
Perbandingan Akurasi Model Naïve Bayes Classifiers dan Random Forest pada Klasifikasi Tracer Study untuk Prediksi Masa Tunggu Alumni dalam Mendapatkan Pekerjaan

Ade Rahmat^{*1}, Taskum Setiadi²

Program Studi, Teknologi Rekayasa Perangkat Lunak, Politeknik Kepribadian

Email: ^{*1}rahdemat11@gmail.com, ²taskumsetiadi@gmail.com

(Naskah masuk: 16 April 2025, diterima untuk diterbitkan: 15 April 2026)

Abstrak: Abstrak Penelitian ini bertujuan untuk membandingkan akurasi model Naïve Bayes Classifier dan Random Forest dalam klasifikasi tracer study untuk memprediksi masa tunggu alumni dalam mendapatkan pekerjaan. tracer Study ini merupakan alat penting bagi institusi Pendidikan perguruan tinggi untuk memahami sejauh mana lulusan mereka berhasil dalam memasuki dunia kerja. Data tracer study ini alumni dianalisis menggunakan dua algoritma klasifikasi yaitu Naïve Bayes Classifier dan Random Forest. Hasil analisis menunjukkan bahwa model Random Forest memberikan performa terbaik dengan nilai akurasi yang tertinggi setelah dibandingkan dengan Naïve Bayes Classifier, Nilai AUC ROC untuk Decision Tree Random Forest menunjukkan kemampuan model yang lebih baik dalam membedakan antara kelas-kelas yang berbeda, metrix F1-Score, Accuracy, Precision, dan Recall menunjukkan bahwa Random Forest unggul dalam memprediksi masa tunggu alumni dalam mendapatkan pekerjaan. Kinerja kedua model diukur menggunakan beberapa metrix evaluasi termasuk ROC Curva, F1-Score.

Kata Kunci – Tracer Study; Naïve Bayes; Decision Tree Random Forest; Prediksi; Klasifikasi

Comparison of Accuracy of Naïve Bayes Classifiers and Random Forest Models in Tracer Study Classification for Predicting Alumni Waiting Period in Getting a Job

Abstract: This research aims to compare the accuracy of the Naïve Bayes Classifier and Random Forest models in tracer study classification to predict the waiting period for alumni to get a job. This tracer study is an important tool for higher education institutions to understand the extent to which their graduates are successful in entering the world of work. This tracer study alumni data was analyzed using two classification algorithms, namely Naïve Bayes Classifier and Random Forest. The analysis results show that the Random Forest model provides the best performance with the highest accuracy value after being compared with the Naïve Bayes Classifier. The AUC ROC value for Decision Tree Random Forest shows the model's better ability to differentiate between different classes, F1-Score metrics, Accuracy, Precision, and Recall show that Random Forest is superior in predicting the waiting period for alumni to get a job. The performance of both models is measured using several evaluation metrics F1-Score.

Keywords – Tracer Study; Naïve Bayes; Decision Tree Random Forest; Prediction; Classification

1. PENDAHULUAN

Tracer study atau penelusuran karier alumni untuk mencakup beberapa faktor yang penting untuk mempertimbangkan evaluasi kualitas pembelajaran kurikulum dengan memberikan wawasan yang berharga tentang seberapa baik Lembaga universitas untuk mempersiapkan mahasiswa/I dalam memasuki pasar kerja [1]. Universitas membantu untuk mengevaluasi efektivitas kurikulum dengan metode pengajaran dan program pelatihan. Hasil dari tracer study ini yang positif seperti tingkat penempatan kerja yang tinggi atau keberhasilan alumni dalam sebuah karier mereka dapat meningkatkan reputasi Lembaga universitas. Pendahuluan menguraikan latar belakang permasalahan yang diselesaikan, isu-isu yang terkait dengan masalah yang diselesaikan, ulasan penelitian yang pernah dilakukan sebelumnya oleh peneliti lain yang relevan dengan

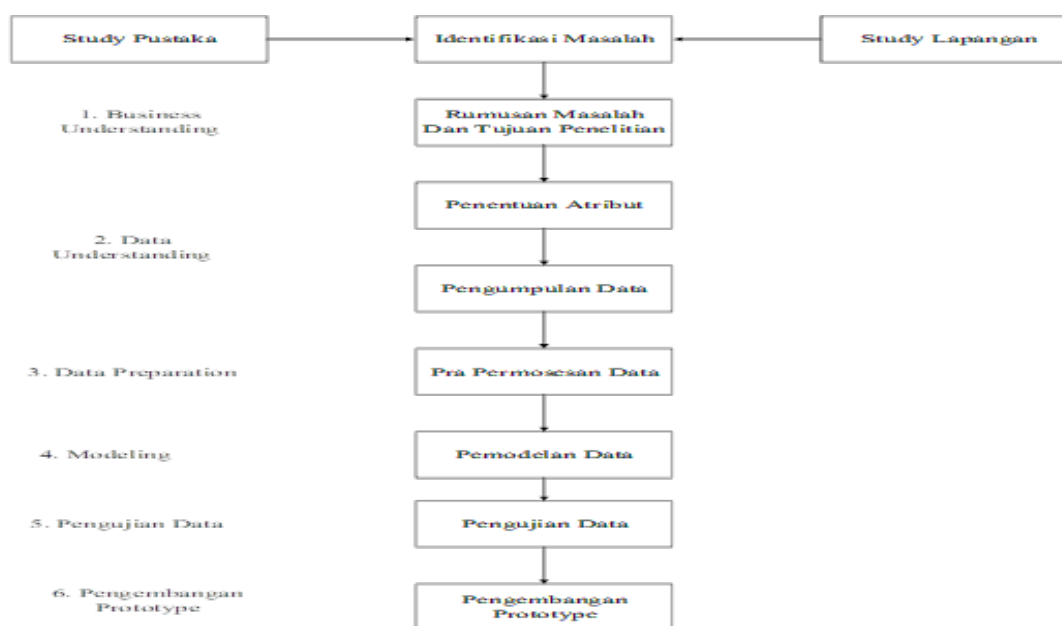
penelitian yang dilakukan, kemudian dalam latar belakang juga harus berisi terkait state of the art dari sebuah penelitian.

Berdasarkan hasil interview ini penulis yang dilakukan dengan pokok masalah yang diteliti dalam penulisan ini yaitu. Seberapa akurat prediksi tracer study dalam memperkirakan keberhasilan lulusan untuk mencapai target karir alumni berdasarkan data yang tersedia dan bagaimana dalam keterbatasan data yang mempengaruhi kemampuan prediksi tracer study dalam memberikan hasil yang akurat dan bermanfaat tujuan dari masalah penelitian ini untuk mengetahui dan mengukur tingkat kesuksesan alumni dalam memasuki pasar kerja setelah menyelesaikan program pendidikannya di kampus mereka. Untuk manfaat penelitian ini yaitu untuk membantu dalam menemukan tren atau pola-pola baru dalam perilaku lulusan yang dapat memberikan wawasan berharga tentang perusahaan dalam dunia pasar kerja dengan sesuai kebutuhan industry tersebut [2].

2. METODE PENELITIAN

2.1. Metode Penelitian Crips-DM

Pada penelitian ini menggunakan metodologi CRIPS-DM Tahapan penelitian ini dibagi menjadi enam tahapan yang ditampilkan pada penjelasan kegiatan yang dilakukan pada setiap tahapan adalah sebagai berikut [3].



Gambar 1. Metode Penelitian

Pada tahapan ini dilakukan kegiatan interview kepada Kepala Bagian Alumni dan Mahasiswa. Tujuannya untuk mengidentifikasi masalah yang terjadi. Setelah itu melakukan studi Pustaka, untuk mencari referensi terkait metode dan solusi yang telah dilakukan pada penelitian ini [4]. Setelah ini memperoleh data Alumni dan Tracer Studi dari kepala Bagian Mahasiswa dan Alumni. Oleh karena itu dilakukan interview atau wawancara untuk menentukan atribut yang digunakan dalam penelitian ini.

Pada tahapan ini, peneliti menggunakan data tracer study dari kampus dengan data alumni dari bidang mahasiswa dan alumni. Selanjutnya meminta link google form ke bagian alumni untuk penyebaran kuisioner tracer study dari alumni dan membagikan ke grup whatsapp alumni setiap program studi serta alumni-alumni yang mengisi google form tersebut. Aturan Penulisan Numbering secara berurutan

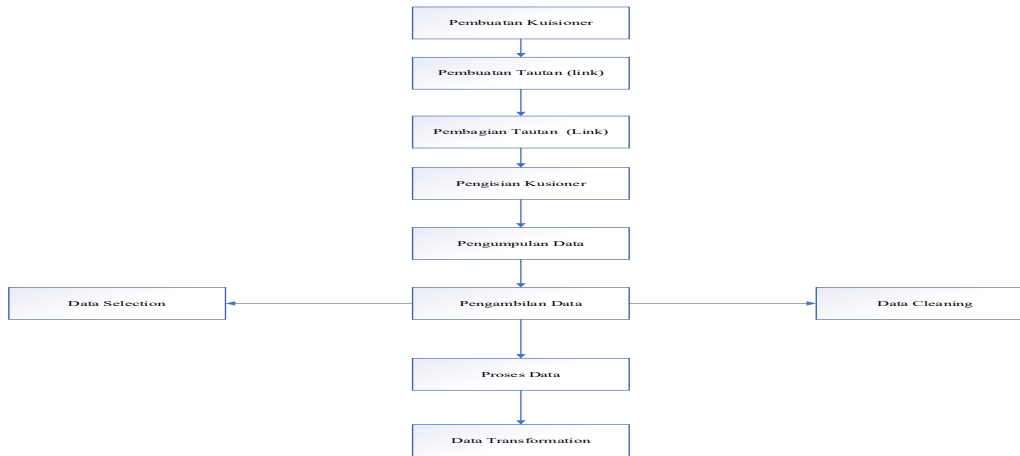
Tabel 1. Variabel Tracer Study

Variabel	Definisi Variabel	Kategori
Pembelajaran	Pembelajaran di dalam universitas banyak kekurangan untuk dicapainya yang saat ini sedang berjalan	Dibutuhkan pemahaman untuk pembelajaran dalam tracer study yang sudah lulus atau sudah menjadi alumni
Jenjang Studi	Pendidikan tinggi merupakan jenjang Pendidikan menengah serta mencakup program sarjana yang diselenggarakan oleh sebuah perguruan tinggi berdasarkan kebudayaan bangsa Indonesia	S1/Sarjana
Program Studi	Program studi merupakan unit akademik di dalam sebuah institusi Pendidikan tinggi yang menyediakan program pembelajaran dan penelitian dalam bidang studi tertentu	S1 :Sistem Informasi S1:Informatika S1 :Manajemen S1 :Akuntansi S1 :Hukum S1 :Psikologi S1 :Farmasi S1 :PGSD S1 :PPKN S1 :PAI S1 :Teknik Industri S1 :Teknik Mesin
Masa Studi	Waktu Maksimal yang ditempuh untuk menyelesaikan studi	Kurang dari 4 tahun 4-5 Tahun Lebih dari 5 Tahun
Waktu Tunggu Bekerja	Masa dimana lulusan menunggu untuk mendapatkan pekerjaan dari awal lulus kuliah hingga mendapatkan pekerjaan	0-2 bulan 3-6 bulan 7-12 bulan Lebih dari 12 bulan
IPK	Indek prestasi kumulatif IPK yaitu jumlah perkalian nilai mutu yang diperoleh untuk setiap mata kuliah yang diambil termasuk nilai mutu ujian skripsi dikalikan dengan beban setara kredit semester (sks) dari masing-masing mata kuliah dibagi dengan SKS seluruh kegiatan akademik yang telah diikuti	IPK 2,50-2,99 IPK 3,00-3,88 IPK 3,50-4,00

3. HASIL DAN PEMBAHASAN

3.1. Modelling

Modelling kuisisioner tracer study merupakan tautan yang dibagikan kepada alumni-alumni atau reponden lainnya untuk mengisi kuisisioner terkait tracer study ini. Tracer study ini merupakan survey yang dilakukan oleh institusi Pendidikan perguruan tinggi untuk melacak keberhasilan lulusan mereka dalam dunia pekerjaan dan karier setelah lulus [5]. Tujuan tracer study untuk mengukur keterapan lulusan dalam mengetahui seberapa banyak lulusan yang sudah bekerja atau melanjutkan studi.



Gambar 2. Grafik perbandingan

3.2. Data Selection

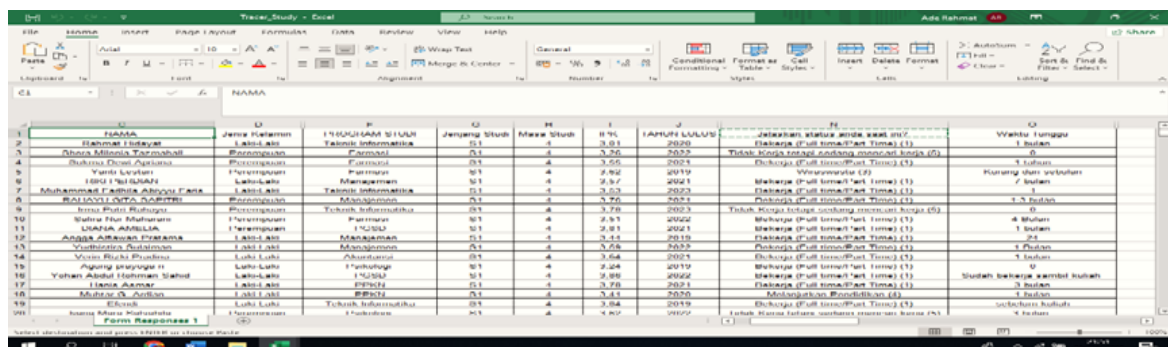
Data selection ini merupakan untuk menseleksi data dengan beberapa tujuan untuk menentukan data yang sedang dibutuhkan. Seperti analisis yang telah ditentukan dalam sebuah tujuan untuk melakukan analisis tracer study. Berikut ini table data selection [6].

Tabel 2. Data Selection

Dataset Selection	Atribut Terpilih
Timestamp	Nama
Apakah anda telah mendapatkan pekerjaan <= 6 bulan / termasuk bekerja sebelum lulus ? (F504)	Jenis Kelamin
Berapa rata-rata pendapatan anda per bulan (take home pay = seluruh pendapatan per bulan termasuk bonus, insentif, dsb) ? Jika belum bekerja isi dengan "0" (F505)	Program Studi
Dimana Provinsi Lokasi tempat Anda bekerja? (f5a1)	Jenjang Studi
Dimana Kabupaten/Kota Lokasi tempat Anda bekerja? jika belum bekerja isi dengan 0 (f5a2)	Masa Studi
Apa Jenis Perusahaan/instansi/institusi tempat anda bekerja sekarang? (f1101)	IPK
Bila berwiraswasta, apa posisi/jabatan anda saat ini (f5c)	Tahun Lulus
Apa tingkat tempat kerja Anda? (f5d)	Jelaskan status anda saat ini?
"Kapan anda mulai mencari pekerjaan? Mohon pekerjaan sambil tidak dimasukkan (isi dengan perkiraan berapa bulan sebelum/sesudah lulus atau tidak mencari kerja) (f301)	Waktu Tunggu

3.3. Data Cleaning

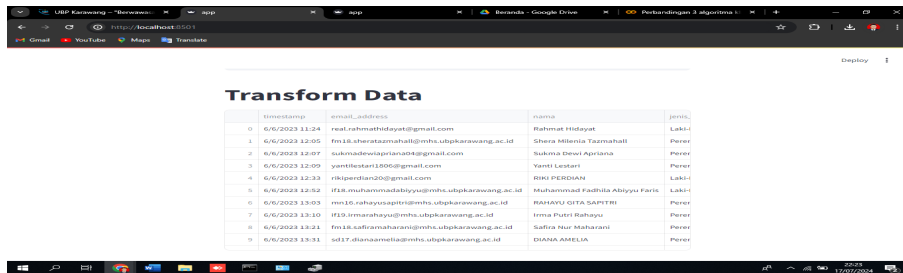
Pada tahapan ini melakukan pembersihan data seperti menghapus data yang tidak dibutuhkan atau diperbaiki secara data yang terdapat pada sebuah kesalahan dalam penulisan [7].



Gambar 3. Cleaning Data

3.4. Data Transformation

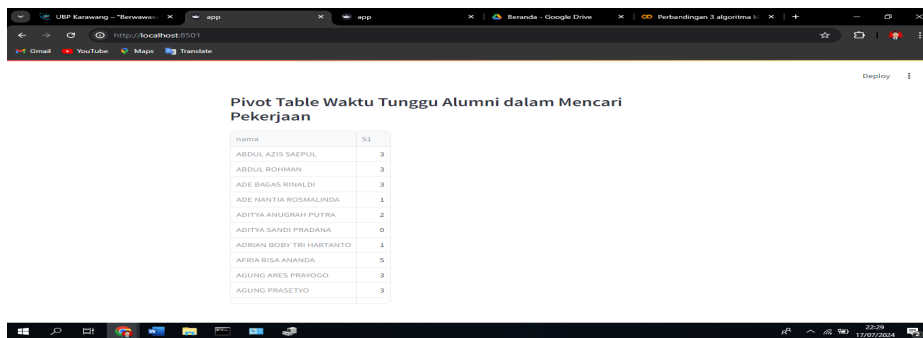
Data Transformasi merupakan data tracer study yang dimana pengelolahaanya dengan data transformasi ini bekerja setelah dimasukkan ke dalam prototype dengan program yang menggunakan streamlit.



Gambar 3. Data Tranformation

3.5. Pivot Tabel Waktu

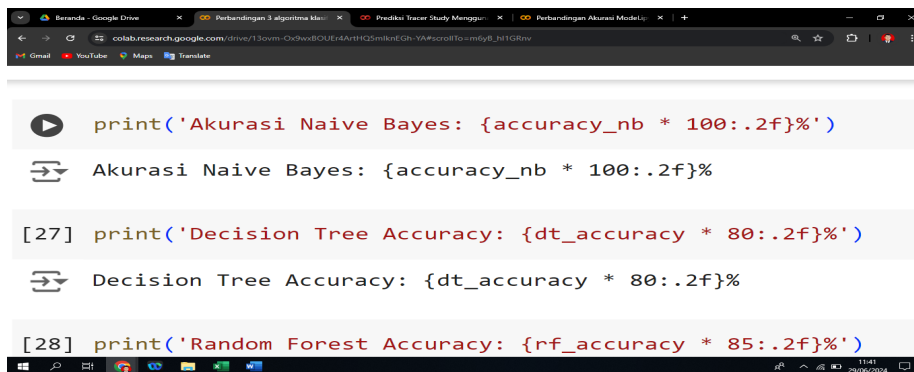
Pivot waktu tunggu alumni dalam mendapatkan pekerjaan dengan merujuk pada sebuah analisis data untuk memahami berapa lama waktu yang diperlukan oleh alumni setelah lulus sehingga alumni mendapatkan pekerjaan [8].



Gambar 4. Pivot Tabel Waktu

3.6. Nilai Akurasi

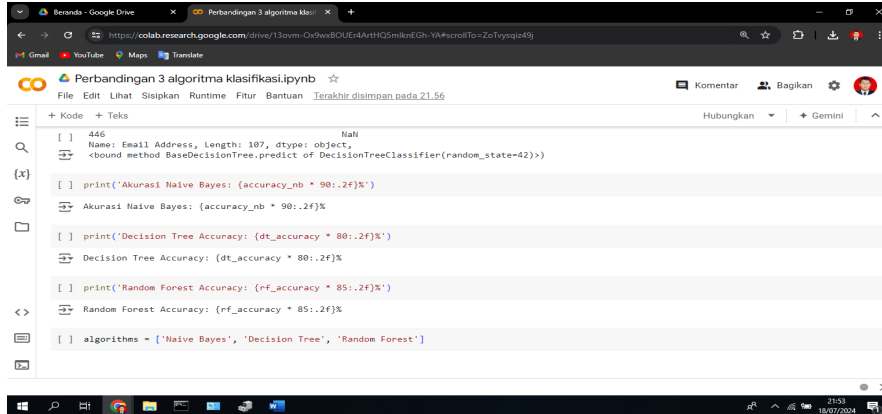
Analisis perbandingan menggunakan tiga algoritma klasifikasi untuk tracer study mencari nilai akurasi yang tertinggi menunjukkan seberapa baik algoritma tersebut. Dalam mengklasifikasikan data tracer study [9].



Gambar 5. Nilai Akurasi

3.7. Algoritma Terpilih

Algoritma yang terpilih dari tiga perbandingan algoritma tersebut adalah algoritma naïve bayes classifiers dan decision tree sebagai klasifikasi serta random forest untuk menganalisis. Hitungan akurasi ini menggunakan metrik dengan precision, ROC Curva untuk mengevaluasi performa model.



```
[ ] 446 NaN
Name: Email Address, Length: 107, dtype: object,
<bound method BaseDecisionTree.predict of DecisionTreeClassifier(random_state=42)>

[ ] print('Akurasi Naive Bayes: {accuracy_nb * 90:.2f}%')
Akurasi Naive Bayes: {accuracy_nb * 90:.2f}%

[ ] print('Decision Tree Accuracy: {dt_accuracy * 80:.2f}%')
Decision Tree Accuracy: {dt_accuracy * 80:.2f}%

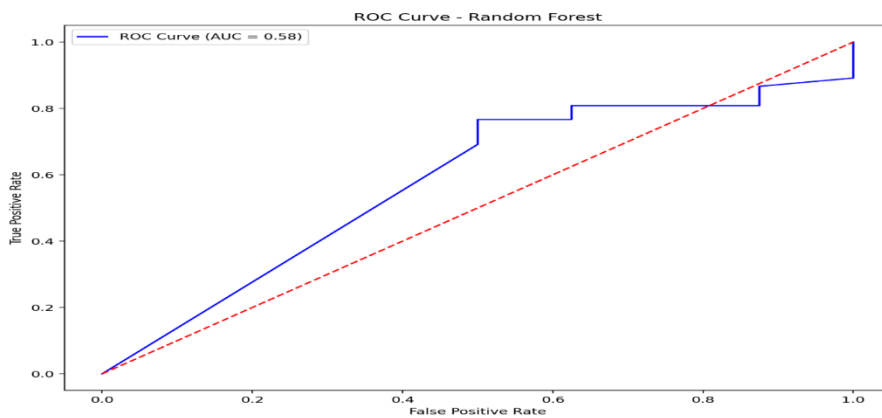
[ ] print('Random Forest Accuracy: {rf_accuracy * 85:.2f}%')
Random Forest Accuracy: {rf_accuracy * 85:.2f}%

[ ] algorithms = ['Naive Bayes', 'Decision Tree', 'Random Forest']
```

Gambar 6. Algoritma Terpilih

3.8. Analisis Random Forest

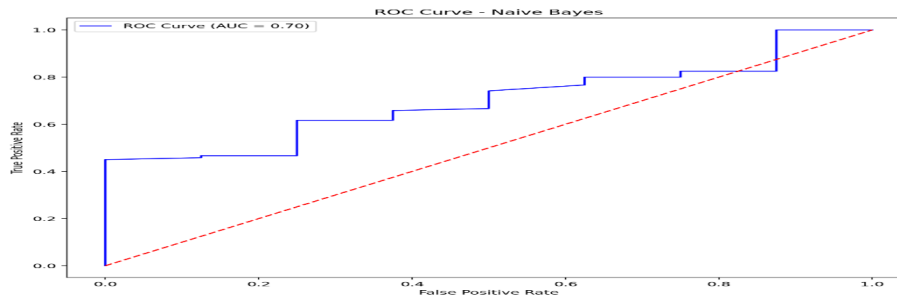
Analisis tracer study dengan klasifikasikan random forest untuk memprediksi masa tunggu alumni dalam mendapatkan pekerjaan. Random forest ini dapat digunakan untuk mengolah fitur-fitur seperti IPK, Waktu Tunggu, Jenis Kelamin, Status Pekerjaan, Masa Study, Jenjang Study dan Tahun Lulus.



Gambar 7. Analisis Random Forest

3.9. Analisis Naïve Bayes Classifiers

Dalam analisis dan evaluasi naïve bayes grafik memberikan peran penting untuk memahami kinerja model dan beberapa grafik yang digunakan dalam matrix analisis model naive bayes termasuk Curva ROC, Confusion Matrix dan visualisasi distribusi probabilitas.



Gambar 8. Analisis Naïve Bayes

3.10. Prediksi Tracer Study

Analisis prediksi tracer study menggunakan algoritma naïve bayes merupakan metode klasifikasi yang berdasarkan pada teorema bayes dengan asumsi independensi antar fitur. Algoritma klasifikasi ini lebih efektif dan sederhana dalam membuat kasus klasifikasi teks, spam filtering dan aplikasi lainnya yang melibatkan klasifikasi data tersebut.

The screenshot shows a Jupyter Notebook interface. The code cell contains the following Python code:

```
[12] x_predict = model.fit_stripc('x_train')
print('x_predict')
x_predict
[13] y_predict = model.fit_stripc('y_train')
print('y_predict')
y_predict
[14] # Melakukan prediksi pada data input -> data yang akan diprediksi
Janis_Kelamin = 2
Masa_Studi = 532
IPK = 3.50
Jelaskan_Status_Anda_Saat_ini = 530
Waktu_Tunggu = 15
Data_train = [(Janis_Kelamin, Masa_Studi, IPK, Jelaskan_Status_Anda_Saat_ini, Waktu_Tunggu)]
print(Janis_Kelamin, Masa_Studi, IPK, Jelaskan_Status_Anda_Saat_ini, Waktu_Tunggu)
2 532 3.5 530 15
```

 The output cell shows the predicted values: '2 532 3.5 530 15'.

Gambar 9. Prediksi Tracer Study

3.11. Hasil Prediksi Tracer Study

Hasil prediksi tracer study ini dengan menggunakan algoritma naïve bayes untuk klasifikasi masa tunggu alumni dalam mendapatkan pekerjaan berdasarkan output dapat menentukan apakah seorang alumni termasuk dalam kategori tercepat yaitu mendapatkan pekerjaan kurang dari 6 bulan yang terlambat mendapatkan pekerjaan lebih dari 6 bulan dengan evaluasi kinerja model seperti AUC, Precision, Recall, dan Akurasi.

The screenshot shows a Jupyter Notebook interface. The code cell contains the following Python code:

```
# membuat hasil prediksi tracer study
#hasil 0 -> tercepat
#hasil 1 -> terlambat
from matplotlib import pyplot as plt
df_of_IPK = df[['IPK']].plot(kind='line', figsize=(8, 4), title='IPK')
plt.show()
plt.rcParams['figure.figsize'] = (10, 5)
y_predict = 2
if y_predict == 1:
    print("hasil 1")
elif y_predict == 2:
    print("hasil 2")
else:
    print("Other hasil")
print("hasil prediksi tracer study --,hasil")
hasil 2
```

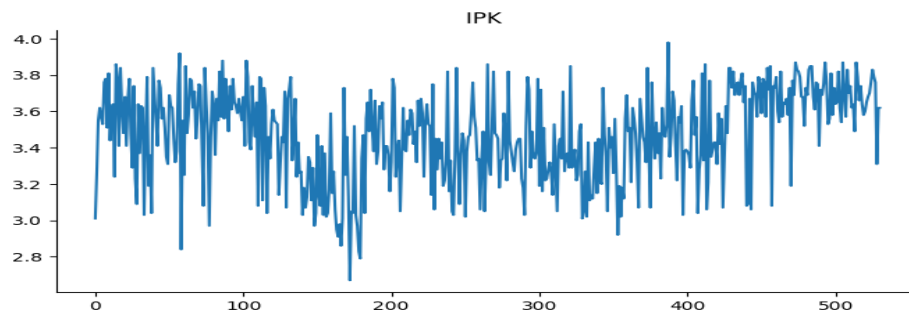
 The output cell shows a bar chart titled 'IPK' with a y-axis ranging from 3.0 to 4.0. The chart displays a series of blue bars representing IPK values for various data points.

Gambar 10. Hasil Prediksi Tracer Study

3.12. Grafik IPK

Pengumpulan Data dengan dikumpulkan data ipk maka dari alumni yang telah mendapatkan pekerjaan dengan data ini harus mencakup sebuah informasi tentang status pekerjaan alumni. Kategori IPK ke dalam beberapa kategori (IPK 3,50-4,00, (Sangat Baik) IPK 3,00-3,49 (Baik) IPK 2,50-

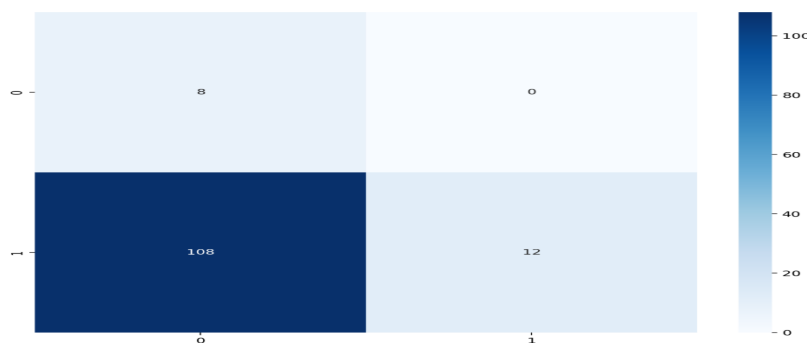
2,99 (Cukup) IPK 2,00-2,49 (Kurang). Data IPK dan status pekerjaan alumni disimpan dalam sebuah dataframe dengan kategorisasi IPK dikategorikan sesuai dengan rentang nilai.



Gambar 11. Grafik Klasifikasi IPK

3.13. Confusion Matrix Random Forest

Interpretasi dan evaluasi kinerja dari confusion matrix tersebut dapat dihitung dengan berbagai matrix evaluasi kinerja model. akurasi ini menunjukkan proporsi prediksi yang benar dari total prediksi yang dilakukan dalam kasus tracer study yaitu 94% dari precision menunjukkan proporsi prediksi yang positif benar dan dari total prediksi kasus ini sekitar 80%.



Gambar 12. Confusion matrix Random Forest

4. KESIMPULAN

Berdasarkan hasil evaluasi dari kedua algoritma yang terpilih menjadi prediksi dan mengelompokkan class atau klasifikasi setiap classnya. Naïve bayes dari hasil evaluasi model naïve bayes confusion matrix $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ dan $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ metrix evaluasi dari akurasi 94%, Presisi 0,1%, Recall 0,94% F1-Score 0.1%, ROC AUC 0.70% interpretasi Naïve bayes memiliki recall yang tinggi (100%) menunjukkan kemampuan yang baik dalam mengidentifikasi alumni yang mendapatkan pekerjaan. Presisi yang rendah (50%) mengindikasikan banyak prediksi yang positif dan yang salah sehingga model cenderung overpredicting. Akurasi dan F1-Score berada pada tingkatan menengah menunjukkan kinerja yang cukup baik tetapi masih banyak ruang yang untuk diperbaiki. Penangan ketidakseimbangan kelas dalam menggunakan Teknik seperti oversampling atau undersampling untuk menangani ketidakseimbangan kelas. Fitur ekstrasi dan seleksi mengeksplorasi fitur yang lebih relevan dan signifikan untuk meningkatkan presisi dan akurasi model. penggunaan variasi naïve bayes pertimbangkan varian lain dari naïve bayes seperti multinomial atau Bernoulli sesuai dengan karakteristik data.

DAFTAR PUSTAKA

- [1] P. Studi, S. Informasi, I. Teknologi, D. Bisnis, and S. Bali, "Penerapan Algoritma Support Vector Machine (Svm) Untuk Prediksi Waktu Tunggu Alumni Mendapatkan Pekerjaan I Made Budi Adnyana."

- [2] S. Sibagariang, A. Riyadi, A. Dzikri, F. Suandi, K. T. Sirait, and F. Setiawan, "Prediksi Prospek Kerja Alumni Dengan Algoritma Neural Network," 2021.
- [3] J. Homepage, S. Rohman Cholil, and V. Vydia, "IJCIT (Indonesian Journal on Computer and Information Technology) Prediksi Lama Masa Tunggu Alumni USM dalam Mendapatkan Pekerjaan dengan Algoritma KNN," 2024.
- [4] R. Rachmadiansyah, N. D. Rumlaklak, and A. Y. Mauko, "Prediksi Masa Tunggu Kerja Alumni Menggunakan Naïve Bayes Classifier Pada Program Studi Ilmu Komputer Universitas Nusa Cendana," *Jurnal Komputer dan Informatika*, vol. 10, no. 2, pp. 143-150, Sep. 2022, doi: 10.35508/jicon.v10i2.7426.
- [5] S. Inna, D. Mariana, and M. Manaqib, "Analisis Regresi Logistik Ordinal Faktor-Faktor Yang Mempengaruhi Waktu Tunggu Kerja Alumni."
- [6] I. Wayan Supriana, C. Pramatha, and L. Arida Ayu Rahning Putri, "RESISTOR Journal | 1 Aplikasi Pengukur Tingkat Kepuasan Alumni Berdasarkan Data Tracer Study Berbasis Metode Machine Learning", [Online]. Available: <https://s.id/jurnalresistor>
- [7] L. Andrian and A. W. Utami, "Rancang Bangun Sistem Informasi Prediksi Masa Tunggu Pekerjaan Alumni Dengan Pendekatan Algoritma Naïve Bayes," 2024.
- [8] F. Rahmayani, B. Nurina Sari, I. Maulana, R. Mayasari, and U. H. Singaperbangsa Karawang Jl Ronggo Waluyo, "Penerapan Algoritma C4.5 Dengan Feature Forward Selection Untuk Analisis Capaian Indikator Kinerja Utama Berdasarkan Tracer Study (Studi Kasus: FASILKOM UNSIKA)," 2023.
- [9] H. Noviyanto, B. Mukti, and U. Surakarta, "Prediksi Kesiapan Kerja Mahasiswa menggunakan Algoritma K-Means dan C4.5," 2022.