**BABAK PENYISIHAN *DATA COMPETITION***

**INFORMATION SYSTEM FESTIVAL 2022**

**UNIVERSITAS MULTIMEDIA NUSANTARA** | **Waktu Pengerjaan: 23 Mei - 14 Juni 2022**

**KETENTUAN/PETUNJUK SOAL**

Babak penyisihan dilakukan mulai dari tanggal 23 Mei - 14 Juni 2022, dengan batas akhir pengumpulan hasil pengerjaan pada tanggal **14 Juni 2022 pukul 23.59 WIB**. Hasil kerja Peserta dapat dikumpulkan ke *email* [**isfestumn@umn.ac.id**](mailto:isfestumn@umn.ac.id)dengan format:

1. Format nama *file* laporan: NamaTim\_Penyisihan\_LaporanDatcom.pdf
2. Format nama *file* pengolahan data: NamaTim\_Penyisihan\_OutputDatcom.(extension)
3. Subjek *email*: (NamaTim) -Pengumpulan Penyisihan DATA COMPETITION ISFEST 2022

Selamat bekerja dan salam sukses!

**SOAL**

Data Competition ISFEST 2022 bertujuan untuk meningkatkan keterampilan para peserta, yakni mahasiswa, di bidang *data science* dalam rangka memenuhi kebutuhan industri untuk profesi *data scientist*. Dalam lomba Data Competition ISFEST 2022, peserta akan menyelesaikan sebuah studi kasus mengenai tren waktu dan faktor yang mempengaruhi kelulusan mahasiswa. Peserta akan melatih keterampilan machine learning mereka lewat data yang telah tersedia menggunakan beberapa variabel pada dokumen transkrip nilai mahasiswa dan hasil evaluasi dosen per mata kuliah. Pada babak penyisihan, peserta akan membuat ***Laporan Pengolahan Data*** yang kemudian dikumpulkan dalam sebuah *file .*pdf kepada panitia. Format penulisan laporan akan menggunakan format yang telah disediakan dan dilampirkan oleh tim panitia.

Penyelesaian studi tepat waktu untuk kelulusan mahasiswa merupakan salah satu indikator yang dapat digunakan untuk mengukur prestasi belajar mahasiswa. Hal ini sejalan dengan penegasan masa studi dalam peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia (Permen. Menristek pasal 16 no.44, tahun 2015). Waktu studi untuk kelulusan mahasiswa tepat waktu menjadi salah satu indikator yang dapat digunakan untuk mengukur performansi studi mahasiswa suatu perguruan tinggi.

**Ruang Lingkup & Aspek:**

* Define
* Laporan nilai mahasiswa tahun 2010 s/d 20120 dan evaluasi pengajaran dosen tahun 2015 s/d 2020 per semester
* Data mahasiswa untuk prediksi tren kelulusan
* Data evaluasi pengajaran dosen untuk prediksi faktor pengaruh kualitas dosen terhadap kelulusan mahasiswa tepat waktu
* Measure
* Tren kelulusan tepat waktu mahasiswa tidak lebih dari 4 (empat) tahun
* Penetapan target kriteria prosentase kepercayaan/confidence terhadap faktor pengaruh adalah sebesar 80%
* Analyze, berikut merupakan beberapa contoh metode yang dapat diaplikasikan:
* Klasifikasi untuk memprediksi kelulusan tepat waktu mahasiswa
* Klastering untuk memprediksi faktor pengaruh dari kualitas dosen
* Model Asosiasi untuk menguji hasil dari model-model prediksi diatas
* Improvement, berikut merupakan beberapa contoh algoritma yang dapat digunakan:
* Algoritma Decision Tree C4.5 untuk prediksi tren mahasiswa lulus tepat waktu
* Algoritma K-Means untuk memprediksi faktor pengaruh dari kualitas dosen
* Algoritma Apriori, untuk mencari pola dan menguji hubungan antar faktor pengaruh yang diprediksikan
* Control
* Statistik faktor pengaruh dari mata kuliah prasyarat
* Analisis hubungan faktor pengaruh kualitas dosen

**Target Kompetisi:**

* Melakukan proses analisis data menggunakan teknik data mining dalam lingkup EDM (Educational Data Mining) dan membandingkan keberhasilan temuan menggunakan beberapa model prediksi untuk mendapatkan tren waktu studi yang dibutuhkan
* Menemukan seberapa besar pengaruh mata kuliah yang diikuti dan atau kualitas dosen mata kulah untuk kelulusan tepat waktu mahasiswa.
* Membuat prediksi tren kelulusan tepat waktu mahasiswa Program Studi Sistem Informasi (sesuai dengan data yang digunakan).
* Melakukan analisis dan faktor pengaruh kualitas dosen melalui pencapaian nilai akhir mahasiswa yang digunakan sebagai prediksi terhadap kelulusan tepat waktu mahasiswa,
* Melakukan analisis dan menetapkan penemuan faktor-faktor yang sangat berpengaruh terhadap waktu studi mahasiswa dan kesimpulan dari prediksi pengaruh kualitas dosen terhadap kelulusan mahasiswa tepat waktu

**Persyaratan & Ketentuan Tambahan:**

* Berbagi kode program atau data secara pribadi di luar tim tidak diizinkan.
* Tidak juga diperbolehkan menggunakan kode program yang sama dari tim lain.
* Penggunaan data pada kompetisi hanya untuk keperluan kompetisi, peserta tidak diperbolehkan menggunakannya dikemudian hari.
* Peserta menghormati semangat kompetisi dan tidak diperbolehkan untuk merubah data yang disediakan.
* Peserta individu dan tim dapat menggunakan beberapa alternatif perangkat lunak pengolah data seperti:
* Python atau R programming language
* RapidMiner data science tools
* SAS cloud analytical services

ataupun perangkat lunak pengolah data lainnya asalkan peserta atau tim memastikan bahwa mereka memiliki lisensi perangkat lunak tersebut sehingga tidak menyalahi peraturan kompetisi.

**Deskripsi Data:**

1. **Transkrip Nilai Mahasiswa**

Data ini merupakan transkrip nilai mahasiswa per mata kuliah yang diikutinya lengkap dengan kode dan deskripsi mata kuliah (MK) dengan susunan atribut-atributnya sebagai berikut:

(1) NIM/ Nomor Induk Mahasiswa,

(2) Angkatan Tahun Mahasiswa,

(3) Periode Semester/ Semester Name,

(4) Mata Kuliah / Course Name terdiri dari kode dan nama mata kuliah yang disesuaikan dengan masing-masing periode tahun ajaran atau semester dengan contoh sebagai berikut:

i. IS100 Sistem Informasi Manajemen

ii. IS240 Probabilitas dan Statistik

iii.IS341 Sistem Basis Data

iv. IS201 Proses Bisnis Korporat

(5) SKS/ Course Weight Total,

(6) Nilai / Score.

\*\* Periode data yang digunakan adalah setiap semester pada tahun ajaran perkuliahan untuk mahasiswa Angkatan tahun 2010 hingga 2018.

\*\* Contoh: atribut Semester=1211 adalah Semester Ganjil tahun 2012

1. **Nilai Evaluasi Dosen**

Data ini merupakan data hasil evaluasi pengajaran dosen yang disampaikan mahasiswa setiap semester pada tahun ajaran perkuliahan dengan susunan atribut sebagai berikut:

(1) Tahun Ajaran,

(2) Semester (contoh “1521” adalah semester genap tahun ajaran 2015)

(3) Mata Kuliah, terdiri dari kode dan nama mata kuliah lengkap

(4) Pertanyaan, merupakan nomor pertanyaan evaluasi dosen

(5) Keterangan, deskripsi dari pertanyaan yang bersangkutan

(6) Nilai, nilai evaluasi yang diberikan mahasiswa per nomor pertanyaan yang bersangkutan.

Peserta dapat merumuskan Klasifikasi Kualitas Dosen berdasarkan Kompetensi sesuai dengan evaluasi pengajaran yang diatur pada Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 tentang Guru dan Dosen.

\*\* Periode data yang digunakan adalah setiap semester pada tahun ajaran perkuliahan tahun 2015, 2016, 2017 dan 2018.

**Hasil Analisis:**

Berikut merupakan beberapa langkah pengolahan data yang dapat dilakukan oleh peserta dalam menghasilkan output analisis data sesuai dengan harapan luaran hasil kompetisi, yakni:

1. Persiapan Data

* Menentukan variabel menjadi mendukung hipotesis
* Melakukan pembersihan & konstruksi data (jika ada)
* Membuat longitudinal data / data tren yang direncanakan
* Membuat data dummy jika membutuhkan tambahan data
* Membuat analisis kumpulan data yang digunakan sebagai pemahaman data
* Menentukan variabel target dan variabel prediktor
* Memisahkan data untuk pelatihan dan data tes untuk memvalidasi model
* Visualisasi distribusi data dan deksripsi statistiknya

1. Pemodelan & Penambangan Data

* Menentukan kerangka waktu analisis untuk pelatihan model dan validasi
* Membuat analisis korelasi dan pemilihan variabel
* Memilih model penambangan data dengan algoritma yang tepat
* Melakukan validasi dengan mengukur performa model seperti akurasi, presisi, sesitivitas, ROC, AUC, Confusion Matrix dan lainnya.
* Visualisasi model prediksi dengan ukuran performansinya

1. Interpretasi Model

* Menjelaskan pentingnya dari setiap variabel
* Mendefinisikan aksi bisnis yang dapat diimplementasikan berdasarkan hasil dari model
* Visualisasi keseluruhan model dengan hasil prediksi yang diperoleh:
* Tren Kelulusan Mahasiswa setiap Tahun Ajaran
* Prediksi Kelulusan Mahasiswa Tepat Waktu
* Prediksi Faktor Pengaruh Mahasiswa Lulus Tepat Waktu

1. Kesimpulan dan Saran

* Berisi poin-poin kesimpulan yang diperoleh dari tahap-tahap analisa dan model prediksi yang dibuat.
* Kesimpulan harus spesifik menjawab kesesuaian antara tujuan kompetisi, pendekatan yang digunakan, dengan solusi yang ditawarkan (modul/fitur).
* Kesimpulan juga berisi ringkasan hasil implementasi dan pengujian.

\*\* Poin-poin ini juga telah tertera dalam template laporan pengerjaan.

**KRITERIA PENILAIAN**

| **No** | **ASPEK YANG DINILAI** | **ASPEK PENILAIAN** | **BOBOT** | **KRITERIA PENILAIAN** |
| --- | --- | --- | --- | --- |
|
| 1 | LAPORAN | Business Understanding | 5% | Penentuan tujuan, permasalahan, dan kebutuhan dari objek analisis |
| Penentuan parameter batasan dan kriteria keberhasilan/kesuksesan analisis |
| Penentuan strategi atau metode untuk menjawab permasalahan kasus secara sistematis |
| Data Understanding | 10% | Mengidentifikasi isi dan variabel data |
| Memeriksa format dan kelengkapan variabel data |
| Menilai dan mengevaluasi kualitas data |
| Data Preparation | 20% | Menambahkan maksimal 3 variabel dummy dari data panitia (optional) |
| Membersihkan data dengan menghapus/mengisi nilai yang hilang, serta menghilangkan data redundan dan outlier |
| Mengubah dan mengonversi entri nilai variabel agar sesuai dan konsisten |
| EDA | 30% | Melakukan eksplorasi data untuk menemukan informasi tersembunyi pada data (berupa pola, anomali, atau asumsi awal penelitian) |
| Membuat, melampirkan, dan menjelaskan deskripsi statistik dari data |
| Memvisualisasikan variabel dan penyebaran dari data |
| Modeling & Evaluation | 20% | Membagi data menjadi training, testing, ataupun validation set |
| Memilih teknik dan algoritma pemodelan yang tepat dan sesuai dengan tujuan pemodelan |
| Membentuk model dengan algoritma yang dipilih |
| Melakukan kalibrasi atau penyesuaian model untuk memperoleh hasil yang optimal (tidak underfit atau overfit) |
| Menilai kualitas dan efektivitas model |
| Mengevaluasi keberhasilan model dalam memenuhi tujuan dan kriteria keberhasilan yang telah ditetapkan |
| Meninjau dan mengoptimalkan rangkaian proses yang telah dilakukan sebelumnya untuk mengatasi kendala dalam pembentukan model |
| Conclusion & Suggestion | 10% | Menjelaskan rencana pemanfaatan dan pengimplementasian model untuk penyelesaian kasus nyata |
| Sistematika Bahasa & Format Laporan | 5% | Menggunakan Bahasa Indonesia yang baik, benar, dan dapat dipahami. |
| Tidak menggunakan bahasa yang mengandung unsur SARA. |
| Kesesuaian format laporan dengan format yang sudah ditentukan oleh panitia. |