

Mata Kuliah: IBDA2032 / Kecerdasan Buatan

PETUNJUK

- Kerjakanlah proyek di bawah ini lalu dikumpulkan dalam format .zip yang berisikan semua file yang diperlukan untuk mengumpulkan nilai (.ipynb, dataset, dll) sesuai dengan rubrik penilaian proyek.
- File yang dikumpulkan harus lengkap, agar dosen pengampu dapat menjalankan kode pemrograman yang dikumpulkan.
- Apabila dosen pengampu tidak dapat menjalankan kode pemrograman karena ada file yang kurang, maka bagian yang tidak dapat dijalankan tersebut tidak mendapatkan nilai.
- Apabila ada bagian yang error yang menyebabkan dosen pengampu tidak dapat menjalankan kode pemrograman, maka bagian selanjutnya dari kode pemrograman tersebut tidak akan dinilai.
- Setelah selesai mengerjakan proposal proyek dan proyek, tulislah komponen persentase kontribusi dari setiap anggota kelompok.

PENGANTAR

Terdapat banyak sekali masalah di sekitar kita yang dapat kita selesaikan dengan *machine learning*. Oleh karena itu peserta kelas IBDA2023 diminta untuk berpartisipasi aktif dengan menggunakan ilmu yang sudah diperoleh di kelas ini untuk melakukan pengolahan data dan menawarkan pemodelan bagi permasalahan tertentu untuk memberkati sesama kita. Kalian bebas menggunakan konsep apapun dari topik yang sudah dipelajari di kelas (regression, classification, clustering, representation). Solusi yang ditawarkan perlu dibuat dalam bentuk laporan pemrograman pada file notebook (ipynb) dan dipresentasikan.

DESKRIPSI TUGAS

1. Proposal (10 poin)

Sebelum melakukan pemodelan, diperlukan pertimbangan yang matang tentang data apa yang dipilih dan pemodelan seperti apa yang diperlukan. Maka 1 minggu sebelum proyek ini dimulai, kumpulkanlah melalui canvas, sebuah proposal mengenai masalah yang ingin diselesaikan beserta data dan pemodelan yang akan kalian gunakan. Proposal akan dinilai secara kualitatif terlebih dahulu untuk memastikan data dari semua kelompok unik. Jika terdapat pilihan data yang sama, hanya proposal dari kelompok yang mengumpulkan terlebih dahulu yang diterima. Data dapat kalian peroleh melalui proses koleksi data secara mandiri (crawling/scrapping/survey/sensor), maupun melalui banyak open data yang tersedia online:

- <https://www.kaggle.com/datasets>
- <https://archive.ics.uci.edu/ml/datasets.php>
- <https://data.worldbank.org/>
- <https://data.un.org/>
- <https://data.gov/>
- <https://data.europa.eu/en>

- <https://data.jakarta.go.id/>
- <https://data.go.id/home>

Proposal ini akan dinilai setelah dipastikan semua kelompok menggunakan data yang berbeda. Jika proposal lengkap dan valid, maka akan mendapatkan nilai 10 poin.

Proposal singkat ini berisi:

- Masalah yang coba diselesaikan
- Sumber data (contoh: url_link/survey/personal repository/SEP) dan deskripsi singkat data
- Cara memperoleh data (contoh: crawling/scraping/raw data ready/gabungan beberapa metode)
- Jenis tugas ML yang akan dikerjakan (contoh: regression/classification/clustering/representation/gabungan beberapa tugas)

2. **Notebook dan Presentasi (90 poin + 30 poin bonus)**

Dengan menggunakan Jupyter Notebook (atau IDE sejenisnya) lakukanlah proses pengolahan data dan pemodelan untuk masalah yang kelompok kalian pilih. Kalian juga diminta menjelaskan secara lebih detail pekerjaan yang dilakukan dengan memanfaatkan fitur **markdown** dari Jupyter Notebook supaya diperoleh tampilan yang lebih mudah dibaca. Kelompok kalian akan mempresentasikan proyek ini (dengan menggunakan file .ipynb dan PPT) di depan kelas pada sebuah jadwal sebelum minggu UAS.

Catatan:

- Kelompok dengan jumlah anggota 4 orang perlu mengumpulkan 120 poin untuk mendapatkan nilai 100%
- Kelompok dengan jumlah anggota 3 orang perlu mengumpulkan 100 poin untuk mendapatkan nilai 100%
- Kelompok dengan jumlah anggota 1 atau 2 orang perlu mengumpulkan 80 poin untuk mendapatkan nilai 100%

Kriteria	Poin Maximal per Kriteria	Opsi Pengumpulan Poin
Proposal	10	<ul style="list-style-type: none"> • Deskripsi masalah yang coba diselesaikan (2.5 poin) • Informasi sumber data dan deskripsi singkat, contoh: url_link/survey/personal repository/SEP (2.5 poin) • Penjelasan singkat cara memperoleh data, contoh: crawling/scraping/raw data ready/gabungan beberapa metode (2.5 poin) • jenis tugas ML yang akan dikerjakan, contoh: regression/classification/clustering/representation/gabungan beberapa tugas (2.5 poin)
1. Pencarian Data	10	<ul style="list-style-type: none"> • Data yang digunakan terdiri dari minimal 100 sampel dan minimal 2 fitur (2 poin) • Data yang digunakan terdiri dari minimal 300 sampel dan minimal 3 fitur (3 poin) • Menggunakan data yang sama sekali belum pernah dibahas di dalam kelas (5 poin)
2. <i>Exploratory Data Analysis</i> (EDA)	5	<ul style="list-style-type: none"> • Memberikan 1 jenis visualisasi dengan penjelasan kesimpulan yang tepat mengenai data yang digunakan (3 poin) • Memberikan 1 jenis visualisasi tambahan dengan penjelasan kesimpulan yang tepat mengenai data yang digunakan (2 poin)
3. <i>Data Preprocessing</i> (<i>Data Cleaning, Feature Engineering</i>)	5	<ul style="list-style-type: none"> • Menggunakan 1 teknik data preprocessing dan menjelaskan alasan pemilihan teknik yang digunakan dengan tepat (3 poin) • Menggunakan 1 teknik data preprocessing tambahan dan menjelaskan alasan pemilihan teknik yang digunakan dengan tepat (2 poin)
4. Pelatihan, penyetelan model (<i>model tuning</i>)	35	<ul style="list-style-type: none"> • Kesesuaian pemilihan algoritma machine learning dengan masalah yang dihadapi (4 poin) • Membagi data menjadi data untuk pelatihan dan data untuk diuji (2 poin) • Menggunakan K-fold cross-validation (2 poin) • Menggunakan Stratified K-Fold Cross-Validation (2 poin) • Keberhasilan algoritma untuk dijalankan dalam proses training (2 poin) • Keberhasilan algoritma untuk dijalankan dalam proses testing (2 poin) • Pelatihan model dengan 1 algoritma (3 poin) • Pelatihan model dengan 1 algoritma tambahan (3 poin) • Pelatihan model dengan 1 algoritma lainnya (3 poin) • Salah satu algoritma yang digunakan belum pernah dijelaskan di dalam kelas (5 poin) • Melakukan pencarian hyperparameter terbaik untuk algoritma yang digunakan (2 poin) • Menggunakan <i>Grid Search</i> untuk mencari kombinasi <i>hyperparameter</i> terbaik dari sebuah model (5 poin)

Kriteria	Poin Maximal per Kriteria	Opsi Pengumpulan Poin
5. Pengukuran Kinerja atau Evaluasi Model	10	<ul style="list-style-type: none"> Melakukan pengukuran kinerja (performance measure) dengan metode yang tepat dan disertai penjelasan yang tepat (10 poin)
6. Visualisasi	5	<ul style="list-style-type: none"> Terdapat visualisasi hasil (misalnya hasil regresi/klasifikasi/<i>clustering</i>) pada data yang dipakai untuk pelatihan (2 poin) Terdapat visualisasi hasil (misalnya hasil regresi/klasifikasi/<i>clustering</i>) pada data yang dipakai untuk pengujian (2 poin) Memberikan penjelasan kesimpulan dari hasil visualisasi (1 poin)
6. Penulisan <i>Markdown</i>	10	<ul style="list-style-type: none"> Penjelasan meliputi permasalahan yang dihadapi (1 poin) Penjelasan meliputi alasan pemilihan data (1 poin) Penjelasan meliputi <i>Exploratory Data Analysis (EDA)</i> (1 poin) Penjelasan meliputi <i>Data Preprocessing</i> (1 poin) Penjelasan meliputi pelatihan dan penyetelan model yang digunakan (2 poin) Penjelasan meliputi pengukuran kinerja/evaluasi model (2 poin) Penjelasan meliputi visualisasi hasil (1 poin) Penjelasan meliputi kesimpulan yang didapatkan dari proyek (1 poin)
7. Presentasi	10	<ul style="list-style-type: none"> Presentasi yang dilakukan tidak melebihi 10 menit (1 poin) Mempresentasikan proyek dengan baik, dimana durasi presentasi dibagi rata pada semua anggota tim (1 poin) Penjelasan meliputi permasalahan yang dihadapi (1 poin) Penjelasan meliputi alasan pemilihan data (1 poin) Penjelasan meliputi <i>Exploratory Data Analysis (EDA)</i> (1 poin) Penjelasan meliputi <i>Data Preprocessing</i> (1 poin) Penjelasan meliputi pelatihan dan penyetelan model yang digunakan (1 poin) Penjelasan meliputi pengukuran kinerja/evaluasi model (1 poin) Penjelasan meliputi visualisasi hasil (1 poin) Penjelasan meliputi kesimpulan yang didapatkan dari proyek (1 poin)
Bonus	30	<ul style="list-style-type: none"> Menggunakan data yang bentuknya bukan angka (12 poin) Menggunakan 1 teknik data preprocessing lainnya dan menjelaskan alasan pemilihan teknik yang digunakan dengan tepat (2 poin) Salah satu algoritma yang digunakan adalah dari jenis Ensemble Learning (2 poin) Menggunakan <code>cross_val_score</code> (3 poin) Menggunakan <code>transformer</code> data <code>sklearn.pipeline</code> (5 poin) Terdapat pengukuran kinerja (performance measure) tambahan dengan metode yang tepat dan disertai penjelasan yang tepat (3 poin) Mampu menunjukkan bahwa kelompok telah menggunakan Git untuk menyimpan minimal 5 versi kode pemrograman (3 poin)

Penilaian masing-masing individu di dalam sebuah kelompok

Anggota kelompok 1: [NIM], [Nama Lengkap]

Anggota kelompok 2: [NIM], [Nama Lengkap]

Anggota kelompok 3: [NIM], [Nama Lengkap]

Anggota kelompok 4: [NIM], [Nama Lengkap]

	Kontribusi menurut anggota kelompok 1	Kontribusi menurut anggota kelompok 2	Kontribusi menurut anggota kelompok 3	Kontribusi menurut anggota kelompok 4	Rata-rata kontribusi anggota kelompok
Anggota kelompok 1	0-100%	0-100%	0-100%	0-100%	0-100%
Anggota kelompok 2	0-100%	0-100%	0-100%	0-100%	0-100%
Anggota kelompok 3	0-100%	0-100%	0-100%	0-100%	0-100%
Anggota kelompok 4	0-100%	0-100%	0-100%	0-100%	0-100%
	Total = 100%	Total = 100%	Total = 100%	Total = 100%	Total = 100%

Untuk kelompok berjumlah 2 orang:

- Anggota Kelompok dengan kontribusi di bawah 35% hanya akan mendapatkan 85% dari nilai yang didapatkan oleh kelompok tersebut
- Anggota Kelompok dengan kontribusi di bawah 20% hanya akan mendapatkan 70% dari nilai yang didapatkan oleh kelompok tersebut

Untuk kelompok berjumlah 3 orang:

- Anggota Kelompok dengan kontribusi di bawah 23% hanya akan mendapatkan 85% dari nilai yang didapatkan oleh kelompok tersebut
- Anggota Kelompok dengan kontribusi di bawah 11.25% hanya akan mendapatkan 70% dari nilai yang didapatkan oleh kelompok tersebut

Untuk kelompok berjumlah 4 orang:

- Anggota Kelompok dengan kontribusi di bawah 12.5% hanya akan mendapatkan 85% dari nilai yang didapatkan oleh kelompok tersebut
- Anggota Kelompok dengan kontribusi di bawah 10% hanya akan mendapatkan 70% dari nilai yang didapatkan oleh kelompok tersebut