

Maksimum nilai: 100

Kontribusi terhadap nilai akhir: 15% (total semua lab)

Dikumpulkan: Sabtu, 7 Oktober 2023, jam 23.55

Cara pengumpulan: upload ke SCELE pada dropbox yang sudah disediakan dengan penamaan file: Lab1-{NamaKelompok}.zip

Sifat: Kelompok.

Tujuan: untuk mempelajari lebih dalam tentang data preprocessing, modeling, dan melakukan uji coba terhadap *tools*.

Deskripsi Tugas

Dalam tugas ini, kelompok Anda diminta untuk melakukan analisa terhadap dataset yang diberikan menggunakan KNIME dan Jupyter Notebook. Laporan harus juga berisi:

- Penjelasan langkah-langkah yang dilakukan
- Screenshot dari langkah yang dilakukan
- Insight yang kalian dapatkan dari analisa data tersebut.
- Berkas yang relevan: source code

Bagian 1

Context: Dalam tugas laboratorium ini, kita akan menggunakan dataset Telco Customer Churn untuk membangun model klasifikasi yang dapat memprediksi apakah seorang pelanggan akan berhenti berlangganan (churn) berdasarkan demografi, informasi akun, dan penggunaan layanan.

Dataset: Dataset Telco Customer Churn terdiri dari 20 variabel input (customerID, gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, dan TotalCharges) dan satu variabel output (Churn). Variabel Churn adalah variabel binary yang menunjukkan apakah pelanggan berhenti berlangganan atau tidak.

Anda diminta untuk melakukan analisa terhadap data ini menggunakan software **Jupyter Notebook**. *Insight* apa yang bisa didapatkan dari dataset **Telco-Customer-Churn.csv**

Tuliskan dalam laporan:

- Problem statement dan latar belakang permasalahan, business value dari Analytics
- Data understanding: eksplorasi data, distribusi data
- Data preprocessing: handling missing values, normalisasi, dlsb.
- Data modelling: buat beberapa model dengan algoritma yang berbeda
- Uji model akhir pada set pengujian dan laporkan metrik kinerjanya. Jelaskan metode resampling yang dilakukan. Metrics apa yang digunakan. Jelaskan mengapa memilih teknik tsb.
- Interpretasikan hasil model dan tarik kesimpulan tentang faktor-faktor yang memengaruhi churn pelanggan.
- Laporkan juga kekurangan dan tantangan yang kalian hadapi dalam Analisa ini.

Bagian 2

Dalam laboratorium ini, kita akan menggunakan dataset "Online Shoppers Purchasing Intention" untuk melakukan klasifikasi. Dataset ini berisi informasi tentang perilaku browsing pembeli online, termasuk apakah mereka melakukan pembelian atau tidak.

Dataset ini memiliki 12.330 sampel dengan 18 atribut termasuk atribut target. Tujuan dari laboratorium ini adalah untuk membangun model klasifikasi yang dapat memprediksi apakah seorang pembeli online akan melakukan pembelian atau tidak berdasarkan perilaku browsing mereka. Atribut target dalam dataset adalah "Revenue", yaitu variabel biner yang menunjukkan apakah pelanggan melakukan pembelian atau tidak.

Selama laboratorium, Anda akan mempelajari cara memuat dan membersihkan data, memilih atribut yang relevan, dan membangun model klasifikasi yang akurat untuk memprediksi apakah pelanggan akan melakukan pembelian atau tidak. Anda juga akan belajar mengukur kinerja model klasifikasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score.

Analisa dataset ini menggunakan software **Python Jupyter Notebook (Scikit Learn)**. Anda diminta untuk melakukan berbagai macam analisa dan visualisasi yang dapat memberikan *insight*. Jelaskan langkah-langkah data pre-processing-nya. Jelaskan juga parameter-parameter yang digunakan.

Tuliskan dalam laporan hal di bawah ini dan langkah-langkah yang kalian lakukan

- Problem statement dan latar belakang permasalahan, business value dari Analytics
- Data understanding: eksplorasi data, distribusi data
- Data preprocessing

- Data modelling
- Results and Analysis: evaluasi model/pola yang dihasilkan
- Laporkan juga kekurangan dan tantangan yang kalian hadapi dalam Analisa ini.

Bagian 3

Pilih sebuah data yang tersedia bebas dari Kaggle (<https://www.kaggle.com/datasets>) atau UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/datasets.html>). Analisa data tersebut menggunakan aplikasi **KNIME**. Tuliskan dalam laporan:

- Problem statement dan latar belakang permasalahan, business value dari Analytics
- Data understanding: eksplorasi data, distribusi data
- Data preprocessing
- Data modelling
- Results and Analysis: evaluasi model/pola yang dihasilkan

Penalti akan diberikan untuk tindakan plagiarisme.

Selamat Bekerja.

Rubrik Penilaian Lab

Criteria	Excellent (4)	Very Good (3)	Good (2)	Satisfactory (1)
1. Problem Definition and Scope				
Clarity of problem statement and significance	Clear and highly significant problem statement with profound understanding	Clear problem statement with evident significance	Problem statement is clear but lacks significance	Problem statement is unclear or lacks significance
2. Data Acquisition and Understanding				
Data collection methods and sources	Diverse and comprehensive data collection from multiple sources	Comprehensive data collection from relevant sources	Adequate data collection from limited sources	Incomplete or irrelevant data collection
Data exploration and preprocessing	Thorough data exploration with advanced preprocessing techniques	Effective data exploration and preprocessing	Basic data exploration and preprocessing	Minimal data exploration and preprocessing
Description of data and its challenges	Detailed description of data's structure, features, and challenges	Clear description of data's key attributes and challenges	Basic description of data's attributes and challenges	Inadequate or missing data description
3. Data Analysis and Modeling				
Selection of analysis techniques and algorithms	Expert selection of appropriate techniques and algorithms	Proficient selection of relevant techniques and algorithms	Adequate selection of techniques and algorithms	Basic or inappropriate selection of techniques
Implementation of preprocessing and modeling	Rigorous implementation of preprocessing and modeling	Sound implementation of preprocessing and modeling	Basic implementation of preprocessing and modeling	Incomplete or ineffective implementation

Criteria	Excellent (4)	Very Good (3)	Good (2)	Satisfactory (1)
Understanding of chosen models and techniques	In-depth understanding of models and techniques demonstrated	Strong understanding of models and techniques demonstrated	Adequate understanding of models and techniques	Superficial understanding of models and techniques
4. Technical Implementation				
Coding practices and tools	Exemplary coding practices and proficient use of tools	Solid coding practices and effective use of tools	Acceptable coding practices with minor issues	Inconsistent or poor coding practices
Reproducibility of results	Highly organized, well-documented, and reproducible code	Well-organized and documented code	Somewhat organized and documented code	Messy or undocumented code, hindered reproducibility
Model validation and evaluation	Thorough model validation and comprehensive evaluation	Solid model validation and adequate evaluation	Basic model validation and evaluation	Inadequate model validation and evaluation
5. Results and Interpretation				
Presentation of results and visualizations	Exceptional presentation with insightful visualizations	Clear presentation with effective visualizations	Adequate presentation with basic visualizations	Poor presentation or lack of visualizations
Interpretation of model performance metrics	Deep insights drawn from model performance metrics	Clear interpretation of model performance metrics	Basic interpretation of model performance metrics	Superficial or incorrect interpretation
Identification of limitations and improvements	Thorough identification of limitations and innovative improvements	Clear identification of limitations and suggestions for improvements	Basic identification of limitations and potential improvements	Inadequate identification of limitations or improvements
6. Communication				

Criteria	Excellent (4)	Very Good (3)	Good (2)	Satisfactory (1)
Quality of written documentation	Highly organized, clear, and grammatically correct documentation	Well-organized and clear documentation	Adequate documentation with minor issues	Disorganized or unclear documentation
7. Creativity and Innovation				
Originality and novel approaches	Highly original and innovative approaches demonstrated	Demonstrated originality and some innovative approaches	Limited originality with basic creativity	Lack of originality or creativity
Exploration of advanced techniques	In-depth exploration of advanced techniques beyond requirements	Solid exploration of advanced techniques	Basic exploration of advanced techniques	Minimal exploration of advanced techniques
Critical thinking and proposing creative solutions	Exceptional critical thinking and creative solutions proposed	Evident critical thinking and creative solutions proposed	Some critical thinking and basic creative solutions	Limited critical thinking and lack of creative solutions