

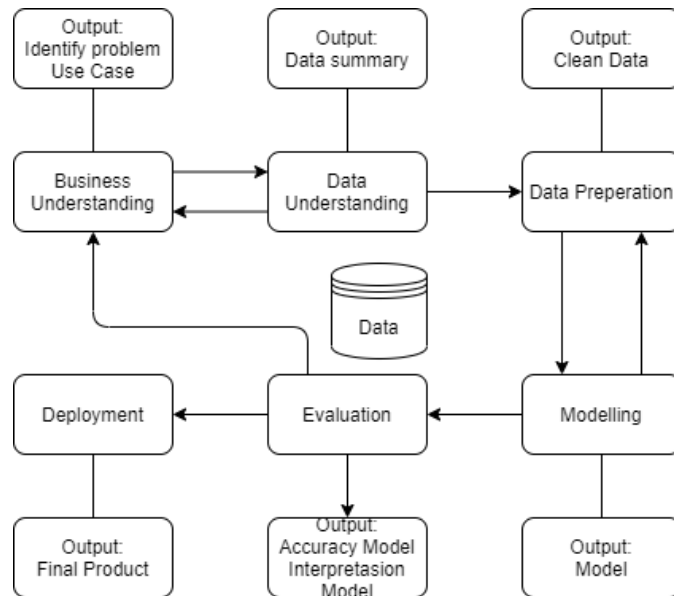
Anindya Dewi Maharani

Universitas Sriwijaya

### **Use-Case Bank Business Credit Scoring Engine Development**

Dalam dunia yang serba digital semua lini kehidupan kini sudah bersinggungan dengan data. Tak terkecuali dalam bidang perbankan, data berperan sangat penting dalam proses berjalannya perbankan. Salah satu data yang terdapat di perbankan adalah data calon nasabah penerima pinjaman. Sekarang dengan memanfaatkan teknologi penentuan kelayakan penerima pinjaman bank dapat dilakukan dengan menerapkan ilmu *Machine Learning*. *Machine Learning* sebagai alat bantu berguna untuk menganalisa kelayakan calon penerima pinjaman bank menjadi lebih cepat dan valid sehingga waktu yang ada dapat lebih dimaksimalkan untuk melakukan hal lain yang lebih penting.

Dalam melakukan *engine development* memiliki runtutan langkah agar produk yang dihasilkan merupakan versi yang seterbai mungkin. Langkah-langkah ini juga akan memudahkan dalam proses pembuatan menjadi lebih rapih dan terstruktur. Langkah ini disebut juga dengan istilah CRISP-DM (*Cross Industry Standard Process For Data Mining*). CRISP-DM pertama kali diperkenalkan pada tahun 1996 sebagai hasil kerjasama dari 5 perusahaan teknologi yaitu ntegral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA. Tahapan yang dilakukan dalam *engine development* tergambar dalam bagan di bawah:



**Gambar 1** Tahapan CRISP-DM

### 1. *Business Understanding*

Tahap ini adalah tahap pertama yang sangat vital. Pada tahap ini dibutuhkan pemahaman mengenai objek bisnis dengan tujuan utama menyelaraskan tujuan model untuk tujuan bisnis sehingga model yang dibangun tepat sasaran.

*Output* dari tahap ini adalah masalah yang teridentifikasi atau *use case*. Identifikasi masalah dilakukan dengan pendekatan logis sehingga menghasilkan pertanyaan mendasar.

Untuk *use-case bank business credit scoring* business understanding dilakukan untuk memahami masalah yang dialami bank. Salah satu produk bank adalah pemberian pinjaman, namun sering kali ada nasabah yang mengalami tunggakan pembayaran. Maka dapat diidentifikasi masalahnya yaitu bank mengalami kesulitan dalam penentuan kelayakan penerima pinjaman untuk nasabah. Solusinya adalah dibutuhkan suatu penentuan *score* secara *objective* untuk menilai kelayakan nasabah dalam menerima pinjaman.

## 2. *Data Understanding*

*Data understanding* harus dilakukan untuk menemukan masalah pada data. Esensinya pada tahap ini berguna untuk memahami data yang dimiliki. Hal-hal yang mungkin ditemukan dalam tahap ini adalah distribusi data yang tidak normal, *outlier* dsb.

Output dari tahap ini adalah data summary mulai dari atribut apa saja yang akan digunakan dan masalah-masalah mengenainya.

Untuk *use-case bank business credit data understanding* dilakukan dengan menyiapkan data yang berisi karakteristik nasabah. Kemudian dilakukan *scanning* apakah data tersebut sudah sesuai dengan yang dibutuhkan. Data tersebut bisa memuat atribut-atribut yang dapat berguna untuk penentuan *score* seperti: Nama\_Nasabah, Jenis\_Kelamin, Umur, Alamat\_Nasabah, Pekerjaan, Jumlah\_Pinjaman, Jangka\_Waktu, Angsuran\_PerBulan, Tipe\_Pinjaman, Jaminan\_Kredit, Norek\_Kredit, Tujuan\_Kredit, Status\_Kredit dan atribut lainnya yang mungkin diperlukan.

## 3. *Data Preperation*

Pada dasarnya tahap ini berguna untuk memperbaiki masalah yang sebelumnya ditemukan pada tahap *Data Understanding*. Langkah-langkah yang mungkin dilakukan dalam tahap ini seperti membuang *outlier*, mengubah tipe data, menghapus duplikasi data dsb.

*Output* dari tahap ini adalah data bersih yang siap pakai untuk selanjutnya dilakukan pemodelan di tahap berikutnya.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DIFFWalking	Sex	AgeCategory	Race	Diabetic	PhysicalActivity	GentHealth	SleepTime	Asthma	KidneyDis	SkinCance	HeartDisease
1	16.6	Yes	No	No	3.0	30.0	No	Female	55-59	White	Yes	Yes	Very good	5.0	Yes	No	Yes	No
2	20.3	No	No	Yes	0.0	0.0	No	Female	80 or older	White	No	Yes	Very good	7.0	No	No	No	No
3	26.6	Yes	No	No	20.0	30.0	No	Male	65-69	White	Yes	Yes	Fair	8.0	Yes	No	No	No
4	23.7	No	No	No	28.0	0.0	Yes	Female	40-44	White	No	Yes	Very good	8.0	No	No	No	No
5	28.9	Yes	No	No	6.0	0.0	Yes	Female	75-79	Black	No	No	Fair	12.0	No	No	No	Yes
6	21.6	No	No	No	15.0	0.0	No	Female	70-74	White	No	Yes	Fair	4.0	Yes	No	Yes	No
7	31.6	Yes	No	No	5.0	0.0	Yes	Female	80 or older	White	Yes	No	Good	9.0	Yes	No	No	No
8	26.5	No	No	No	0.0	0.0	No	Female	80 or older	White	No, borde	No	Fair	5.0	No	Yes	No	No
9	34.3	Yes	No	No	30.0	0.0	Yes	Male	60-64	White	Yes	No	Poor	15.0	Yes	No	No	Yes
10	28.7	Yes	No	No	0.0	0.0	No	Female	55-59	White	No	Yes	Very good	5.0	No	No	No	No
11	28.4	Yes	No	No	0.0	0.0	Yes	Male	75-79	White	Yes	Yes	Very good	8.0	No	No	No	No
12	28.2	No	No	No	7.0	0.0	Yes	Female	80 or older	White	No	No	Good	7.0	No	No	No	No
13	29.3	Yes	No	No	0.0	30.0	Yes	Female	60-64	White	No	No	Good	5.0	No	No	No	No
14	29.2	No	No	No	1.0	0.0	No	Female	50-54	White	No	Yes	Very good	6.0	No	No	No	No
15	26.3	No	No	No	5.0	2.0	No	Female	70-74	White	No	No	Very good	10.0	No	No	No	No
16	22.6	Yes	No	No	0.0	30.0	Yes	Male	70-74	White	No, borde	Yes	Good	8.0	No	No	No	No
17	29.9	Yes	No	No	0.0	0.0	Yes	Female	75-79	Black	Yes	No	Fair	5.0	No	Yes	No	No
18	18.1	No	No	No	0.0	0.0	No	Male	80 or older	White	No	Yes	Excellent	8.0	No	No	Yes	No
19	21.2	No	No	No	0.0	0.0	No	Female	80 or older	Black	No, borde	No	Good	8.0	No	No	No	No
20	28.9	No	No	No	2.0	5.0	No	Female	70-74	White	Yes	No	Very good	7.0	No	No	No	No
21	26.2	Yes	No	No	0.0	15.0	No	Female	45-49	White	No	Yes	Very good	6.0	No	No	No	No
22	25.8	Yes	No	No	0.0	30.0	No	Male	80 or older	White	Yes	Yes	Fair	8.0	No	No	No	No

**Gambar 2** Contoh hasil cleaning data

Untuk *use-case bank business credit scoring data preparation* dilakukan dengan membersihkan data mentah yang kemudian akan diolah. Langkah pembersihan data terbagi menjadi 4:

- Data Cleaning

Melalui tahap ini kolom-kolom yang kosong dihapus, data yang inkonsisten diperbaiki, dan *noisy* data dihaluskan.

- Data Integration

Dalam tahap ini data yang memiliki perbedaan representasi disatukan menjadi satu format. Misalnya perbedaan penulisan format jenis kelamin.

- Data Transformation

Data kemudian akan dinormalisasikan dan pemilihan atribut. Tidak selalu semua atribut dalam data digunakan cukup pilih atribut yang kira-kira dibutuhkan dalam penentuan *credit score* saja. Data juga akan dinormalisasikan sehingga data tidak ada yang berlebihan. Transformasi data juga perlu dilakukan agar bentuknya sesuai dengan yang dibutuhkan saat proses modelling. Tahap ini juga melakukan penentuan label dimana dalam kasus ini labelnya adalah atribut `Status_Kredit`.

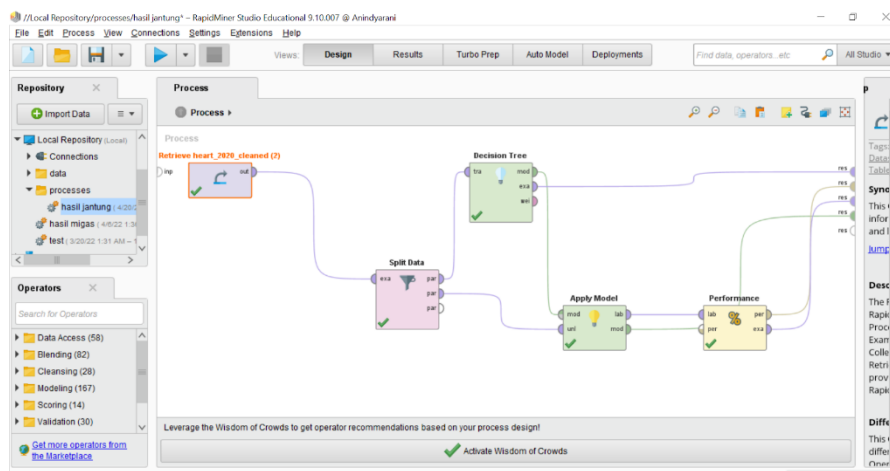
- Data Reduction

Volume data yang besar tidak selamanya menguntungkan. Melakukan modelling dengan volume data yang besar akan menyulitkan dalam proses analisis dan tidak efisien dalam penyimpanan data. Maka perlu dilakukan *data reduction* agar proses modelling menjadi efisien dan menekan biaya penyimpanan dan analisis data.

#### 4. Modeling

Untuk *use-case bank business credit scoring modelling* dimulai dengan memilih algoritma yang cocok dengan hasil identifikasi masalah. Algoritma dalam *machine learning* secara garis besar terbagi dalam 3 kategori yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. Selanjutnya dilakukan proses pemodelan data sesuai dengan algoritma yang telah dipilih.

Output dari tahap ini adalah berupa model data yang memuat pengetahuan-pengetahuan baru untuk selanjutnya akan diuji pada tahap *evaluation*.



**Gambar 3** Conroh Proses Modelling

## 5. *Evaluation*

Sesuai dengan namanya, tahap *evaluation* secara garis besar mengevaluasi kembali model data yang telah dihasilkan. Salah satu evaluasinya dilakukan dengan cara mengukur akurasi model, kemudian dilakukan interpretasi model. Pada tahap ini dikaji ulang apakah model yang dihasilkan sudah memenuhi hasil identifikasi masalah.

Output dari tahap ini adalah berupa akurasi dan interpretasi dari model yang dihasilkan untuk diolah pada tahap selanjutnya yaitu *deployment*.

accuracy: 91.50% +/- 0.05% (micro average: 91.50%)

	true No	true Yes	class precision
pred. No	291298	26071	91.79%
pred. Yes	1124	1302	53.67%
class recall	99.62%	4.76%	

**Gambar 4** Contoh Output *Evaluation*

Untuk *use-case bank business credit scoring evaluation* dilakukan dengan memilih teknik evaluasi model. Salah satu metode evaluasi yang cukup populer adalah cross validation. Hasil *evaluation* kemudian dikaji ulang apakah sudah sesuai dengan tujuan awal, jika dirasa belum sesuai maka kembali lagi ke proses *business understanding*.

## 6. *Deployment*

Tahap ini adalah akhir dari *use-case bank business credit scoring*. Pengetahuan-pengetahuan yang sebelumnya telah dihasilkan akan dipresentasikan sedemikian rupa untuk dapat digunakan oleh pengguna. Tahap ini adalah satu-satunya tahap yang melibatkan pengguna karena penting untuk mengetahui bagaimana pengguna bereaksi dengan sistem yang dibuat.

*Output* dari tahap ini adalah *product* akhir dari semua perencanaan yang telah dilakukan dari awal dan tentu saja tidak menutup kemungkinan akan ada perubahan ketika sudah tidak relevan lagi dengan masalah yang ada.

Di atas adalah tahapan-tahapan CRISP-DM untuk *use-case bank business credit scoring deployment* yang terdiri dari 6 tahapan yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation*, dan *deployment*.

## Daftar Pustaka

- Mauritsius, T., & Binsar, F. (2020). *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. <https://mmsi.binus.ac.id/2020/09/18/cross-industry-standard-process-for-data-mining-crisp-dm/#:~:text=Cross-Industry Standard Process for Data Mining> atau CRISP-DM, dikembangkan oleh ratusan organisasi dan
- Oliver, A. (2021). *Bikin Data Lebih Mudah Dibaca, yuk, Kenalan dengan Data Preprocessing*. <https://glints.com/id/lowongan/data-preprocessing-adalah/#.YyaOd3ZBw2w>
- Wiejoyo. (2020). Aplikasi Scoring System Untuk Penentuan Keputusan Kredit Pada BPR DRJ Berbasis Web. *Jurnal Infra*, 8(2), 1–20.  
<http://publication.petra.ac.id/index.php/teknik-informatika/article/view/10499>