

# Final Project Stage 3



# Brief & Challenge

The final project is a comprehensive exercise designed to help students **apply their learning to a real-world scenario** and to leverage their knowledges to **solve a problem using 4 prepared datasets.**

They must **extract insights** and create **solutions with potential business or industrial impact.**

[\(Link Datasets\)](#)

## Dataset 1: E-commerce Sales Data

The goal is to derive insights related to customer segmentation, sales forecasting, recommendation systems, or optimizing product pricing.

## Dataset 2: Banking Dataset

It's ideal for building predictive models that assist in Credit Risk Assessment, Fraud Detection, Churn Prediction, etc.

## Dataset 3: Travel and Tourism Data

It can be used for predict Popular Destinations and Travel Trends, Traveler Behavior, Seasonal and Geographic Demand Patterns, etc.

## Dataset 4: Human Resource Dataset

It's commonly used for recruitment rate, average salary in company, Employee Performance, etc.



# Teknis Pengerjaan

**Estimasi Waktu Pengerjaan**

**3 - 5 jam**

**Jumlah Soal**

**5 Soal**

**Total Point**

**200 poin**

Pekerjaan dilakukan secara berkelompok, sesuai kelompok  
Final Project

Masing-masing anggota kelompok tetap perlu submit ke LMS  
(jadi bukan perwakilan)

Upload hasil pengerjaan melalui LMS. Masukkan semua file ke  
dalam 1 file dengan format ZIP.

Nama File: Stage 3 - <Nama Kelompok>.zip

## Stage 3: Model Validation and Evaluation

### 100 Poin

1. **Final Model Testing:** Apply the final trained models to the test dataset. Evaluate the results on unseen data to measure how well the model generalizes.
2. **Discuss Model Evaluation:** Analyze the test results using evaluation metrics. Compare these with the validation results to check for overfitting or underfitting.
3. **Error Analysis:** Perform error analysis to identify areas where the model struggles. For example, review cases where predictions are most inaccurate. Discuss with the group what may be causing these errors.
4. **Adjustments (Poin++):** *If necessary, adjust the model or parameters based on the insights from the error analysis. Re-run evaluations if any significant changes are made.*
5. **Final Report Drafting:** Begin drafting a report on model performance, including a summary of the model validation, key metrics, and areas for improvement



### Common Pitfalls

Wrong evaluation metrics, lack of cross-validation.

#### Tools:

- Scikit-learn, Tableau for comparing results.

#### Output: 100 Poin

- A model evaluation report covering metrics such as accuracy, precision, recall, F1-score, or MSE (for regression).

## 1. Final Model Testing 20 Poin

### Deskripsi

Terapkan model yang telah dilatih pada dataset uji. Dataset uji adalah data yang belum pernah digunakan dalam pelatihan atau validasi, sehingga penting untuk mengukur seberapa baik model dapat menggeneralisasi pada data baru. Evaluasi hasil prediksi untuk mendapatkan pemahaman tentang performa model di dunia nyata.

### Tugas:

### Langkah-langkah:

1. Gunakan model terbaik yang telah dilatih untuk memprediksi hasil pada test dataset.
2. Bandingkan hasil prediksi model dengan nilai sebenarnya (ground truth) dalam dataset uji.
3. Catat hasil metrik evaluasi untuk dianalisis lebih lanjut.

### Tools:

- **Scikit-learn (Python):** Untuk mengimplementasikan prediksi dan metrik evaluasi.
- **Google Colab / Jupyter Notebook:** Untuk menjalankan kode dan menampilkan hasil.

## 2. Discuss Model Evaluation 25 Poin

### Deskripsi

Analisis hasil pengujian model menggunakan metrik evaluasi yang sesuai, seperti accuracy, precision, recall, F1-score, atau RMSE, tergantung pada jenis tugas. Bandingkan hasil dari dataset uji dengan hasil validasi untuk memeriksa adanya overfitting atau underfitting. Pastikan evaluasi model ini dapat **menjawab metrik kesuksesan yang telah ditetapkan di awal proyek**, sehingga performa model sesuai dengan tujuan yang ingin dicapai.

### Tugas:

### Langkah-langkah:

1. Evaluasi hasil dari dataset uji menggunakan metrik yang relevan.
2. Bandingkan performa pada dataset uji dengan dataset validasi.
3. Diskusikan dengan tim apakah model mengalami overfitting, underfitting, atau apakah sudah dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat.

### Tools:

- **Scikit-learn (Python):** Untuk menghitung metrik seperti precision\_score, recall\_score, f1\_score, dan mean\_squared\_error.
- **Google Colab / Jupyter Notebook:** Untuk memvisualisasikan hasil evaluasi.

### 3. Error Analysis 25 Poin

#### Deskripsi

Lakukan analisis kesalahan untuk mengidentifikasi area di mana model masih kesulitan. Analisis kesalahan dapat mencakup peninjauan kasus di mana prediksi model paling tidak akurat, serta mendiskusikan dengan kelompok penyebab utama dari kesalahan tersebut (misalnya, apakah ada pola yang tidak ditangkap oleh model atau data yang tidak cukup representatif).

#### Tugas:

#### Langkah-langkah:

1. Identifikasi kasus di mana model membuat prediksi yang keliru atau hasil metriknya rendah.
2. Diskusikan apakah kesalahan ini disebabkan oleh kurangnya data, noise dalam data, atau model yang tidak sesuai.
3. Tinjau kembali fitur atau preprocessing data yang mungkin mempengaruhi kinerja model.

#### Tools:

- **Pandas (Python):** Untuk meninjau prediksi yang salah dan melakukan analisis lebih mendalam.
- **Matplotlib / Seaborn:** Untuk membuat visualisasi kesalahan dan melihat pola yang muncul dari prediksi yang salah.

## 4. Adjustments (Poin++) 20 Poin

### Deskripsi

Jika diperlukan, lakukan penyesuaian terhadap model atau hyperparameter berdasarkan hasil dari analisis kesalahan. Misalnya, jika ditemukan bahwa model kurang mampu menangani fitur tertentu, pertimbangkan untuk menambahkan fitur baru atau menggunakan model yang lebih kompleks. Setelah penyesuaian dilakukan, lakukan kembali evaluasi untuk melihat dampak dari perubahan yang dibuat.

### Tugas:

### Langkah-langkah:

1. Berdasarkan analisis kesalahan, diskusikan apakah perlu ada penyesuaian model atau preprocessing data.
2. Lakukan penyesuaian hyperparameter atau tambahkan fitur baru jika diperlukan.
3. Jalankan kembali evaluasi model setelah penyesuaian untuk mengukur perbaikan performa.

### Tools:

- **Scikit-learn (Python):** Untuk melakukan tuning hyperparameter dan melatih ulang model.
- **TensorFlow / Keras:** Jika menggunakan model neural networks dan perlu melakukan penyesuaian arsitektur atau parameter.



## 5. Final Report Drafting 10 Poin

### Deskripsi

Mulailah menyusun laporan tentang performa model yang mencakup ringkasan hasil validasi model, metrik kunci yang dihasilkan, serta area yang masih perlu diperbaiki. Laporan harus mencakup semua tahapan mulai dari pemilihan model, pelatihan, evaluasi, dan penyesuaian. Dokumentasikan juga setiap tantangan yang dihadapi serta solusi yang diambil.

### Tugas:

### Langkah-langkah:

1. Ringkas hasil evaluasi model, termasuk metrik kunci seperti accuracy, precision, recall, atau RMSE.
2. Sertakan analisis kesalahan dan setiap langkah penyesuaian model yang telah dilakukan.
3. Berikan saran untuk perbaikan lebih lanjut atau bagaimana model dapat diterapkan dalam kasus nyata.
4. Pastikan laporan bersifat ringkas dan mudah dipahami oleh pembaca yang non-teknis.

### Tools:

- **Google Docs / Microsoft Word:** Untuk menyusun laporan secara kolaboratif dengan tim.
- **GitHub:** Untuk menyimpan laporan dan dokumentasi proyek.
- **Markdown:** Untuk membuat dokumentasi laporan di GitHub dalam format yang rapi.

# Mekanisme Pengumpulan

1. Perhatikan setiap **OUTPUT** yang harus diselesaikan setiap stage nya
2. Buatlah report progress setiap STAGE dengan TEMPLATE berikut:  
**<https://bit.ly/report-progress-stage-finpro-BC>**
3. Masukkan semua hasil diskusi se **DETAIL-DETAIL** nya pada REPORT PROGRESS STAGE
4. Submit doc REPORT PROGRESS STAGE dalam **format PDF**
5. Setiap peserta tidak harus membuat satu REPORT
6. REPORT dibuat per **kelompok** tapi **SETIAP PESERTA WAJIB SUBMIT di LMS**



# Thank you!