

Analisis Hubungan dan Prediksi Depresi Mahasiswa Berdasarkan Faktor Akademik dan Gender

Miranti Verdiana^{*1}, Eko Dwi Nugroho², Leslie Anggraini³, Radhinka Bagaskara⁴, Winda Yulita⁵, Aidil Afriansyah⁶, Muhammad Habib Algifari⁷

^{1,2}Program Studi Teknik Informatika, Institut Teknologi Sumatera

E-mail: ^{*1}miranti.verdiana@if.itera.ac.id, ²eko.nugroho@if.itera.ac.id, ³leslie.anggraini@if.itera.ac.id, ⁴radhinka.bagaskara@if.itera.ac.id, ⁵winda.yulita@if.itera.ac.id, ⁶aidil.afriansyah@if.itera.ac.id, ⁷muhammad.algifari@if.itera.ac.id

Abstrak

Penelitian ini bertujuan untuk menganalisis tingkat depresi pada mahasiswa dengan mempertimbangkan faktor gender dan sejumlah indikator akademik. Data yang digunakan terdiri dari 27.901 responden mahasiswa dari berbagai wilayah, dengan variabel yang mencakup usia, jenis kelamin, tekanan akademik, kepuasan studi, jam kerja/belajar, IPK (CGPA), serta status depresi. Metode analisis yang digunakan meliputi uji chi-square untuk menguji hubungan antara gender dan status depresi, analisis korelasi point-biserial antara variabel numerik dan depresi, serta pembangunan model prediksi menggunakan regresi logistik. Hasil analisis menunjukkan bahwa tidak terdapat hubungan yang signifikan antara gender dan status depresi ($p = 0,774$), yang mengindikasikan bahwa depresi terjadi lintas jenis kelamin. Sebaliknya, tekanan akademik memiliki korelasi paling kuat dengan status depresi ($r = 0,47$), disusul oleh jam kerja/belajar ($r = 0,209$) dan kepuasan studi ($r = -0,168$). Model Logistic Regression yang dibangun menggunakan empat variabel utama menunjukkan performa cukup baik, dengan akurasi sebesar 75,5% dan recall sebesar 82,1% dalam mendeteksi mahasiswa dengan status depresi. Hasil ini menegaskan bahwa faktor akademik, khususnya tekanan akademik, memegang peran signifikan dalam memengaruhi kondisi mental mahasiswa. Oleh karena itu, diperlukan strategi dukungan akademik yang lebih terarah sebagai upaya pencegahan depresi di lingkungan pendidikan tinggi.

Kata Kunci— Depresi Mahasiswa, Tekanan Akademik, Gender, Logistic Regression, Prediksi Mental Health.

Abstract

This study aims to analyze the level of depression among university students by examining gender and several academic indicators. The dataset includes responses from 27,901 students across various regions, with variables covering age, gender, academic pressure, study satisfaction, work/study hours, CGPA, and depression status. The analytical methods applied in this study include the chi-square test to evaluate the association between gender and depression status, point-biserial correlation to examine relationships between numeric variables and depression, and logistic regression to develop a prediction model. The chi-square test results revealed no significant relationship between gender and depression ($p = 0.774$), indicating that depression affects both genders. In contrast, academic pressure exhibited the strongest correlation with depression status ($r = 0.47$), followed by work/study hours ($r = 0.209$) and study satisfaction ($r = -0.168$). The Logistic Regression model constructed using the four most relevant variables demonstrated satisfactory performance, achieving 75.5% accuracy and 82.1% recall in identifying students experiencing depression. These findings highlight the critical role of academic-related factors—particularly academic pressure in influencing students' mental health. Therefore, targeted academic support strategies are essential to mitigate depression risks in higher education environments.

Keywords— Student Depression, Academic Pressure, Gender, Logistic Regression, Mental Health Prediction

1. PENDAHULUAN

Kesehatan mental mahasiswa merupakan isu yang semakin mendapat perhatian dalam beberapa dekade terakhir, terutama dengan meningkatnya tekanan akademik dan tuntutan sosial yang dihadapi oleh kalangan pelajar perguruan tinggi. Organisasi Kesehatan Dunia (WHO) menyebutkan bahwa

depresi menjadi salah satu penyebab utama gangguan kesehatan mental di usia produktif, termasuk di kalangan mahasiswa [1]. Mahasiswa sering kali dihadapkan pada tantangan akademik, tekanan karier, ketidakpastian masa depan, serta transisi kehidupan yang cepat, yang berkontribusi terhadap risiko gangguan psikologis seperti depresi dan kecemasan [2].

Di Indonesia, isu depresi di kalangan mahasiswa juga menjadi perhatian karena dapat memengaruhi performa akademik, relasi sosial, dan kualitas hidup secara keseluruhan. Hasil survei dari beberapa perguruan tinggi menunjukkan bahwa tekanan akademik, beban tugas, dan kurangnya dukungan emosional menjadi pemicu utama munculnya gejala depresi [3]. Selain itu, banyak kasus gangguan mental tidak terdeteksi dengan baik karena keterbatasan sistem pemantauan serta stigma terhadap kesehatan mental.

Permasalahan khusus yang ingin dikaji dalam penelitian ini adalah identifikasi hubungan antara faktor-faktor akademik dan karakteristik demografis (seperti gender dan usia) terhadap tingkat depresi mahasiswa. Penelitian ini juga bertujuan untuk membangun model prediksi status depresi mahasiswa berdasarkan variabel-variabel signifikan dengan metode regresi logistik. Dengan adanya model prediktif ini, diharapkan institusi pendidikan dapat melakukan deteksi dini dan memberikan intervensi yang tepat kepada mahasiswa yang berisiko tinggi mengalami depresi.

Penelitian ini bersifat kuantitatif dengan pendekatan deskriptif dan prediktif. Data diperoleh dari *kaggle* berupa dataset survei mahasiswa yang mencakup 27.901 baris data. Analisis dilakukan dengan menggunakan uji *chi-square* untuk menganalisis hubungan antara gender dan depresi, analisis korelasi *point-biserial* untuk variabel numerik, serta pembangunan model klasifikasi menggunakan *logistic regression*. Temuan dari studi ini diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem deteksi dini depresi berbasis data di lingkungan pendidikan tinggi.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dan inferensial, dengan tujuan untuk menganalisis hubungan antara faktor-faktor akademik dan status depresi mahasiswa, serta membangun model prediksi menggunakan metode regresi logistik. Tahapan penelitian meliputi:

2.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari Kaggle dan dikembangkan oleh Adil Shamim. Dataset ini memuat data survei mahasiswa yang bertujuan untuk mengeksplorasi hubungan antara tekanan akademik, karakteristik demografis, dan kondisi psikologis dengan status depresi. Terdiri dari 27.901 catatan mahasiswa, dataset ini mencakup informasi seperti jenis kelamin, usia, kota asal, tekanan akademik, jumlah jam belajar atau bekerja, tingkat kepuasan terhadap studi, nilai IPK, dan status depresi yang dialami mahasiswa. Dataset ini sangat berguna sebagai bahan kajian dalam upaya memahami faktor-faktor yang mempengaruhi kesehatan mental mahasiswa, khususnya dalam mengidentifikasi risiko depresi. Penggunaan data ini diharapkan dapat mendukung pengembangan sistem prediksi dan deteksi dini depresi berbasis data, yang nantinya akan bermanfaat untuk menciptakan lingkungan akademik yang lebih sehat dan suportif bagi mahasiswa.

2.2. Analisis Statistik Deskriptif

Analisis statistik deskriptif pada penelitian ini dilakukan untuk menggambarkan distribusi dan karakteristik variabel-variabel yang diteliti dari data berjumlah 27.901 mahasiswa. Tahap ini penting sebagai dasar untuk memahami profil data sebelum dilakukan analisis inferensial dan pembangunan model prediksi [4].

Ukuran pemusatan data yang digunakan mencakup rata-rata (*mean*), median dan modus. Rata-rata dihitung untuk variabel numerik seperti tekanan akademik, usia, jam kerja/belajar, dan kepuasan studi, untuk mengetahui nilai tengah yang mewakili distribusi data. Median digunakan untuk memetakan posisi nilai tengah dalam distribusi, sementara modus diaplikasikan terutama pada variabel kategorik, seperti gender dan kota asal mahasiswa, untuk melihat kategori terbanyak [5].

Ukuran penyebaran yang digunakan meliputi range (jangkauan), varians, dan standar deviasi. Variabel numerik seperti tekanan akademik dan jam kerja/belajar dianalisis untuk menilai seberapa besar variasi yang terjadi pada data responden [6]. Untuk mendeteksi kemungkinan outlier atau pencilan data yang signifikan, digunakan Interquartile Range (IQR), yang dihitung sebagai selisih antara kuartil ketiga (Q_3) dan kuartil pertama (Q_1) [7].

Visualisasi berupa *boxplot* ditampilkan untuk menggambarkan penyebaran data, mendeteksi *outlier*, dan membandingkan distribusi antar kelompok (misalnya antara mahasiswa depresi dan tidak depresi).

Hasil analisis deskriptif ini memberikan informasi awal bahwa mayoritas responden adalah mahasiswa dengan tekanan akademik bervariasi, usia yang mayoritas berada di rentang dewasa muda, serta proporsi mahasiswa yang mengalami depresi yang perlu perhatian khusus. Tahap ini menjadi dasar penting untuk mendukung pengujian hipotesis dan pengembangan model prediktif berbasis regresi logistik pada tahap selanjutnya.

2.3. Uji Hubungan dan Korelasi

Setelah pemetaan karakteristik data melalui analisis statistik deskriptif, tahap selanjutnya adalah menguji hubungan dan korelasi antara variabel-variabel penelitian, khususnya antara variabel demografis, akademik, dan status depresi mahasiswa. Uji ini bertujuan untuk mengidentifikasi variabel yang memiliki hubungan signifikan dengan depresi sehingga dapat digunakan dalam pembangunan model prediktif [4][5].

2.3.1. Uji Hubungan antara Variabel Kategorik

Untuk menguji hubungan antara variabel kategorik seperti gender dan status depresi, digunakan Uji Chi-Square (χ^2). Uji ini menguji hipotesis nol bahwa tidak ada hubungan antara dua variabel kategorik. Rumus umum Chi-Square adalah:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

di mana O_i adalah frekuensi observasi dan E_i adalah frekuensi ekspektasi. Jika p-value hasil uji $< 0,05$, maka hipotesis nol ditolak, yang menunjukkan adanya hubungan signifikan [7].

2.3.2. Uji Korelasi antara Variabel Numerik dan Depresi

Untuk mengukur kekuatan dan arah hubungan antara variabel numerik (tekanan akademik, kepuasan studi, jam kerja/belajar, usia) dengan status depresi (variabel biner), digunakan korelasi point-biserial (r_{pb}). Rumus korelasi point-biserial adalah:

$$r_{pb} = \frac{\bar{X}_1 - \bar{X}_0}{s_X} \sqrt{\frac{n_1 n_0}{n(n-1)}}$$

di mana \bar{X}_1 dan \bar{X}_0 adalah rata-rata kelompok 1 (depresi) dan 0 (tidak depresi), s_X adalah standar deviasi variabel numerik, n_1 dan n_0 adalah jumlah data untuk masing-masing kelompok, dan n adalah jumlah total data [4].

2.4. Integrasi Hasil Uji ke Model Prediksi

Setelah diperoleh hasil analisis deskriptif dan pengujian hubungan serta korelasi antar variabel, tahapan selanjutnya adalah mengintegrasikan variabel-variabel yang signifikan ke dalam pembangunan model prediksi. Metode yang digunakan dalam penelitian ini adalah *Logistic Regression*, yang sesuai untuk memodelkan probabilitas kejadian pada variabel target yang bersifat kategorikal, dalam hal ini status depresi mahasiswa (*Depression* = 0 atau 1). Variabel-variabel prediktor yang dipilih meliputi tekanan akademik, usia, jam kerja/belajar, dan kepuasan studi, yang telah menunjukkan hubungan signifikan dengan status depresi pada tahap analisis sebelumnya [5], [8].

Sebelum dilakukan pemodelan, data yang telah melalui tahap prapemrosesan dibagi menjadi data latih (80%) dan data uji (20%) dengan metode *holdout validation* untuk menghindari *overfitting*. Variabel kategorikal diubah ke format numerik menggunakan label *encoding*, sedangkan variabel numerik distandarisasi untuk menjaga skala yang seragam [9]. Model *Logistic Regression* dibangun dengan menggunakan persamaan berikut:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

di mana p adalah probabilitas mahasiswa mengalami depresi, X_i adalah variabel prediktor, dan β_i adalah koefisien regresi yang akan diestimasi [7]. Model ini dikalibrasi menggunakan data latih dan diuji performanya dengan data uji. Evaluasi dilakukan dengan menghitung metrik akurasi, *precision*, *recall*, *F1-score*, serta menggunakan *confusion matrix* untuk memetakan klasifikasi benar dan salah.

Interpretasi koefisien regresi dilakukan dengan menghitung *odds ratio* untuk menilai seberapa besar pengaruh variabel terhadap peluang terjadinya depresi [8].

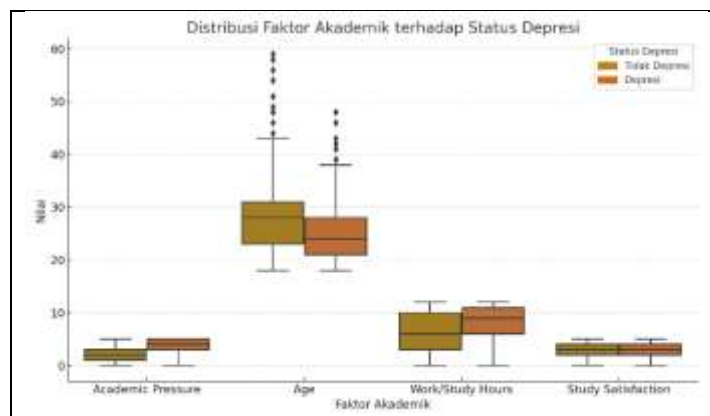
3. HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil analisis data dan pembahasan berdasarkan metodologi yang telah ditetapkan pada penelitian ini. Analisis dilakukan secara menyeluruh untuk memahami pola distribusi data, mengidentifikasi hubungan antar variabel, dan membangun model prediksi status depresi mahasiswa. Tujuan utama dari penyajian hasil ini adalah untuk menjawab permasalahan penelitian, apakah terdapat hubungan signifikan antara variabel akademik dan demografis terhadap status depresi mahasiswa, serta seberapa efektif model dalam memprediksi kondisi depresi tersebut.

Setiap langkah analisis tidak hanya berfokus pada hasil kuantitatif, tetapi juga didukung oleh interpretasi yang mendalam serta penguatan dengan literatur terbaru. Penyajian data dalam bentuk tabel, visualisasi grafik, dan metrik evaluasi model memberikan pemahaman yang komprehensif dan mendalam terhadap pola hubungan variabel, dengan implikasi praktis bagi pengembangan kebijakan kesehatan mental di lingkungan perguruan tinggi.

3.1. Hasil Analisis Statistik Deskriptif

Hasil analisis deskriptif menunjukkan bahwa dari 27.901 mahasiswa yang menjadi responden, distribusi variabel numerik seperti tekanan akademik, usia, jam kerja/belajar, dan kepuasan studi bervariasi. Rata-rata tekanan akademik berada pada tingkat sedang hingga tinggi, sebarannya dapat dilihat pada Gambar 1.



Gambar 1. Boxplot Distribusi Faktor Akademik terhadap Status Depresi

Gambar 1 memberikan visualisasi distribusi nilai faktor akademik terhadap status depresi mahasiswa, dengan kategori Tidak Depresi dan Depresi. Boxplot tersebut menunjukkan distribusi empat faktor akademik utama terhadap status depresi mahasiswa, yaitu tekanan akademik, usia, jam kerja/belajar, dan kepuasan studi. Berdasarkan visualisasi, terlihat bahwa mahasiswa yang mengalami depresi cenderung memiliki median tekanan akademik yang lebih tinggi dibandingkan dengan mahasiswa yang tidak mengalami depresi. Hal ini juga disertai dengan sebaran data yang lebih lebar pada kelompok depresi, yang mengindikasikan adanya mahasiswa dengan tingkat tekanan akademik yang sangat tinggi. Pada variabel usia, kelompok mahasiswa yang mengalami depresi umumnya memiliki median usia yang lebih rendah, sementara kelompok tidak depresi menunjukkan sebaran yang lebih luas, bahkan terdapat outlier mahasiswa berusia lebih tua. Pola ini memberikan indikasi bahwa mahasiswa yang lebih muda mungkin memiliki kerentanan lebih besar terhadap depresi.

Selanjutnya, distribusi pada variabel jam kerja/belajar menunjukkan bahwa kelompok depresi memiliki median jam kerja/belajar yang sedikit lebih tinggi dibandingkan kelompok tidak depresi. Hal ini memperkuat dugaan bahwa jam kerja/belajar yang panjang berpotensi meningkatkan risiko depresi. Sementara itu, pada variabel kepuasan studi, median nilai kelompok depresi cenderung lebih rendah dibandingkan kelompok tidak depresi, dengan sebaran yang relatif kecil. Hal ini menunjukkan bahwa mahasiswa yang memiliki kepuasan studi rendah lebih berisiko mengalami depresi.

Secara keseluruhan, boxplot ini memberikan indikasi bahwa tekanan akademik yang tinggi, usia yang lebih muda, jam kerja/belajar yang panjang, dan kepuasan studi yang rendah merupakan faktor-

faktor yang mungkin berhubungan dengan tingginya prevalensi depresi pada mahasiswa. Namun, temuan ini bersifat deskriptif dan memerlukan pengujian statistik lanjutan untuk mengonfirmasi kekuatan hubungan antar variabel.

3.2. Hasil Uji Hubungan Gender dan Depresi

Berikut adalah tabel kontingensi untuk uji Chi-Square antara Gender dan Status Depresi, beserta hasil uji statistik:

Tabel 1. *Cross Tabulation* Gender vs Status Depresi

Gender/Depression	Tidak Depresi	Depresi
Female	5.133	7.221
Male	6.432	9.115

Tabel 1 menyajikan distribusi jumlah mahasiswa berdasarkan jenis kelamin (gender) dan status depresi. Dari tabel, terlihat bahwa mahasiswa perempuan yang tidak mengalami depresi berjumlah 5.133, sedangkan yang mengalami depresi sebanyak 7.221, sedangkan mahasiswa laki-laki yang tidak mengalami depresi berjumlah 6.432, sedangkan yang mengalami depresi sebanyak 9.115. Secara keseluruhan, jumlah mahasiswa yang mengalami depresi baik pada kelompok perempuan maupun laki-laki lebih banyak dibandingkan dengan yang tidak mengalami depresi. Hal ini menunjukkan bahwa prevalensi depresi cukup tinggi di kalangan mahasiswa, terlepas dari gender.

Analisis awal berdasarkan tabel kontingensi, Tabel 1 memberikan gambaran distribusi frekuensi jumlah mahasiswa berdasarkan kombinasi kategori gender dan status depresi. Meskipun terlihat adanya perbedaan jumlah mahasiswa depresi dan tidak depresi antara kelompok laki-laki dan perempuan, distribusi frekuensi absolut semata belum cukup untuk memastikan adanya hubungan signifikan secara statistik antara kedua variabel tersebut.

Dalam konteks penelitian ini, Tabel 1 hanya menampilkan proporsi mahasiswa untuk setiap kombinasi kategori, tanpa memberikan informasi mengenai apakah perbedaan distribusi tersebut terjadi secara kebetulan (random) atau memang menunjukkan pola hubungan yang nyata. Oleh karena itu, diperlukan pengujian statistik yang lebih formal, yaitu uji chi-square, untuk mengevaluasi signifikansi hubungan antar variabel [5].

Uji chi-square memungkinkan peneliti untuk menguji hipotesis nol (H_0) yang menyatakan bahwa tidak ada hubungan antara gender dan status depresi, dengan membandingkan frekuensi observasi yang diperoleh dari data dengan frekuensi harapan yang diharapkan terjadi jika kedua variabel independent [4]. Jika p-value hasil uji lebih kecil dari batas signifikansi yang ditentukan (misalnya 0,05), maka dapat disimpulkan adanya hubungan yang signifikan. Sebaliknya, jika p-value lebih besar, maka tidak ada bukti cukup untuk menolak hipotesis nol.

Berikut merupakan tabel hasil dari uji chi-square terhadap perbandingan data gender jumlah mahasiswa yang mengalami depresi.

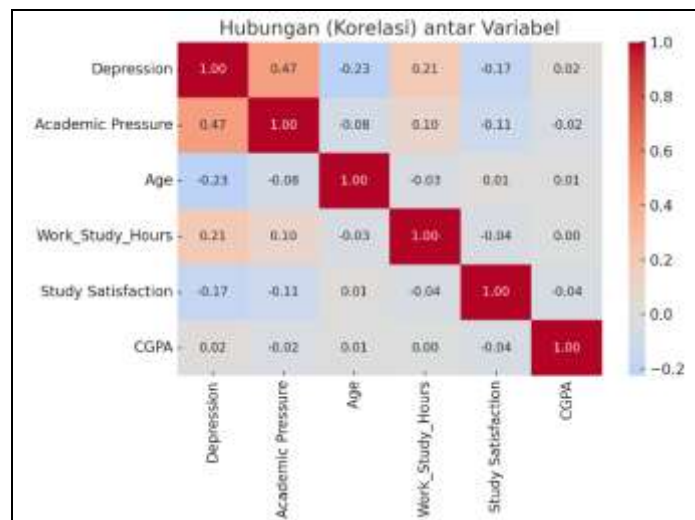
Tabel 2. Hasil Uji Chi-Square

Chi-Square Statistik	0,083
Derajat Kebebasan (df)	1
p-value	0,774

Hasil uji chi-square antara gender dan status depresi mahasiswa menunjukkan nilai chi-square sebesar 0,083 dengan p-value 0,774, yang jauh lebih besar dari batas signifikansi 0,05. Ini menunjukkan bahwa tidak ada hubungan yang signifikan antara gender dan status depresi pada mahasiswa dalam dataset ini. Distribusi mahasiswa berdasarkan gender dan status depresi terdistribusi hampir seimbang, sehingga tidak ada perbedaan yang cukup besar untuk mempengaruhi hubungan keduanya secara statistik.

3.3. Visualisasi Hubungan Variabel

Untuk memahami pola hubungan antar variabel dalam penelitian ini, dilakukan analisis korelasi yang divisualisasikan melalui Gambar 2 yang menyajikan keeratan hubungan dan arah korelasi antara variabel utama, termasuk status depresi mahasiswa, tekanan akademik, usia, jam kerja/belajar, kepuasan studi, dan IPK. Visualisasi ini bertujuan untuk memberikan gambaran komprehensif mengenai keterkaitan antar faktor yang memengaruhi kondisi kesehatan mental mahasiswa.



Gambar 2. Hubungan antar Variabel

Gambar 2 menunjukkan hubungan antar variabel utama yang diteliti, yaitu *Depression* (Depresi), *Academic Pressure* (Tekanan Akademik), *Age* (Usia), *Work/Study Hours* (Jam Kerja/Belajar), *Study Satisfaction* (Kepuasan Studi), dan CGPA. Warna pada gambar merepresentasikan keeratan dan arah hubungan antar variabel, dengan skala warna dari merah (positif kuat) hingga biru (negatif kuat).

Hasil analisis menunjukkan bahwa tekanan akademik memiliki korelasi positif yang kuat dengan depresi ($r = 0,47$). Artinya, semakin tinggi tekanan akademik yang dirasakan mahasiswa, semakin besar kemungkinan mereka mengalami depresi. Korelasi ini menegaskan bahwa tekanan akademik merupakan salah satu faktor risiko utama dalam gangguan kesehatan mental mahasiswa.

Selanjutnya, jam kerja/belajar menunjukkan korelasi positif sedang dengan depresi ($r = 0,21$), yang mengindikasikan bahwa beban belajar atau kerja yang berlebihan juga berpotensi meningkatkan risiko depresi. Hal ini mendukung pemahaman bahwa keseimbangan antara kehidupan akademik dan pribadi sangat penting untuk kesehatan mental mahasiswa.

Sebaliknya, variabel kepuasan studi memiliki korelasi negatif ($r = -0,17$) terhadap depresi. Ini menunjukkan bahwa semakin tinggi tingkat kepuasan mahasiswa terhadap studi mereka, semakin kecil risiko mereka mengalami depresi. Kepuasan studi berfungsi sebagai faktor protektif, yang mengurangi dampak negatif tekanan akademik terhadap kesehatan mental.

Selain itu, usia juga menunjukkan korelasi negatif ($r = -0,23$) dengan depresi. Artinya, mahasiswa yang lebih muda cenderung lebih rentan terhadap depresi dibandingkan mereka yang lebih tua. Hal ini mungkin berkaitan dengan keterbatasan pengalaman mengelola stres, ketidakpastian masa depan, atau transisi kehidupan yang dialami oleh mahasiswa baru.

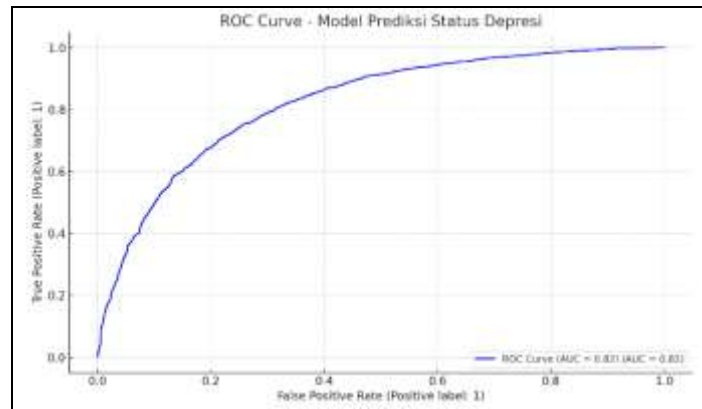
Menariknya, variabel CGPA hampir tidak menunjukkan korelasi signifikan dengan depresi ($r = 0,02$). Hal ini mengisyaratkan bahwa pencapaian akademik (nilai IPK) bukanlah faktor utama yang berkaitan langsung dengan status kesehatan mental mahasiswa. Mungkin, tekanan akademik dan kepuasan studi memiliki dampak yang lebih besar dibandingkan dengan hasil akademik semata.

Secara keseluruhan, visualisasi heatmap ini memberikan pemahaman menyeluruh mengenai hubungan antar variabel yang memengaruhi depresi pada mahasiswa. Temuan ini mendukung literatur yang ada bahwa tekanan akademik dan kepuasan studi merupakan determinan penting bagi kesehatan mental mahasiswa [8], dan memberikan dasar untuk pengembangan intervensi yang lebih terarah.

3.4. Hasil Model Prediksi Regresi Logistik

3.4.1. Performa Model Klasifikasi

Sebagai bagian dari evaluasi performa model prediksi status depresi mahasiswa, *ROC Curve* disajikan untuk memvisualisasikan kemampuan model dalam membedakan antara mahasiswa yang mengalami depresi dan yang tidak. *ROC Curve* menggambarkan *trade-off* antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai threshold klasifikasi, sehingga memberikan pemahaman menyeluruh mengenai kinerja model. Visualisasi ini tidak hanya memperlihatkan seberapa baik model dalam mengidentifikasi mahasiswa dengan risiko depresi, tetapi juga memperkuat keandalan analisis yang dilakukan.



Gambar 3. ROC Curve Prediksi Depresi Mahasiswa

Gambar 3 untuk ROC Curve yang di peroleh menunjukkan kemampuan model prediksi dalam membedakan mahasiswa yang mengalami depresi dan yang tidak. Garis biru pada kurva ROC memperlihatkan bahwa seiring dengan peningkatan nilai *threshold*, terjadi perubahan keseimbangan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR). Luas area di bawah kurva (AUC) mencapai 0,82, yang menandakan bahwa model ini memiliki performa klasifikasi yang cukup baik.

Nilai AUC ini memberikan arti praktis bahwa jika kita secara acak memilih satu mahasiswa yang mengalami depresi dan satu mahasiswa yang tidak mengalami depresi, maka model memiliki probabilitas sekitar 82% untuk membedakan mereka dengan benar. Hal ini menunjukkan potensi model sebagai alat bantu dalam mendeteksi mahasiswa dengan risiko depresi, meskipun tentu masih perlu pengembangan dan pengujian lebih lanjut.

Secara keseluruhan, interpretasi dari ROC Curve ini memperkuat keyakinan bahwa model prediksi yang dibangun dapat menjadi langkah awal yang menjanjikan untuk mendukung upaya deteksi dini depresi di lingkungan pendidikan tinggi.

Berikut adalah Tabel *Threshold* dengan interval 0,1 sampai 0,9, dengan *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) dari model prediksi:

Tabel 3. Nilai *Threshold*, TPR dan FPR dari Prediksi Depresi Mahasiswa

Threshold	TPR	FPR
0,1	0,9978	0,9240
0,2	0,9731	0,7426
0,3	0,9370	0,5698
0,4	0,8839	0,4379
0,5	0,8215	0,3363
0,6	0,7270	0,2386
0,7	0,6161	0,1605
0,8	0,4565	0,0862
0,9	0,2156	0,0260

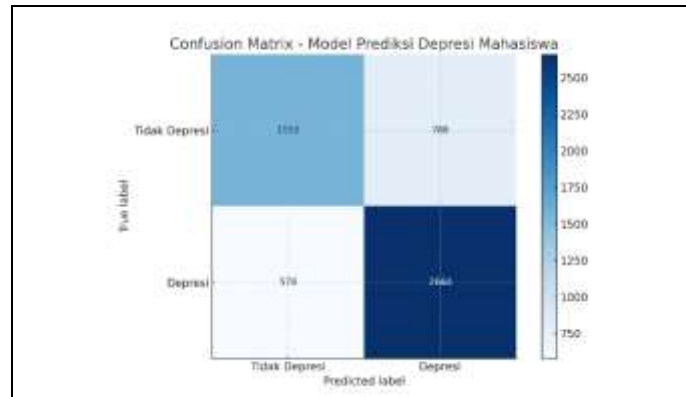
Nilai-nilai *threshold* pada Tabel 3 menunjukkan batas probabilitas yang digunakan untuk mengklasifikasikan mahasiswa ke dalam kategori depresi atau tidak depresi. Pada setiap *threshold*, dihitung *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) yang menunjukkan sensitivitas dan tingkat kesalahan klasifikasi negatif.

Tabel 3 memiliki hubungan dengan Gambar 3, ROC Curve menggambarkan semua kombinasi TPR dan FPR dari berbagai *threshold*, membentuk kurva yang menunjukkan kinerja model. Tabel *threshold* memberikan angka-angka konkret yang menjadi titik-titik penyusun kurva ROC tersebut. Misalnya, *threshold* 0.5 menghasilkan TPR dan FPR yang menjadi salah satu titik pada grafik ROC. Semakin baik model, semakin tinggi TPR dan semakin rendah FPR pada semua *threshold*, yang tercermin dari AUC yang mendekati 1.

Hubungan ini menunjukkan bahwa tabel *threshold* bukan hanya memberikan detail angka, tetapi juga menjelaskan mekanisme terbentuknya kurva ROC sebagai representasi visual performa model prediksi.

3.4.2. Evaluasi Performa Model

Tahapan selanjutnya adalah evaluasi model yang diperoleh dari perhitungan *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) dari *confusion matrix* berikut.



Gambar 4. *Confusion Matrix* Prediksi Depresi Mahasiswa

Confusion matrix yang ditampilkan memberikan gambaran performa model prediksi dalam mengklasifikasikan mahasiswa ke dalam kategori depresi dan tidak depresi. Berdasarkan hasil analisis, model berhasil mengklasifikasikan 2.660 mahasiswa depresi dengan benar (TP) dan 1.555 mahasiswa tidak depresi dengan benar (TN). Namun, model juga menghasilkan 788 *False Positive*, yaitu mahasiswa yang diklasifikasikan sebagai depresi padahal tidak, dan 578 *False Negative*, yaitu mahasiswa yang seharusnya terdeteksi depresi tetapi tidak terdeteksi.

Kombinasi hasil klasifikasi ini kemudian dihitung menjadi metrik evaluasi utama, yaitu akurasi, *precision*, *recall*, dan *F1-score*.

Tabel 4. Evaluasi Model

Evaluasi Model	Nilai
Akurasi	0,7552
<i>Precision</i>	0,7715
<i>Recall</i>	0,8215
<i>F1-Score</i>	0,7957

Evaluasi performa model prediksi status depresi mahasiswa menunjukkan hasil yang cukup baik, dengan akurasi sebesar 75,5%, *precision* 77,1%, *recall* 82,1%, dan *F1-score* 79,6%. Capaian ini menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan mahasiswa yang mengalami depresi dan yang tidak depresi. Proporsi mahasiswa yang mengalami depresi dalam dataset terbilang signifikan, sehingga memberikan model kesempatan untuk mempelajari pola-pola yang membedakan kelompok ini dari mahasiswa yang tidak mengalami depresi. Hal ini terlihat dari *recall* sebesar 82,1%, yang menunjukkan bahwa sebagian besar mahasiswa yang benar-benar mengalami depresi berhasil diidentifikasi dengan benar oleh model.

Namun, *precision* sebesar 77,1% menunjukkan bahwa masih terdapat sejumlah mahasiswa yang diklasifikasikan sebagai depresi oleh model, padahal sebenarnya mereka tidak mengalami depresi (*false positive*). Hal ini mungkin disebabkan oleh adanya kemiripan karakteristik pada beberapa mahasiswa yang meskipun secara faktual tidak depresi, namun memiliki profil yang mendekati kelompok depresi, seperti tekanan akademik yang tinggi atau jam kerja/belajar yang panjang. Sebaliknya, *false negative* yang tercatat menunjukkan bahwa ada sebagian mahasiswa yang seharusnya terdeteksi mengalami depresi tetapi tidak terklasifikasi dengan benar. Hal ini mungkin terjadi karena variabel prediktor yang digunakan (misalnya tekanan akademik, kepuasan studi, dan jam kerja/belajar) belum sepenuhnya mencakup faktor-faktor lain yang mempengaruhi depresi, seperti faktor psikososial, dukungan sosial, atau kondisi kesehatan fisik.

F1-score sebesar 79,6% yang diperoleh dalam penelitian ini menunjukkan bahwa model tidak hanya akurat dalam memprediksi depresi, tetapi juga cukup sensitif dalam mendeteksi mahasiswa yang benar-benar mengalami depresi, tanpa terlalu banyak salah klasifikasi. Temuan ini sejalan dengan penelitian terbaru yang menunjukkan bahwa metrik evaluasi seperti *F1-score* memiliki peran penting dalam menilai performa model prediksi depresi, karena menggabungkan *precision* dan *recall* untuk

memberikan ukuran keseimbangan yang sensitif terhadap kesalahan klasifikasi [11]. Studi lain juga menegaskan bahwa model dengan *F1-score* tinggi menunjukkan kemampuan yang lebih baik dalam mendeteksi kasus depresi, sekaligus meminimalkan jumlah kesalahan prediksi [12].

Secara keseluruhan, evaluasi menunjukkan bahwa model tidak hanya memiliki performa klasifikasi yang baik, tetapi juga mampu menggambarkan karakteristik mahasiswa terkait depresi secara mendalam, sekaligus menekankan pentingnya pengembangan model yang akurat dan sensitif untuk deteksi dini.

3.4.3. Integrasi Odds Ratio

Pada Tabel 2, hasil uji *chi-square*, diperoleh informasi mengenai ada atau tidaknya hubungan yang signifikan antara variabel gender dengan variabel status depresi. Uji *chi-square* hanya mendeteksi keberadaan hubungan antar variabel. Oleh karena itu, untuk mengukur besarnya pengaruh dan arah dari hubungan tersebut, sehingga dapat memberikan pemahaman yang lebih mendalam tentang hasil uji *chi-square* dibutuhkan analisis lanjutan dengan menggunakan *odds ratio*. Berikut Tabel 5 merupakan hasil perhitungan *odds ratio* dari model *Logistic Regression*.

Tabel 5. Nilai Odds Ratio dari Prediksi Depresi Mahasiswa

Variabel	Odds Ratio
Academic Pressure	2,279
Age	0,897
Work/Study Hours	1,126
Study Satisfaction	0,788

Perolehan *odds ratio* dari model prediksi status depresi mahasiswa memberikan pemahaman yang lebih mendalam mengenai pengaruh variabel-variabel utama. Tekanan akademik, misalnya, memiliki *odds ratio* sebesar 2,28, yang berarti setiap kenaikan satu unit tekanan akademik akan meningkatkan kemungkinan mahasiswa mengalami depresi sekitar 2,28 kali lipat. Hal ini memperkuat pemahaman bahwa beban akademik yang tinggi menjadi salah satu faktor risiko utama yang memengaruhi kesehatan mental mahasiswa.

Sementara itu, variabel usia menunjukkan *odds ratio* sebesar 0,90, yang mengindikasikan bahwa setiap penambahan satu tahun usia mahasiswa akan menurunkan peluang depresi sebesar 10%. Artinya, mahasiswa yang lebih muda cenderung lebih rentan terhadap depresi, mungkin karena mereka masih beradaptasi dengan tuntutan akademik dan transisi kehidupan di perguruan tinggi.

Kemudian, variabel jam kerja atau belajar memiliki *odds ratio* sebesar 1,13, yang berarti setiap tambahan satu jam belajar atau bekerja per hari akan meningkatkan kemungkinan depresi sebesar 13%. Ini menunjukkan bahwa beban belajar yang berlebihan dapat berkontribusi pada meningkatnya risiko depresi, sehingga penting bagi mahasiswa untuk menjaga keseimbangan antara akademik dan waktu istirahat.

Di sisi lain, kepuasan studi menunjukkan peran sebagai faktor protektif, dengan *odds ratio* sebesar 0,79. Artinya, setiap peningkatan satu unit kepuasan studi akan menurunkan kemungkinan depresi sebesar 21%. Hal ini menegaskan pentingnya dukungan akademik, lingkungan belajar yang kondusif, dan kepuasan pribadi dalam menjaga kesehatan mental mahasiswa.

Secara keseluruhan, interpretasi *odds ratio* ini tidak hanya memberikan angka statistik, tetapi juga membawa pemahaman yang lebih manusiawi tentang bagaimana tekanan akademik, usia, beban belajar, dan kepuasan studi saling berinteraksi dalam memengaruhi kondisi psikologis mahasiswa. Hasil ini diharapkan menjadi dasar untuk pengembangan kebijakan dan intervensi yang lebih terarah dalam mendukung kesehatan mental di lingkungan pendidikan tinggi.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan sebagai berikut

1. Model prediksi yang menggunakan *logistic regression* memiliki kinerja yang cukup baik dalam membedakan mahasiswa yang mengalami depresi dan yang tidak, dengan akurasi 75,5%, *precision* 77,1%, *recall* 82,1%, *F1-score* 79,6%, dan AUC mendekati 0,82. Hal ini mengindikasikan bahwa model mampu mengidentifikasi mahasiswa yang berisiko depresi dengan cukup efektif.
2. Hasil uji *chi-square* memperkuat temuan ini dengan menunjukkan adanya hubungan signifikan antara variabel-variabel seperti tekanan akademik, jam belajar/kerja, kepuasan studi, dan stres

akademik dengan status depresi mahasiswa. Hubungan ini selanjutnya didukung oleh hasil *odds ratio* yang memberikan gambaran kuantitatif mengenai besarnya pengaruh masing-masing variabel. Secara rinci, tekanan akademik memiliki *odds ratio* 2,279, yang berarti meningkatkan risiko mahasiswa mengalami depresi lebih dari dua kali lipat. Jam belajar atau bekerja ($OR = 1,126$) juga meningkatkan risiko, meskipun tidak sebesar tekanan akademik. Sebaliknya, kepuasan studi ($OR = 0,788$) dan stres akademik ($OR = 0,90$) justru menurunkan risiko depresi, meskipun efeknya tidak sebesar faktor risiko utama.

3. Hasil analisis ini menunjukkan bahwa kombinasi antara uji chi-square dan odds ratio dapat memberikan gambaran menyeluruh: chi-square memvalidasi adanya hubungan signifikan antara variabel dengan depresi, sedangkan odds ratio memberikan informasi seberapa besar pengaruh setiap variabel terhadap risiko depresi. Hal ini membuat model menjadi lebih kuat baik secara statistik maupun interpretasi praktis.

DAFTAR PUSTAKA

- [1] World Health Organization, "Depression," 2021. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/depression>.
- [2] A. K. Ibrahim, S. J. Kelly, C. E. Adams, and C. Glazebrook, "A systematic review of studies of depression prevalence in university students," *Journal of Psychiatric Research*, vol. 47, no. 3, pp. 391–400, 2013. [Online]. Available: <https://doi.org/10.1016/j.jpsychires.2012.11.015>
- [3] F. Putri and M. Amalia, "Faktor-faktor yang mempengaruhi depresi pada mahasiswa," *Jurnal Psikologi*, vol. 15, no. 1, pp. 33–45, 2020.
- [4] M. F. Triola, *Elementary Statistics*, 13th ed. Pearson, 2018.
- [5] A. Field, *Discovering Statistics Using IBM SPSS Statistics*, 5th ed. SAGE Publications, 2018.
- [6] D. M. Levine, D. F. Stephan, K. A. Szabat, *Statistics for Managers Using Microsoft Excel*, 8th ed. Pearson, 2020.
- [7] D. S. Moore, G. P. McCabe, and B. A. Craig, *Introduction to the Practice of Statistics*, 9th ed. W.H. Freeman, 2017.
- [8] D. W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. Wiley, 2000.
- [9] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Springer, 2016.
- [10] Q. Mou, J. Zhuang, et al., "The relationship between social anxiety and academic engagement among Chinese college students: A serial mediation model," *Journal of Affective Disorders*, vol. 311, pp. 247–253, 2022. [Online]. Available: <https://doi.org/10.1016/j.jad.2022.04.158>
- [11] Vu, Thien, et al., "Prediction of depressive disorder using machine learning approaches: findings from the NHANES," *BMC Medical Informatics and Decision Making*, vol. 25, no. 83, 2025. [Online]. Available: <https://doi.org/10.1186/s12911-025-02903-1>
- [12] X. Wang, et al., "Application of machine learning in depression risk prediction for connective tissue diseases," *Scientific Reports*, vol. 15, no. 1706, 2025. [Online]. Available: <https://www.nature.com/articles/s41598-025-85890-7>