

Tugas Akhir Data Science

Students' Social Media Addiction

Presented By:
Kelompok 9



Our Team



M. Ramadhani S.
(2012210016)



Salwana Billa
(2012210707)

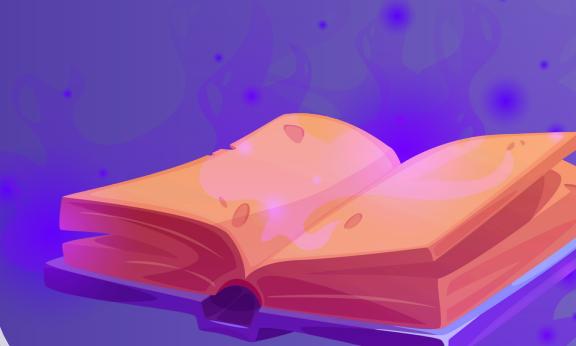


M. Fahrizal A. P.
(2012310502)



Outline

**Study
Case**



Pengolahan



Kesimpulan





Study Case

Dataset Media Sosial & Hubungan Mahasiswa berisi catatan anonim tentang perilaku media sosial mahasiswa dan hasil kehidupan terkait. Data ini mencakup berbagai negara dan jenjang akademik, dengan fokus pada dimensi-dimensi utama seperti intensitas penggunaan, preferensi platform, dan dinamika hubungan. Setiap baris mewakili respons survei satu mahasiswa, menawarkan gambaran lintas sektor yang cocok untuk analisis statistik dan aplikasi pembelajaran mesin.

Dataset berisi variabel:

- Student_ID
- Age
- Gender
- Academic_LevelCountry
- Avg_Daily_Usage_Hours
- Most_Used_Platform
- Affects_Academic_Performance
- Sleep_Hours_Per_Night
- Mental_Health_Score
- Relationship_Status
- Conflicts_Over_Social_Media
- Addicted_Score



Tujuan



Memahami Karakteristik Penggunaan Media Sosial

Menggali pola penggunaan media sosial siswa berdasarkan demografi (usia, gender, tingkat akademik, negara) dan kebiasaan (rata-rata jam penggunaan harian, platform paling sering digunakan).



Menilai Dampak Penggunaan Media Sosial

Menyelidiki hubungan antara penggunaan media sosial dengan aspek lain kehidupan siswa, seperti kinerja akademik, pola tidur, kesehatan mental, dan hubungan sosial (status hubungan, konflik).



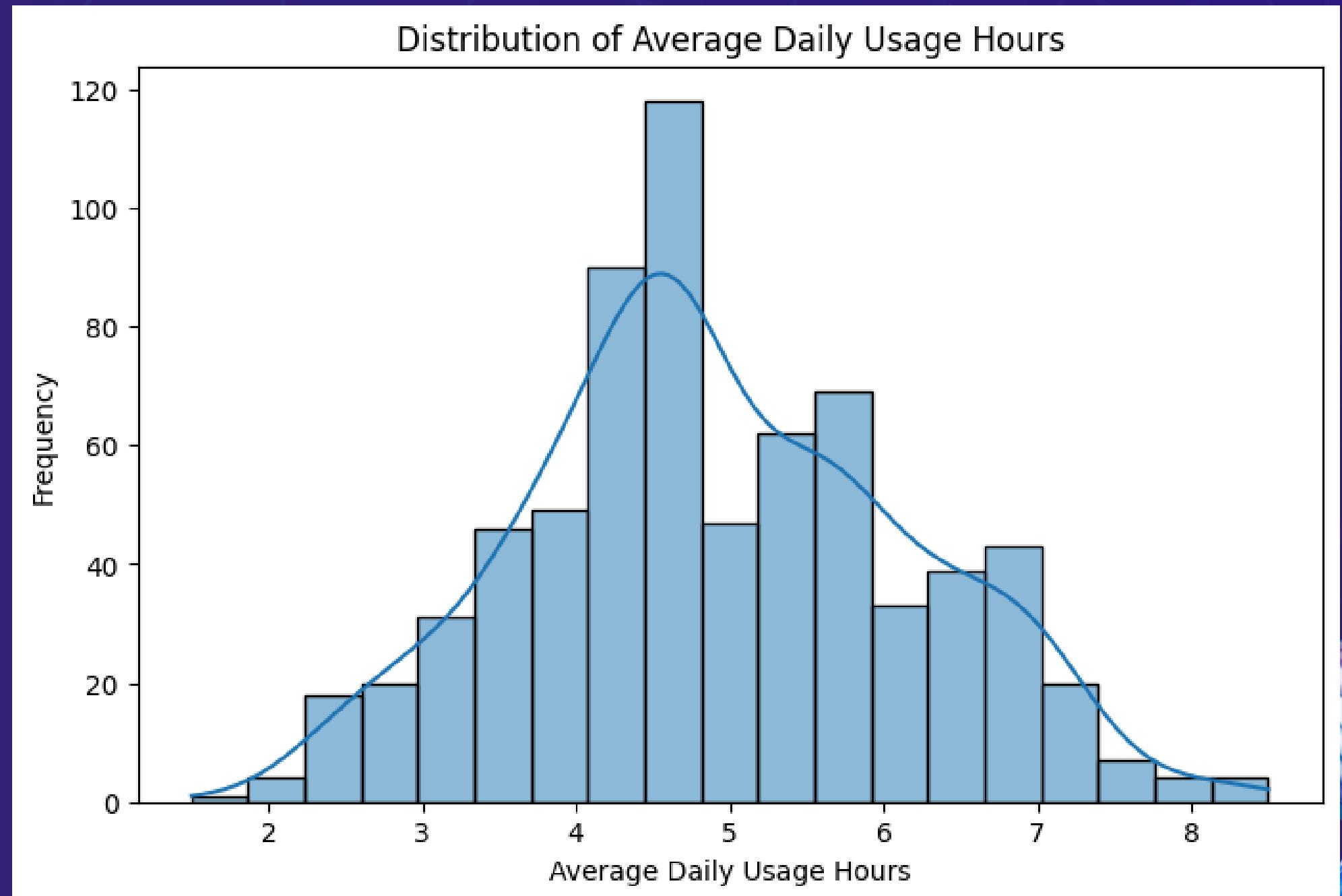
Membangun Model Prediktif

Mengembangkan model machine learning (dalam hal ini, Random Forest Regressor) yang dapat memprediksi skor kecanduan media sosial berdasarkan fitur-fitur yang tersedia dalam dataset.

Explanatory Data Analysis (EDA)

Dari plot tersebut, kita dapat mengamati:

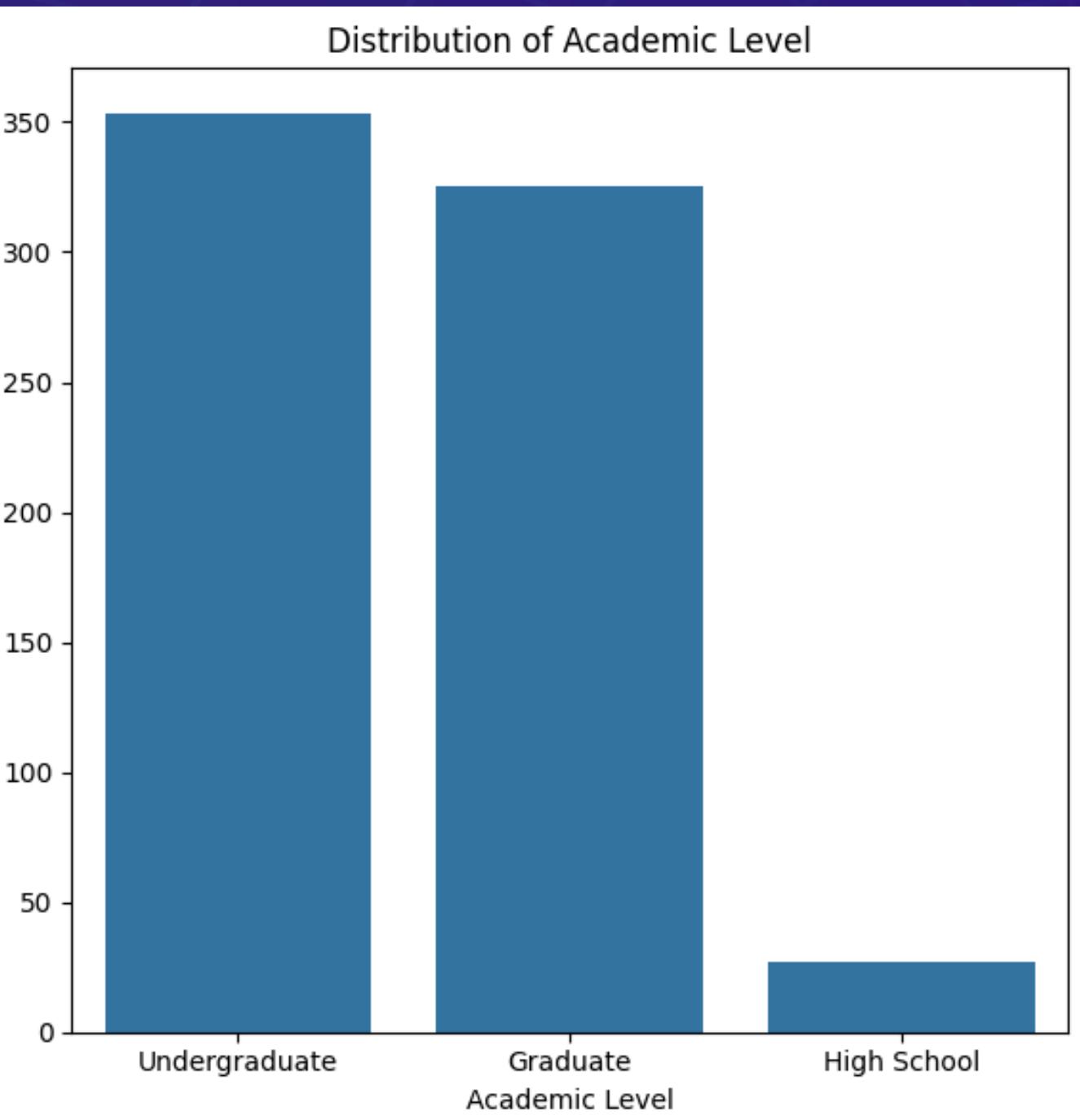
- Mayoritas siswa tampaknya menghabiskan waktu antara sekitar 2 hingga 6 jam per hari di media sosial, karena puncak distribusi berada dalam rentang ini.
- Distribusi terlihat agak miring ke kanan (positively skewed), yang berarti ada sejumlah siswa yang menghabiskan waktu jauh lebih lama di media sosial dibandingkan mayoritas siswa. Ekor distribusi memanjang ke arah nilai yang lebih tinggi.
- Ada frekuensi yang lebih rendah pada penggunaan jam yang sangat rendah atau sangat tinggi.



Explanatory Data Analysis (EDA)

Kesimpulan dari Plot Ini:

Plot distribusi tingkat akademik memberikan gambaran tentang komposisi sampel siswa dan mahasiswa dalam dataset berdasarkan jenjang pendidikan mereka. Informasi ini penting untuk memahami konteks data dan dapat digunakan nanti dalam analisis lebih lanjut untuk melihat apakah ada perbedaan pola penggunaan media sosial atau skor kecanduan di antara siswa dari tingkat akademik yang berbeda. Mayoritas data berasal dari siswa jenjang universitas (Undergraduate dan Graduate).

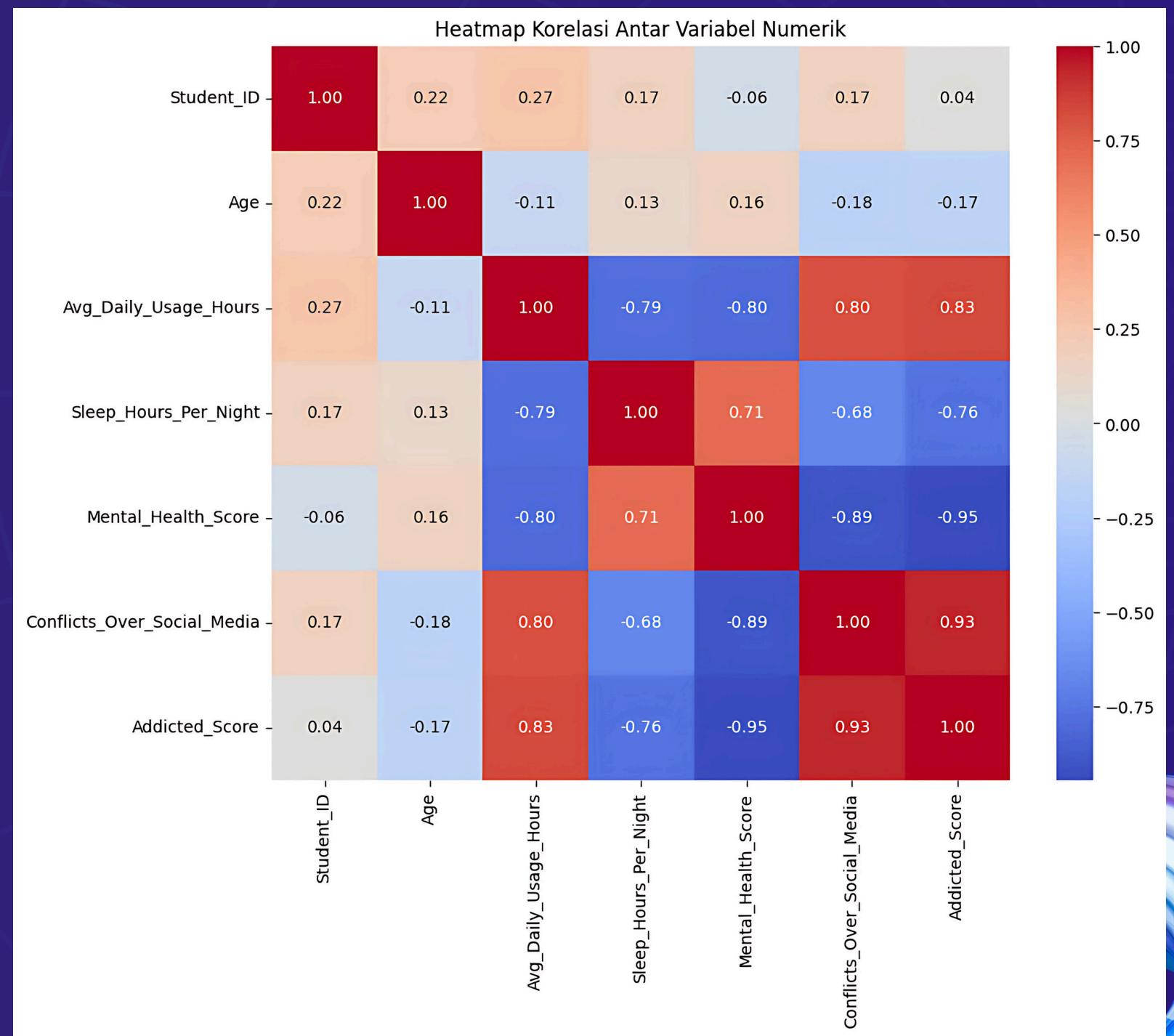


Explanatory Data Analysis (EDA)

Pengamatan dari Heatmap:

Terlihat korelasi positif yang kuat dengan beberapa variabel:

- **Avg_Daily_Usage_Hours**: Ini adalah korelasi yang paling diharapkan. Semakin tinggi rata-rata jam penggunaan harian, semakin tinggi skor kecanduan.
- **Conflicts_Over_Social_Media**: Semakin banyak konflik terkait media sosial, semakin tinggi skor kecanduan.
- **Mental_Health_Score**: Ada korelasi negatif sedang dengan skor kesehatan mental (semakin tinggi skor kecanduan, semakin rendah skor kesehatan mental), meskipun korelasinya tidak sekuat dua variabel sebelumnya.
- **Sleep_Hours_Per_Night**: Ada korelasi negatif sedang dengan jam tidur (semakin tinggi skor kecanduan, semakin sedikit jam tidur).



Preprocessing

- **Hasil:** Ditemukan bahwa tidak ada nilai yang hilang (missing values) di seluruh kolom dataset (`df.isnull().sum()` menunjukkan 0 untuk semua kolom).
- **Intrepetasi:** Ini adalah hasil yang ideal. Karena tidak perlu melakukan langkah-langkah penanganan nilai yang hilang seperti imputasi atau penghapusan baris/kolom. Kualitas data awal cukup baik terkait kelengkapan.

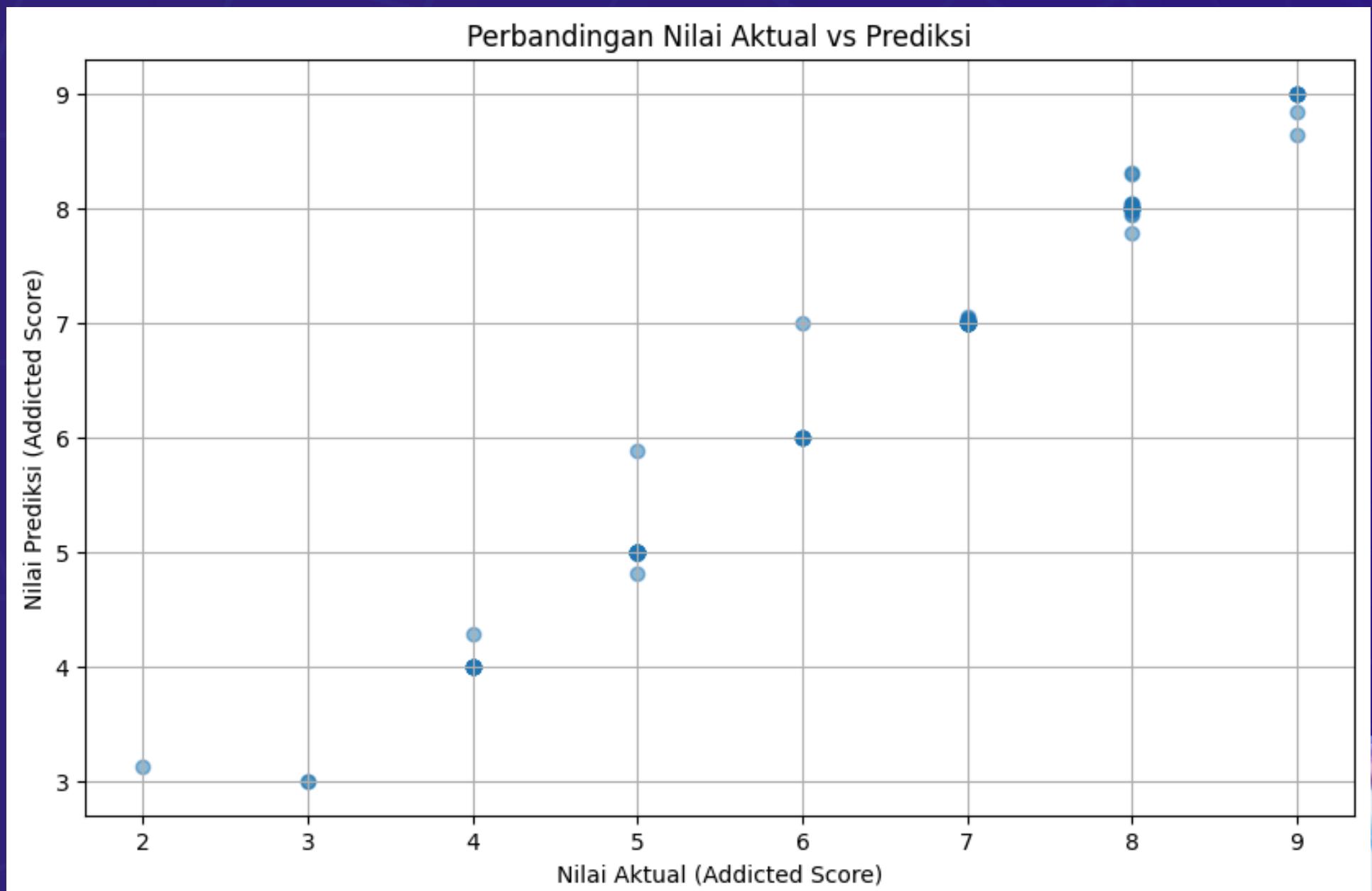
```
▶ print("Jumlah nilai yang hilang di setiap kolom:")
print(df.isnull().sum())

→ Jumlah nilai yang hilang di setiap kolom:
Student_ID           0
Age                  0
Gender               0
Academic_Level      0
Country              0
Avg_Daily_Usage_Hours 0
Most_Used_Platform   0
Affects_Academic_Performance 0
Sleep_Hours_Per_Night 0
Mental_Health_Score   0
Relationship_Status   0
Conflicts_Over_Social_Media 0
Addicted_Score        0
dtype: int64
```

Modelling

Kesimpulan dari Plot Ini:

- Scatter plot perbandingan nilai aktual vs prediksi secara visual mengkonfirmasi bahwa model Random Forest Regressor yang diolah memiliki kemampuan prediksi yang sangat kuat pada data pengujian. Dibuktikan dengan titik-titik yang rapat di sekitar garis ideal menunjukkan bahwa model berhasil menangkap hubungan antara fitur-fitur input dan skor kecanduan media sosial dengan akurasi yang tinggi.
- Analisis ini berhasil mengidentifikasi beberapa faktor kunci yang sangat terkait dengan skor kecanduan media sosial pada siswa, terutama durasi penggunaan harian dan dampak pada hubungan sosial. Model Random Forest Regressor yang dibangun terbukti sangat efektif dalam memprediksi skor kecanduan berdasarkan fitur-fitur yang tersedia.



Modelling

Mean Absolute Error (MAE): 0.04

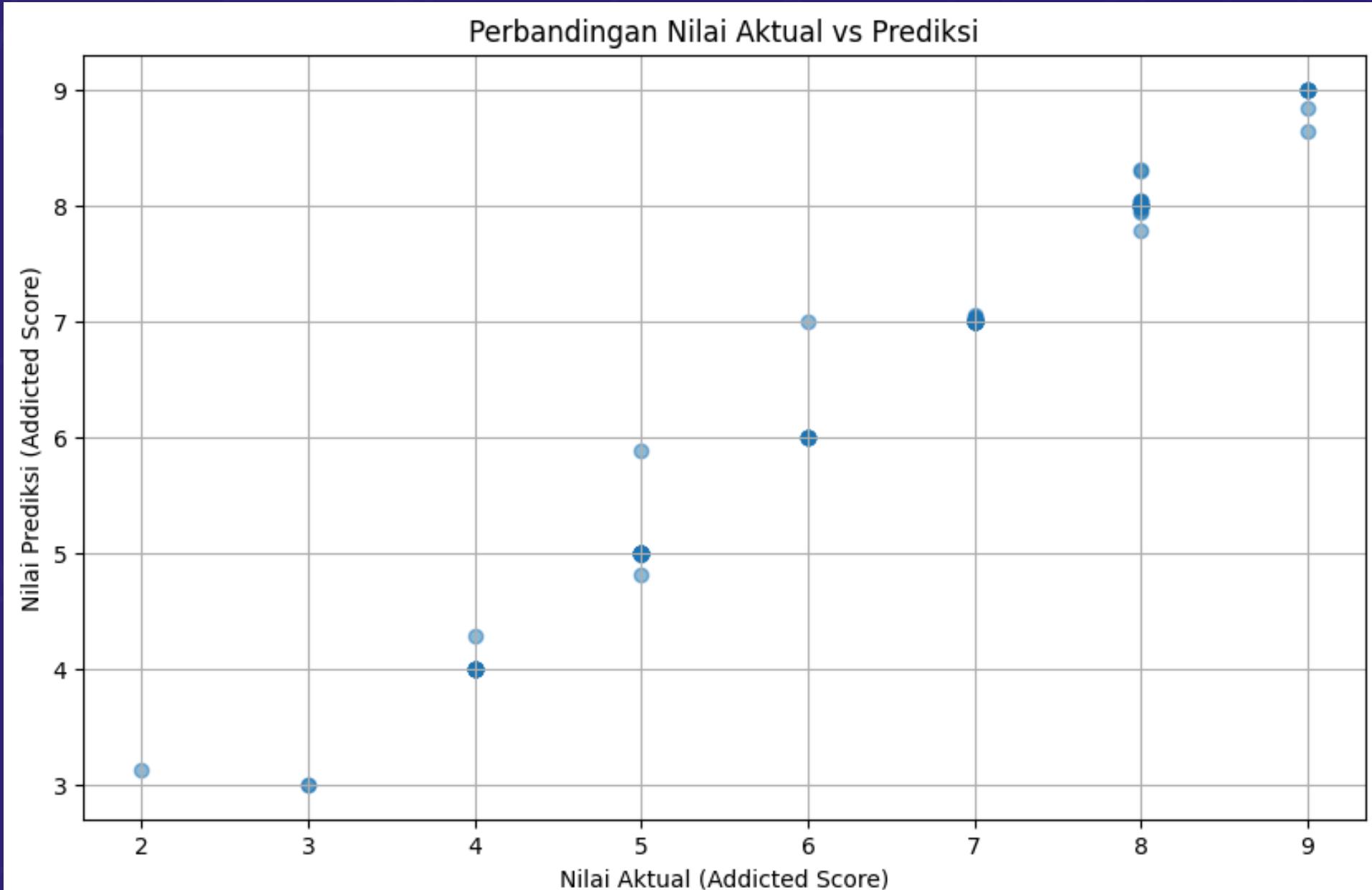
Mean Squared Error (MSE): 0.03

Root Mean Squared Error (RMSE): 0.16

R-squared (R²): 0.99

Nilai R-squared = 0.99 (sangat tinggi), menunjukkan bahwa model menjelaskan 99% variabilitas dalam Addicted_Score, yang mengindikasikan kinerja prediksi yang sangat baik.

Nilai MAE dan RMSE yang rendah juga mengkonfirmasi bahwa prediksi model sangat dekat dengan nilai aktual.



Kesimpulan

1. Dataset ini memiliki kualitas yang baik. Tidak ada nilai yang hilang sama sekali, dan pemeriksaan outlier menggunakan Z-score juga menunjukkan tidak adanya nilai ekstrem yang signifikan pada kolom numerik.
2. Analisis distribusi rata-rata jam penggunaan harian menunjukkan bahwa mayoritas siswa menghabiskan waktu antara 2 hingga 6 jam per hari di media sosial, meskipun ada sebagian siswa yang menggunakan media sosial jauh lebih lama (ditunjukkan oleh kemiringan distribusi ke kanan).
3. Rata-rata jam penggunaan media sosial harian (Avg_Daily_Usage_Hours) dan jumlah konflik terkait media sosial (Conflicts_Over_Social_Media) memiliki korelasi positif yang kuat dengan skor kecanduan media sosial (Addicted_Score).
4. Terdapat juga korelasi negatif moderat antara skor kecanduan dengan skor kesehatan mental (Mental_Health_Score) dan jam tidur per malam (Sleep_Hours_Per_Night). Ini mengindikasikan bahwa siswa dengan skor kecanduan yang lebih tinggi cenderung memiliki kesehatan mental yang lebih buruk dan jam tidur yang lebih sedikit.
5. Model Random Forest Regressor yang dibangun terbukti sangat efektif dalam memprediksi skor kecanduan berdasarkan fitur-fitur yang tersedia.

Thank You