

PANDUAN PROJECT AKHIR PENAMBANGAN DATA

Semester Genap 2024/2025

TAHAPAN PEMAPARAN IDE PROJECT

- Mekanisme pemaparan hasil project dilakukan sebagai berikut:
 - a) Pemaparan ide project :
 - Memaparkan ide project dan tanya jawab melalui presentasi offline di perkuliahan minggu ke-12 dengan durasi maks 15 menit
- Pada tahapan ini mahasiswa diminta untuk menjelaskan kemajuan project yang memuat:
 - a) [Skor 20] Penjelasan ide secara umum
 - b) [Skor 20] Penjelasan diagram alur (berisi tahapan-tahapan yang dikerjakan)
 - c) [Skor 20] Penjelasan *pre-processing* yang akan dikerjakan
 - d) [Skor 20] Rencana output yang akan diperoleh dari project yang akan dikerjakan
 - e) [Skor 20] Referensi/publikasi acuan yang digunakan (prosiding/jurnal internasional)

LAPORAN PROJECT PENAMBANGAN DATA SEMESTER GENAP 2024/2025

IDENTITAS PROYEK	
Judul	Prediksi Ranking QS WUR Perguruan Tinggi dan Faktor Penentunya
Topik	Pendidikan
Identitas Penyusun	1. Berlianti Debby Maharani (23031554006) 2. Nafila Hanum Al Hasaniy (23031554109) 3. Nisrina Afaf (23031554165) 4. Dafanov Dixie Einkinderen (23031554201)
Kelas	2023D

1. PENDAHULUAN

1.1. Latar Belakang (min. 250 kata)

QS World University Rankings (QS WUR) mencerminkan tren yang sedang berlangsung dalam pendidikan tinggi global, menekankan prestise kelembagaan, hasil penelitian, dan kolaborasi internasional yang dilakukan oleh lembaga pemeringkatan Quacquarelli Symonds (QS). Peringkat dipengaruhi oleh berbagai faktor, termasuk reputasi akademik, reputasi riwayat karir alumni, rasio fakultas/mahasiswa, dan kutipan per fakultas. Pendekatan komprehensif ini memungkinkan pemahaman berdasarkan kinerja universitas di berbagai wilayah. QS World University Rank menggunakan enam indikator utama untuk mengevaluasi universitas: reputasi akademik, reputasi pemberi kerja, rasio fakultas/mahasiswa, kutipan per fakultas, rasio fakultas internasional, dan rasio mahasiswa internasional[1][2]

Indikator-indikator ini dirancang untuk memberikan pandangan komprehensif tentang kinerja universitas, meskipun QS WUR telah dikritik karena subjektivitas dan fokus yang kecil – yang mungkin tidak menangkap semua aspek keunggulan universitas[2] *Sustainability practice* telah dikaitkan dengan peningkatan kinerja akademik dan peringkat, menunjukkan bahwa universitas yang terlibat dalam *sustainability initiative* dapat memperoleh keunggulan kompetitif [3] Universitas menggunakan peringkat QS untuk meningkatkan daya saing global mereka dengan menarik talenta top dan mendapatkan pendanaan. Peringkat tersebut mempengaruhi keputusan strategis yang berkaitan dengan kolaborasi internasional, perekrutan fakultas, dan pengembangan infrastruktur [4][5] Lembaga dengan peringkat yang lebih baik lebih berpotensi untuk mendapatkan hibah dan sumbangan. Hal ini semakin meningkatkan kemampuan dan reputasi mereka[6]

Berbagai model prediktif telah dikembangkan untuk meramalkan peringkat universitas berdasarkan data multivariate. Teknik seperti linear regression, random forest regression, dan XGBoost telah digunakan. XGBoost menunjukkan kinerja unggul dalam memprediksi peringkat [7] Penelitian ini memprediksi nilai rank di masa depan dan juga melihat faktor - faktor yang mempengaruhi perubahan rank tersebut dengan menggunakan algoritma linear regression, random forest regression, dan XGBoost.

1.2. Rumusan Masalah dan Tujuan

Rumusan Masalah :

1. Apa faktor penentu utama yang mempengaruhi variasi peringkat universitas berdasarkan QS World University Rankings (QS WUR)?
2. Sejauh mana indikator yang digunakan oleh QS WUR secara efektif merangkum kinerja akademik yang komprehensif dan keberlanjutan universitas?
3. Apakah metodologi prediktif seperti regresi linier, random forest, dan XGBoost dapat digunakan untuk memproyeksikan peringkat universitas prospektif?
4. Metodologi prediktif mana yang menghasilkan tingkat akurasi tertinggi dalam memperkirakan peringkat QS WUR berdasarkan kumpulan data multivariat?

Tujuan Proyek:

1. Menganalisa metrik yang digunakan dalam QS World University Rankings dan pengaruhnya terhadap fluktuasi peringkat universitas.
2. Menilai sejauh mana inisiatif keberlanjutan dapat memfasilitasi peningkatan peringkat universitas.
3. Mengembangkan dan mengevaluasi model prediktif (regresi linier, random forest, XGBoost) yang bertujuan untuk memprediksi peringkat QS WUR di masa depan.

4. Menentukan metodologi mana yang paling berpengaruh dalam menghasilkan tingkat akurasi tertinggi dalam peringkat QS WUR.

2. Metodologi

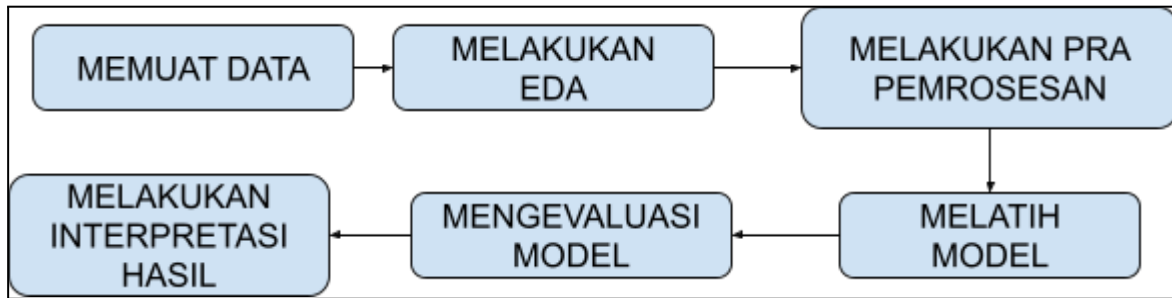
2.1. Eksplorasi Dataset

Pemahaman dataset yang dimiliki

Dataset :

1. **RANK_2025**: Peringkat keseluruhan universitas dalam QS World University Rankings tahun 2025
2. **RANK_2024**: Peringkat keseluruhan universitas dalam QS Rankings tahun 2024
3. **Institution_Name**: Nama universitas atau institusi
4. **Location**: Negara tempat institusi berada
5. **Region**: Wilayah global (misalnya: Eropa, Asia, Amerika Utara)
6. **SIZE**: Klasifikasi ukuran institusi (misalnya: S, M, L, XL)
7. **FOCUS**: Jenis fokus institusi (misalnya: Komprehensif, Terfokus)
8. **RES.**: Intensitas riset (misalnya: Sangat Tinggi, Tinggi)
9. **STATUS**: Status institusi (misalnya: Negeri, Swasta)
10. **Academic_Reputation_Score**: Skor berdasarkan survei reputasi akademik global
11. **Academic_Reputation_Rank**: Peringkat berdasarkan reputasi akademik
12. **Employer_Reputation_Score**: Skor berdasarkan survei reputasi dari pemberi kerja secara global
13. **Employer_Reputation_Rank**: Peringkat berdasarkan reputasi dari pemberi kerja
14. **Faculty_Student_Score**: Skor yang mencerminkan rasio dosen terhadap mahasiswa
15. **Faculty_Student_Rank**: Peringkat berdasarkan rasio dosen-mahasiswa
16. **Citations_per_Faculty_Score**: Skor yang mencerminkan dampak riset (jumlah sitasi per dosen)
17. **Citations_per_Faculty_Rank**: Peringkat berdasarkan jumlah sitasi per dosen
18. **International_Faculty_Score**: Skor yang merepresentasikan keragaman dosen internasional
19. **International_Faculty_Rank**: Peringkat berdasarkan keberadaan dosen internasional
20. **International_Students_Score**: Skor yang merepresentasikan keragaman mahasiswa internasional
21. **International_Students_Rank**: Peringkat berdasarkan rasio mahasiswa internasional
22. **International_Research_Network_Score**: Skor berdasarkan kolaborasi riset global
23. **International_Research_Network_Rank**: Peringkat berdasarkan kemitraan riset internasional
24. **Employment_Outcomes_Score**: Skor yang mencerminkan keberhasilan dan daya saing lulusan
25. **Employment_Outcomes_Rank**: Peringkat berdasarkan hasil kerja lulusan
26. **Sustainability_Score**: Skor yang mencerminkan inisiatif dan kinerja keberlanjutan
27. **Sustainability_Rank**: Peringkat berdasarkan ukuran keberlanjutan
28. **Overall_Score**: Skor komposit akhir yang digunakan untuk menentukan peringkat universitas

2.2. Langkah Penelitian



Gambar 1. Tahapan Mining

1. Pada penelitian ini, proses pertama adalah memuat data yang berbentuk Comma Separated Value (CSV) dari kaggle.
2. Kemudian lakukan Eksplorasi data untuk mengetahui isi dataset. Seperti melakukan cek apakah terdapat missing value pada data serta memvisualisasikan perbandingan antar 2 variabel tertentu.
3. Dalam tahap preprocessing, kami membersihkan data dengan menghapus nilai kosong pada kolom Overall_Score dan mengisi kolom Status dengan modus serta beberapa kolom skor dengan nilai rata-rata. Kolom-kolom peringkat yang kosong diisi dengan nilai 999. Selanjutnya, tipe data Rank_2024 dan Rank_2025 dikonversi menjadi numerik dan digunakan untuk membuat kolom Rank_Change. Nilai "Not Classified" pada kolom Region dikoreksi menjadi "Europe" untuk Eastern Mediterranean University. Data duplikat dicek, dan simbol seperti, "=" dan "+" dihapus dari kolom peringkat sebelum dikonversi menjadi numerik. Skor dinormalisasi menggunakan MinMaxScaler, dan peringkat dibalik skalanya agar searah dengan skor. Kolom kategorikal diubah ke format numerik menggunakan One Hot Encoding. Distribusi Rank_Change divisualisasikan dengan histogram, dan Overall_Score dipastikan bertipe numerik untuk analisis selanjutnya.
4. Dalam proses modelling ini, peneliti membersihkan data dari nilai hilang, memisahkan fitur dan target, lalu melakukan standarisasi fitur sebelum mereduksi dimensi menggunakan PCA untuk mempertahankan 90% varians. peneliti menganalisis kontribusi fitur terhadap tiap komponen utama PCA, kemudian membagi data menjadi train dan test. Tiga model regresi (Linear Regression, Random Forest, dan XGBoost) dilatih dan dievaluasi menggunakan R^2 dan MSE, serta divisualisasikan dengan grafik prediksi, residual, feature importance, dan partial dependence plot untuk interpretasi. Hasil evaluasi dibandingkan untuk menentukan model terbaik.
5. Proses terakhir adalah melakukan evaluasi terhadap model Menggunakan RMSE dan R^2 score

3. Hasil dan Analisis

3.1 Exploratory Data Analysis

1. Dataset yang Digunakan.

Dataset berisi berisi 28 kolom dan 1503 baris. Dari 1503 baris tersebut kemudian diambil sebanyak 503 baris. Tabel 1 menampilkan nama fitur, tipe data, dan deksripsi tiap fitur yang digunakan.

Tabel 1. Nama kolom, tipe data tiap kolom, dan missing value

Nama Fitur	Tipe Data	Deskripsi Fitur
Rank_2025	object	Rank Universitas pada Tahun 2025
Rank_2024	object	Rank Universitas pada Tahun 2025
Institution_Name	object	Nama Institusi
Location	object	Negara asal universitas
Region	object	Wilayah Universitas
Size	object	Klasifikasi ukuran institusi
FOCUS	object	Tipe fokus universitas
Res	object	Intensitas Research
STATUS	object	Status Institusi
Academic_Reputation_Score	float64	Skor berdasarkan reputasi akademik scara global
Academic_Reputation_Rank	object	Rank berdasarkan reputasi employer secara global
Employer_Reputation_Score	float64	Skor berdasarkan reputasi employer secara global
Employer_Reputation_Rank	object	Rank berdasarkan reputasi employer
Faculty_Student_Score	float64	Skor refleksi murid berdasarkan rasio fakultas
Faculty_Student_Rank	float64	Rank berdasarkan rasio mahasiswa per fakultas
Citations_per_Faculty_Score	float64	Refleksi skor dampak penelitian (sitasi per fakultas)
International_Faculty_Score	float64	Skor Keberagaman Representasi Mahasiswa Internasional berdasarkan Fakultas
International_Faculty_Rank	object	Rank Berdasarkan Adanya Keberagaman Fakultas Internasional
International_Students_Score	float64	Skor representasi keberagaman mahasiswa internasional
International_Students_Rank	object	Peringkat berdasarkan rasio mahasiswa internasional
International_Research_Network_Score	float64	Skor berdasarkan kolaborasi penelitian global
International_Research_Network_Rank	object	Rank berdasarkan kerjasama penelitian
Employment_Outcomes_Score	float64	Skor yang mencerminkan kemampuan kerja dan keberhasilan lulusan
Employment_Outcomes_Rank	object	Rank berdasarkan kinerja lulusa universitas
Sustainability_Score	float64	Skor mencerminkan keberlanjutan inisiatif dan performance
Sustainability_Rank	object	Rank berdasarkan perhitungan keberlanjutan

Overall_Score	object	Skor gabungan akhir digunakan untuk menentukan peringkat universitas
---------------	--------	--

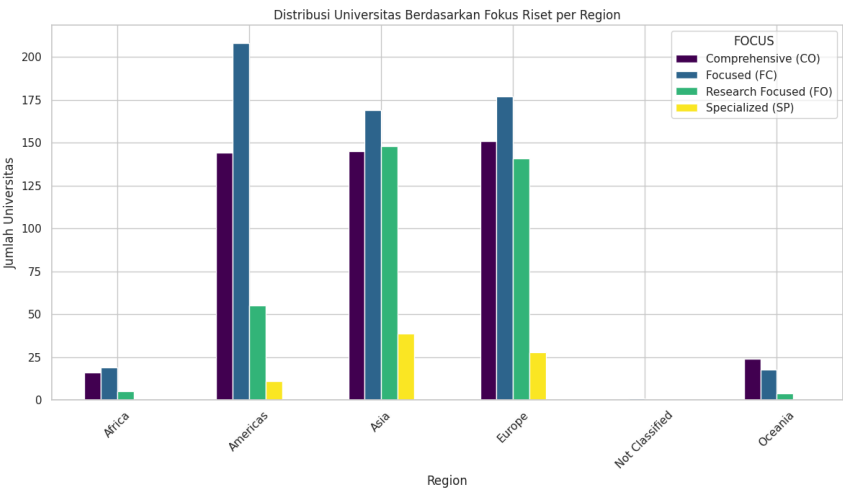
2. Handle Missing Value

Terdapat missing value pada data yakni :

Fitur	Missing Value
RANK_2024	21
STATUS	37
International_Faculty_Score	100
International_Faculty_Rank	100
International_Students_Score	58
International_Students_Rank	58
International_Research_Network_Score	1
International_Research_Network_Rank	1
Sustainability_Score	19
Sustainability_Rank	19
Overall_Score	902

3. Feature Engineering

Proses selanjutnya yakni feature engineering, setelah dicek nilai unik pada tiap fitur ditemukan data yang janggal seperti berikut :



Gambar 2. Distribusi Universitas berdasarkan Fokus dan Region

Terdapat universitas yang tidak terklasifikasi dalam benua sehingga perlu dieksplorasi lebih lanjut.

```
[88] df[df['Region'] == 'Not Classified']
```

	RANK_2025	RANK_2024	Institution_Name	Location	Region	SIZE	FOCUS	RES.	STAT
616	611-620	1201-1400	Eastern Mediterranean University	Northern Cyprus	Not Classified	L	FC	MD	

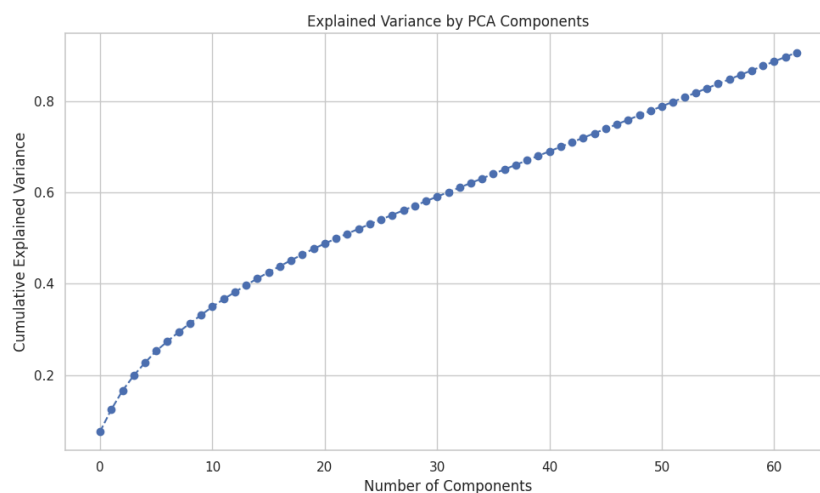
Gambar 3. Universitas yang Belum Terklasifikasi

Setelah dianalisis ternyata Eastern Mediterranean University (Universitas Mediterania Timur) terletak di Famagusta, Siprus Utara yang belum terklasifikasi sehingga diperlukan untuk mengganti fitur Region dari universitas tersebut dengan 'Europe'. Jika diakses dari websitenya, universitas tersebut merupakan universitas terbaik di Eropa. EMU menduduki peringkat ke-656 universitas terbaik di Eropa pada tahun 2024. Hal itu juga melandasi alasan kami untuk mengklasifikannya ke "Europe".

4. One Hot Encoding

Dataset dilakukan one hot encoding agar kolom kategorik bisa dijadikan kolom numerik untuk dilakukan modeling lebih lanjut.

5. Perlakuan PCA terhadap dataset



Gambar 4. Elbow Method reduksi dimensi

Grafik menunjukkan cumulative explained variance atau varians kumulatif yang dijelaskan terhadap jumlah komponen PCA yang digunakan. Sumbu X menunjukkan jumlah komponen PCA dari PC1 hingga PC63. Sumbu Y menunjukkan proporsi varians kumulatif yang dijelaskan

oleh komponen tersebut. Dari grafik juga terlihat bahwa semakin banyak komponen yang digunakan maka semakin besar varians yang dapat dijelaskan. Dengan tren naik yang mulai melandai setelah sekitar 30 komponen. Pada titik tersebut varians kumulatif yang dijelaskan sudah cukup tinggi yakni sekitar 60 - 70% yang menandakan bahwa sebagian besar informasi dalam data dapat direpresentasikan dengan jumlah komponen ini.

Tabel 2. Hasil PCA

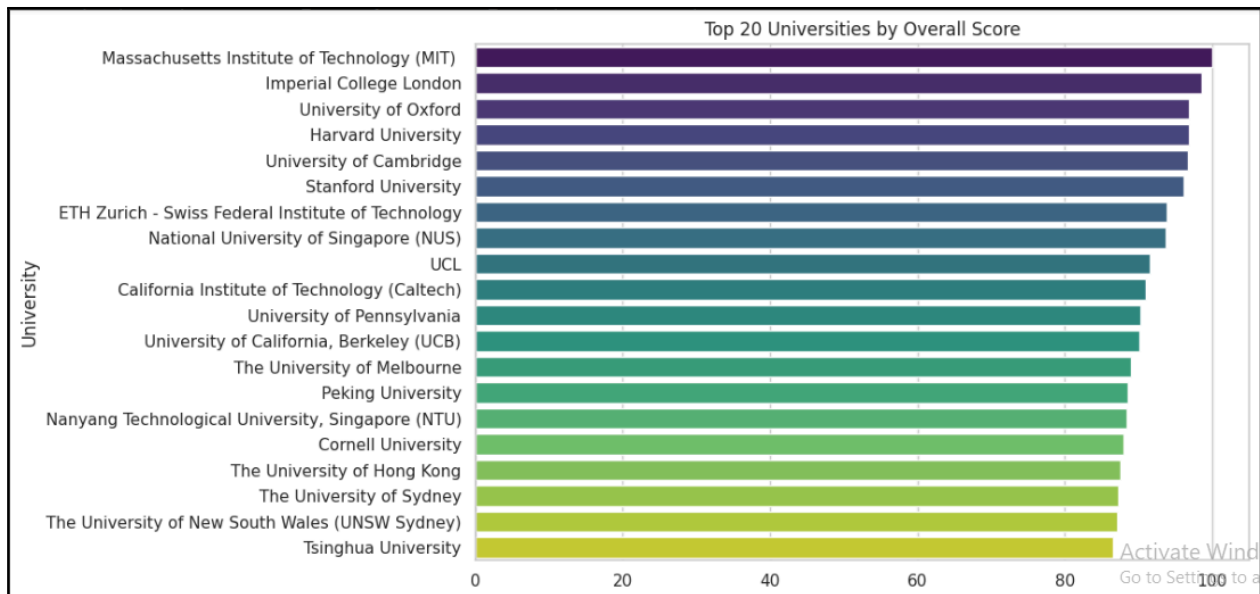
PCx	Fitur	Kumulatif Varian
PC1	Academic_Reputation_Score, Sustainability_Rank, dan Academic_Reputation_Rank	0.2902, 0.2870, 0.2864.
PC2	International_Faculty_Score, International_Faculty_Rank, dan Employer_Reputation_Rank	0.2574, 0.2468, 0.2164
PC3	International_Students_Score, International_Students_Rank, International_Faculty_Rank.	0.3006, 0.2980, 0.2670
PC4	Citations_per_Faculty_Score, Citations_per_Faculty_Rank, dan Region_Europe	0.4058, 0.3941, 0.3282
PC5	Region_Americas, Region_Europe, dan Faculty_Student_Rank.	0.3022, 0.2728, 0.2621

Pada tabel tersebut terdapat nama fitur - fitur yang terdapat pada tiap PC, diambil PC63 dari dimensi awal 103.

3.2 Insight dan hasil mining dari project

Dari dataset QS WUR 2025 kemudian didapat insight.

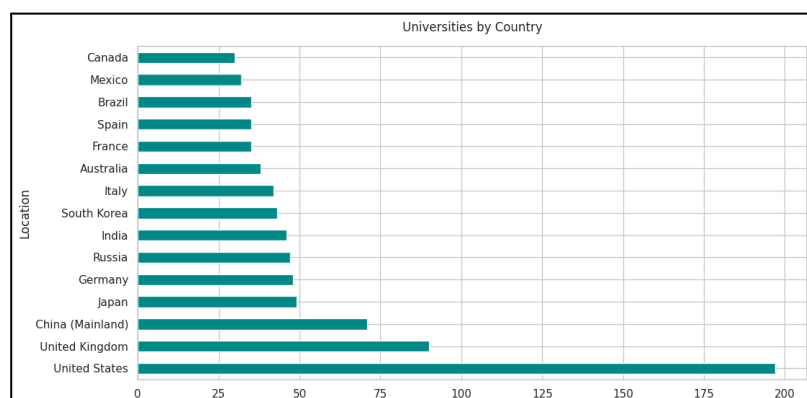
1. Universitas dengan Skor Keseluruhan Tertinggi



Gambar 5. Peringkat Universitas dengan skor tertinggi

Massachusetts Institute of Technology (MIT) menempati posisi teratas dengan skor keseluruhan yang hampir sempurna. Hal ini menegaskan konsistensi kualitas dan keunggulan MIT di kancah pendidikan global. Beberapa universitas yang berasal dari negara-negara berbahasa Inggris seperti Amerika Serikat dan Inggris mendominasi peringkat 10 besar. Di antaranya termasuk Harvard University, Stanford University, University of Oxford, dan University of Cambridge, yang menunjukkan bahwa kedua negara tersebut masih menjadi pusat utama pendidikan tinggi global. Sementara itu, universitas dari kawasan Asia dan Oseania juga menunjukkan kemajuan signifikan. National University of Singapore (NUS) dan Tsinghua University dari Tiongkok berhasil masuk ke jajaran 20 besar dunia, yang mengindikasikan peningkatan kualitas dan daya saing pendidikan tinggi di kawasan Asia. Hal yang sama juga terlihat dari performa UNSW Sydney dan University of Sydney dari Australia yang berhasil bersaing di peringkat global.

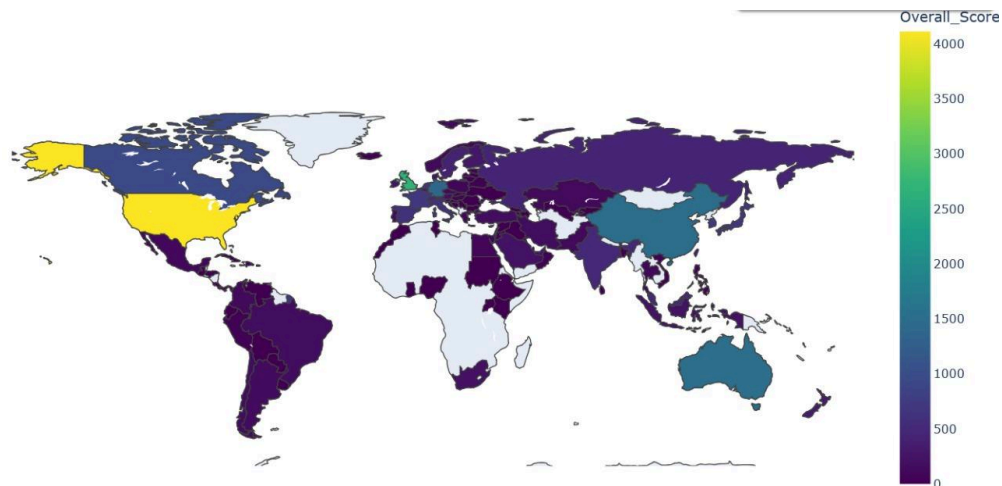
2. Jumlah Universitas tiap Negara



Gambar 6. Jumlah universitas tiap negara (bar chart)

United State atau Amerika Serikat merupakan negara terbanyak jumlah universitasnya, dengan lebih dari 175 universitas. Jumlah ini sangat tinggi jika dibandingkan dengan negara lainnya yang masih dibawah 100 universitas. Hal ini menandakan bahwa Amerika Serikat unggul dalam fasilitas perguruan tinggi secara global.

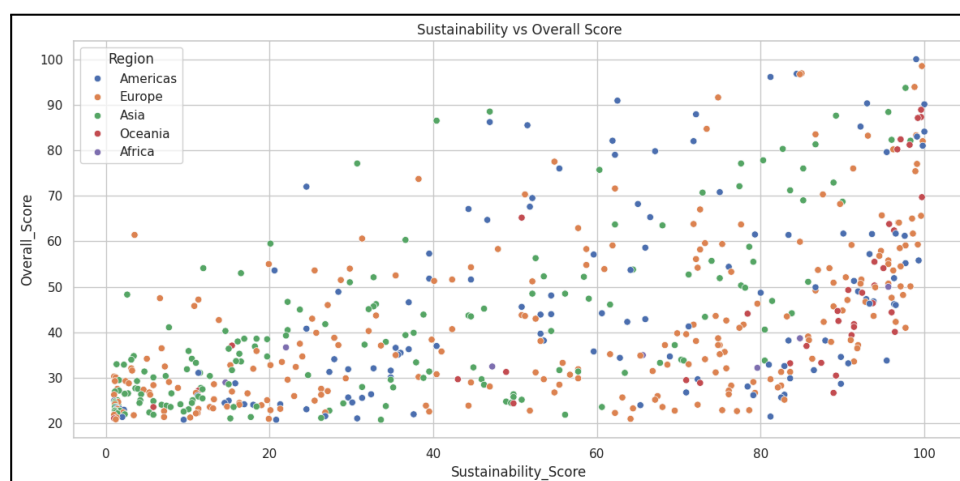
3. Peta Persebaran Skor Keseluruhan berdasarkan Negara



Gambar 7. Peta persebaran skor Universitas berdasarkan negara

Dari peta tersebut, terlihat bahwa skor terbesar keseluruhan skor berdasarkan pewarnaan negara - negara di peta tersebut. Terdapat Amerika Serikat dengan warna kuning terang memiliki skor yang sangat tinggi <4000. Sedangkan jika dibandingkan dengan negara Indonesia dengan total skor keseluruhan adalah 232,6. Hal ini adalah bukti kesenjangan pendidikan di lintas negara.

4. Hubungan antar Variabel

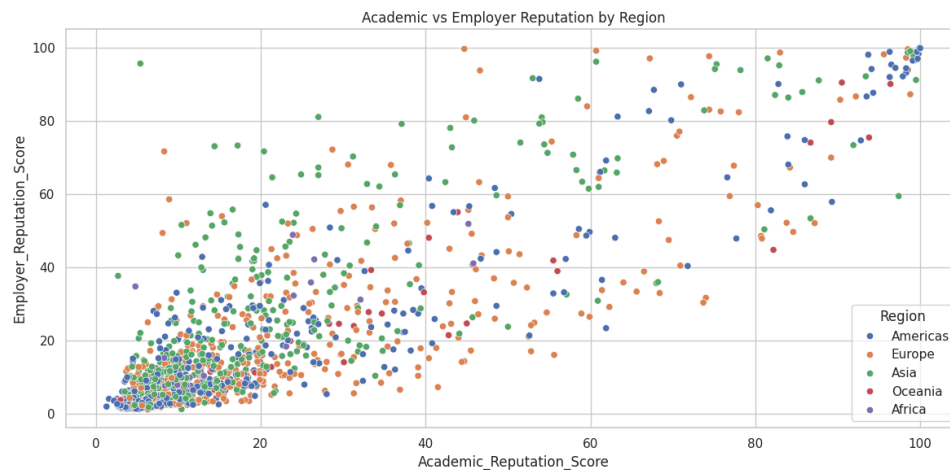


Gambar 8. Scatter plot korelasi antara fitur Overall_Score dan Sustainability_Score

Scatter plot menunjukkan hubungan antara dua variabel yakni Sustainability_Score (sumbu X) dan Overall_Score (sumbu Y). Titik - titik warna didasarkan oleh Benua yang sudah tertulis pada legenda. Adanya tren positif pada scatter plot ini sehingga ini mengindikasikan bahwa

semakin tinggi skor keseluruhan (Sustainability_Score) maka semakin tinggi pula skor keseluruhan (Overall_Score). Hal ini menunjukkan bahwa keberlanjutan signifikan terhadap keseluruhan skor. Titik dari wilayah Amerika dan Europe cenderung terkonsentrasi di area dengan skor keberlanjutan yang lebih tinggi, kemudian untuk wilayah Asia dan Oceania menunjukkan adanya distribusi yang lebih menyebar dengan beberapa negara yang memiliki skor tinggi dan beberapa rendah. Disisi lain, wilayah Afrika cenderung memiliki skor keberlanjutan dan skor keseluruhan yang lebih rendah.

5. Hubungan Akademik dengan Reputasi Employer berdasarkan Wilayahnya.

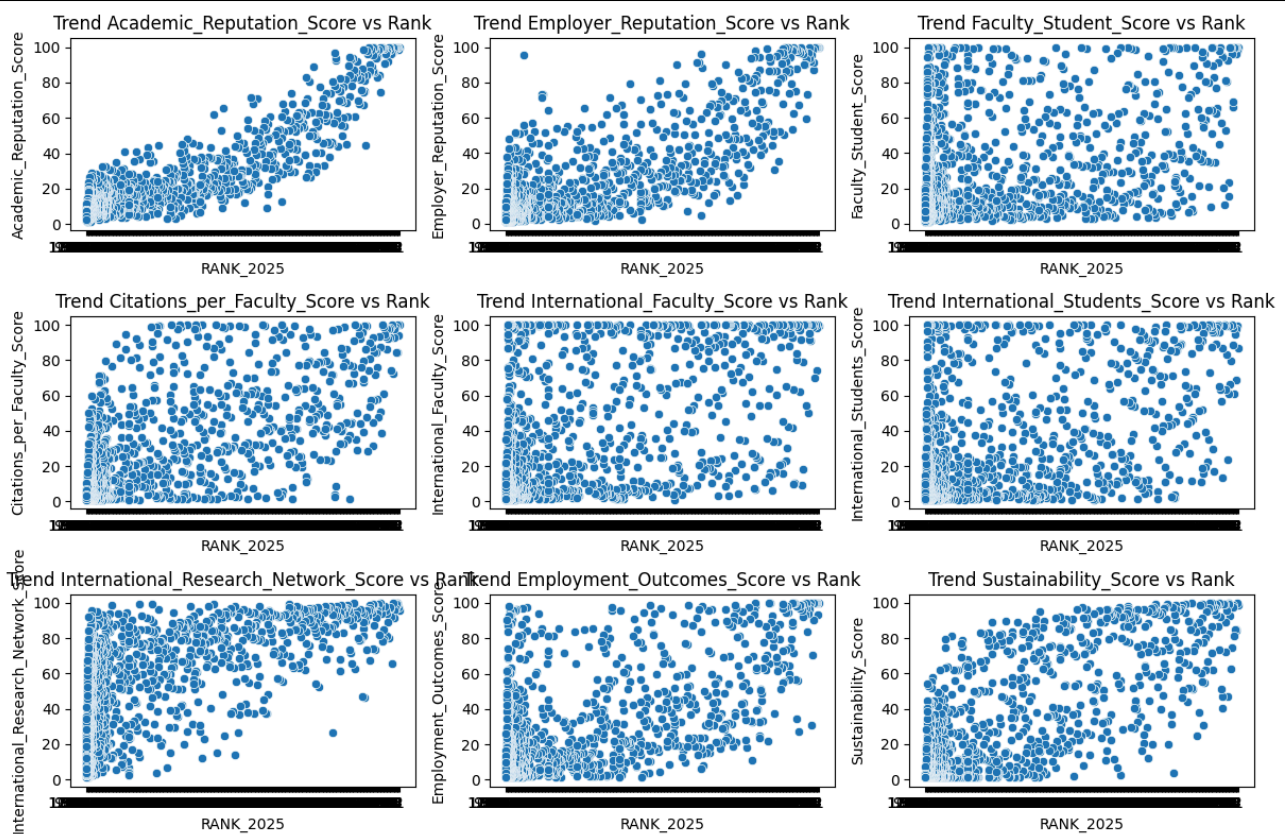


Gambar 9. Scatter plot korelasi fitur antara Employer_Reputation_Score dan Academic_Reputation_Score

Dari scatter plot, Sumbu X adalah skor reputasi akademik dengan rentang nilai 0 - 100 kemudian Sumbu Y adalah skor reputasi employer dengan rentang nilai yang sama. Warna pada titik mencerminkan masing-masing benua sesuai dengan legenda. Setiap titik mewakili satu entitas dengan skor reputasi akademik dan reputasi employer tertentu. Dari scatter plot ini terlihat bahwa adanya korelasi positif antara reputasi akademik dengan reputasi di mata employer. Terlihat juga sebagian besar titik biru tersebar di rentang nilai yang tinggi (80 - 100) untuk kedua variabel. Sedangkan sebagian besar titik berwarna ungu berada di rentang nilai yang rendah. Hal ini menandakan bahwa nilai reputasi yang tinggi mayoritas masih dipegang oleh Benua Amerika namun Benua Afrika masih terdapat tantangan dalam reputasi akademik dan employer di wilayah ini.

6. Relasi kolom Rank 2025 dengan beberapa Fitur lain

Gambar tersebut merupakan scatter plot yang dapat menunjukkan tren berbagai skor. Dan fitur terhadap peringkat institusi (Rank_2025). Setiap plot menggambarkan hubungan antara skor tertentu dengan peringkat institusi sehingga memberikan insight tentang faktor - faktor yang berkontribusi pada peringkat tersebut.



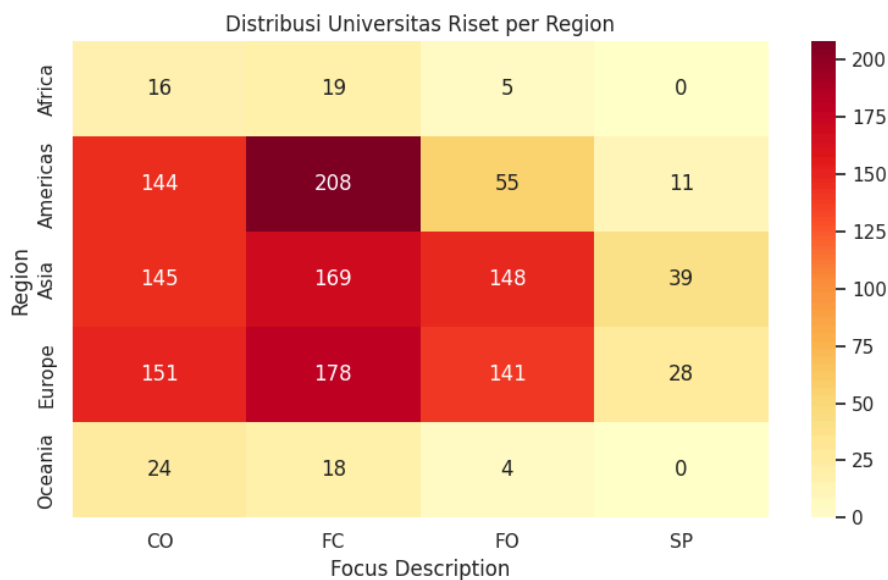
Gambar 10. Scatter plot korelasi fitur antara Employer_Reputation_Score dan Academic_Reputaion_Score

1. Skor Reputasi Akademik dengan Ranking Universitas : Adanya korelasi positif yang sangat kuat antara skor reputasi akademik dengan Ranking Universitas, sehingga institusi dengan peringkat lebih baik (nilai ranking rendah) cenderung memiliki skor reputasi akademik yang tinggi. Reputasi akademik menjadi faktor utama yang dapat mempengaruhi peringkat institusi.
2. Skor Reputasi Employer dengan Ranking Universitas : Terdapat korelasi positif antara skor reputasi employer dengan rank 2025 sehingga institusi dengan peringkat yang tertinggi memiliki reputasi yang baik di mata employer. Hubungan di dunia kerja dan employer juga sangat penting dalam menentukan peringkat.
3. Skor Mahasiswa per Fakultas dengan Rank Universitas : Tidak adanya korelasi yang jelas meskipun ada sedikit potensi bahwa institusi dengan peringkat yang lebih baik memiliki nilai skor mahasiswa tiap fakultas yang bervariasi. Kemudian rasio dosen terhadap mahasiswa tidak selalu menjadi indikator utama peringkat meskipun masih sedikit relevan. Namun dari dataset ini pola korelasi belum terlihat jelas.
4. Sitasi per Fakultas dengan Rank Universitas : Adanya korelasi positif sedang antara skor sitasi per fakultas dengan ranking universitas sehingga institusi dengan peringkat yang lebih baik cenderung memiliki jumlah sitasi per dosen yang lebih tinggi hal ini juga mengindikasikan adanya kualitas riset yang baik.

5. Skor Fakultas Internasional dengan Rank Universitas menunjukkan pola penyebaran cukup acak tanpa adanya korelasi yang jelas diantara dua fitur sehingga proporsi dosen internasional tidak secara langsung berkorelasi dengan peringkat institusi.
6. Skor Mahasiswa Internasional dengan Rank Universitas : Adanya pola penyebaran yang acak tanpa adanya korelasi yang jelas sehingga jumlah mahasiswa internasional tidak menjadi faktor utama dalam menentukan peringkat.
7. Skor Jaringan Penelitian Internasional vs Ranking Universitas : Adanya korelasi positif yang cukup kuat antara skor jaringan penelitian internasional dengan ranking perusahaan sehingga institusi yang memiliki jaringan riset internasional yang kuat cenderung memiliki peringkat lebih baik sehingga kolaborasi riset internasional adalah faktor penting dalam reputasi dan peringkat.
8. Skor Hasil Kerja Lulusan dengan Ranking Perusahaan : Pola penyebaran yang acak sehingga beberapa berkonsentrasi di skor yang tinggi untuk universitas dengan ranking yang baik.
9. Skor Sustainability dengan Ranking : Adanya korelasi positif tingkat sedang antara skor keberlanjutan dengan ranking 2025 sehingga institusi yang lebih berperingkat cenderung memiliki skor keberlanjutan yang lebih tinggi sehingga fokus keberlanjutan mulai menjadi faktor penting dalam penilaian institusi.

7. Distribusi Universitas Riset berdasarkan Wilayah

Heatmap pada Gambar 11 menunjukkan bahwa jumlah universitas riset yang tersebar di berbagai region di dunia berdasarkan kategori fokus riset yang berbeda. Sumbu vertikal adalah region dan sumbu horizontal adalah kategori fokus riset tiap universitas. Warna heatmap menunjukkan intensitas jumlah universitas riset dari warna kuning muda hingga dengan jumlah rendah hingga merah tua dengan jumlah tinggi. Amerika dan Eropa memiliki jumlah universitas riset terbanyak di hampir semua kategori fokus riset dengan intensitas warna tua yang dominan. Hal ini menandakan bahwa riset didominasi pada wilayah tersebut. Sedangkan Asia menunjukkan distribusi yang cukup merata dan signifikan terutama pada kategori FC dan FO. Kemudian Afrika dan Oceania memiliki jumlah universitas riset yang lebih relatif lebih sedikit dengan lebih terang. Kemudian SP (kemungkinan riset spesifik yang memiliki jumlah universitas riset paling sedikit diantara semua region).



Gambar 11 Heatmap Universitas Fokus Riset

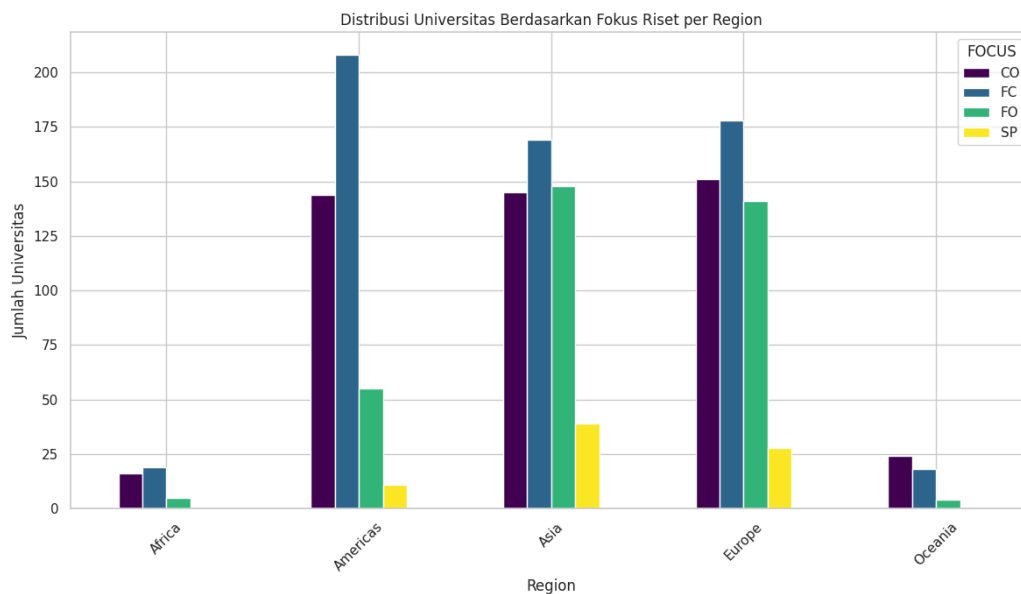
8. Kernel Density Estimation dalam Reputasi Employer dengan Hasil Kinerja Lulusan



Gambar 12 KDE Reputasi Employer dengan Hasil Kinerja Lulusan

Gambar tersebut adalah visualisasi dari fitur skor reputasi employer dengan hasil kerja lulusan. Sumbu Horizontal adalah skor reputasi employer sedangkan sumbu vertikal adalah skor hasil kerja lulusan. Warna terang merupakan area dengan kepadatan data yang lebih tinggi. Dari gambar ini terdapat konsentrasi data yang sangat tinggi di area skor rendah hingga menengah untuk kedua variabel yang dapat menunjukkan banyak institusi dengan skor reputasi employer dan hasil kerja lulusan yang relatif rendah hingga sedang. Juga terdapat area dengan kepadatan yang lebih rendah namun tersebar hingga skor tinggi yang menandakan bahwa beberapa institusi memiliki reputasi employer dan hasil kerja lulusan yang sangat baik. Pola ini menampilkan bahwa adanya korelasi positif antara reputasi employer dan hasil kerja lulusan sehingga apabila terdapat institusi dengan reputasi employer yang baik maka cenderung juga memiliki hasil kerja lulusan yang baik. Namun terdapat konsentrasi data di skor rendah hingga menengah yang juga menunjukkan bahwa ada banyak institusi masih perlu meningkatkan kedua aspek tersebut.

9. Distribusi Universitas berdasarkan Fokus Riset tiap Wilayah



Gambar 13 Bar Plot Jumlah Universitas tiap Wilayah dan Fokusnya

Gambar tersebut merupakan visualisasi berupa bar chart yang menunjukkan jumlah universitas di berbagai region dunia berdasarkan kategori fokus riset. Region yang dianalisis yakni Afrika, Amerika, Asia, eropa, dan Oceania. Kategori fokus riset yang ditampilkan adalah CO, FC, FO, dan SP. (singkatan kategori riset yang spesifik / pada bidang tertentu) dan setiap kategori fokus riset yang diwakili oleh warna berbeda pada bar chart. Amerika dan Eropa mendominasi jumlah universitas riset terutama pada kategori FC, CO, dan FO dengan jumlah universitas yang sangat tinggi. Asia juga menunjukkan distribusi yang kuat dan emrata di hampir semua kategori terutama pada CO, FC, dan FO. Khususnya kategori SP di Asia cukup rendah yakni <50 universitas yang menandakan fokus riset spesifik masih kurang. Afrika dan Oceania memiliki jumlah universitas riset yang jauh lebih sedikit di semua kategori yang menunjukkan potensi pengembangan riset yang masih besar pada kedua region ini. Serta kategori SP memiliki jumlah universitas yang paling sedikit di semua region menandakan area riset yang spesifik kurang berkembang.

Model	R2 Score	MSE
XGBoost	0.9468	18.0774
Random Forest	0.9240	25.8504
Linear Regression	0.8597	47.7023

Tabel 3. Kinerja tiap model terhadap dataset

Tabel ini menunjukkan bahwa adanya performa tiga model prediksi yang digunakan untuk suatu tugas (mungkin prediksi skor atau ranking universitas). R2 Score adalah koefisien determinasi semakin mendekati angka 1 maka semakin baik. MSE dalam Mean Squared Error maka semakin kecil semakin baik modelnya. XGBoost adalah model terbaik dengan nilai R2 tertinggi (0.9469) dan MSE terendah (18.08) yang menunjukkan prediksi yang sangat akurat

dan kesalahan yang sangat kecil. Random Forest juga menunjukkan performa yang sangat baik dengan R2 diatas 0.92 dan MSE yang lebih tinggi daripada XG Boost tapi masih jauh lebih baik daripada Linear Regression. Linear regression memiliki performa yang paling rendah diantara ketiganya.

3. Daftar Pustaka

Sitasi disusun dan ditulis berdasarkan sistem nomor sesuai dengan urutan pengutipan, mengikuti format APA. Hanya pustaka yang disitasi pada usulan penelitian yang dicantumkan dalam Daftar Pustaka. Pustaka yang disitasi maksimal 8 tahun terakhir sebanyak minimal 10 pustaka.

- [1] Polezhaeva, T. Yu. (2024). World University Rankings: Peculiarities and Differences in Methodology. *Theory and Practice of Social Development*, 2, 60–65. <https://doi.org/10.24158/tipor.2024.2.7>
- [2] Liu, X., Chen, G., Wen, S., & Han, X. (2022). Analysis and Prediction of QS World University Rankings based on Data Mining Technology. <https://doi.org/10.1145/3551708.3556207>
- [3] Irungu, W. R., & Liu, Z. (2024). Exploring the nexus between university sustainability practices and academic performance: An empirical analysis of the QS sustainability ranking and four world university rankings. *PLOS ONE*, 19(10), e0306286. <https://doi.org/10.1371/journal.pone.0306286>
- [4] Estrada-Real, A., & Cantu-Ortiz, F. J. (2022). A data analytics approach for university competitiveness: the QS world university rankings. *International Journal On Interactive Design And Manufacturing (Ijidem)*, 16(3), 871–891. <https://doi.org/10.1007/s12008-022-00966-2>
- [5] Bublyk, M., Vysotska, V., Kolyasa, L. I., & Vlasenko, O. (n.d.). World Universities Strategic Analysis Based on Data from the QS World University Rankings. *International Workshop on Intelligent Information Technologies & Systems of Information Security*, 354–375
- [6] Subbarayalu, A. V. (2024). Climbing the Global Ladder. *Advances in Higher Education and Professional Development Book Series*, 273–306. <https://doi.org/10.4018/979-8-3693-6849-7.ch011>
- [7] Liu, X., Chen, G., Wen, S., & Han, X. (2022). Analysis and Prediction of QS World University Rankings based on Data Mining Technology. <https://doi.org/10.1145/3551708.3556207>
- [8] Eastern Mediterranean University. (n.d.). Accreditations, recognitions, rankings & memberships. Retrieved May 28, 2025, from <https://www.emu.edu.tr/en/about-emu/accreditations-recognitions-rankings-memberships/> 597

[9]Gürsoy, S., & Kunt, N. (2019). Acculturation of university students in Northern Cyprus. *Culture & Psychology*, 25(2), 146–160. <https://doi.org/10.1177/1354067X18808213>

[10] Jobbehdar Nourafkan, N., Jumah, B., Asif, R., & Dehghanmongabadi, A. (2020). Conflicts between students living on-campus dormitories: The case of dormitories at Eastern Mediterranean University campus. *Smart Journal of Management*, 3(4), 31–43. <https://doi.org/10.33215/SJOM.V3I4.397>