

Analisis Komparatif Model Pembelajaran Mesin Untuk Klasifikasi Kesehatan dan Kehadiran Mahasiswa

Muhammad Hidayat

2025-02-28

Contents

Abstract	1
1 Pendahuluan	2
2 Eksplorasi dan Pengolahan Data	3
2.1 Memuat data	3
2.2 Transformasi data	5
2.3 Eksplorasi Data	6
2.4 Persiapan data pelatihan	10
3 Pembelajaran Mesin	13
3.1 Model Pembelajaran Mesin	13
3.2 Metrik Pengukuran Performa Model	14
4 Hasil dan Pembahasan	15
5 Kesimpulan	18
6 Daftar Pustaka	18
A Informasi Sesi R	20

Abstract

The health of students during classes is pivotal to their success in their day-to-day studies. This report aims to explore the possible factors that correlate to risk levels such as stress levels and attendance status to help optimize the data to be used in machine learning, as well as evaluate three machine learning models – Logistic Regression, Random Forest, and Support Vector Machines – under a complete dataset and a dataset optimized using only sufficiently correlated attributes in classifying health risk levels using the public Student Health and Attendance Data dataset. The models were assessed by their accuracy score, their duration during prediction, and their overall composite score to determine their effectiveness. The Random Forest model using the optimized dataset outperforms all other models with perfect accuracy and very fast calculation time, giving most credit to attribute selection in the optimized dataset. The study highlights the performance of Random Forest models of predicting health risk levels in this dataset as well as the benefits gained through careful selection of attributes in its models. Future research should focus on digging into the psychological aspect of stress levels and attendance status, seeking potential improvements across all models, and exploring the use of the Gradient Boosting Classifier model.

Keywords: Student Health Risk Classification, Student Health Risk Indicators, Exploratory Data Analysis, Machine Learning, Logistic Regression, Random Forest, Support Vector Machines

1 Pendahuluan

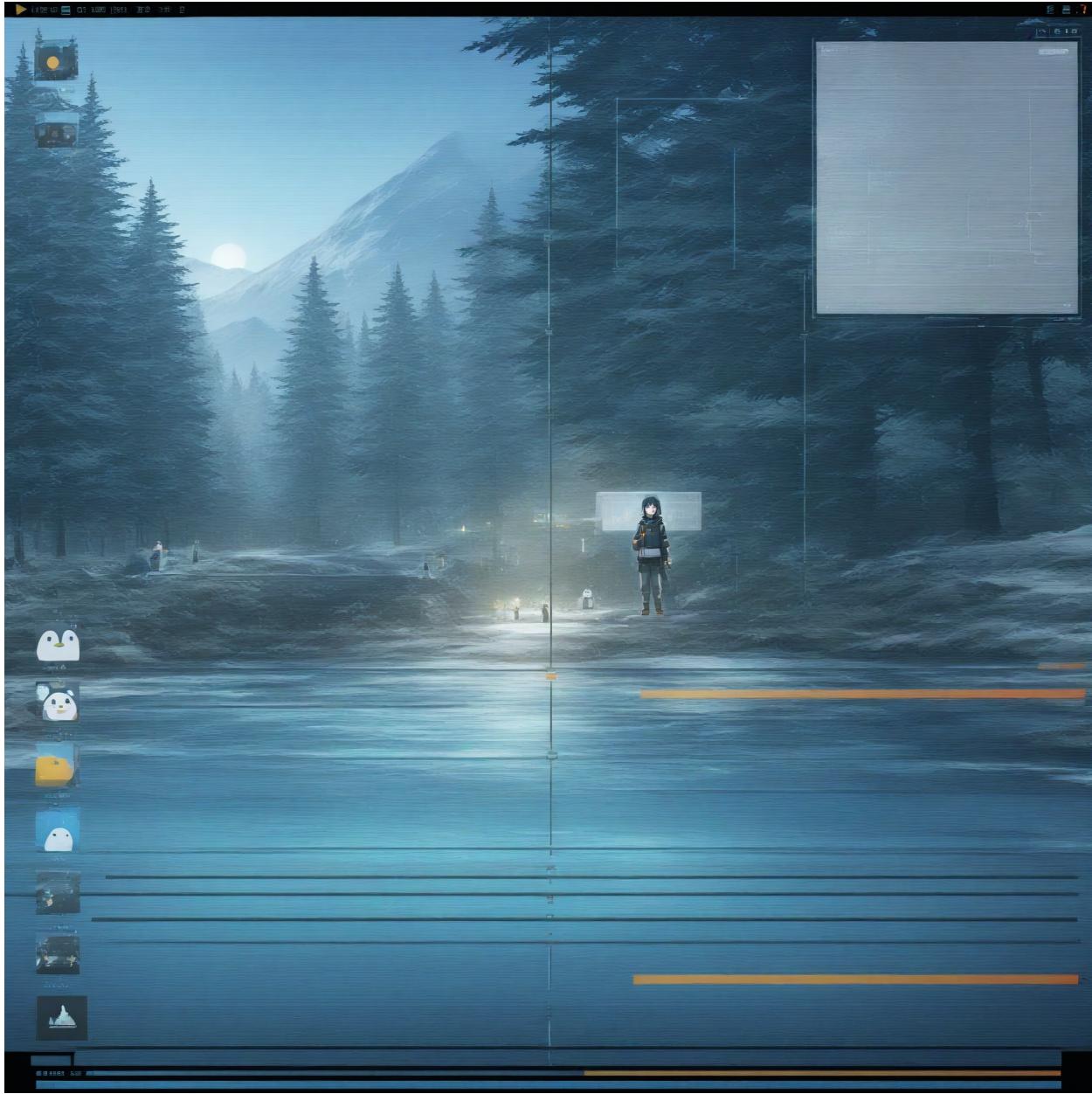


Figure 1: Ilustrasi berjudul “Random Forest Logistics on Support Vector” yang dibuat oleh AI menggunakan PixAI

Dalam perkuliahan, kesehatan mahasiswa harus dijaga untuk pembelajaran optimal. Menurut penelitian dari Pratama and Nur Zaimah (2024), melakukan aktivitas fisik dapat mengurangi tingkat kecemasan, depresi, serta meningkatkan suasana hati dan fungsi kognitif. Selain kesehatan fisik, kesehatan mental juga perlu diperhatikan dikarenakan kontribusinya terhadap perkembangan dan kesuksesan pola pikir mahasiswa yang bukan hanya berpengaruh secara individual, namun juga dalam kelompok.

Analisis terhadap data kesehatan mampu mendeteksi pola keseharian mahasiswa melalui berbagai faktor dan dapat menentukan tingkat risiko kesehatan mereka. Proses pengambilan keputusan berbasis data dapat menjadi acuan institusi pendidikan untuk meningkatkan kesejahteraan peserta didiknya.

Dengan perkembangan teknologi yang begitu pesat akhir-akhir ini, kecerdasan buatan telah melampaui batas yang semula sulit diraih. Dari membuat makalah lengkap yang terstruktur hingga gambar yang menyerupai lukisan, seperti contoh gambar pada Figure 1, kecerdasan buatan telah menjadi alat untuk membantu manusia mencapai tujuannya. Semua ini berasal dari data yang dipilih dan diolah, yang kemudian dijadikan bahan latihan untuk komputer agar mampu mengambil keputusan yang tepat. Inilah inti dari pembelajaran mesin, atau disebut *machine learning*.

Penggunaan teknik dan model pembelajaran mesin terhadap data kesehatan mahasiswa diharapkan dapat membantu mengidentifikasi tanda-tanda risiko pada masing-masing individu. Pembelajaran mesin terutama bermanfaat untuk mengenali pola pada data dengan banyak variabel dengan beragam karakteristik, baik linear, polinomial, eksponensial, maupun fluktuatif. Hasil pengenalan ini nantinya membentuk suatu model yang akan digunakan untuk memprediksi kasus-kasus baru yang belum pernah ditemui sebelumnya. Dengan informasi baru yang didapat, intervensi maupun pencegahan dapat dilakukan oleh pihak terkait.

Tujuan dari laporan ini adalah sebagai latihan penulis dalam mengeksplorasi data, memprosesnya sebagai bahan untuk pembelajaran mesin, mengoptimalkan data yang digunakan untuk pembelajaran mesin, dan menentukan algoritma dan parameter yang tepat untuk melatih mesin agar mampu memprediksi hasil dengan akurat dan efisien. Semua tahapan ini dijalankan menggunakan bahasa pemrograman R.

Laporan ini terbagi menjadi beberapa bagian, selain dari bagian ini. Bagian 2 akan membahas tentang data yang digunakan, pengolahan data untuk eksplorasi data, dan pengolahan data untuk pembelajaran mesin. Bagian 3 akan membahas tentang model pembelajaran mesin yang digunakan, parameter yang ditetapkan, serta langkah evaluasi model. Bagian 4 akan membahas tentang performa model dan menetapkan model terbaik. Bagian 5 berisikan kesimpulan dari hasil temuan, keterbatasan, serta saran untuk penelitian ke depannya.

2 Eksplorasi dan Pengolahan Data

Dalam laporan ini, penulis mengacu pada dataset kesehatan dan kehadiran mahasiswa yang disediakan oleh Ziya (2024).

2.1 Memuat data

Dataset disediakan dalam bentuk file CSV, dan proses membaca file dapat dilakukan dengan mudah. Dalam dataset tersebut, terdapat dua file CSV yang identik. Berikut perintah bash untuk mengecek kesamaan file.

```
cd "Student Health and Attendance Data/"  
shasum *  
  
## 4ab8799345531cc7978909b88ae6bb6c5209e2fc student_monnitoring_data.csv  
## 4ab8799345531cc7978909b88ae6bb6c5209e2fc student_tracking_data.csv  
  
Kedua file sama, jadi cukup satu yang digunakan, yaitu student_monnitoring_data.csv.  
data <- read.csv("Student Health and Attendance Data/student_monnitoring_data.csv")
```

Lihat Tabel 1 untuk 10 data pertama dan Tabel 2 untuk 10 data terakhir dari dataset. Hanya kolom tertentu yang ditampilkan karena keterbatasan ruang.

Tabel 3 dan 4 memberikan gambaran bentuk data apa adanya dari file CSV.

Langkah selanjutnya adalah memastikan tidak ada baris kosong dalam dataset. Banyak baris kosong dapat diketahui dengan perintah berikut:

Table 1: Data awal (10 pertama)

Student.ID	Date	Class.Time	Attendance.Status	Risk.Level
1	2024-12-01	9:00-15:00	Late	Low
1	2024-12-02	8:00-16:00	Late	Medium
1	2024-12-03	11:00-14:00	Late	High
1	2024-12-04	11:00-16:00	Late	Low
1	2024-12-05	9:00-13:00	Absent	High
1	2024-12-06	8:00-14:00	Present	High
1	2024-12-07	11:00-15:00	Absent	High
1	2024-12-08	8:00-15:00	Absent	High
1	2024-12-09	11:00-13:00	Absent	High
1	2024-12-10	9:00-16:00	Late	Low

Table 2: Data awal (10 terakhir)

	Student.ID	Date	Class.Time	Attendance.Status	Risk.Level
14991	500	2024-12-21	8:00-16:00	Present	High
14992	500	2024-12-22	9:00-12:00	Present	High
14993	500	2024-12-23	8:00-12:00	Absent	High
14994	500	2024-12-24	10:00-12:00	Present	Medium
14995	500	2024-12-25	11:00-12:00	Absent	High
14996	500	2024-12-26	9:00-16:00	Late	Low
14997	500	2024-12-27	9:00-15:00	Absent	High
14998	500	2024-12-28	11:00-14:00	Absent	High
14999	500	2024-12-29	11:00-14:00	Late	Medium
15000	500	2024-12-30	9:00-16:00	Present	High

Table 3: Ringkasan data awal (1)

Student.ID	Stress.Level..GSR.	Sleep.Hours	Anxiety.Level	Mood.Score
Min. : 1.0	Min. :0.500	Min. :5.000	Min. : 1.000	Min. : 1.000
1st Qu.:125.8	1st Qu.:1.640	1st Qu.:6.000	1st Qu.: 3.000	1st Qu.: 3.000
Median :250.5	Median :2.760	Median :7.000	Median : 6.000	Median : 5.000
Mean :250.5	Mean :2.763	Mean :6.997	Mean : 5.547	Mean : 5.472
3rd Qu.:375.2	3rd Qu.:3.900	3rd Qu.:8.000	3rd Qu.: 8.000	3rd Qu.: 8.000
Max. :500.0	Max. :5.000	Max. :9.000	Max. :10.000	Max. :10.000

Table 4: Ringkasan data awal (2)

	Date	Class.Time	Attendance.Status	Risk.Level
	Length:15000	Length:15000	Length:15000	Length:15000
	Class :character	Class :character	Class :character	Class :character
	Mode :character	Mode :character	Mode :character	Mode :character

```

sum(is.na(data))

## [1] 0

```

Tidak ada data kosong dalam dataset, sehingga tidak perlu dibersihkan.

2.2 Transformasi data

Untuk mempermudah eksplorasi dan pengolahan data lebih lanjut, data di atas diproses dengan mengubah tipe data:

- Date (Date)
- Class.Time → Start.Time (POSIXct), End.Time (POSIXct) “POSIX C Time”. Kolom Class.Time akan dihapus karena tidak akan digunakan lagi kedepannya.
- Attendance.Status → (Factor), dengan ketentuan nilai AttendanceStatus → {1(Present), 2(Late), 3(Absent)}.
- Risk.Level → (Factor), dengan ketentuan nilai RiskLevel → {1(Low), 2(Medium), 3(High)}

Selain itu, terdapat data yang ditambahkan:

- Class.Duration (Numeric), jeda waktu antara Start.Time dan End.Time dalam satuan jam. Operasi pengurangan dengan data tersebut sebenarnya memanggil fungsi “difftime” dan secara bawaan menghasilkan keluaran dalam satuan jam.
- Start.Hour dan End.Hour (Numeric), waktu mulai dan selesai kelas mahasiswa dalam satuan jam.

```

new_data <- data |>
  mutate(
    Student.ID = Student.ID |> as.integer(),
    Date = Date |> as.Date(format = "%Y-%m-%d"),
    Start.Time = Class.Time |> as.character() |> strsplit("-") |> sapply(\(x) x[1]),
    Start.Time = paste(Date, Start.Time) |> as.POSIXct(format = "%Y-%m-%d %H:%M"),
    Start.Hour = Start.Time |> format("%H") |> as.numeric(),
    End.Time = Class.Time |> as.character() |> strsplit("-") |> sapply(\(x) x[2]),
    End.Time = paste(Date, End.Time) |> as.POSIXct(format = "%Y-%m-%d %H:%M"),
    End.Hour = End.Time |> format("%H") |> as.numeric(),
    Class.Duration = (End.Time - Start.Time) |> as.numeric(),
    Stress.Level..GSR. = Stress.Level..GSR. |> as.numeric(),
    Sleep.Hours = Sleep.Hours |> as.numeric(),
    Anxiety.Level = Anxiety.Level |> as.numeric(),
    Mood.Score = Mood.Score |> as.numeric(),
    Attendance.Status = Attendance.Status |>
      as.factor() |>
      recode_factor(
        "1" = "Present",
        "2" = "Late",
        "3" = "Absent"
      ),
    Risk.Level = Risk.Level |>
      as.factor() |>
      recode_factor(
        "1" = "Low",
        "2" = "Medium",
        "3" = "High"
      ),
    .keep = "none"
  )

```

Berikut penjelasan dari perubahan di atas:

Table 5: Data yang diubah (10 pertama)

Student.ID	Start.Time	End.Hour	Class.Duration	Attendance.Status	Risk.Level
1	2024-12-01 09:00:00	15	6	Late	Low
1	2024-12-02 08:00:00	16	8	Late	Medium
1	2024-12-03 11:00:00	14	3	Late	High
1	2024-12-04 11:00:00	16	5	Late	Low
1	2024-12-05 09:00:00	13	4	Absent	High
1	2024-12-06 08:00:00	14	6	Present	High
1	2024-12-07 11:00:00	15	4	Absent	High
1	2024-12-08 08:00:00	15	7	Absent	High
1	2024-12-09 11:00:00	13	2	Absent	High
1	2024-12-10 09:00:00	16	7	Late	Low

Table 6: Data yang diubah (10 terakhir)

	Student.ID	Start.Time	End.Hour	Class.Duration	Attendance.Status	Risk.Level
14991	500	2024-12-21 08:00:00	16	8	Present	High
14992	500	2024-12-22 09:00:00	12	3	Present	High
14993	500	2024-12-23 08:00:00	12	4	Absent	High
14994	500	2024-12-24 10:00:00	12	2	Present	Medium
14995	500	2024-12-25 11:00:00	12	1	Absent	High
14996	500	2024-12-26 09:00:00	16	7	Late	Low
14997	500	2024-12-27 09:00:00	15	6	Absent	High
14998	500	2024-12-28 11:00:00	14	3	Absent	High
14999	500	2024-12-29 11:00:00	14	3	Late	Medium
15000	500	2024-12-30 09:00:00	16	7	Present	High

- “Attendance Status” dan “Risk Level” diubah menjadi tipe Factor karena merupakan data kualitatif/kategorik bertingkat. Tingkatan diatur sebagaimana di atas.
- “Date” diubah menjadi tipe Date karena merupakan data tanggal.
- “Class Time” dipecah menjadi dua, yaitu “Start Time” dan “End Time”. Data Character dari “Class Time” mula-mula dipecah menjadi dua pada tanda “-” menggunakan fungsi “strsplit()” sehingga “9:00-15:00” berubah menjadi c(“9:00”, “15:00”). Masing-masing kemudian digabung dengan Character dari Date sehingga membentuk format “2024-12-01 9:00” yang nantinya dikonversi menjadi tipe waktu POSIXct.
- “Start Hour” dan “End Hour” mengambil jam dari “Start Time” dan “End Time” berturut-turut.
- Data yang lain diubah ke tipe Numeric karena sudah berupa bilangan.
- Proses transformasi data menggunakan sintaks baru untuk piping, yaitu |>, yang dikenalkan sejak R versi 4.1.

Berikut Tabel 5 untuk 10 data pertama dan Tabel 6 untuk 10 data terakhir dari dataset yang diubah. Sekali lagi, hanya kolom tertentu yang ditampilkan karena keterbatasan ruang.

Tabel 7, 8, dan 9 memberikan gambaran bentuk data setelah diolah.

2.3 Eksplorasi Data

Sebelum menerapkan metodologi berikutnya, ada baiknya suatu data dipahami terlebih dahulu melalui proses eksplorasi data. Eksplorasi data bertujuan untuk memahami data dengan lebih baik sebelum diproses dalam tahapan analisis lebih lanjut (Ramdani & Utami, 2024). Dalam sains data, tahapan ini juga dikenal sebagai *Exploratory Data Analysis* (EDA), yaitu seperangkat alat yang digunakan untuk memahami sifat, struktur, dan distribusi data, serta melihat hubungan antara beberapa atribut dalam dataset.

Table 7: Ringkasan data yang diubah (1)

Student.ID	Date	Start.Time	End.Time
Min. : 1.0	Min. :2024-12-01	Min. :2024-12-01 08:00:00	Min. :2024-12-01 12:00:00.00
1st Qu.:125.8	1st Qu.:2024-12-08	1st Qu.:2024-12-08 10:00:00	1st Qu.:2024-12-08 14:00:00.00
Median :250.5	Median :2024-12-15	Median :2024-12-15 21:30:00	Median :2024-12-16 02:00:00.00
Mean :250.5	Mean :2024-12-15	Mean :2024-12-15 21:30:06	Mean :2024-12-16 01:59:48.24
3rd Qu.:375.2	3rd Qu.:2024-12-23	3rd Qu.:2024-12-23 09:00:00	3rd Qu.:2024-12-23 14:00:00.00
Max. :500.0	Max. :2024-12-30	Max. :2024-12-30 11:00:00	Max. :2024-12-30 16:00:00.00

Table 8: Ringkasan data yang diubah (2)

Start.Hour	End.Hour	Class.Duration	Sleep.Hours	Stress.Level..GSR.
Min. : 8.000	Min. :12	Min. :1.000	Min. :5.000	Min. :0.500
1st Qu.: 8.000	1st Qu.:13	1st Qu.:3.000	1st Qu.:6.000	1st Qu.:1.640
Median : 9.000	Median :14	Median :5.000	Median :7.000	Median :2.760
Mean : 9.502	Mean :14	Mean :4.495	Mean :6.997	Mean :2.763
3rd Qu.:11.000	3rd Qu.:15	3rd Qu.:6.000	3rd Qu.:8.000	3rd Qu.:3.900
Max. :11.000	Max. :16	Max. :8.000	Max. :9.000	Max. :5.000

Table 9: Ringkasan data yang diubah (3)

Anxiety.Level	Mood.Score	Attendance.Status	Risk.Level
Min. : 1.000	Min. : 1.000	Present:4966	Low :3271
1st Qu.: 3.000	1st Qu.: 3.000	Late :5032	Medium:3349
Median : 6.000	Median : 5.000	Absent :5002	High :8380
Mean : 5.547	Mean : 5.472		
3rd Qu.: 8.000	3rd Qu.: 8.000		
Max. :10.000	Max. :10.000		

Table 10: Kehadiran per hari (5 hari pertama)

Date	Present	Late	Absent	Total
2024-12-01	164	177	159	500
2024-12-02	170	170	160	500
2024-12-03	138	192	170	500
2024-12-04	163	175	162	500
2024-12-05	167	168	165	500

Table 11: Ringkasan statistik kehadiran mahasiswa

Attendance.Status	mean	median	stdev	min	range	max	skew	kurt
Absent	166.73	167.0	9.79	149	37	186	0.17	-0.61
Late	167.73	167.0	11.76	144	48	192	-0.16	-0.39
Present	165.53	166.5	10.64	138	51	189	-0.23	0.00

Terdapat beberapa hal yang ingin penulis ketahui dalam data ini, di antaranya:

- Berapa banyak mahasiswa yang hadir, terlambat, maupun tidak hadir dalam satu hari perkuliahan?
- Bagaimana hubungan antara kehadiran mahasiswa dengan tingkat risiko kesehatan?
- Bagaimana persebaran dan korelasi dari atribut numerik terhadap tingkat risiko kesehatan?

2.3.1 Banyak mahasiswa per hari berdasarkan kehadiran

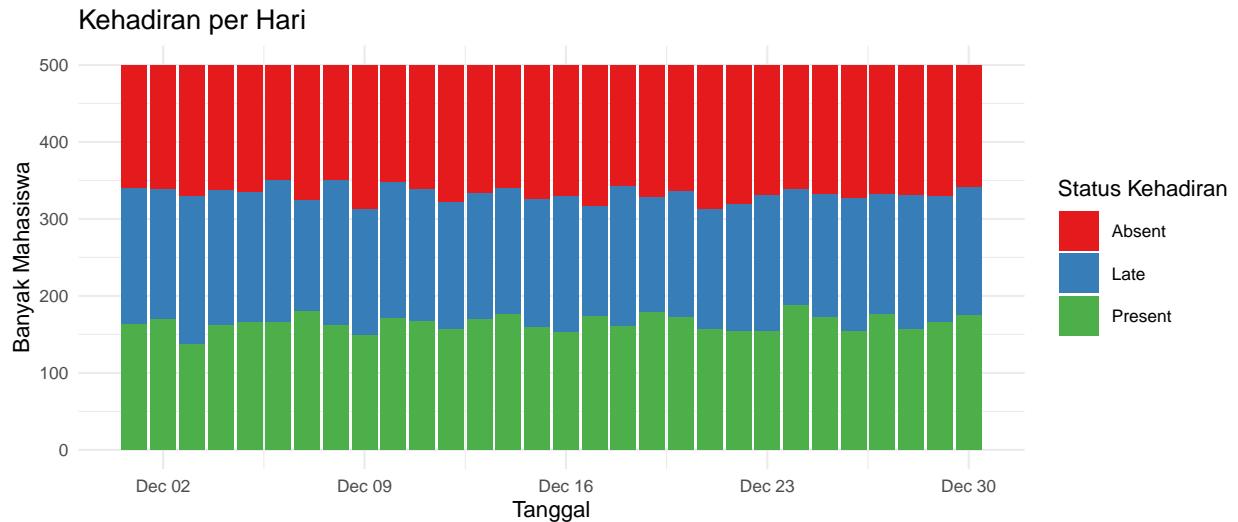


Figure 2: Grafik Kehadiran Per Hari

Tingkat kehadiran, keterlambatan, dan absen relatif imbang hari ke hari. Banyak mahasiswa secara keseluruhan per hari pun konsisten. Ini berarti tidak ada banyak hubungan antara status kehadiran dan risiko kesehatan dengan pergerakan hari ke hari. Selengkapnya, lihat Grafik 2 dan Tabel 11.

2.3.2 Tingkat kehadiran mahasiswa berdasarkan tingkat risiko kesehatan

Grafik 3 menunjukkan tingkat risiko kesehatan tinggi jika mahasiswa tidak hadir pada kuliah hari itu. Ini berarti terdapat hubungan erat antara status kehadiran dan risiko kesehatan.

Table 12: Tabel kehadiran berdasarkan tingkat risiko

Attendance.Status	Low	Medium	High
Present	1585	1722	1659
Late	1686	1627	1719
Absent	0	0	5002

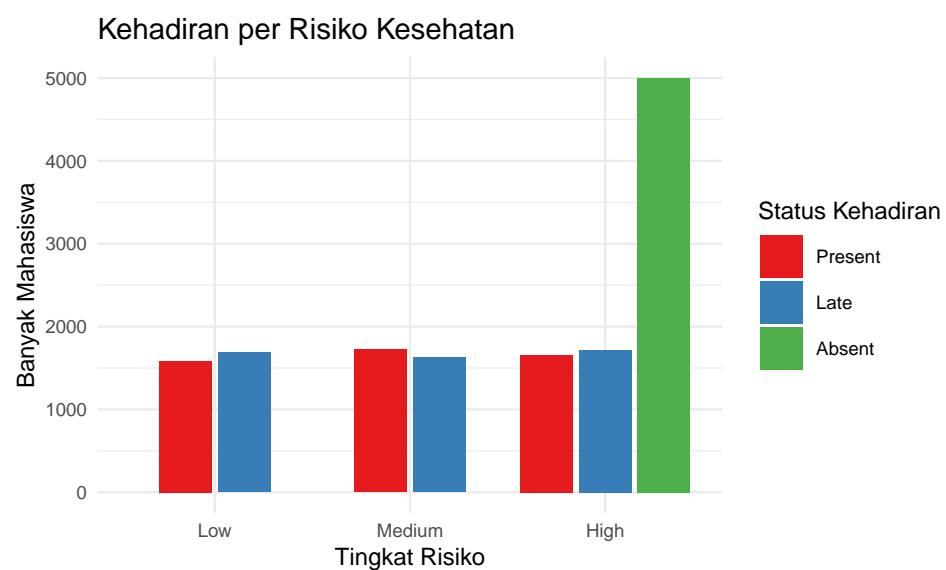


Figure 3: Grafik per Risiko Kesehatan

2.3.3 Faktor-faktor yang mempengaruhi tingkat risiko kesehatan

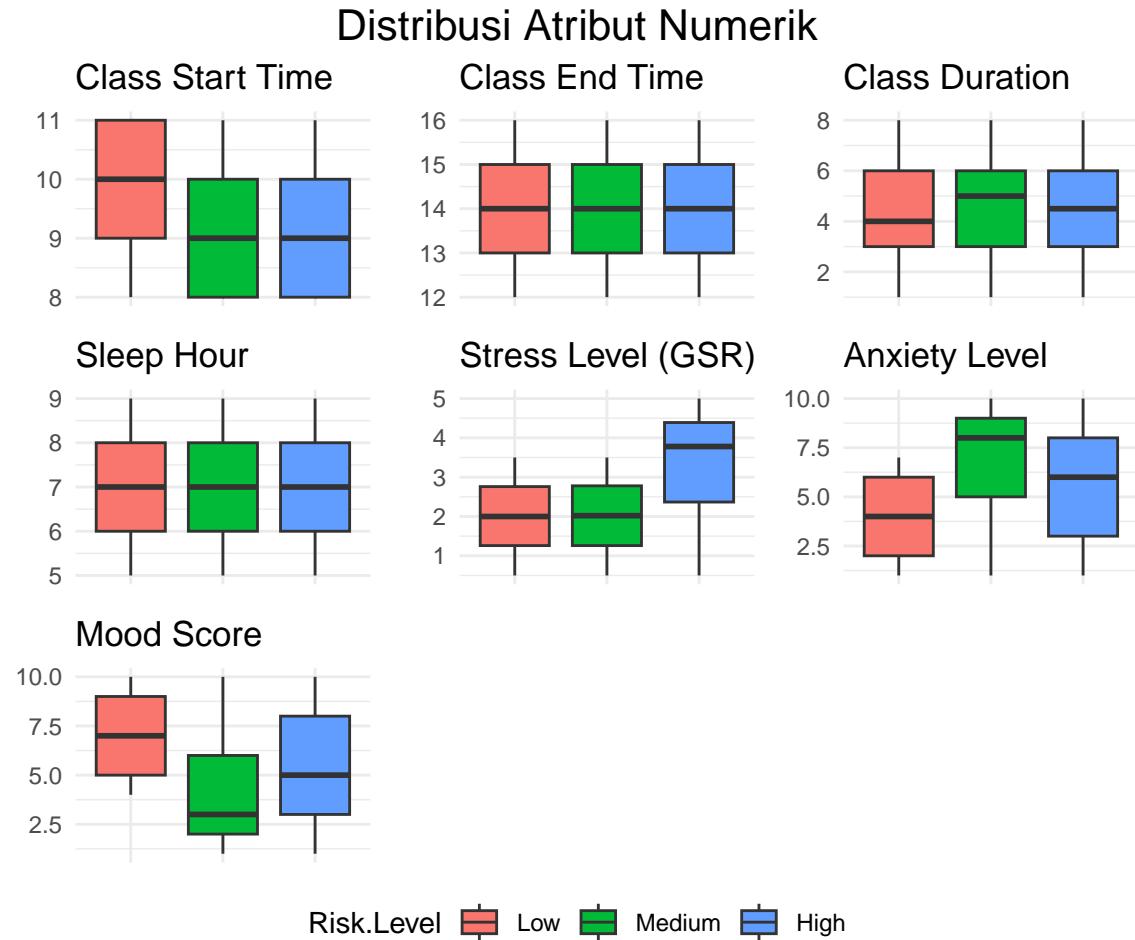


Figure 4: Grafik Distribusi Atribut Numerik

Dari distribusi nilai pada Grafik 4, terlihat bahwa waktu mulai kuliah dan tingkat stres memiliki hubungan terhadap tingkat risiko secara linear. Sementara itu, tingkat suasana hati dan kekhawatiran menunjukkan hubungan non-linear, dimana tingkat risiko menengah membentuk titik ekstrem di antara tingkat risiko yang lain. Untuk atribut yang lain, mereka menunjukkan hubungan konstan, atau hampir tidak ada hubungan dengan tingkat risiko.

Hubungan non-linear pada tingkat suasana hati dan kekhawatiran perlu dikaji lebih lanjut. Topik seperti ini lebih mengarah ke ranah psikologi, yang berada di luar cakupan laporan ini.

Grafik 5 menunjukkan bahwa wktus kehadiran dan tingkat stres mahasiswa berperan besar terhadap tingginya risiko kesehatan mahasiswa. Sementara itu, tingkat kekhawatiran (*anxiety*) dan suasana hati (*mood*) juga berpengaruh, meskipun tidak sebesar sebelumnya. Atribut-atribut ini akan menentukan langkah yang diambil dalam pembelajaran mesin.

2.4 Persiapan data pelatihan

Untuk hasil yang konsisten dalam percobaan ini, digunakan seed tertentu untuk melakukan operasi. Ini memastikan tiap operasi random dalam proses ini tetap sama tiap kali dijalankan dari awal.

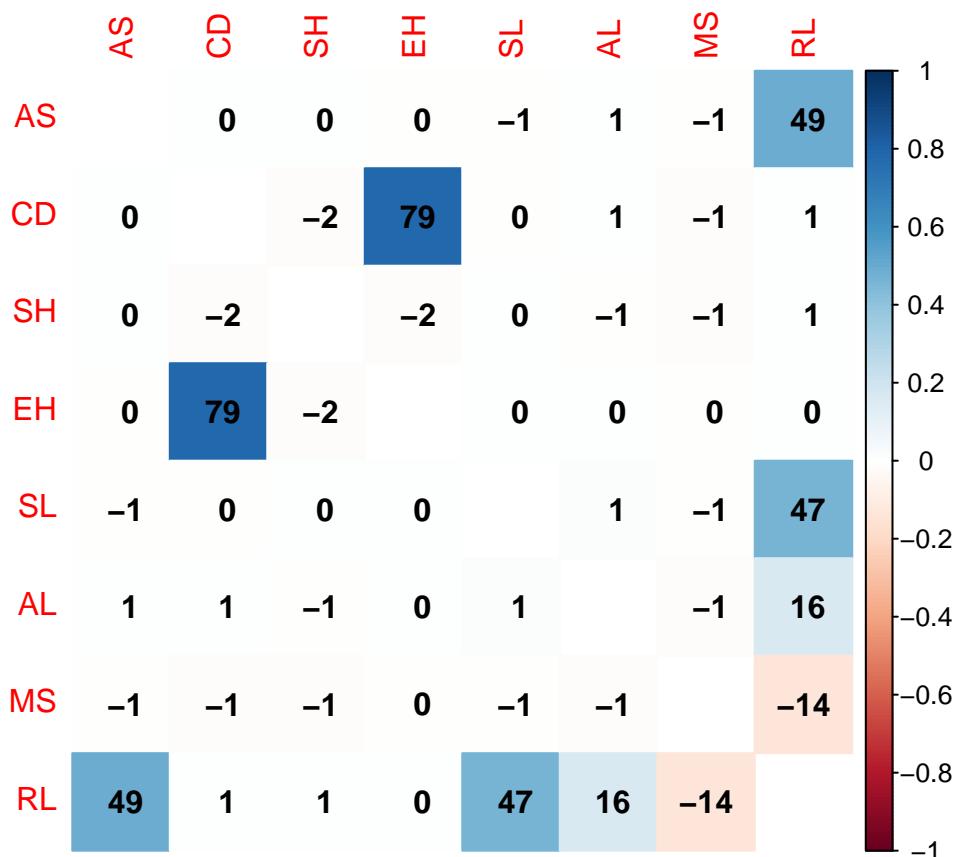


Figure 5: Grafik matriks korelasi antara variabel numerik

Table 13: Tabel Korelasi Antar Variabel

	AS	CD	SH	EH	SL	AL	MS	RL
AS	1.00	0.00	0.00	0.00	-0.01	0.01	-0.01	0.49
CD	0.00	1.00	-0.02	0.79	0.00	0.01	-0.01	0.01
SH	0.00	-0.02	1.00	-0.02	0.00	-0.01	-0.01	0.01
EH	0.00	0.79	-0.02	1.00	0.00	0.00	0.00	0.00
SL	-0.01	0.00	0.00	0.00	1.00	0.01	-0.01	0.47
AL	0.01	0.01	-0.01	0.00	0.01	1.00	-0.01	0.16
MS	-0.01	-0.01	-0.01	0.00	-0.01	-0.01	1.00	-0.14
RL	0.49	0.01	0.01	0.00	0.47	0.16	-0.14	1.00

```
# 1740812400 adalah 28 Februari 2025, 23:59:59 WIB
set.seed(1740812399)
```

Untuk memudahkan algoritma pembelajaran mesin dalam menginterpretasi nilai, data yang ada harus dinormalisasikan dan diubah menjadi bentuk angka. Data diubah menjadi bentuk angka menggunakan fungsi `mutate_all` dari library `dplyr` (Wickham et al., 2023). Kemudian, data dimasukkan ke fungsi normalisasi `preProcess` yang disediakan oleh library ML `caret` (Kuhn, 2024). Normalisasi dilakukan sehingga semua data berada pada rentang $[-1, 1]$ (parameter `rangeBounds`), dengan rata-rata di sekitar 0 (metode “center”), dan deviasi standar di sekitar 1 (metode “scale”).

Karena tingkat kesehatan merupakan variabel yang akan diuji, tipe datanya dibiarkan dalam bentuk *factor*. Jika tidak, fungsi-fungsi di bawah tidak akan bisa berjalan.

```
new_data_just_numbers <- new_data |>
  mutate_all(as.numeric)
new_data_just_numbers$Risk.Level <- new_data$Risk.Level

normalized_data_values <- new_data_just_numbers |>
  preProcess(method = c("center", "scale"), rangeBounds = c(-1, 1))

normalized_data <- predict(normalized_data_values, new_data_just_numbers)
```

Terakhir, untuk menguji efisiensi pelatihan dan pengujian, dibentuk dua buah data:

1. Data lengkap: Data yang mencakup semua variabel.
2. Data pilihan: Data yang hanya melibatkan kolom-kolom yang berpengaruh terhadap tingkat risiko kesehatan, dengan nilai korelasi terhadap tingkat risiko di atas 0.1, yaitu status kehadiran, tingkat stres, tingkat kekhawatiran, dan suasana hati.

```
numeric_data <- normalized_data

selected_data <- normalized_data |>
  select(
    Risk.Level,
    Attendance.Status,
    Stress.Level..GSR.,
    Anxiety.Level,
    Mood.Score
  )
```

Pembagian data ke dalam dua kelompok. Pembagian menggunakan partisi yang sama untuk data lengkap dan data terpilih (Kuhn, 2024).

```

train_index <- createDataPartition(new_data$Risk.Level, p = 0.8, list = FALSE)
numeric_train_data <- numeric_data[train_index, ]
numeric_test_data <- numeric_data[-train_index, ]
selected_train_data <- selected_data[train_index, ]
selected_test_data <- selected_data[-train_index, ]

```

3 Pembelajaran Mesin

Metodologi penelitian yang digunakan di sini adalah eksperimen, yaitu telaah empirik sistematis yang meminimumkan varian dari semua variabel yang berpengaruh terhadap masalah yang diteliti dengan melakukan manipulasi, kontrol, dan operasi secara cermat dan teliti (Hikmawati, 2020).

Dalam penelitian ini:

- Variabel bebasnya adalah model yang digunakan serta pemilihan data untuk model.
- Variabel terikatnya adalah performa model dalam pengujian terhadap data yang berbeda.
- Kontrol dilakukan dengan menjalankan pembagian data, proses pelatihan, dan evaluasi dengan menggunakan *seed* yang telah ditetapkan. Data yang digunakan untuk pelatihan dan pengujian disamakan untuk semua model.

3.1 Model Pembelajaran Mesin

Untuk permasalahan klasifikasi seperti ini, tiga model akan diuji, yaitu Multinomial Logistic Regression, Random Forest, dan Support Vector Machines. Masing-masing akan dilatih pada data lengkap dan data pilihan. Logistic Regression (LR) merupakan model yang berlandaskan pada persamaan linear antar variabel dengan memperkirakan koefisien-koefisien pada persamaan linear. Awalnya dikembangkan untuk klasifikasi kategori biner (dua nilai), namun dapat diterapkan untuk kategori dengan tiga nilai atau lebih menggunakan teknik lanjutan seperti Multinomial Logistic Regression (MLR) (Paliling et al., 2023). Detail tentang koefisien yang dihasilkan tidak akan dijelaskan dalam laporan ini. MLR diimplementasikan dalam library **nnet** menggunakan fungsi **multinom** (Ripley, 2025).

```

logit_model_numeric <- multinom(
  Risk.Level ~ .,
  data = numeric_train_data, trace = FALSE
)
logit_model_selected <- multinom(
  Risk.Level ~ .,
  data = selected_train_data, trace = FALSE
)

```

Random Forest (RF) adalah kombinasi dari masing-masing tree yang baik kemudian dikombinasikan ke dalam satu model. Random Forest bergantung pada sebuah nilai vector random dengan distribusi yang sama pada semua pohon yang masing-masing cabang memiliki kedalaman yang maksimal. RF merupakan pengembangan dari Decision Tree, dimana data dibagi menjadi himpunan bagian berdasarkan variabel inputnya (2022). Algoritma RF diimplementasikan oleh library **randomForest** (Breiman et al., 2024). Banyak cabang yang dibentuk adalah 54, dengan 3 kali sampling untuk data lengkap dan 2 kali untuk data pilihan.

```

rf_ntree <- as.numeric(floor(sqrt(count(numeric_train_data)) / 2))
rf_mtry_numeric <- floor(sqrt(length(numeric_train_data)))
rf_mtry_selected <- floor(sqrt(length(selected_train_data)))

rf_model_numeric <- randomForest(
  Risk.Level ~ .,
  data = numeric_train_data,
  ntree = rf_ntree,
)

```

```

    mtry = rf_mtry_numeric
  )

rf_model_selected <- randomForest(
  Risk.Level ~ .,
  data = selected_train_data,
  ntree = rf_ntree,
  mtry = rf_mtry_selected
)

```

Support Vector Machines (SVM) adalah model klasifikasi yang menggunakan teknik kernel untuk memetakan data ke dalam ruang berdimensi tinggi, sehingga dapat menangani masalah klasifikasi non-linear (Saepudin et al., 2024). Algoritma RF diimplementasikan oleh library e1071 pada fungsi SVM (Meyer et al., 2024). Kernel yang digunakan di sini adalah kernel radial, yang lebih dapat diandalkan untuk data yang tidak linear, seperti *anxiety level* dan *mood score*.

```

svm_model_numeric <- svm(
  Risk.Level ~ .,
  data = numeric_train_data, kernel = "radial"
)
svm_model_selected <- svm(
  Risk.Level ~ .,
  data = selected_train_data, kernel = "radial"
)

```

3.2 Metrik Pengukuran Performa Model

Metrik pengukuran performa model yang digunakan adalah akurasi prediksi dan lama waktu prediksi. Akurasi berperan penting agar dapat mendekripsi faktor dengan tepat. Metrik akurasi diukur dari atribut “Accuracy” dalam *confusion matrix* (Kuhn, 2024). Sementara itu, waktu juga diperlukan agar mampu membaca data dengan jumlah besar dengan cepat. Performa waktu diukur menggunakan fungsi `system.time` yang mengukur lama waktu prediksi (R Core Team, 2024).

```

# Evaluate models with predictions
evaluate_model <- function(model, testData, model_name) {
  # Measure the time taken to make the predictions
  prediction_time <- system.time({
    predictions <- predict(model, newdata = testData)
  })["elapsed"] # Extract the elapsed time

  # Calculate the confusion matrix
  cm <- confusionMatrix(predictions, testData$Risk.Level)

  # Extract the accuracy
  accuracy <- cm$overall["Accuracy"]

  # Return a list including the model name, accuracy,
  # model object, and prediction time
  list(
    ModelName = model_name,
    ConfusionMatrix = cm,
    Accuracy = accuracy,
    PredictionTime = prediction_time,
    ModelObject = model
  )
}

```

Table 14: Tabel Akurasi dan Waktu Prediksi Model

Model	Accuracy	PredictionTime	NormAccuracy	NormPredictionTime	CompositeScore
MLR (All Fields)	797266	142000	797266	843750	801914
MLR (Selected Fields)	797266	27000	797266	1000000	817539
RF (All Fields)	1000000	105000	1000000	894022	989402
RF (Selected Fields)	1000000	35000	1000000	989130	998913
SVM (All Fields)	971991	763000	971991	0	874792
SVM (Selected Fields)	992998	164000	992998	813859	975084

```

}

# Initialize results data frame
results <- list(
  evaluate_model(
    logit_model_numeric, numeric_test_data, "MLR (All Fields)"
  ),
  evaluate_model(
    logit_model_selected, selected_test_data, "MLR (Selected Fields)"
  ),
  evaluate_model(
    rf_model_numeric, numeric_test_data, "RF (All Fields)"
  ),
  evaluate_model(
    rf_model_selected, selected_test_data, "RF (Selected Fields)"
  ),
  evaluate_model(
    svm_model_numeric, numeric_test_data, "SVM (All Fields)"
  ),
  evaluate_model(
    svm_model_selected, selected_test_data, "SVM (Selected Fields)"
  )
)

```

4 Hasil dan Pembahasan

Untuk evaluasi, R^2 dan $RMSE$ tidak bisa digunakan untuk soal klasifikasi, karena lebih cocok untuk masalah regresi. Data yang dihasilkan dari *confusion matrix* lebih cocok untuk dalam kasus ini. Hasil dapat dilihat pada tabel 14 serta grafik 6, 7, dan 8.

Model Random Forest menghasilkan akurasi sempurna, baik menggunakan data lengkap maupun data pilihan. SVM tidak jauh tertinggal, namun menghasilkan akurasi lebih baik menggunakan data dengan atribut yang telah disaring sebelumnya. MLR menghasilkan akurasi terendah dari semuanya, dan tidak ada perbedaan antara penggunaan data lengkap ataupun data pilihan.

Untuk waktu, terlihat jelas bahwa penggunaan data pilihan mengurangi waktu prediksi secara signifikan terhadap ketiga model. Karena kompleksitasnya, SVM memakan waktu lebih lama daripada model lainnya, meskipun demikian SVM mengalami penurunan waktu yang signifikan dengan menggunakan data yang telah dipilah untuk optimasi.

Untuk mengukur performa keseluruhan, nilai waktu terlebih dahulu distandardkan di sekitar maksimum dan minimum sehingga waktu tersingkat memberi nilai lebih. Kemudian nilai keduanya digabung dengan komposisi akurasi 90% dan waktu 10%. Hasilnya, model Random Forest mengungguli model yang lain berkat

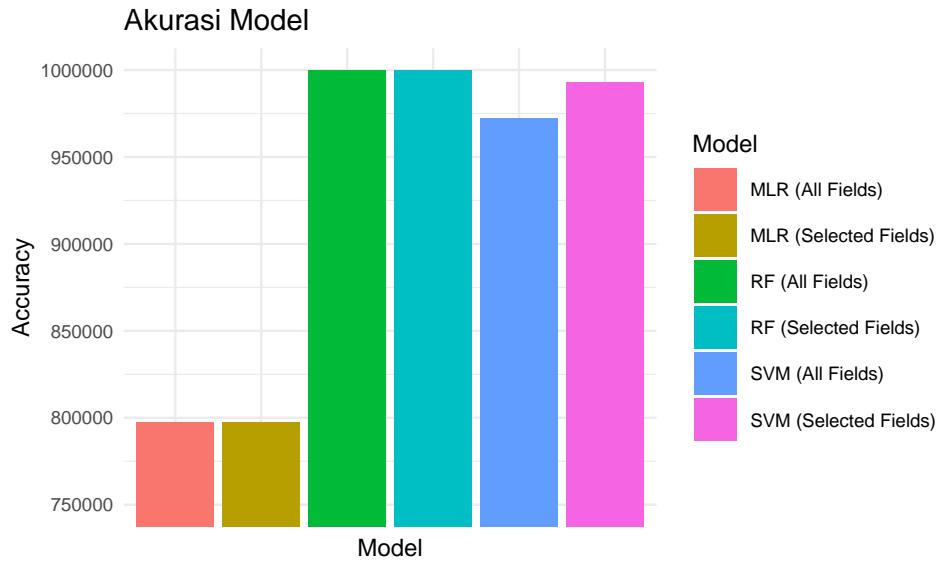


Figure 6: Akurasi Model

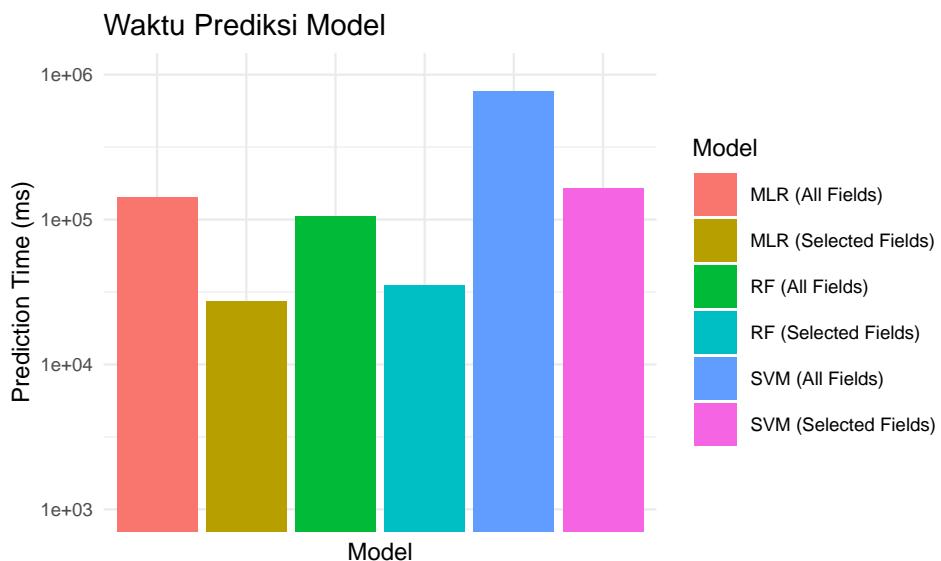


Figure 7: Waktu Prediksi Model

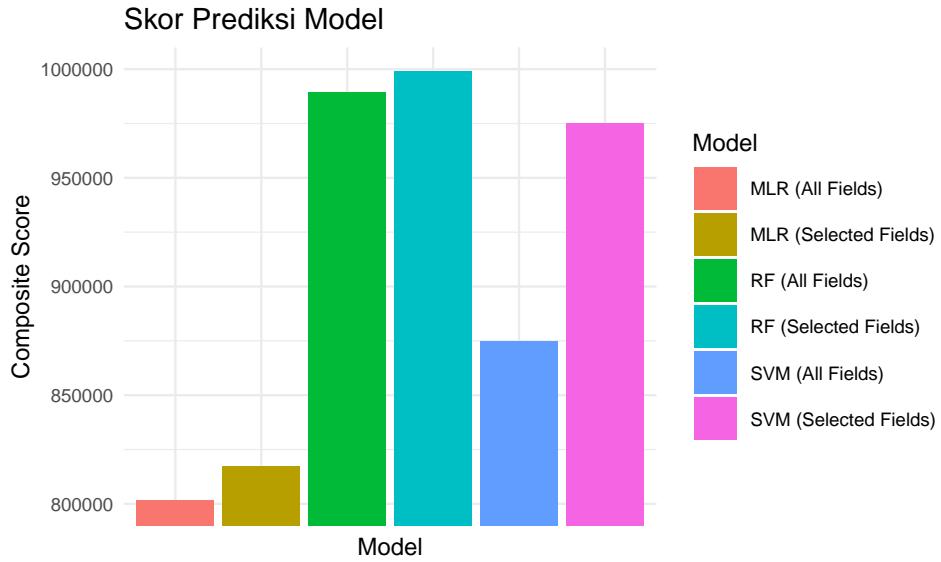


Figure 8: Skor Prediksi Model

akurasinya yang tinggi. Model RF dengan data pilihan tampil lebih baik dengan optimasi waktunya.

```
## [1] "Best Model: RF (Selected Fields)"
```

Berdasarkan hasil di atas, model dan pilihan data terbaik adalah RF (Selected Fields) dengan skor akhir 9.9891304×10^5 , dan nilai *precision*, *recall*, dan skor *F1* yang sempurna. Model ini mampu mengidentifikasi semua data tes dengan tepat.

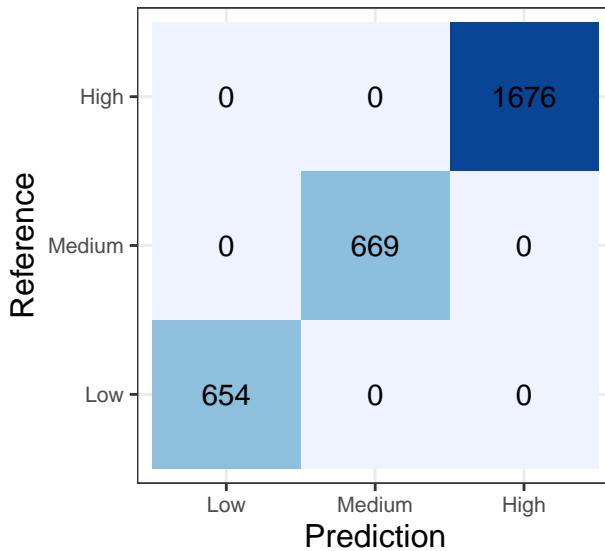


Figure 9: Confusion Matrix Model Terbaik

Table 15: Tabel Prediksi Per Tingkat Risiko

	Sens	Spec	+ p-v	- p-v	Prec	Recl	F1	Prev	DetR	DetP	Accr
Class: Low	1	1	1	1	1	1	1	0.218073	0.218073	0.218073	1
Class: Medium	1	1	1	1	1	1	1	0.223074	0.223074	0.223074	1
Class: High	1	1	1	1	1	1	1	0.558853	0.558853	0.558853	1

Table 16: Ringkasan Performa Model Terbaik

	x
Accuracy	1.0000000
Kappa	1.0000000
AccuracyLower	0.9987707
AccuracyUpper	1.0000000
AccuracyNull	0.5588530
AccuracyPValue	0.0000000
McnemarPValue	N/A

5 Kesimpulan

Laporan ini bertujuan untuk mengeksplorasi data kesehatan dan kehadiran mahasiswa untuk mengungkap keterkaitan antar variabel, membentuk model pembelajaran mesin untuk mengidentifikasi tingkat risiko kesehatan mahasiswa, dan bagaimana hasil eksplorasi data berperan dalam optimalisasi data yang digunakan dalam pembelajaran mesin.

Hasil eksplorasi data menunjukkan bahwa terdapat hubungan kuat pada status kehadiran, dan tingkat stres, serta hubungan moderat pada tingkat kekhawatiran dan suasana hati terhadap tingkat risiko kesehatan. Keempat atribut ini yang nantinya berperan dalam optimalisasi data yang digunakan sebagai acuan model.

Berdasarkan faktor akurasi dan performa waktu, hasil pelatihan dan pengujian model pembelajaran mesin menunjukkan bahwa Random Forest dengan data pilihan merupakan model terbaik untuk dataset ini. Secara keseluruhan, performa RF mengungguli model lainnya baik dalam hal akurasi maupun waktu. Penggunaan data pilihan, yaitu data dengan atribut yang memiliki tingkat korelasi yang memadai, mampu meningkatkan akurasi dan mengurangi lama prediksi, terutama pada model SVM.

Laporan ini masih meninggalkan banyak ranah untuk pengembangan dan perbaikan. Pertama, analisis ke ranah psikologi perlu dilakukan untuk memahami lebih lanjut hubungan dari tingkat stres, kekhawatiran, dan suasana hati terhadap kehadiran dan risiko kesehatan. Kedua, penyebab dari kekurangan tiap model masih perlu digali lebih lanjut dengan memperhatikan karakteristik dari tiap model dan data yang digunakan untuk meningkatkan performa terhadap semua model. Dan ketiga, penelitian lain menyuguhkan soal penggunaan metode lain yang disebut *Gradient Boosting* atau XGBoost yang diklaim lebih efektif untuk data yang lebih kompleks, sehingga metode ini layak menjadi pertimbangan untuk penelitian kedepannya.

6 Daftar Pustaka

- Airlangga, Gregorius. (2024). Comparative analysis of machine learning models for chronic disease indicator classification using u.s. Chronic disease indicators dataset. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4, 1034–1042. <https://doi.org/10.57152/malcom.v4i3.1403>
- Breiman, Leo, Cutler, Adele, Liaw, Andy, & Wiener, Matthew. (2024). *randomForest: Breiman and cutlers random forests for classification and regression*. <https://www.stat.berkeley.edu/~breiman/RandomForests/>
- Dahl, David B., Scott, David, Roosen, Charles, Magnusson, Arni, & Swinton, Jonathan. (2019). *Xtable: Export tables to LaTeX or HTML*. <http://xtable.r-forge.r-project.org/>
- Fadhil, Mutasimbillah, & Matang. (2024). KESEHATAN MENTAL MAHASISWA DAN PENGARUHNYA TERHADAP PERAN MAHASISWA DALAM BERMASYARAKAT BERBANGSA DAN BERNE-

- GARA. *Maha Widya Bhuwana: Jurnal Pendidikan, Agama Dan Budaya*, 7, 29–32. <https://journal.stahnmpukuturan.ac.id/index.php/bhuwana/article/view/344>
- Fox, John, & Weisberg, Sanford. (2019). *An R companion to applied regression* (Third). Sage. <https://www.john-fox.ca/Companion/>
- Fox, John, Weisberg, Sanford, & Price, Brad. (2022). *carData: Companion to applied regression data sets*. <https://r-forge.r-project.org/projects/car/>
- Fox, John, Weisberg, Sanford, & Price, Brad. (2024). *Car: Companion to applied regression*. <https://r-forge.r-project.org/projects/car/>
- Grolemund, Garrett, & Wickham, Hadley. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, 40(3), 1–25. <https://www.jstatsoft.org/v40/i03/>
- Gross, Juergen, & Ligges, Uwe. (2015). *Nortest: Tests for normality*. <https://CRAN.R-project.org/package=nortest>
- Harrell, Frank E., Jr. (2025). *Hmisc: Harrell miscellaneous*. <https://hbiostat.org/R/Hmisc/>
- Hikmawati, Fenti. (2020). *Metodologi penelitian*. Rajawali Pers. <https://etheses.uinsgd.ac.id/31676/1/Metodologi%20Penelitian.pdf>
- Kamila, Siti Alifia. (2024). *MEMBANGUN KESEIMBANGAN: MENJAGA KESEHATAN MENTAL DALAM GELOMBANG TANTANGAN MAHASISWA*. i-WIN (International Waqaf Ilmu Nusantara) Library. https://waqafilmunusantara.com/wp-content/uploads/2024/08/Siti-Alifia-Kamila_Artikel.docx.pdf
- Kassambara, Alboukadel. (2023). *Ggpubr: ggplot2 based publication ready plots*. <https://rpkg.datanovia.com/ggpubr/>
- Khair, Afra Fitri Aulia, & Usiono. (2025). Pengaruh pola hidup terhadap kesehatan mahasiswa. *Jurnal Ilmiah Nusantara*, 2, 48–54. <https://ejurnal.kampusakademik.co.id/index.php/jinu/article/download/3536/3183/14261>
- Komsta, Lukasz, & Novomestky, Frederick. (2022). *Moments: Moments, cumulants, skewness, kurtosis and related tests*. <https://www.r-project.org>
- Kuhn, Max. (2024). *Caret: Classification and regression training*. <https://github.com/topepo/caret/>
- Kuhn, & Max. (2008). Building predictive models in r using the caret package. *Journal of Statistical Software*, 28(5), 1–26. <https://doi.org/10.18637/jss.v028.i05>
- Liaw, Andy, & Wiener, Matthew. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22. <https://CRAN.R-project.org/doc/Rnews/>
- Lüdecke, Daniel, Makowski, Dominique, Patil, Indrajeet, Waggoner, Philip, Ben-Shachar, Mattan S., Wiernik, Brenton M., Arel-Bundock, Vincent, & Bacher, Etienne. (2025). *Insight: Easy access to model information for various model objects*. <https://easystats.github.io/insight/>
- Lüdecke, Daniel, Waggoner, Philip, & Makowski, Dominique. (2019). insight: A unified interface to access information from model objects in R. *Journal of Open Source Software*, 4(38), 1412. <https://doi.org/10.21105/joss.01412>
- Meyer, David, Dimitriadou, Evgenia, Hornik, Kurt, Weingessel, Andreas, & Leisch, Friedrich. (2024). *e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071)*, TU wien. <https://CRAN.R-project.org/package=e1071>
- Müller, Kirill, & Wickham, Hadley. (2023). *Tibble: Simple data frames*. <https://tibble.tidyverse.org/>
- Olszewski, Jessica Larocco. (2024). *Introduction to classification model comparison methods – OMSCS 7641: Machine learning*. Georgia Tech; Georgia Institute of Technology. <https://sites.gatech.edu/omschs7641/2024/02/07/introduction-to-classification-model-comparison-methods/>
- Paliling, Yusrian, Fathurahman, M., & Wahyuningsih, Sri. (2023). Multinomial logistic regression to model the combination of phdi status and hdi status of districts/cities in kalimantan island. *Jurnal Matematika Statistika Dan Komputasi*, 19, 460–472. <https://doi.org/10.20956/j.v19i3.22299>
- Pratama, Rivan Saghita, & Nur Zaimah, Annisa Fauziyah. (2024). PENGARUH AKTIVITAS FISIK TERHADAP KESEHATAN MENTAL DAN KINERJA AKADEMIK MAHASISWA PENDIDIKAN BAHASA INGGRIS. *Jurnal Lingkar Pembelajaran Inovatif*, 5. <https://oaj.jurnalhst.com/index.php/jli/article/view/4725>
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ramdani, Fatwa, & Utami, Ika Qutsiati. (2024). *Pengantar data science*. Bumi Aksara.

- Ripley, Brian. (2025). *Nnet: Feed-forward neural networks and multinomial log-linear models*. http://www.stats.ox.ac.uk/pub/MASS4/
- Saepudin, Asep, Faqih, Ahmad, & Dwilestari, Gifthera. (2024). Perbandingan algoritma klasifikasi support vector machine, random forest dan logistic regression pada ulasan shopee. *Jurnal Tekno Kompak*, 18, 178–178. https://doi.org/10.33365/jtk.v18i1.3764
- Salsabila MR. (2022). *6 jenis machine learning wajib diketahui dalam data science*. DQLab. https://dqlab.id/6-jenis-machine-learning-wajib-diketahui-dalam-data-science
- Sarkar, Deepayan. (2008). *Lattice: Multivariate data visualization with r*. Springer. http://lmdvr.r-forge.r-project.org
- Sarkar, Deepayan. (2024). *Lattice: Trellis graphics for r*. https://lattice.r-forge.r-project.org/
- Spinu, Vitalie, Grolemund, Garrett, & Wickham, Hadley. (2024). *Lubridate: Make dealing with dates a little easier*. https://lubridate.tidyverse.org
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with s* (Fourth). Springer. https://www.stats.ox.ac.uk/pub/MASS4/
- Wei, Taiyun, & Simko, Viliam. (2024a). *Corrplot: Visualization of a correlation matrix*. https://github.com/taiyun/corrplot
- Wei, Taiyun, & Simko, Viliam. (2024b). *R package 'corrplot': Visualization of a correlation matrix*. https://github.com/taiyun/corrplot
- Wickham, Hadley. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org
- Wickham, Hadley. (2023a). *Forcats: Tools for working with categorical variables (factors)*. https://forcats.tidyverse.org/
- Wickham, Hadley. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. https://stringr.tidyverse.org
- Wickham, Hadley. (2023c). *Tidyverse: Easily install and load the tidyverse*. https://tidyverse.tidyverse.org
- Wickham, Hadley, Averick, Mara, Bryan, Jennifer, Chang, Winston, McGowan, Lucy D'Agostino, François, Romain, Grolemund, Garrett, Hayes, Alex, Henry, Lionel, Hester, Jim, Kuhn, Max, Pedersen, Thomas, Lin, Miller, Evan, Bache, Stephan Milton, Müller, Kirill, Ooms, Jeroen, Robinson, David, Seidel, Dana, Paige, Spinu, Vitalie, ... Yutani, Hiroaki. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. https://doi.org/10.21105/joss.01686
- Wickham, Hadley, Chang, Winston, Henry, Lionel, Pedersen, Thomas Lin, Takahashi, Kohske, Wilke, Claus, Woo, Kara, Yutani, Hiroaki, Dunnington, Dewey, & van den Brand, Teun. (2024). *ggplot2: Create elegant data visualisations using the grammar of graphics*. https://ggplot2.tidyverse.org
- Wickham, Hadley, François, Romain, Henry, Lionel, Müller, Kirill, & Vaughan, Davis. (2023). *Dplyr: A grammar of data manipulation*. https://dplyr.tidyverse.org
- Wickham, Hadley, & Henry, Lionel. (2025). *Purrr: Functional programming tools*. https://purrr.tidyverse.org/
- Wickham, Hadley, Hester, Jim, & Bryan, Jennifer. (2024). *Readr: Read rectangular text data*. https://readr.tidyverse.org
- Wickham, Hadley, Vaughan, Davis, & Girlich, Maximilian. (2024). *Tidyr: Tidy messy data*. https://tidyr.tidyverse.org
- Zhao, Heng, Chen, Yixing, & Fu, Xianghua. (2023). Comparison of machine learning based on category theory. *Journal of Web Engineering*, 22. https://doi.org/10.13052/jwe1540-9589.2213
- Zhu, Hao. (2024). *kableExtra: Construct complex table with kable and pipe syntax*. http://haozhu233.github.io/kableExtra/
- Ziya. (2024). *Student health and attendance data*. Kaggle. https://www.kaggle.com/datasets/ziya07/student-health-and-attendance-data

A Informasi Sesi R

```
sessionInfo()

## R version 4.4.2 (2024-10-31)
## Platform: x86_64-redhat-linux-gnu
```

```

## Running under: Fedora Linux 40 (Toolbox Container Image)
##
## Matrix products: default
## BLAS/LAPACK: FlexiBLAS OPENBLAS-OPENMP; LAPACK version 3.12.0
##
## locale:
## [1] LC_CTYPE=en_GB.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=en_GB.UTF-8       LC_COLLATE=en_GB.UTF-8
## [5] LC_MONETARY=en_GB.UTF-8   LC_MESSAGES=en_GB.UTF-8
## [7] LC_PAPER=en_GB.UTF-8      LC_NAME=C
## [9] LC_ADDRESS=C              LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_GB.UTF-8 LC_IDENTIFICATION=C
##
## time zone: Asia/Makassar
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats      graphics   grDevices datasets  utils      methods   base
##
## other attached packages:
## [1] e1071_1.7-16      randomForest_4.7-1.2 nnet_7.3-20
## [4] caret_7.0-1       lattice_0.22-6      kableExtra_1.4.0
## [7] xtable_1.8-4      corrplot_0.95      ggpubr_0.6.0
## [10] insight_1.0.2     nortest_1.0-4      car_3.1-3
## [13] carData_3.0-5     moments_0.14.1    Hmisc_5.2-2
## [16] lubridate_1.9.4   forcats_1.0.0      stringr_1.5.1
## [19] dplyr_1.1.4       purrr_1.0.4       readr_2.1.5
## [22] tidyverse_1.3.1    tibble_3.2.1      ggplot2_3.5.1
## [25] tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
## [1] pROC_1.18.5        gridExtra_2.3      rlang_1.1.5
## [4] magrittr_2.0.3      compiler_4.4.2     systemfonts_1.2.1
## [7] vctrs_0.6.5         reshape2_1.4.4     pkgconfig_2.0.3
## [10] fastmap_1.2.0      backports_1.5.0    labeling_0.4.3
## [13] CoprManager_0.5.7  rmarkdown_2.29     prodlim_2024.06.25
## [16] tzdb_0.4.0          xfun_0.50        recipes_1.1.1
## [19] broom_1.0.7         parallel_4.4.2    cluster_2.1.8
## [22] R6_2.6.1           RColorBrewer_1.1-3 stringi_1.8.4
## [25] parallelly_1.42.0  rpart_4.1.24     Rcpp_1.0.14
## [28] bookdown_0.42       iterators_1.0.14 knitr_1.49
## [31] future.apply_1.11.3 base64enc_0.1-3   Matrix_1.7-2
## [34] splines_4.4.2       timechange_0.3.0  tidyselect_1.2.1
## [37] rstudioapi_0.17.1   abind_1.4-8      yaml_2.3.10
## [40] timeDate_4041.110  codetools_0.2-20  listenv_0.9.1
## [43] plyr_1.8.9          withr_3.0.2      evaluate_1.0.3
## [46] foreign_0.8-88     future_1.34.0   survival_3.8-3
## [49] proxy_0.4-27        xml2_1.3.6      pillar_1.10.1
## [52] checkmate_2.3.2     foreach_1.5.2    stats4_4.4.2
## [55] generics_0.1.3      hms_1.1.3       munsell_0.5.1
## [58] scales_1.3.0        globals_0.16.3  class_7.3-23
## [61] glue_1.8.0           tools_4.4.2      data.table_1.16.4
## [64] ModelMetrics_1.2.2.2 gower_1.0.2    ggsignif_0.6.4
## [67] grid_4.4.2          ipred_0.9-15   colorspace_2.1-1

```

```
## [70] nlme_3.1-167          htmlTable_2.4.3      Formula_1.2-5
## [73] cli_3.6.4                viridisLite_0.4.2    svglite_2.1.3
## [76] lava_1.8.1                gtable_0.3.6        rstatix_0.7.2
## [79] digest_0.6.37              farver_2.1.2       htmlwidgets_1.6.4
## [82] htmltools_0.5.8.1         lifecycle_1.0.4     hardhat_1.4.1
## [85] MASS_7.3-64
```