

# INITIAL ASSESSMENT DATA ANALYTICS

Oleh Nur Kholifah

Pada esai ini, saya akan menjelaskan mengenai langkah-langkah pada *business credit scoring* dan *dashboard* monitoringnya. *Credit scoring* merupakan analisis prediktif untuk membuat laporan kredit konsumen yang akan digunakan untuk menentukan kelayakan kredit dari konsumen (Irby, 2022). *Credit scoring* akan menghasilkan suatu nilai tunggal, yakni *credit score* atau resiko pinjaman. Semakin tinggi *credit score*, semakin besar kemungkinan pelanggan membayar kembali pinjaman mereka. Suatu bisnis menggunakan sistem *credit scoring* untuk membuat keputusan pembiayaan, memasarkan ke pelanggan baru, memelihara akun, dan membuat proyeksi keuangan. Kelebihan dari *credit scoring* adalah kecepatannya dalam memproses dan hasilnya yang cukup akurat. Kecepatan dan akurasi dari *credit scoring* menjadikan *credit score* sebagai alat manajemen resiko dasar di berbagai sektor seperti perbankan, telekomunikasi, asuransi, dan ritel (Altair, 2022a).

Langkah-langkah dalam *business credit scoring* adalah (Algoritma Data Science School, 2021):

1. *Business understanding*

Pada tahap ini pelaku bisnis perlu mengetahui mengenai kebutuhan dari sistem *business credit scoring engine*. Dari hal tersebut nantinya akan diketahui data yang diperlukan untuk membuat sistem *business credit scoring*. Data yang baik sangat menentukan hasil dari pemodelan prediktif yang akan didapatkan. Untuk menjamin hal itu, diperlukan langkah selanjutnya yakni, *cleaning data*.

2. *Cleaning data*

Terdapat dua tahapan dalam proses *cleaning data*, yakni ***check missing value*** dari sini kita akan mengetahui data yang kita peroleh sudah lengkap atau belum atau mungkin ada data yang hilang. Terdapat beberapa cara yang dapat digunakan untuk meng-*handle missing value*. Beberapa diantaranya yaitu, menggunakan cara *take out variable* atau membuang variabel yang memiliki jumlah observasi *missing values* sangat besar dan cara *complete case* atau hal yang dilakukan setelah membuang kolom variabel yang memiliki jumlah *missing value* yang cukup banyak akan dilanjutkan dengan membuang baris-baris yang memiliki *missing value* karena jumlah observasinya tidak terlalu banyak. Tahap selanjutnya dari *cleaning data* adalah **menyesuaikan tipe data**. Pada tahap ini, variabel-variabel yang tipe datanya belum sesuai perlu disesuaikan. Contohnya untuk data-data yang kategorik yang sebelumnya memiliki tipe data karakter akan diubah menjadi tipe faktor.

3. *Exploratory data*

Pada tahap ini akan digunakan visualisasi dari data kategorik maupun data numerik untuk mengetahui sebaran dari data tersebut. Dari hasil visualisasi tersebut, variabel yang dirasa tidak membantu dalam modeling akan dibuang.

4. *Cross validation*

Pada tahap ini, data yang ada akan dibagi menjadi dua yakni data train dan data test. Data train biasanya jumlahnya lebih banyak dari data test, data train akan digunakan untuk modeling dan data test akan dijadikan sebagai evaluasi.

5. Modelling

Pada tahap modeling, dapat dilakukan perbandingan atau meng-*compare* beberapa model yang dapat digunakan untuk data yang ada. Misalnya menggunakan perbandingan dengan mengimplementasikan model *random forest* dan XGBoost akan dicari tahu model yang mempunyai performa yang lebih baik.

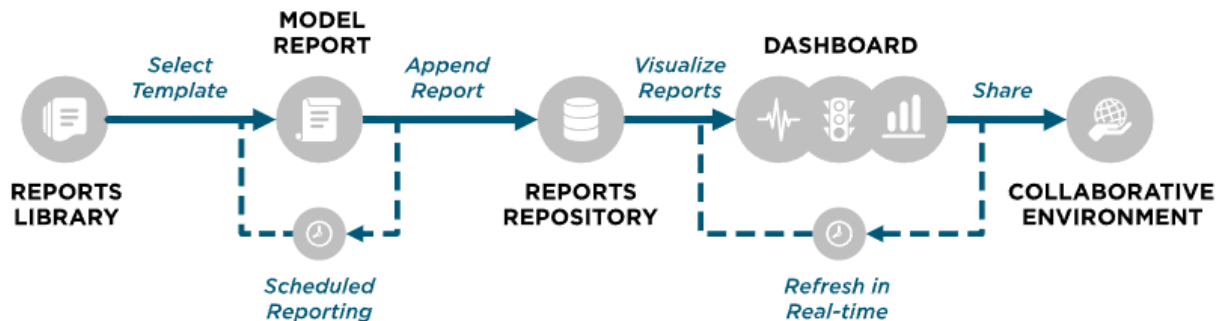
6. Evaluasi model

Akan dibandingkan matriks evaluasi dari model yang telah diimplementasikan sebelumnya. Matriks yang akan difokuskan adalah *recall* karena diinginkan meminimalisir hasil prediksi berdasarkan data aktualnya. Akan dicari model nilai dengan *recall* terbesar yang akan digunakan untuk membuat *credit scoring*.

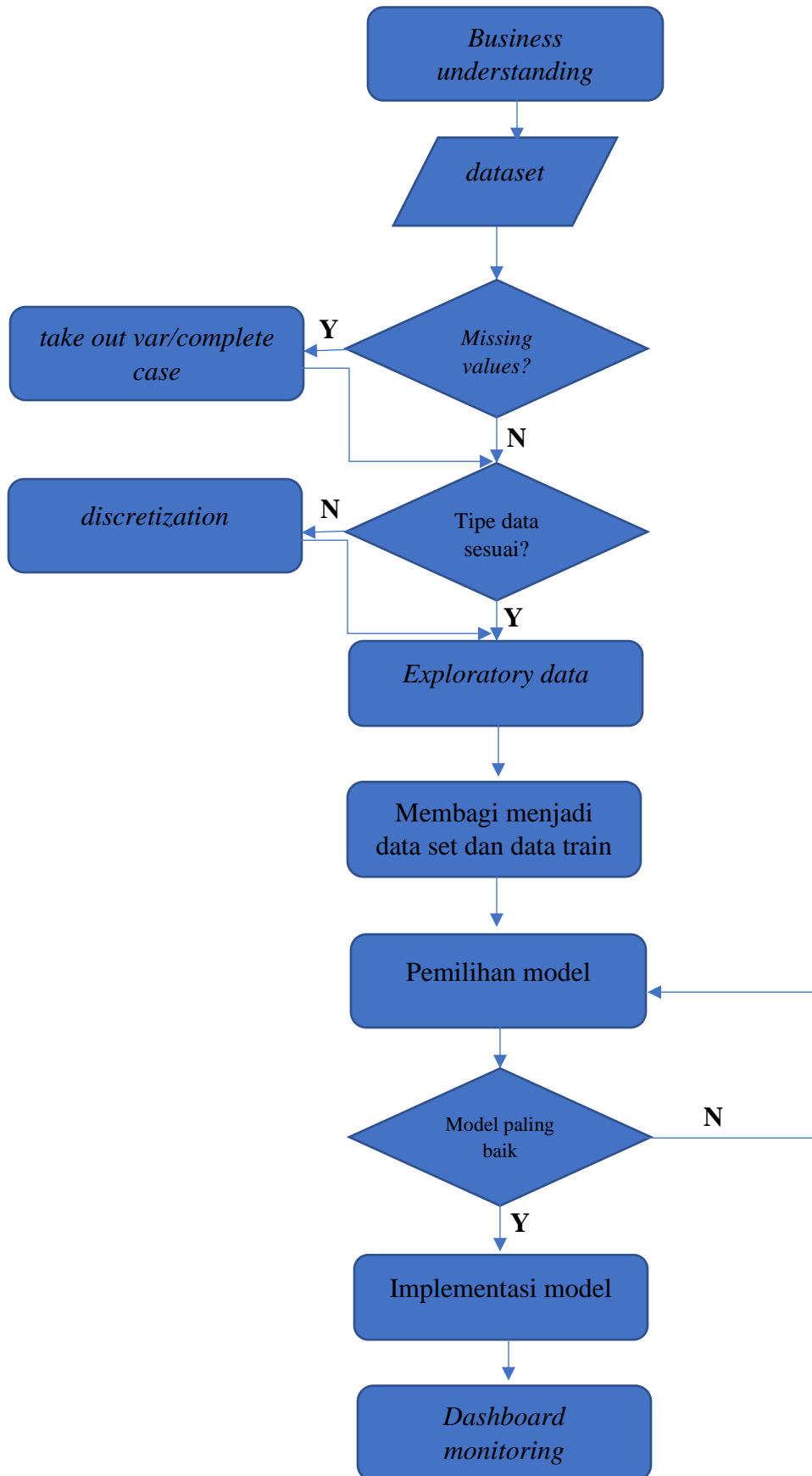
7. Interpretasi model

Berdasarkan tahapan-tahapan sebelumnya akan didapatkan informasi mengenai variabel-variabel yang paling berpengaruh dan paling penting dalam model yang digunakan. Dari model yang ada dapat dimasukkan data tiap nasabah untuk mengetahui apakah nasabah termasuk ke dalam *good credit* atau *bad credit* sehingga dapat dijadikan untuk bahan pertimbangan apakah dapat diberikan pinjaman.

Selanjutnya, *dashboard monitoring* adalah bagian dari *post implementation* dari *credit scoring*. Di dalamnya mencakup monitoring model secara *real time* untuk menentukan apakah model terus sejalan dengan kinerja yang diharapkan (Altair, 2022b). *Dashboard monitoring* dapat menampilkan beberapa grafik serta metrik dari kinerja model yang digunakan.



Gambar 1. Proses monitoring model



Gambar 2. Flowchart data scoring system

## DAFTAR PUSTAKA

Algoritma Data Science School (Director). (2021, July 12). *CREDIT SCORING / Algoritma Show / Episode #1 / Algoritma 2021*. <https://youtu.be/L7564DMRcdY>

Altair. (2022a, May 16). *Credit Scoring Series Part One: Introduction to Credit Scoring*.  
<https://www.altair.com/newsroom/articles/credit-scoring-series-part-one-introduction-to-credit-scoring>

Altair. (2022b, August 5). *Credit Scoring Series Part Nine: Scorecard Implementation, Deployment, Production, and Monitoring*.  
<https://www.altair.com/newsroom/articles/credit-scoring-series-part-nine-scorecard-implementation-deployment-production-and-monitoring>

Irby, L. (2022, April 7). *What Is Credit Scoring?* <https://www.thebalancemoney.com/what-is-credit-scoring-5194416>