

Indonesia University Rank

1. Pendahuluan

Proyek Indonesia University Rank bertujuan untuk membuat model pembelajaran mesin yang mampu mengklasifikasikan universitas di Indonesia berdasarkan peringkat mereka. Dengan menggunakan algoritma Random Forest dan teknik hyperparameter tuning, proyek ini berupaya memberikan wawasan tentang bagaimana universitas dapat dikelompokkan berdasarkan peringkat mereka dengan mempertimbangkan berbagai faktor seperti kota dan region universitas tersebut.

2. Tujuan Proyek

Tujuan utama dari proyek ini adalah untuk:

1. Membangun model klasifikasi yang dapat mengelompokkan universitas ke dalam tiga kategori peringkat: Top 10, Top 50, dan Lainnya.
2. Menggunakan teknik machine learning untuk mengidentifikasi fitur-fitur penting yang memengaruhi peringkat universitas.
3. Mengoptimalkan performa model dengan menggunakan hyperparameter tuning untuk mencapai akurasi terbaik.

3. Metodologi

Proses yang diikuti dalam proyek ini melibatkan beberapa tahapan penting:

- a. Pengumpulan dan Pembersihan Data
 - Data yang digunakan adalah data peringkat universitas di Indonesia.
 - Langkah pertama adalah memuat data dari file CSV dan membersihkan data dari nilai yang hilang (missing values) atau data yang tidak relevan.
- b. Penambahan Fitur
 - Penambahan fitur dilakukan dengan membuat kolom baru untuk region berdasarkan kota universitas, yang memungkinkan analisis lebih dalam terhadap wilayah geografis.
- c. Encoding Variabel Kategoris
 - Kolom seperti University, Town, dan Region diencoding menjadi variabel numerik menggunakan One-Hot Encoding agar dapat diproses oleh algoritma machine learning.
- d. Pembagian Data
 - Data dibagi menjadi dua set: Training set (80%) dan Test set (20%) untuk melatih dan menguji model.
- e. Pelatihan Model

- Model Random Forest digunakan untuk melatih data yang telah diproses.
- Setelah model dilatih, prediksi dilakukan pada test set dan hasilnya dievaluasi menggunakan akurasi.

f. Hyperparameter Tuning

- Untuk mendapatkan model terbaik, dilakukan GridSearchCV untuk melakukan tuning hyperparameter seperti jumlah pohon dalam Random Forest dan kedalaman maksimum pohon (max depth).

g. Evaluasi

- Evaluasi model dilakukan menggunakan metrik akurasi untuk mengukur seberapa baik model dalam mengklasifikasikan peringkat universitas.

h. Penyimpanan Model

- Model yang telah dilatih dan dievaluasi disimpan dalam format pickle agar dapat digunakan kembali di masa depan tanpa perlu melatih ulang.

4. Hasil dan Diskusi

- Model Random Forest yang dilatih mencapai akurasi sekitar [X%] (sesuai hasil evaluasi model).
- Fitur-fitur seperti Region dan Town memiliki pengaruh signifikan terhadap prediksi peringkat universitas.
- Melalui tuning hyperparameter, model mengalami peningkatan performa, menghasilkan model yang lebih akurat.

5. Kesimpulan

Proyek Indonesia University Rank berhasil membangun model machine learning yang cukup akurat dalam mengklasifikasikan universitas di Indonesia berdasarkan peringkat. Model ini memanfaatkan algoritma Random Forest yang setelah dilakukan tuning hyperparameter, memberikan akurasi yang memuaskan. Dengan menyimpan model, sistem ini dapat digunakan kembali untuk memperkirakan peringkat universitas di masa mendatang.

6. Rekomendasi

Untuk peningkatan di masa depan, beberapa hal yang bisa dipertimbangkan adalah:

- Menambah lebih banyak fitur, seperti jumlah mahasiswa, akreditasi, atau publikasi riset.
- Menggunakan model lain seperti Gradient Boosting atau Support Vector Machine untuk membandingkan performa.
- Memperluas dataset dengan memasukkan data universitas yang lebih beragam dari tahun-tahun yang berbeda.

7. Kode Sumber

Seluruh kode yang digunakan dalam proyek ini dapat ditemukan di repositori GitHub:
<https://github.com/haekalsetiawan/Indonesia-University-Ranking-ML.git>