

Judul : "Analisis Regresi Logistik Ordinal untuk Memprediksi Tingkat Pendidikan Berdasarkan Jenis Kelamin di Jawa Tengah."

Anggota Kelompok : 1. Aditya Prayoga (123210098) (IF-D)
2. Rizal Hanifa Pratama (123210114) (IF-D)

1. Why? (Business Understanding & Analytic Approach)

Ketidaksetaraan akses pendidikan antara pria dan wanita membawa dampak serius pada kesempatan dan kesejahteraan sosial di Jawa Tengah. Untuk mengatasi permasalahan ini kami mencoba melakukan analisis data pendidikan berdasarkan jenis kelamin, dapat dihasilkan wawasan yang mendalam untuk membentuk dasar keputusan yang kokoh dan merancang kebijakan pendidikan yang lebih efektif. Pendekatan ini memungkinkan pihak berwenang untuk memahami secara detail pola ketidaksetaraan dan meresponsnya dengan tindakan yang lebih terarah, menjadikan pendidikan sebagai instrumen utama dalam mencapai kesetaraan gender dan kesejahteraan sosial yang berkelanjutan.

Task :

1. Mendefinisikan business understanding
2. Menentukan analytic approach, pada topik ini kami menggunakan analytic approach diagnostic
3. Menganalisa data yang dibutuhkan
4. Mengumpulkan data (primer/sekunder/keduanya)
5. Membersihkan data berupa missing value serta menangani outlier yang ada di data
6. Menentukan jenis model yang digunakan. Jenis model yang kami pilih ialah model regresi
7. Menentukan algoritma yang digunakan, Algoritma yang digunakan adalah algoritma *Ordinal Logistic Regression*
8. Mendeskripsikan hasil yang telah didapat

2. Data Requirement & Data Collection

Data yang dibutuhkan untuk projek data science ini adalah data riil tentang tingkat pendidikan dan jenis kelamin dari penduduk Jawa Tengah. Data yang dibutuhkan untuk projek data science ini setidaknya harus memiliki kolom-kolom berikut:

- Jenis kelamin: Kolom ini berisi informasi tentang jenis kelamin penduduk, yaitu laki-laki atau perempuan.
- Tingkat pendidikan: Kolom ini berisi informasi tentang tingkat pendidikan penduduk, yaitu SD, SMP, SMA, Diploma, atau Sarjana.

Kami menggunakan dataset dari sumber resmi Pemerintah Provinsi Jawa Tengah. Dataset ini dapat diakses melalui tautan <https://data.jatengprov.go.id/dataset/data-kependudukan-dispermasdesdukcapil>. Dataset ini melibatkan informasi tingkatan pendidikan akhir masyarakat Jawa Tengah berdasarkan jenis kelamin dan lokasi (kota/kabupaten). Totalnya terdapat 945 entri data yang akan di gunakan dalam analisis ini, mencakup seluruh data yang tersedia.

Secara umum, seluruh data yang tersedia akan digunakan untuk proyek data science ini. Hal ini dilakukan untuk mendapatkan gambaran yang lengkap tentang pengaruh jenis kelamin terhadap tingkat pendidikan di Jawa Tengah.

3. Modelling & Evaluation

Modelling

- Jenis model yang dipilih:

Berdasarkan task yang dipilih, yaitu untuk mengidentifikasi pengaruh jenis kelamin terhadap tingkat pendidikan, maka jenis model yang dipilih adalah model regresi logistik ordinal. Model regresi logistik ordinal adalah metode statistika yang digunakan untuk memprediksi variabel kategorik dengan urutan. Dalam hal ini, variabel kategorik yang diprediksi adalah tingkat pendidikan yang memiliki urutan dari terendah ke tertinggi, yaitu SD, SMP, SMA, Diploma, dan Sarjana. Variabel prediktor yang digunakan adalah jenis kelamin yang memiliki dua kategori, yaitu laki-laki dan perempuan.

- Alasannya:

Model regresi logistik ordinal dipilih karena model ini cocok untuk memprediksi variabel kategorik dengan urutan. Model ini juga telah banyak digunakan dalam berbagai penelitian sebelumnya untuk memprediksi berbagai variabel kategorik, seperti tingkat pendidikan, tingkat pendapatan, dan tingkat kepuasan pelanggan.

Algoritma Regresi logistik ordinal digunakan untuk memprediksi tingkat pendidikan seseorang berdasarkan variabel prediktor jenis kelamin. Jenis kelamin adalah variabel kategorik dengan dua kategori, yaitu laki-laki dan perempuan. Algoritma ini menghasilkan model yang dapat memprediksi kemungkinan seseorang memiliki tingkat pendidikan tertentu berdasarkan jenis kelaminnya. Model ini dapat digunakan untuk memahami pengaruh jenis kelamin terhadap tingkat pendidikan.

Referensi :

[Regresi Logistik Ordinal \(Studi Kasus Faktor Yang Mempengaruhi](#)

Rencana Evaluasi:

1. Pembagian Data:

Pisahkan dataset menjadi data pelatihan (training) dan data pengujian (testing) dengan proporsi yang sesuai.

2. Pemilihan Metrik Evaluasi:

Menggunakan metrik yang sesuai seperti evaluasi proporsional (proportional odds) atau metrik ordinal lainnya.

3. Pelatihan Model:

Latih model regresi logistik ordinal menggunakan data pelatihan.

4. Evaluasi pada Data Pengujian:

Evaluasi kinerja model pada data pengujian menggunakan metrik yang telah dipilih.

5. Analisis Kesalahan:

Analisis kesalahan model untuk memahami di mana model cenderung salah dan melihat pola-pola tertentu.

6. Validasi Silang (Cross-Validation):

Gunakan validasi silang untuk menguji keberlanjutan model dan memastikan kinerjanya di berbagai set data.

7. Interpretasi Koefisien:

Jika memungkinkan, interpretasikan koefisien regresi logistik ordinal untuk memahami dampak jumlah pria dan wanita terhadap tingkat pendidikan.

8. Optimasi Hyperparameter:

Jika perlu, lakukan optimasi hyperparameter untuk meningkatkan kinerja model.

9. Dokumentasi dan Presentasi Hasil:

Dokumentasikan temuan dan hasil evaluasi model, dan presentasikan dengan jelas kepada pemangku kepentingan.

10. Iterasi Model (Jika Diperlukan):

Berdasarkan hasil evaluasi, lakukan iterasi pada model jika diperlukan untuk meningkatkan kinerja dan interpretabilitasnya.

4. Analytics Goal

Tujuan dari analisis ini adalah untuk memprediksi tingkat pendidikan seseorang berdasarkan jenis kelaminnya. Hasil analisis ini dapat digunakan untuk memahami pengaruh jenis kelamin terhadap tingkat pendidikan. Fokus utama analisis adalah mengeksplorasi bagaimana jumlah pria dan wanita dalam setiap kategori pendidikan berkontribusi terhadap tingkat pendidikan secara keseluruhan. Analisis ini dapat membantu mengidentifikasi daerah-daerah yang mungkin memerlukan perhatian khusus dalam upaya meningkatkan akses dan kesetaraan pendidikan berdasarkan gender di tingkat Jawa Tengah.

*definisi bisnis yang dimaksud di bagian ini bukan wajib bidang bisnis, namun berarti berbagai bidang (sesuai yang telah dijelaskan/ dicoba implementasinya pada mini project)