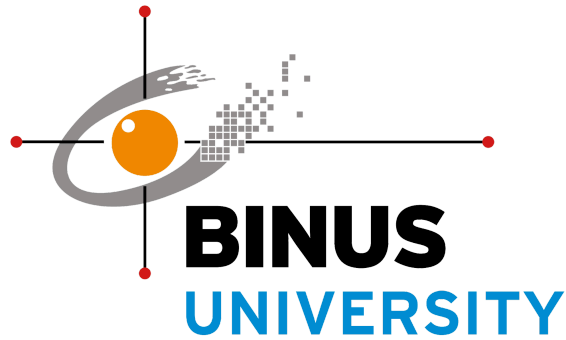


SSM - Laporan Kelompok 5



Topik :

Klasifikasi Prestasi Akademik Berdasarkan Tipe Kepribadian MBTI Mahasiswa Data Science B27

Nama Anggota :

Angela Valerie Christy - 2702244624

Lea Audrey - 2702261914

Naira Faizanoor - 2702241465

Vionita Lesia - 2702238312

1. Introduction

1.1 Background

Kelompok kami memilih topik “Klasifikasi Prestasi Akademik Berdasarkan Tipe Kepribadian MBTI Mahasiswa Data Science B27” karena melihat bahwa setiap mahasiswa memiliki karakter dan pendekatan belajar yang unik. MBTI (Myers-Briggs Type Indicator) adalah alat yang cukup dikenal dan sering digunakan untuk memahami kepribadian seseorang, termasuk di lingkungan BINUS, di mana mahasiswa baru diwajibkan mengikuti tes ini saat masa orientasi. MBTI mengelompokkan individu ke dalam 16 tipe kepribadian, namun demi keperluan analisis dan pemerataan data, kelompok kami menyederhanakannya menjadi dua kategori besar: introvert dan ekstrovert. Tujuan utama dari penelitian ini adalah untuk melihat apakah terdapat pola hubungan antara tipe kepribadian tersebut dengan tingkat prestasi akademik mahasiswa.

1.2 Problem

Di dalam satu kelas, gaya belajar setiap mahasiswa bisa sangat berbeda. Ada yang lebih nyaman bekerja sendiri, ada pula yang lebih aktif saat berdiskusi. Meskipun berbeda, setiap mahasiswa memiliki peluang yang sama untuk berprestasi. Oleh karena itu, kelompok kami ingin menjawab pertanyaan: apakah tipe kepribadian mahasiswa, khususnya berdasarkan MBTI, dapat digunakan untuk mengelompokkan tingkat prestasi akademik mereka. Dengan mempertimbangkan bahwa MBTI merupakan data yang sudah dimiliki mahasiswa sejak awal studi, kelompok kami merasa pendekatan ini relevan. Selain tipe kepribadian, kami juga mempertimbangkan variabel lain seperti gaya belajar, kebiasaan belajar, dan keaktifan mahasiswa.

1.3 Measurement

Kelompok kami mengumpulkan data melalui kuesioner yang mencakup tipe kepribadian MBTI, IPK semester terakhir, serta informasi pendukung seperti durasi belajar, frekuensi, keaktifan, dan strategi ujian. IPK dikategorikan menjadi tiga: tinggi (≥ 3.50), sedang ($3.00-3.49$), dan rendah (< 3.00). Setelah data terkumpul, dilakukan pembersihan terhadap nilai kosong, duplikasi, dan outlier. Data kategorikal diubah ke bentuk numerik menggunakan label encoding dan one-hot encoding. Dengan data yang sudah rapi dan terstruktur, proses dilanjutkan ke tahap eksplorasi dan pembangunan model klasifikasi untuk menjawab permasalahan penelitian.

2. Sampling Design

Survei ini ditujukan kepada 192 mahasiswa Program Studi Data Science angkatan B27 sebagai populasi target. Untuk menentukan ukuran sampel yang representatif, digunakan pendekatan dengan tingkat kepercayaan 95%, alpha (α) 5%, margin of error 10%, dan proporsi (\hat{p}) sebesar 0,5. Nilai proporsi 0,5 dipilih karena tidak ada

informasi awal tentang distribusi populasi, sehingga dianggap sebagai pendekatan konservatif. Berdasarkan nilai-nilai tersebut, diperoleh jumlah minimum responden yang dibutuhkan agar hasil survei valid dan dapat diandalkan, yaitu 65 responden.

- **Sample size (n)** = $\frac{Z(a/2)^2 \times \hat{p} \times (1-\hat{p})}{e^2} = \frac{(1.96)^2 \times 0.5 \times (1-0.5)}{(0.1)^2} = 96.04 \approx 97$ responden
- **Penyesuaian sample** = $\frac{n}{1 + (\frac{n-1}{N})} = \frac{97}{97 + (\frac{97-1}{192})} = 64.67 \approx 65$ responden

Dengan tingkat kepercayaan 95% dan margin of error 10%, jumlah sampel minimum yang dibutuhkan dari populasi 192 mahasiswa adalah 65 responden. Survei berhasil mengumpulkan 72 responden, melebihi jumlah minimum, sehingga hasilnya dianggap valid dan representatif. Pemilihan sampel dilakukan dengan simple random sampling, yaitu metode acak sederhana di mana setiap individu memiliki peluang yang sama untuk terpilih. Metode ini dipilih karena objektif dan adil, memastikan hasil survei mencerminkan kondisi populasi secara keseluruhan.

2.1 Life Cycle of A Survey From a Design Perspective

Penjelasan lebih lanjut mengenai Life Cycle of A Survey From a Design Perspective dapat diakses melalui link berikut:

 [Life Cycle of A Survey From a Design Perspective](#)

2.2 Life Cycle of A Survey From a Quality Perspective

Penjelasan lebih lanjut mengenai Life Cycle of A Survey From a Quality Perspective dapat diakses melalui link berikut:

 [Life Cycle of A Survey From a Quality Perspective](#)

3. Questionnaire Design

Pengumpulan data dalam penelitian ini dilakukan melalui penyebaran kuesioner. Kuesioner terdiri dari tiga bagian utama:

1. Pendataan Diri

Mengumpulkan informasi dasar seperti identitas, tipe kepribadian MBTI, dan IPK terakhir.

2. Manajemen Diri dan Strategi Belajar

Mengukur bagaimana mahasiswa mengatur proses belajar mereka, khususnya saat menghadapi ujian.

3. Interaksi dan Belajar Bersama

Mengeksplorasi preferensi dan pola interaksi dalam belajar kelompok, peran dalam diskusi, serta dampak MBTI terhadap dinamika sosial dalam proses pembelajaran.

Pertanyaan menggunakan skala Likert untuk memudahkan proses kuantifikasi dan pemodelan dengan algoritma klasifikasi. Beberapa pertanyaan lainnya bersifat kategorikal atau numerik, seperti data MBTI dan IPK. **Detail pertanyaan kuesioner dapat diakses melalui link berikut:** <https://forms.gle/AxpJKpskyYy3H8LN8>

4. Data Collection

4.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah melalui penyebaran kuesioner daring (online questionnaire). Instrumen ini dipilih karena praktis dan efisien dalam menjangkau responden yang tersebar, serta memungkinkan pengumpulan data dalam jumlah besar dalam waktu relatif singkat.

4.2 Proses Pengumpulan Data

Proses pengumpulan data dimulai dengan menyusun daftar mahasiswa Data Science B27 berdasarkan data kehadiran dan daftar nama kelas. Daftar ini digunakan sebagai *sampling frame*, yaitu himpunan individu yang memiliki peluang untuk menjadi sampel penelitian. Penyebaran kuesioner dilakukan melalui dua jalur utama:

1. **Grup WhatsApp mahasiswa B27**, yang mencakup sebagian besar mahasiswa aktif.
2. **Pesan pribadi (direct message)** kepada mahasiswa yang belum merespons pada tahap awal.

Penyebaran melalui grup bertujuan menjangkau responden secara umum, sedangkan pendekatan pribadi digunakan untuk meningkatkan tingkat respons dan memastikan jumlah sampel tercapai. Seluruh proses pengumpulan data berlangsung selama kurang lebih satu minggu.

4.3 Upaya Meningkatkan Tingkat Respons dan Dampaknya

Tingkat respons ditingkatkan melalui pengingat berkala, baik lewat pesan pribadi maupun pengumuman ulang di grup WhatsApp. Selain itu, disiapkan daftar cadangan responden sebagai antisipasi jika sampel belum tercukupi. Kombinasi penyebaran massal, komunikasi personal, dan *backup list* ini berhasil menjaga tingkat respons tetap tinggi dan representatif.

4.4 Jumlah Data Response

Populasi target dalam penelitian ini adalah seluruh mahasiswa Program Studi Data Science kelas B27. Dari 192 mahasiswa yang menjadi sasaran, sebanyak 72 orang memberikan respons, sementara 120 lainnya tercatat sebagai non-responden.

5. Measurement Quality

5.1 Validity

Item	r (korelasi)	p-value
MBTI	0.173	0.14499
Manajemen Waktu	0.535	0.00000
Kesulitan Fokus	0.364	0.00167
MBTI Bantu Belajar	0.564	0.00000
Jadwal Mingguan	0.526	0.00000
Strategi Terencana	0.625	0.00000
Belajar Jauh Hari	0.588	0.00001
Gaya Belajar MBTI	0.481	0.00002
Paham Kekuatan MBTI	0.601	0.00000
Peningkatan IPK	0.525	0.00000
Tahu Metode	0.478	0.00002
Termotivasi	0.429	0.00017
Pulihkan Motivasi	0.421	0.00023
Belajar Kelompok	0.424	0.00021
Bantu Pahami	0.380	0.00099
Aktif Diskusi	0.457	0.00006
Nyaman Sendiri	-0.033	0.78509
Peran Kelompok	0.492	0.00001
MBTI & Interaksi	0.510	0.00000
Dukungan Sosial	0.513	0.00000
Strategi Ujian Belajar jauh hari dengan jadwal...	0.331	0.00449
Strategi Ujian Belajar saat mendekati ujian	-0.051	0.66754
Strategi Ujian Sistem kebut semalam (SKS)	-0.060	0.61726
Strategi Ujian Tidak punya strategi khusus	-0.243	0.03956

Uji validitas dilakukan dengan korelasi Pearson untuk menilai apakah variabel survei relevan terhadap strategi belajar dan IPK. Variabel dianggap valid jika $p\text{-value} < 0.05$. Dari 24 variabel, 20 dinyatakan valid, sementara 4 tidak valid (MBTI, Nyaman Sendiri, serta dua strategi ujian: belajar mendekati ujian dan SKS) karena $p\text{-value} \geq 0.05$.

0.05 dan korelasi rendah atau negatif. Secara keseluruhan, sebagian besar variabel terbukti relevan dalam mengukur aspek yang diteliti.

5.2 Reliability

```
import pingouin as pg
numeric_cols = df.select_dtypes(include='number').columns.tolist()

item_cols = [col for col in numeric_cols if col.lower() != 'kategori_ipk']

# Hitung reliabilitas (Cronbach's Alpha)
alpha_value, _ = pg.cronbach_alpha(data=df[item_cols])
print(f"Cronbach's Alpha: {round(alpha_value, 3)}")

if alpha_value >= 0.5:
    print("Interpretasi: Reliabel")
else:
    print("Interpretasi: Tidak reliabel")

Cronbach's Alpha: 0.714
Interpretasi: Reliabel
```

Uji reliabilitas dilakukan untuk menilai konsistensi internal instrumen survei dalam mengukur konstruk yang sama. Metode yang digunakan adalah Cronbach's Alpha, dengan ambang batas minimum 0,5. Hasil uji menunjukkan nilai 0.714, yang menandakan bahwa survei memiliki

konsistensi internal yang baik dan variabel-variabelnya tergolong reliabel.

6. Preliminary Analysis

6.1 Exploratory Data Analysis

a. Jumlah Missing Value

df.isnull().sum()	
Timestamp	0
Nama	0
NIM	0
MBTI	0
IPK	0
Strategi Ujian	0
Manajemen Waktu	0
Kesulitan Fokus	0
MBTI Bantu Belajar	0
Jadwal Mingguan	0
Strategi Terencana	0
Belajar Jauh Hari	0
Gaya Belajar MBTI	0
Gaya Belajar MBTI	0
Paham Kekuatan MBTI	0
Peningkatan IPK	0
Tahu Metode	0
Termotivasi	0
Pulihkan Motivasi	0
Belajar Kelompok	0
Bantu Pahami	0
Aktif Diskusi	0
Nyaman Sendiri	0
Peran Kelompok	0
MBTI & Interaksi	0
Dukungan Sosial	0

Dalam proses pengumpulan data, kami memastikan kelengkapan data dalam dataset dengan merancang seluruh form isian sebagai form yang wajib diisi (*required*). Hal ini secara efektif mencegah adanya celah untuk nilai-nilai yang kosong atau hilang (*missing values*), sehingga hasil pemeriksaan awal terhadap dataset kami mengonfirmasi bahwa tidak terdapat satu pun *missing value* di dalamnya.

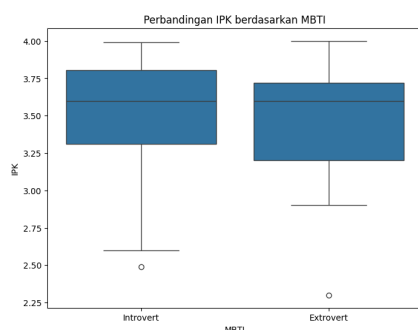
b. Dataset Information

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 72 entries, 0 to 71
Data columns (total 25 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   Timestamp             72 non-null    object
 1   Nama                  72 non-null    object
 2   NIM                   72 non-null    int64
 3   MBTI                  72 non-null    object
 4   IPK                   72 non-null    float64
 5   Strategi Ujian        72 non-null    object
 6   Manajemen Waktu       72 non-null    int64
 7   Kesulitan Fokus       72 non-null    int64
 8   MBTI Bantu Belajar    72 non-null    int64
 9   Jadwal Mingguan       72 non-null    int64
10   Strategi Terencana    72 non-null    int64
11   Belajar Jauh Hari     72 non-null    int64
12   Gaya Belajar MBTI     72 non-null    int64
13   Paham Kekuatan MBTI  72 non-null    int64
14   Peningkatan IPK       72 non-null    int64
15   Tahu Metode           72 non-null    int64
16   Termotivasi           72 non-null    int64
17   Pulihkan Motivasi     72 non-null    int64
18   Belajar Kelompok      72 non-null    int64
19   Bantu Pahami          72 non-null    int64
20   Aktif Diskusi         72 non-null    int64
21   Nyaman Sendiri        72 non-null    int64
22   Peran Kelompok        72 non-null    int64
23   MBTI & Interaksi      72 non-null    int64
24   Dukungan Sosial       72 non-null    int64
dtypes: float64(1), int64(20), object(4)
memory usage: 14.2+ KB
```

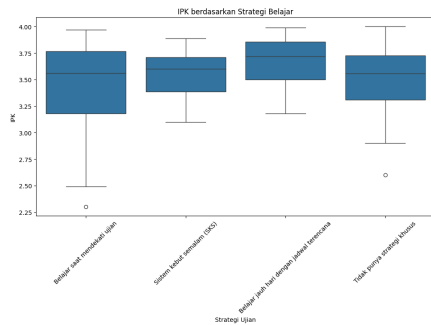
Dataset ini berisi 72 baris dan 25 kolom, mencakup informasi seperti Timestamp, Nama, NIM, MBTI, dan IPK. Mayoritas kolom berisi data numerik berupa skor dari respons survei, seperti "Manajemen Waktu", "Kesulitan Fokus", dan "Belajar Kelompok". Beberapa kolom lain menyoroti hubungan MBTI dengan kebiasaan belajar, seperti "MBTI Bantu Belajar" dan "Gaya Belajar MBTI". Seluruh kolom terisi lengkap tanpa missing value.

c. Boxplot



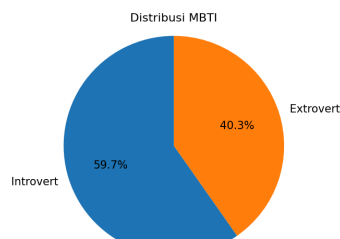
Boxplot menunjukkan bahwa baik introvert maupun extrovert memiliki rentang IPK serupa (2.25–4.00) dengan median hampir sama, menandakan tidak ada pengaruh

signifikan tipe kepribadian terhadap prestasi akademik. Meski introvert sedikit lebih rendah di batas bawah, keduanya mencapai IPK tertinggi 4.00. Bentuk boxplot yang mirip dan sebaran simetris memperkuat kesimpulan ini. Outlier ditemukan di kedua kelompok, menunjukkan bahwa MBTI, khususnya introvert dan ekstrovert bukan penentu utama capaian akademik mahasiswa.



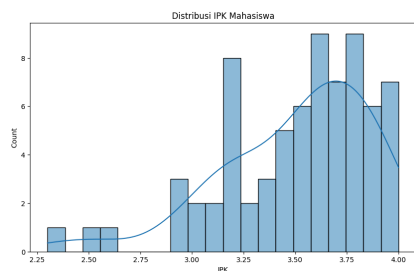
rentang nilai yang lebih lebar dan kemunculan outlier. Hal ini mengindikasikan bahwa pola belajar yang terburu-buru dapat meningkatkan risiko pencapaian akademik yang rendah.

d. Distribusi MBTI



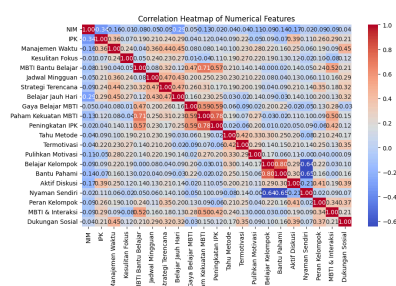
Berdasarkan hasil diagram di atas, data menunjukkan komposisi responden didominasi oleh kepribadian introvert (59.7%) dibandingkan extrovert (40.3%). Hal ini menunjukkan bahwa dalam kelompok mahasiswa Data Science B27 yang menjadi responden kami, mayoritasnya adalah tipe kepribadian introvert.

e. Distribusi IPK



Sebaran IPK mahasiswa menunjukkan mayoritas berada di rentang 3.00–3.50, mencerminkan performa akademik yang umumnya baik. Terdapat beberapa nilai ekstrem, yakni IPK tertinggi 4.00 dan terendah 2.25. Menariknya, tidak ada mahasiswa dengan IPK antara 2.50–2.75, menciptakan celah dalam distribusi. Pola ini mengindikasikan bahwa sebagian besar responden berhasil mempertahankan IPK di atas standar minimum, dengan sedikit penyimpangan pada nilai ekstrem.

f. Heat Map



Heatmap ini menunjukkan hubungan antar variabel dengan IPK menggunakan skala warna. Motivasi Akademik (0.62) dan Fokus Belajar

(0.58) memiliki korelasi positif kuat terhadap IPK, terlihat dari warna merah gelap—artinya, semakin tinggi keduanya, semakin tinggi IPK. Manajemen Waktu dan Disiplin juga menunjukkan korelasi positif moderat. Sebaliknya, variabel seperti MBTI (0.09) dan Dukungan Sosial memiliki korelasi sangat lemah, menandakan pengaruh minim terhadap IPK.

g. Cek Normalitas

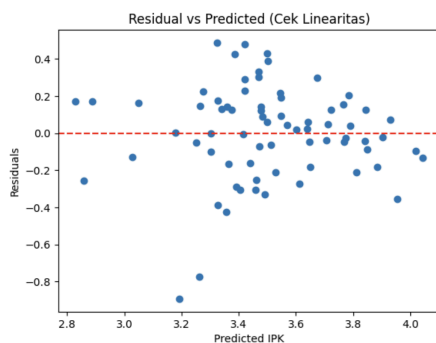
```
from scipy.stats import shapiro
residuals = model.resid
stat, p = shapiro(residuals)
print(f"Shapiro-Wilk Test p-value: {p}")

Shapiro-Wilk Test p-value: 0.01691700898257117
```

Hasil uji normalitas residual menggunakan Shapiro-Wilk menunjukkan p-value sebesar 0.0169, menunjukkan bahwa data belum sesuai

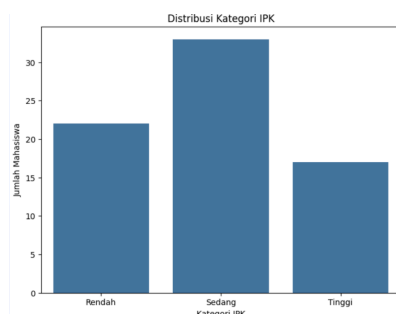
untuk dianalisis dengan model regresi linear karena tidak memenuhi syarat normalitas. Hal ini membuat hasil analisis, seperti pengaruh masing-masing faktor terhadap IPK, menjadi kurang akurat. Oleh karena itu, kami akan mencoba pendekatan lain seperti mengelompokkan IPK ke dalam kategori.

h. Q-Q Plot Linearitas



Plot menunjukkan adanya pola pada residual yang menandakan bahwa asumsi linearitas pada model regresi linear belum sepenuhnya terpenuhi. Ini menunjukkan kemungkinan hubungan antara variabel prediktor dengan IPK tidak sepenuhnya linear, sehingga model mungkin perlu ditinjau kembali atau pada kasus ini kami memutuskan untuk mengubah menjadi model klasifikasi.

i. Distribusi Kategori IPK



Grafik batang menunjukkan sebaran mahasiswa berdasarkan tiga kategori IPK: Rendah (< 3.4), Sedang ($3.4-3.8$), dan Tinggi (≥ 3.8). Mayoritas mahasiswa berada di kategori IPK Rendah, disusul oleh IPK Tinggi, sementara kategori IPK Sedang justru paling sedikit. Pola ini mengindikasikan distribusi IPK yang condong ke nilai rendah, namun tetap ada proporsi signifikan mahasiswa dengan IPK tinggi, dan

hanya sedikit yang berada di tengah-tengah rentang tersebut.

6.2 Data Preprocessing

a. Encoding Variabel Kategorikal

```
df_encode={"MBTI": {"Extrovert":1,"Introvert":0}}
df=df.replace(df_encode)
```


Label encoding digunakan untuk mengubah nilai kategori pada kolom MBTI menjadi angka (Introvert = 0, Extrovert = 1), agar data dapat dikenali oleh model yang hanya memproses data numerik.

b. One-Hot Encoding pada Variabel Kategorikal (Strategi Ujian)

```
from sklearn.preprocessing import OneHotEncoder
enc = OneHotEncoder()
df_prof_enc = pd.DataFrame(enc.fit_transform(df[['Strategi Ujian']]).toarray(), columns=enc.get_feature_names_out())
df = df.reset_index()
df = pd.concat([df, df_prof_enc], axis=1)

# drop kolom non fitur
df = df.drop(columns=['Timestamp', 'Nama', 'NIM', 'Strategi Ujian', 'index'])
```

Langkah ini dilakukan untuk mengubah data kategori pada kolom “Strategi Ujian” menjadi format numerik melalui teknik one-hot encoding. Hasilnya adalah beberapa kolom baru yang mewakili tiap jenis strategi ujian dalam bentuk nilai biner (0 atau 1). Proses ini dilakukan agar variabel “Strategi Ujian” dapat dikenali oleh model dengan mengubahnya ke format numerik.

7. Model

Berikut link colab mulai dari preliminary analysis hingga model:

[SSM_Kelompok 5.ipynb](#)

Penelitian ini bertujuan mengklasifikasikan tingkat prestasi akademik mahasiswa Program Studi Data Science kelas B27 berdasarkan tipe kepribadian MBTI dan variabel pendukung lainnya, menggunakan algoritma Random Forest Classification. Model dikembangkan untuk memetakan tiga kelas prestasi, 0 (Rendah), 1 (Sedang), 2 (Tinggi).

Akurasi Model: 0.6

Laporan Klasifikasi:

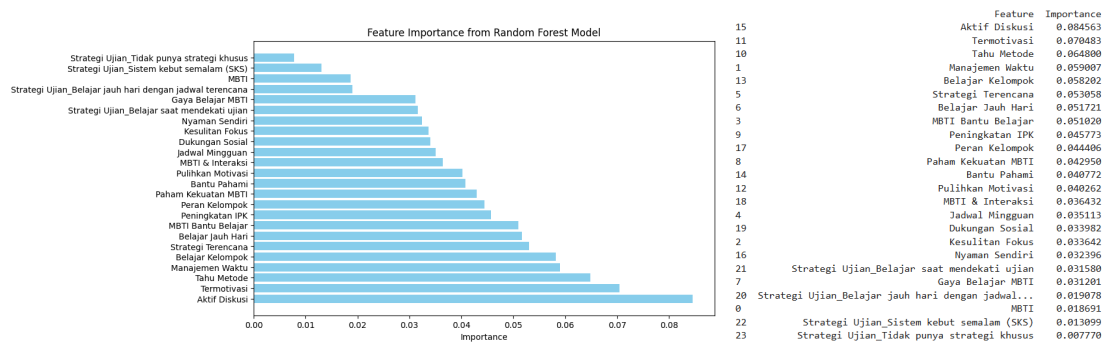
	precision	recall	f1-score	support
0	1.00	0.40	0.57	5
1	0.55	1.00	0.71	6
2	0.50	0.25	0.33	4
accuracy			0.60	15
macro avg	0.68	0.55	0.54	15
weighted avg	0.68	0.60	0.56	15

Model mencapai akurasi 60% dari 15 data uji. Nilai F1-score macro average sebesar 0.54 dan weighted average sebesar 0.56 menunjukkan performa moderat, namun belum seimbang antar kelas.

Analisis Per Kelas:

- **Kelas 0 (Rendah):** Precision tinggi (1.00), namun recall rendah (0.40), menunjukkan model hanya memprediksi saat sangat yakin namun sering gagal mendeteksi kasus lainnya.
- **Kelas 1 (Sedang):** Performa terbaik, dengan recall sempurna (1.00) dan F1-score 0.71. Model sangat akurat dalam mengenali kategori ini.
- **Kelas 2 (Tinggi):** Performa terendah, recall hanya 0.25 dan F1-score 0.33. Model masih kesulitan mengenali mahasiswa berprestasi tinggi.

Untuk memahami lebih jauh faktor-faktor yang paling berkontribusi terhadap prediksi model, dilakukan analisis terhadap nilai feature importance dari Random Forest.



Dari hasil analisis, lima fitur dengan kontribusi terbesar dalam pengambilan keputusan model adalah Aktif Diskusi (0.0846), Termotivasi (0.0704), Tahu Metode (0.0648), Manajemen Waktu (0.0590), dan Belajar Kelompok (0.0582). Ini menunjukkan bahwa perilaku belajar yang aktif, termotivasi, dan terorganisir berperan penting dalam membedakan kategori prestasi akademik. Sebaliknya, fitur seperti MBTI (0.0186) memiliki nilai importance rendah, yang berarti kontribusinya dalam klasifikasi prestasi akademik tergolong minimal.

8. Conclusion

8.1 Kesimpulan

Penelitian ini bertujuan untuk mengklasifikasikan tingkat prestasi akademik mahasiswa Data Science B27 berdasarkan tipe kepribadian MBTI, dengan menggunakan algoritma Random Forest Classification. Model yang dibangun menghasilkan akurasi sebesar 60%, dengan performa terbaik pada kategori prestasi akademik “Sedang”. Fitur yang paling berpengaruh dalam klasifikasi meliputi Aktif Diskusi, Termotivasi, Tahu Metode, Manajemen Waktu, dan Belajar Kelompok. Sementara itu, fitur MBTI memiliki pengaruh relatif rendah, meskipun tetap memberikan kontribusi dalam proses klasifikasi. Temuan ini mengindikasikan bahwa perilaku belajar dan kemampuan manajemen diri memiliki peran yang lebih signifikan terhadap prestasi akademik dibanding tipe kepribadian secara langsung.

Ke depannya, kelompok kami melihat peluang peningkatan hasil melalui penambahan jumlah responden guna mengurangi ketimpangan kelas, serta penerapan teknik normalisasi dan seleksi fitur yang lebih optimal. Selain itu, pengujian terhadap algoritma klasifikasi lain juga dapat menjadi alternatif untuk memperoleh hasil yang lebih akurat dan andal.

8.2 Analisa Lanjutan

Sebagai rencana lanjutan, akan dilakukan eksplorasi terhadap algoritma klasifikasi lain seperti XGBoost dan Support Vector Machine (SVM) untuk memperoleh hasil yang lebih stabil. Ke depannya, kami juga berencana untuk menggali lebih dalam pertanyaan-pertanyaan dalam kuesioner yang secara langsung berkaitan dengan pencapaian prestasi akademik. Hal ini bertujuan untuk memastikan bahwa setiap variabel yang digunakan benar-benar relevan dan memiliki potensi kontribusi yang signifikan terhadap model klasifikasi, serta

dapat memperkuat hubungan antara faktor-faktor tipe kepribadian dan hasil akademik mahasiswa.