak akan digunakan sebagai kelas dari hasil prediksi pada data uji (Samudra, et al., 2020).

Algoritma K-Nearest Neighbor memiliki kelebihan, yaitu kuat terhadap data dengan banyak noise dan data dalam jumlah besar. Salah satu kelemahan algoritma K-Nearest Neighbor adalah sensitif terhadap nilai K yang tidak tepat (Cholil, et al., 2020). Klasifikasi K-Nearest Neighbor dapat digunakan untuk klasifikasi teks, pengenalan pola, sistem rekomendasi, dll.

2.8 EVALUASI KLASIFIKASI

Evaluasi klasifikasi adalah sebuah proses untuk menilai seberapa baik metode klasifikasi bekerja (Alpaydin, 2014). Proses evaluasi ini mencakup beberapa metrik yang memberikan gambaran kemampuan metode untuk memahami pola data. Salah satu metrik tersebut adalah *confusion matrix*, *confusion matrix* adalah sebuah tabel yang membandingkan hasil prediksi dengan kelas sebenarnya. Bentuk *confusion matrix* dapat dilihat pada Tabel 1.

Tabel 1. Confusion Matrix

| Nilai Aktual | Prediksi | |
|--------------|---------------------------|----------------------------|
| | Positif | Negatif |
| Positif | <i>True Positive (TP)</i> | <i>False Positive (FP)</i> |

| Negatif | <i>False Negative (FP)</i> | <i>True Negative (TN)</i> |
|---------|----------------------------|---------------------------|
|---------|----------------------------|---------------------------|

Penjelasan dari Tabel 1:

True Positive (TP)= Banyak data kelas positif, dan diprediksi dalam kelas positif juga

True Negative (TN)= Banyak data kelas negatif, dan diprediksi dalam kelas negatif juga

False Positive (FP)= Banyak data kelas positif, namun diprediksi dalam kelas negatif

False Negative (FP)= Banyak data kelas negatif, namun diprediksi dalam kelas positif

Selain *confusion matrix*, ada beberapa metrik evaluasi yang sangat penting, seperti *accuracy* yang menghitung presentase prediksi yang benar terhadap kelas data asli, *precision* yang mengukur seberapa tepat metode dalam memprediksi kelas positif, *recall* yang menunjukkan seberapa baik metode mendeteksi kelas positif, dan *F1-score* yaitu nilai gabungan dari *precision* dan *recall*, yang memberikan gambaran tentang kualitas mode secara menyeluruh (Alpaydin, 2014).

2.9 PYTHON

Python merupakan salah satu bahasa pemrograman yang bisa digunakan untuk mengembangkan aplikasi, analisis data, dan lainnya (Romzi & Kurniawan, 2020). Python awalnya dikembangkan oleh Guido Van Rossum di Belanda pada tahun 1990 sebagai hobi (Romzi & Kurniawan, 2020). Karena sederhana, ringkas, dan punya pustaka yang luas, python dipakai secara luas dalam bidang pendidikan dan industri. Menurut Luiz (2013), Python umum digunakan dalam hal Pemrograman sistem, Pemrograman berbasis gambar, Pemrograman jaringan, Pemrograman basis data, Pemrograman ilmiah, Pemrograman Game, Penambahan data, Pemrograman robot, Excel, dll.

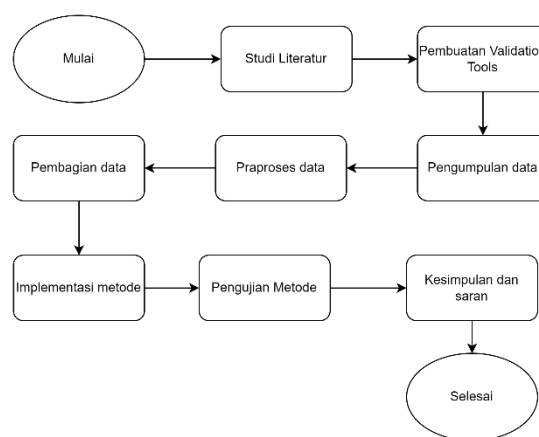
2.10 GOOGLE COLAB

Google colab merupakan platform yang disediakan google untuk pemrograman python berbasis cloud (Guntara, 2023). Google colab telah menyediakan banyak pustaka (*library*) yang bisa digunakan. (Guntara, 2023). Pustaka yang disediakan oleh Google Colab, yaitu, Numpy, Pandas, dll (Guntara, 2023). Google Colab dapat digunakan untuk *machine learning*

dengan fitur impor dataset, pengklasifikasian, dan evaluasi metode (Google, 2023).

Google colab dijalankan oleh server dari Google (Guntara, 2023). Google Colab dapat dijalankan tanpa kendala dengan internet yang lancar (Guntara, 2023). Google colab memiliki beberapa kelebihan, yaitu, tidak perlu konfigurasi, akses gratis, dan mudah dibagikan (Google, 2023).

3. METODOLOGI PENELITIAN



Gambar 1 Diagram Alir proses Penelitian

Pada Gambar 1 tahap pertama adalah Studi Pustaka, memungkinkan peneliti untuk mencari, memilih, dan memahami literatur terkait yang relevan dengan topik penelitian. Langkah berikutnya adalah Pembuatan *Validation Tools*, di mana peneliti melakukan simulasi proses penelitian untuk memvalidasi metode yang akan digunakan. Setelah itu, Pengumpulan Data dilakukan melalui *Tracer Study* Filkom untuk mendapatkan informasi pekerjaan alumni dan nilai mata kuliah dari sistem informasi akademik mahasiswa. Proses Praproses Data kemudian dilakukan, di mana data yang diperoleh dirapikan, diubah menjadi bentuk yang sesuai, dan dikelompokkan sesuai dengan kategori pekerjaan alumni. Data kemudian dimasukkan ke Google Colab dalam langkah Memasukkan Data agar dapat digunakan dalam program Python. Pembagian Data dilakukan untuk memisahkan data menjadi data latih dan data uji menggunakan Python. Langkah selanjutnya adalah Implementasi Metode, di mana data yang telah dipersiapkan diuji menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor dalam program Python. Terakhir, Pengujian Metode

dilakukan untuk mengukur keakuratan metode menggunakan data uji, dengan berbagai metrik evaluasi seperti akurasi, *precision*, *recall*, *f1score*, dan *confusion matrix*. Proses ini menghasilkan output dalam bentuk persentase yang mewakili kinerja metode yang telah diuji.

4. PROSES IMPLEMENTASI METODE

4.1 PENGUMPULAN DATA

Data didapatkan setelah serangkaian perizinan pengambilan data dilakukan. Data yang didapatkan ada 2 buah yaitu data nilai dan data *tracer study* alumni. Bentuk data yang didapatkan merupakan data berformat excel.

4.2 PRAPROSES DATA

4.2.1 DATA CLEANING

Data cleaning atau tahap pembersihan data dilakukan pada data *tracer study* yang telah didapatkan. Data yang dihapus adalah data lulusan yang belum mendapatkan pekerjaan dan telah mendapat pekerjaan namun diluar kesempatan karir pada profil PTI. Proses data cleaning dilakukan menggunakan Microsoft excel dengan diseleksi secara manual. Data sebelum di bersihkan dapat dilihat pada Gambar 2.

| Bekerja/Wiraswasta/Studi Laf | Tempat Bekerja | Posisi Aktuas | Jabatan |
|--|---------------------------------------|---------------------|-------------------------------|
| Bekerja (fulltime / part time) | ELC Tuban | | Guru |
| Dosen | SMK PGRI 3 Kota Blitar | | Dosen |
| Tenaga Profesional/Karyawan | Indobuyer | | jr. Machine learning engineer |
| Dosen | Universitas Brawijaya | | Dosen LB |
| S2 sambilan Kerja/Karya sambilan S2 | BADAN USAHA AKADEMIK UIR / UIR MEDCOM | | STAF IT |
| Wiraswasta | Surya nirmana kontraktor | | Wiraswasta |
| Bekerja (fulltime / part time) | PT Bank Sinarmas Tbk | | Product designer |
| Tenaga Profesional/Karyawan | Global Coding Indonesia | | Web developer |
| Wiraswasta | Indi Software | | Wiraswasta |
| Tidak kerja tetapi sedang mencari | | null | |
| Tenaga Profesional/Karyawan | PT. HOME CENTER INDONESIA (INFORMA) | | DIGITAL MARKETING WEB |
| Bekerja (fulltime / part time) | PT. MENJANGAN MAS | | |
| Tenaga Profesional/Karyawan | PT. ADVISIA Teknologi Global | | Engineering Support Lead |
| Bekerja (fulltime / part time) | SMAN 1 Manganti | | Guru |
| Bekerja (fulltime / part time) | PT. UNINET BHAKTI NUSA (Kompas.com) | | |
| Dosen | Universitas Brawijaya | | Dosen LB |
| Bekerja (fulltime / part time) | PT. Roket Software Indonesia | | Software QA |
| Tenaga Profesional/Karyawan | PT. Global Coding Indonesia | | Manager, HRD |
| Bekerja (fulltime / part time) | Edubit Indonesia | | Programmer |
| Bekerja ke suatu perusahaan / ke SMK Muhammadiyah 1 Sangatta | | Guru honorar | |
| Bekerja ke suatu perusahaan / ke SMK Muhammadiyah 1 Sangatta | | Head of Engineering | Backend Developer |
| Tidak kerja tetapi sedang mencari | | null | |
| Tenaga Profesional/Karyawan | PT. Bumi Amerta Teknologi Mandiri | | Full-Stack Developer |

Gambar 2 Data Tracer Study Sebelum Dibersihkan

Proses pembersihan data dilakukan dengan menghapus data mahasiswa yang belum memiliki pekerjaan. Proses penghapusan data dilakukan secara manual dengan memilih data mahasiswa yang kolom tempat bekerja atau jabatannya null. Data yang memenuhi kriteria untuk dimasukkan kedalam dataset ada sebanyak 80 data. Hasil pembersihan data dapat dilihat pada Gambar 3.

| Bekerja/Wiraswasta/Studi Laf | Tempat Bekerja | Posisi Aktuas | Jabatan |
|---|--|-------------------------|----------------------------------|
| Bekerja (fulltime / part time) | ELC Tuban | | Guru |
| Guru | SMK PGRI 3 Kota Blitar | | Guru |
| Dosen | Universitas Brawijaya | | Dosen |
| Bekerja (fulltime / part time) | Universitas Brawijaya | | Dosen LB |
| Dosen | SMAN 1 Manganti | | Guru |
| Dosen | Universitas Brawijaya | | Dosen LB |
| Bekerja ke suatu perusahaan / ke SMK Muhammadiyah 1 Sangatta | | Guru honorar | |
| Guru | SMK Cendekia Bangsa Kepanjen | | Information Technology Guru |
| Bekerja ke suatu perusahaan / ke - Pondok Pesantren Tereasabidusul Uluu Warohmi Pimpinan Pondok Pesantren | | | GURU MENGAJI |
| Bekerja (fulltime / part time) | SMKS PGRI Wlingi | | Guru |
| Bekerja (fulltime / part time) | Coding Bee Academy | | Mentor IT |
| Bekerja ke suatu perusahaan / ke - Pondok Pesantren Tereasabidusul Uluu Warohmi Pimpinan Pondok Pesantren | | | Guru/honor |
| Bekerja ke suatu perusahaan / ke SMKN 1 KOTABARU | | Kepala Sekolah | Guru |
| Guru | Smk kesehatan bhakti Indonesia medika kota mojokerto | | pendidik dan IT sekolah |
| Bekerja (fulltime / part time) | MTS Negeri 1 Mojokerto | | Guru |
| Bekerja (fulltime / part time) | Yayasan Pelita Hidayah | | Guru |
| Bekerja (fulltime / part time) | Bani Masjim | | Guru |
| Bekerja (fulltime / part time) | SMK Negeri Wrongonan | | Guru |
| Guru | SMK Queen Al-Falah | | Guru MM dan TKJ |
| Guru | MAN 2 TULUNGAGUNG | | Guru Informatika dan Staff Humas |
| Guru | Brightchamps | | Coding Educator |
| Bekerja ke suatu perusahaan / ke SMK Negeri Puncokari | | Kepala Program Keahlian | Guru |
| S2 sambilan Kerja/Karya sambilan S2 | PT Brightchamps Edutech Indonesia | | Coding Educator |
| Bekerja ke suatu perusahaan / ke SMK Muhammadiyah 03 Ngimbang | | Kepala sekolah | Guru |

Gambar 3 Data Tracer Study Setelah Dibersihkan

4.2.2 DATA INTEGRATION

Pada proses ini dibuatkan file untuk penggabungan antara data nilai mata kuliah dengan pekerjaan. Data alumni yang nilainya diambil adalah data alumni yang sudah memiliki pekerjaan pada data *tracer study*. Hasil data integration dapat dilihat pada gambar 4.

| Statistik | Teknik Pengolahan Audio dan Vlc | Teknologi Pembelajaran Kru | Teori Belajar dan Pembelajaran | Tugas Akhir/ Skripsi | Pekerjaan |
|-----------|---------------------------------|----------------------------|--------------------------------|----------------------|---------------------|
| B | - | - | B | - | B |
| B | - | - | A | - | B+ |
| B | - | - | B | - | Guru |
| A | - | - | B+ | - | Guru |
| A | - | - | A | - | Developer |
| B | - | - | A | - | Guru |
| B | - | - | B+ | - | A |
| B+ | - | - | A | - | Programmer |
| C+ | - | - | A | - | Developer |
| D | - | - | A | - | Developer |
| B+ | - | - | B+ | - | Developer |
| C | - | - | B | - | Wiraswasta |
| C | - | - | C | - | Developer |
| C | - | - | C+ | - | Guru |
| B | - | - | B | - | Developer |
| B | - | - | B+ | - | Fullstack Developer |
| D+ | - | - | A | - | Developer |
| C | - | - | C | - | Guru |
| E | - | - | D | - | Developer |
| D | - | - | C+ | - | Wiraswasta |
| B | - | - | B | - | Developer |
| A | - | - | C+ | - | Guru |
| A | - | - | C | - | Guru |
| C+ | - | - | C+ | - | Guru |

Gambar 4 Data Integration

4.2.3 Data Transformation Agar Bisa diproses oleh algoritma Naïve Bayes, data nilai mata kuliah diubah menjadi data numerik dengan ketentuan yang ada. Dan data pekerjaan dijadikan kelas antara 1 atau 2. Data pekerjaan diubah menjadi 1 atau 2 agar mempermudah algoritma dalam mengelaskan data. Hasil data transformation daro data nilai dan pekerjaan dapat dilihat pada Gambar 5

| Sistem Op | Statistika | Teknik Pe | Teknologi | Teori Bela | Tugas Akh | Pekerjaan |
|-----------|------------|-----------|-----------|------------|-----------|-----------|
| 3 | 3 | 0 | 0 | 3 | 4 | 1 |
| 3 | 3 | 0 | 0 | 4 | 3.5 | 1 |
| 3.5 | 3 | 0 | 0 | 3 | 4 | 1 |
| 4 | 4 | 0 | 0 | 3.5 | 3 | 2 |
| 4 | 4 | 0 | 0 | 4 | 4 | 1 |
| 2.5 | 3 | 0 | 0 | 4 | 4 | 2 |
| 2 | 3 | 0 | 0 | 3.5 | 4 | 2 |
| 3 | 2.5 | 0 | 0 | 4 | 4 | 2 |
| 3 | 2.5 | 0 | 0 | 4 | 3.5 | 2 |
| 3 | 1 | 0 | 0 | 3.5 | 3 | 2 |
| 3.5 | 3.5 | 0 | 0 | 3 | 4 | 2 |
| 4 | 2 | 0 | 0 | 2.5 | 4 | 2 |
| 3.5 | 2 | 0 | 0 | 2 | 3.5 | 1 |
| 3 | 0 | 0 | 0 | 2.5 | 3 | 1 |
| 2 | 2 | 0 | 0 | 3 | 3.5 | 2 |
| 3.5 | 3 | 0 | 0 | 3.5 | 3.5 | 2 |
| 3.5 | 1.5 | 0 | 0 | 4 | 3.5 | 2 |
| 3 | 2 | 0 | 0 | 2.5 | 4 | 1 |

Gambar 5 Data Transformation

4.3 MEMASUKKAN DATA

Setelah tahap praproses data selesai, dataset yang sudah jadi kemudian dimasukkan ke dalam google colab. Setelah data diupload, kemudian library dipanggil untuk membantu pemrosesan data.

4.4 PEMBAGIAN DATA

Setelah data dimasukkan ke dalam google colab dan dijadikan variabel. Kemudian data dibagi atau dipisah menjadi data training dan data testing.

4.5 IMPLEMENTASI METODE

Implementasi dilakukan dengan memasukkan data training ke dalam classifier. Pada tahap ini, ada 2 skenario, yaitu implementasi data pada metode Naïve Bayes, dan K-Nearest Neighbor.

5. HASIL PENGUJIAN DAN ANALISIS

5.1 PENGUJIAN NAÏVE BAYES

Proses pengujian Naïve Bayes dilakukan 5 kali dengan 5 variasi jumlah data training dan data testing yang berbeda yakni 50%:50%, 60%:40%, 70%:30%, 80%:20%, dan 90%:10. Analisis dilakukan dengan membandingkan hasil pengujian metode berupa nilai *accuracy*, *precision*, *recall*, dan *f1-score* pada tiap pembagian data training dan data testing. Hasil perbandingan tiap skenario telah disatukan dalam tabel dan dapat dilihat pada Tabel 2

Tabel 2 Hasil Pengujian Naive Bayes

| Training: Testing | Accur acy | Preci sion | Recal l | F1scor e |
|----------------------|--------------|---------------|-------------|-------------|
| 50%:50% | 63.41 % | 64.38 % | 63.69 % | 63.0 6% |
| 60%:40% | 63.63 % | 71.75 % | 66.11 %% | 61.9 2% |
| 70%:30% | 60% | 67.86 % | 65% | 59.4 2% |
| 80%:20% | 64.70 % | 76.92 % | 70% | 63.5 7% |
| 90%:10% | 66.66 % | 78.57 % | 70% | 64.9 4% |

5.2 PENGUJIAN K-NEAREST NEIGHBOR

Proses pengujian K-Nearest Neighbor dilakukan 5 kali dengan 5 variasi jumlah data training dan data testing yang berbeda yakni 50%:50%, 60%:40%, 70%:30%, 80%:20%, dan 90%:10. Analisis dilakukan dengan membandingkan hasil pengujian metode berupa nilai *accuracy*, *precision*, *recall*, dan *f1-score* pada tiap pembagian data training dan data testing. Hasil perbandingan tiap skenario telah disatukan dalam tabel dan dapat dilihat pada Tabel 3.

Tabel 3 Hasil Pengujian K-Nearest Neighbor

| Training: Testing | Accur acy | Preci sion | Recal l | F1scor e |
|----------------------|--------------|---------------|-------------|-------------|
| 50%:50% | 63.41 % | 66.36 % | 62.86 % | 61.1 % |
| 60%:40% | 66.66 % | 66.66 % | 66.66 %% | 65.5 3% |
| 70%:30% | 64% | 62.15 % | 61.67 % | 61.8 % |
| 80%:20% | 52.94 % | 51.43 % | 51.43 % | 51.4 3% |
| 90%:10% | 55.55 % | 53.57 % | 52.5 % | 50% |

5.3 Analisis Perbandingan Hasil Metode

Pada langkah ini, perbandingan hasil tiap metode dilakukan untuk mengetahui metode mana yang memiliki tingkat akurasi untuk tiap metode terhadap data. Perbandingan dilakukan dengan melihat besar nilai akurasi, *precision*, *recall*, dan *f1-score* dari tiap metode. Hasil perbandingan klasifikasi kedua metode dapat dilihat pada Tabel 4.

Tabel 4 Perbandingan Hasil Metode

| Nilai Evaluasi | Metode | |
|-------------------|-------------|--------------------|
| | Naive Bayes | K-Nearest Neighbor |
| Skenario | 90%:10% | 60%:40% |
| Accuracy | 66.66% | 66.66% |
| Precision | 78.57% | 66.66% |

| | | |
|-----------------|--------|--------|
| <i>Recall</i> | 70% | 66.66% |
| <i>F1-Score</i> | 64.94% | 65.53% |

Pada Tabel 4 ditunjukkan bahwa nilai akurasi kedua metode bernilai sama yaitu sebesar 66.66%. Nilai akurasi Naive Bayes tertinggi ada pada skenario 90% data *training* dan 10% data *testing*. Nilai akurasi K-Nearest Neighbor tertinggi ada pada skenario 60% data *training* dan 40% data *testing*. Hasil K-Nearest Neighbor lebih efektif jika jumlah data testing banyak, sedangkan naive bayes memiliki keunggulan lebih stabil meskipun data testing tidak banyak (Devita, et al., 2018). Untuk nilai *precision* dan *recall*, metode Naive Bayes menunjukkan nilai yang lebih tinggi yaitu 70% nilai *recall* dan 78.57% nilai *precision*.

6. KESIMPULAN

Berdasarkan temuan-temuan pada penelitian ini, maka dapat ditarik kesimpulan bahwa:

1. Proses analisis dimulai dengan membuat *dataset* yang berisi nilai mata kuliah dan pekerjaan para alumni. Sebelum *dataset* bisa dianalisis, *dataset* perlu dipraproses terlebih dahulu. Praproses dilakukan dengan membersihkan data, menggabungkan data, dan mentransformasi data. Setelah praproses selesai dan *dataset* siap digunakan, *dataset* dimasukkan ke dalam Google Colab untuk dapat dianalisis. Setelah data dimasukkan, data kemudian dibagi menjadi data latih dan data uji. Data kemudian akan dimasukkan kedalam metode Naive Bayes dan K-Nearest Neighbor untuk didapatkan hasil prediksinya. Setelah itu hasil prediksi tiap metode dianalisis dengan membandingkannya dengan data uji. Proses analisis dilakukan dengan melihat nilai akurasi data, *precision*, *recall*, *f1-score*, dan *confusion matrix*.
2. Berdasarkan hasil analisis pengujian, ditemukan faktor apa saja yang mempengaruhi nilai akurasi dari tiap metode. Pada metode Naive Bayes akurasi meningkat ketika data latih semakin banyak. Sedangkan pada metode K-Nearest Neighbor akurasi meningkat ketika data uji semakin banyak.
3. Pada metode naive bayes, nilai akurasi, *precision*, *recall*, dan *f1-score* terbesar terdapat pada pembagian data 90% data *Training* dan 10% data *testing*. Sedangkan pada metode K-Nearest Neighbor nilai tertingginya terdapat pada pembagian data 60% data *training* dan 40% data *testing*.

7. SARAN

Melihat hasil penelitian ini, terdapat beberapa saran yang dapat dilakukan untuk penelitian selanjutnya terkait topik serupa, yaitu:

1. Menggunakan dataset dengan data yang jumlahnya lebih banyak
2. Menggunakan metode lain untuk menambah pengetahuan.

8. DAFTAR PUSTAKA

- Alpaydin, E., 2014. *Introduction to Machine Learning*. 3 ed. Massachusetts: Massachusetts Institute of Technology.
- Annisa, R. & Sasongko, A., 2020. Prediksi Nilai Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes. *Jurnal Sains dan Teknologi (JST)*, 9(1), pp. 1-10.
- Bengkulu, K. B., 2016. *Kantor Bahasa Provinsi Bengkulu*. [Online] Available at: <https://kantorbahasabengkulu.kemdikbud.go.id/makna-kata-pekerjaan-profesi-dan-jabatan/>
- BINUS , 2022. *Teknik pre-processing dan classification dalam data scienc*. [Online] Available at: <https://mie.binus.ac.id/2022/08/26/teknik-pre-processing-dan-classification-dalam-data-science/>
- Brawijaya, U., 2015. *DPKA UB*. [Online] Available at: <https://dpka.ub.ac.id/id/profil-singkat/>
- Cholil, S. R., Handayani, T., Prathivi, R. & Ardianita, T., 2020. Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *Indonesian Journal on Computer and Information Technology*, 6(2), pp. 118-127.

- Devita, R. N., Herwanto, H. W. & Wibawa, A. P., 2018. Perbandingankinerjametode Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, 5(4), pp. 427-434.
- Ekawati, S., Basir, F. & Karmila, 2021. Pengaruh Kemampuan Awal dan Minat Belajar Terhadap Hasil Belajar Mahasiswa pada Mata Kuliah Kalkulus Dasar Universitas Cokroaminoto Palopo. *Pedagogy: Jurnal Pendidikan Matematika*, 6(2), pp. 188-196.
- Google, 2023. *Google Colab*. [Online] Available at: <https://colab.research.google.com/>
- Guntara, R. G., 2023. Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1), pp. 55-60.
- Hidayah, K., 2020. *Lembaga Administrasi Negara Republik Indonesia*. [Online] Available at: <https://lan.go.id/?p=12800>
- Kaligis, R., Sompie, B., Tjakra, J. & Walangitan, D., 2013. Pengaruh Implementasi Program Keselamatan dan Kesehatan Kerja (K3) Terhadap Produktivitas Kerja. *Jurnal Sipil Statik*, 1(3), pp. 219-255.
- Luiz, M., 2013. *Learning Python*. 5th ed. Sebastopol: O'Reilly Medi.
- Rachmadiansyah, Rumlaklak, N. D. & Mauko, A. Y., 2022. Prediksi Masa Tunggu Kerja Alumni Menggunakan Naïve Bayes Classifier pada Program Studi Ilmu Komputer Universitas Nusa Cendana. *J-ICON*, 10(2), pp. 143-150.
- Rahayu, P., 2019. PENGARUH ERA DIGITAL TERHADAP PERKEMBANGAN BAHASA ANAK. *Al-Fathin*, Volume 2, pp. 47-59.
- Rahman, A., Rahmat, F., Fariqi, M. Y. & Adi, S., 2020. Metode Naive Bayes untuk Menganalisis Akurasi Sentimen Komentar di Youtube. *Jurnal EECCIS*, 14(1), pp. 31-34.
- Romzi, M. & Kurniawan, B., 2020. Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma. *JTIM: Jurnal TeknikInformatikaMahakarya*, 3(2), pp. 37-44.
- Samudra, J. A., Anraenia, S. & Hermana, 2020. Penerapan Metode K-Nearest Neighbor untuk Memprediksi Tingkat Kelulusan Mahasiswa Berbasis Web pada Fakultas Ilmu Komputer UMI. *Buletin Sistem Informasi dan Teknologi Islam*, 1(4), pp. 230-237.
- UB, P. F., 2021. *FAKULTAS ILMU KOMPUTER UB*. [Online] Available at: <https://filkom.ub.ac.id/s1-pti/>
- UNAIR, 2023. *UNAIR NEWS*. [Online] Available at: <https://unair.ac.id/pahami-perbedaan-mata-kuliah-wajib-dan-mata-kuliah-pilihan/>
- Walidain & Birrul, 2022. Perencanaan Karir Mahasiswa Semester Akhir. *Jurnal Pemikiran Pendidikan*, 12(2), pp. 58-63.