

Capstone Project Module 3

Setelah menyelesaikan seluruh materi modul 3, siswa diharapkan sudah memiliki pengetahuan fundamental terkait pemodelan machine learning dan kegunaannya. Untuk dapat mengimplementasikan materi yang telah dipelajari, setiap siswa wajib mengerjakan dan menyelesaikan *capstone project* modul 3.

Pengerjaan *capstone project* modul 3 ini akan melatih siswa dalam melakukan analisis data serta membangun suatu model machine learning yang sesuai dengan data dan kebutuhan. Tentunya, setiap siswa harus mengerjakan *project* ini secara *end-to-end*, meliputi dari perumusan masalah, tujuan yang ingin dicapai, limitasi model yang dibuat, hingga kesimpulan dan rekomendasi yang dapat diberikan terkait dengan *project* yang sudah dikerjakan oleh masing-masing siswa.

Poin Penilaian

Untuk dapat dinyatakan lulus pada **Capstone Project Module 3** ini, setiap siswa dibebaskan untuk memilih algoritma *machine learning* yang sesuai dengan permasalahan yang dihadapi, dengan catatan sebagai berikut:

Siswa harus dapat menjelaskan cara kerja algoritma yang digunakan sebagai *final model* dan juga mendapatkan skor pada *evaluation metrics* **setinggi** mungkin.

Berikut adalah poin-poin penilaian untuk **Capstone Project Module 3** beserta penjelasan detailnya:

- | | |
|---|------|
| - Documentation & Video | : 25 |
| - Business Problem & Data Understanding | : 10 |
| - Data Cleaning, Feature selection, & Feature Engineering | : 15 |
| - Analytics (Algorithm, Evaluation Metrics) | : 40 |
| - Conclusion and Recommendation/Improvement | : 10 |

Documentation & Video (25%)

- Siswa wajib membuat dokumentasi *capstone project* secara lengkap pada file **Jupyter Notebook**.
- Siswa wajib membuat sebuah video penjelasan berdurasi maksimal 15 menit mengenai *capstone project* yang sudah dikerjakan. Konten dari video ini dapat mencakup permasalahan bisnis yang dihadapi, tujuan yang ingin dicapai, analytic approach yang dilakukan (regresi atau klasifikasi), *evaluation metrics* apa yang difokuskan berdasarkan permasalahan bisnis, metode preprocessing yang dilakukan, penjelasan cara kerja algoritma yang digunakan, hingga kesimpulan dan rekomendasi terkait dari *project* yang sudah diselesaikan.

Business Problem & Data Understanding (10%)

- Siswa memahami bagaimana implementasi penggunaan model ML nantinya, seperti siapa yang akan pakai model ML-nya, dan kapan digunakannya.
- Siswa harus memahami siapa (*stakeholder*) yang mengalami masalah, masalahnya apa (*specific, measurable*), mengapa masalahnya penting untuk dipecahkan (*specific, measurable*), dan apa tujuan/target dari penyelesaian masalah tersebut (*specific, measurable, achievable*). Disarankan untuk melakukan *research* lebih lanjut terhadap *problem* dan domainnya bila belum menguasai.
- Siswa harus memahami data apa yang dimiliki (setiap kolom dan barisnya merepresentasikan apa), sumber datanya dari mana, siapa pembuat datanya, kapan dibuat datanya, serta memahami apakah data yang dimiliki bisa digunakan dan berhubungan untuk menyelesaikan masalah (selengkap mungkin sesuai ketersediaan informasinya).

Data Cleaning, Feature selection, & Feature Engineering (15%)

- Siswa harus bisa memastikan data yang dimiliki sudah bersih dan siap digunakan di tahap selanjutnya (*Analytical Approach*) dengan menggunakan metode yang sudah diajarkan.
- Siswa juga harus bisa menjelaskan alasan mengapa melakukan setiap metode atau tahapan dalam proses pembersihan dan penyiapan data ini. Metode yang dimaksud meliputi metode-metode untuk *feature selection* maupun *feature engineering*.

Analytics (Algorithm, Evaluation Metrics) (40%)

- Analisa data: Siswa mampu menyampaikan *insight* (temuan) hasil analisis sesuai dengan masalah yang telah didefinisikan. Temuan yang disampaikan membantu menyelesaikan masalah yang ditetapkan di awal.
- Pemodelan:
 - Siswa mampu menjelaskan cara kerja algoritma model *machine learning* (ML) yang digunakan.
 - Siswa memahami kapan model ML yang dibangun hasilnya dapat dipercaya dan kapan tidak dapat dipercaya (limitasi model).
 - Siswa harus memahami data-data seperti apa model bisa dipercaya (akurat) dan data-data seperti apa model masih belum dapat dipercaya (masih kurang akurat).
- *Evaluation metrics*:
 - Siswa mampu menjelaskan cara kerja *evaluation metrics* yang digunakan, misalnya: Regresi (MAE, MSE, RMSE, MAPE, MPSE, MSLE, R-squared), Klasifikasi (accuracy, recall, precision, F1 Score, ROC AUC, PR Score), dan lainnya.
 - Siswa mampu memilih *evaluation metrics* dan mampu menjelaskan alasan pemilihannya, serta keterhubungan *evaluation metrics* dengan aspek bisnis. Selain itu, siswa dapat membuat *metric* baru tambahan sendiri bila memungkinkan dan diperlukan.

Conclusion and Recommendation/Improvement (10%)

- Siswa mampu memberikan konklusi akhir untuk solusi yang dibuat dari masalah yang telah didefinisikan. Solusi yang ditawarkan mampu diukur pencapaiannya (measurable).
- Siswa mampu memberikan rekomendasi kapan model baik untuk digunakan dan tidaknya (kapan model dapat dipercaya dan kurang dapat dipercaya).
- Siswa memahami dampak dari implementasi solusi terhadap proses bisnis (measurable).
- Siswa memahami batasan dari proyek mereka baik dari datanya, modelnya, maupun waktunya, sehingga bisa memberikan rekomendasi apa saja yang dapat dilakukan ke depannya untuk memperbaiki *project*-nya atau meningkatkannya menjadi lebih baik lagi.

Dataset

Berikut adalah daftar dataset yang dapat digunakan untuk **Capstone Project Module 3**:

- [Saudi Arabia Used Cars](#)
- [California Housing Price](#)
- [Bike Sharing](#)
- [Apartment Data](#)
- [Customer Lifetime Value](#)
- [Hotel Booking Demand](#)
- [Telco Customer Churn](#)
- [Bank Marketing](#)
- [Ecommerce Customer Churn](#)
- [Travel Insurance](#)

Teknis Pengerjaan

Setiap siswa akan mendapatkan dataset sesuai dengan yang dipilih. Namun, dataset yang didapat hanyalah data *training*. Di mana, siswa dapat menggunakan data *training* ini untuk membangun model. Nantinya, model yang dikumpulkan akan diuji dengan menggunakan data *test* yang dimiliki oleh tim penilai. Pada akhirnya, kualitas prediksi yang dihasilkan oleh model akan menjadi salah satu komponen penilaian.

Sebagai catatan tambahan, siswa tidak diwajibkan untuk menyertakan *Exploratory Data Analysis* (EDA) pada file **Jupyter Notebook**. Namun, jika ingin menyertakan pun tidak menjadi masalah, tetapi tidak masuk ke dalam poin penilaian.

Batas Waktu Pengumpulan

Setiap siswa diberikan batas waktu pengumpulan **Capstone Project Module 3** maksimal **7 hari kerja** setelah ujian modul 3. Dalam hal ini, diharapkan siswa dapat memenuhi seluruh komponen penilaian yang sudah dijabarkan di bagian 'Poin Penilaian'.

Berikut adalah teknis pengumpulan **Capstone Project Module 3**:

1. *Upload* video penjelasan *project* pada *cloud storage* masing-masing. Jangan lupa untuk memberikan hak akses untuk publik.
2. Upload beberapa file ke akun GitHub masing-masing, dengan rincian:
 - Buat file *readme* pada *repository* GitHub yang berisikan **Project Introduction**.
 - File **Jupyter Notebook** sebagai dokumentasi tertulis pengerjaan *project*.
 - File model yang sudah disimpan melalui Pickle.
3. Menginformasikan tautan data tersebut ke form yang telah disediakan oleh operasional.