

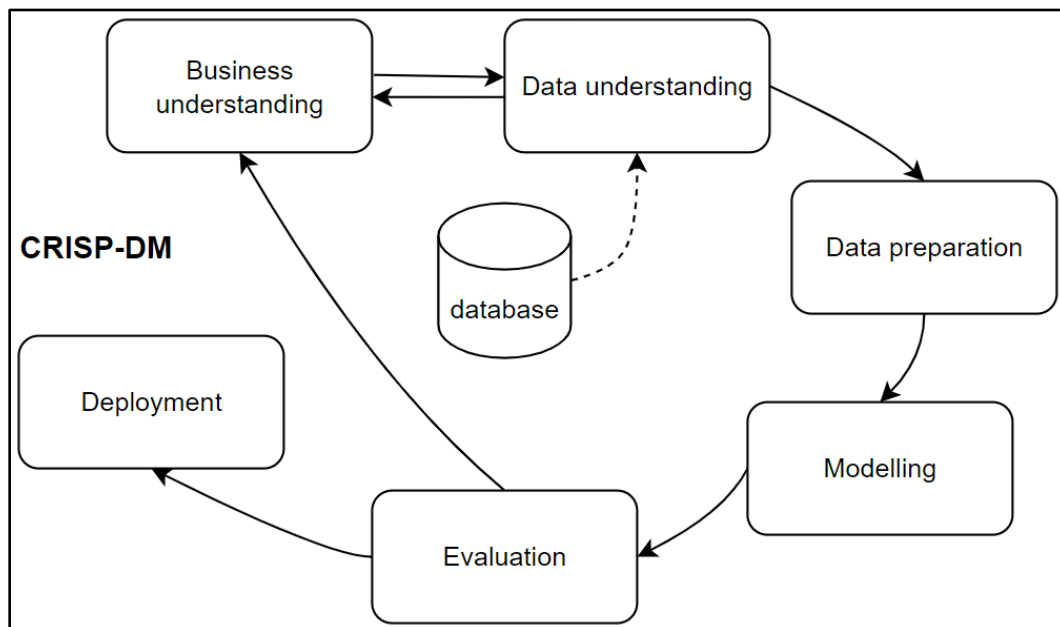
Credit Scoring pada Industri Bank

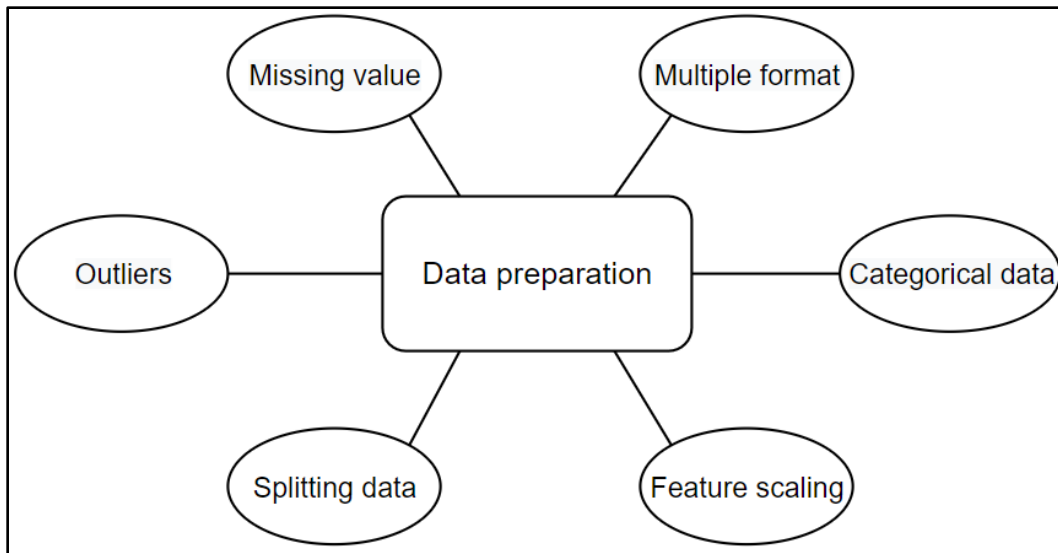
Link github: <https://github.com/christophermarcelino/machine-learning-zenius>

Industri menawarkan layanan pinjaman sebagai salah produknya. Dalam mengeluarkan pinjaman, bank perlu memperhatikan siapa peminjamnya, apakah peminjam tersebut mampu mengembalikan pinjaman sehingga mampu memberikan keuntungan pada bank. Pada masa lampau, keputusan semacam itu dilakukan oleh para ahli bank. Namun, dengan makin berkembangnya laju ekonomi yang berdampak pada meningkatnya jumlah transaksi, cara konvensional tersebut tidak efisien. *Data science*, dalam kasus ini, memiliki kemampuan untuk membantu pengambilan keputusan terkait masalah tersebut yang disebut dengan credit scoring.

Credit scoring merupakan cara menilai seseorang berdasarkan parameter yang diberikan. Dengan memanfaatkan parameter seperti gaji, umur, dsb kita bisa menilai apakah seorang peminjam layak diberikan pinjaman. Algoritma yang dipakai dalam kasus ini adalah *logistic regression* karena kita akan memberikan nilai angka pada target, lalu hasil nilai tersebut dipakai untuk menentukan kelayakan seorang peminjam.

Esai ini akan menjelaskan tahapan pengembangan *credit scoring* berdasarkan CRISP-DM.





Business Understanding

Pada tahap ini, *data scientist* berusaha untuk memahami alur bisnis yang terjadi, contohnya apa konsep dari pinjaman, bunga, gagal bayar, dan sebagainya.

Data Understanding

Setelah memahami proses bisnis yang terjadi, *data scientist* akan menentukan data-data apa saja yang penting untuk dimasukkan ke dalam model *machine learning*. Data bisa didapatkan dari data internal bank dan peminjam secara langsung.

Dalam menentukan data-data tersebut, *data scientist* perlu memahami korelasi data dengan implikasi di dunia nyata, contohnya gaji peminjam perlu dipertimbangkan karena gaji menunjukkan kemampuan peminjam mengembalikan pinjaman.

Tabel berikut ini akan menunjukkan data-data apa saja yang penting dalam kasus *credit scoring* pinjaman bank.

No	Data	Deskripsi
1	Gaji	Menunjukkan kemampuan bayar peminjam
2	Usia	Menunjukkan produktivitas dalam bekerja
3	Status tempat tinggal	Menunjukkan apakah peminjam memiliki beban lain dalam membayar kontrak rumah
4	Jumlah kredit	Menunjukkan risiko peminjaman
5	Durasi pinjaman	Menunjukkan seberapa ringan beban bayar
6	Aset lain-lain	Menunjukkan jaminan
7	Tipe pekerjaan	Menunjukkan potensi kesanggupan bayar
8	Tempat bekerja	Menunjukkan potensi kesanggupan bayar
9	Durasi menjadi nasabah	Menunjukkan loyalitas peminjam
10	Riwayat gagal bayar	Menunjukkan potensi gagal bayar selanjutnya
11	Berapa kali telah meminjam	Menunjukkan kesanggupan peminjam membayar balik

Banyak hal menjadi pertimbangan terhadap seorang peminjam, maka pastinya ada lebih banyak data/parameter yang dipertimbangkan selain dari yang disebutkan pada tabel di atas.

Data Preparation

Tidak semua data yang tersedia bersih, bisa jadi terdapat *missing value*, format yang berantakan, dan sebagainya. Pada tahap ini, *data scientist* membersihkan data (*data cleaning*) tersebut. Proses ini memakan waktu paling lama karena tingginya tingkat kompleksitas dan diperlukannya ketelitian yang tinggi. Hal ini disebabkan oleh buruknya data akan mengakibatkan model memberikan hasil yang buruk (*garbage in, garbage out*).

Selain membersihkan data, *data scientist* juga perlu mengubah data kategorikal menjadi data numerik (*one hot encoding*) karena mesin hanya memahami data numerik.

Setelah data diolah, data perlu dibagi menjadi data train dan data test. Rumus pembagian yang umum digunakan adalah 80/20.

Data scientist juga bisa melakukan feature scaling untuk menyamakan rentang nilai data sehingga tidak terjadi bias pada model.

Modelling

Pada kasus ini, kita akan menggunakan *logistic regression* sebagai algoritma utama.

Algoritma bisa kita dapatkan di [sklearn.linear_model.LogisticRegression](#).

Pelatihan model bisa dilakukan dengan fungsi *fit*, lalu eksekusi pada data bisa dilakukan dengan fungsi *predict*.

Evaluation

Setelah tahap pelatihan, model perlu diuji performanya. Terdapat beberapa metrik evaluasi, seperti *Adjusted R Square*, MSE, MAE, dan sebagainya. Evaluasi bisa disesuaikan oleh *data scientist* sejauh apa evaluasi perlu dicek.

Deployment

Setelah model telah diyakini sudah layak pakai, maka kita perlu men-*deploy* model ke *production*. Ada bermacam-macam jalur komunikasi data yang bisa diimplementasikan, salah satunya adalah men-*generate* data dalam bentuk JSON (*Javascript Object Notation*) untuk dikomunikasikan dengan API (*Application Programming Language*).

Impacts

Sistem *credit scoring* untuk menilai kelayakan peminjam memungkinkan proses bisnis berjalan lebih cepat dan efisien. Pengambilan keputusan tidak lagi memerlukan diskusi konvensional para ahli yang cukup membuang tenaga dan waktu. Kita cukup memasukkan data peminjam ke dalam

model, lalu model akan mengembalikan pada kita nilai/skor peminjam tersebut atau Baik/Buruk sesuai pengkondisian yang kita tentukan.

Peraturannya sendiri bisa diatur berdasarkan kesepakatan pemangku bisnis, misalnya peminjam dengan skor di bawah 20 tidak memenuhi kriteria peminjaman, maka pengajuan peminjaman akan ditolak.

Penggunaan *machine learning* untuk *credit scoring* akan memberikan manfaat sebagai berikut:

1. Kerja lebih efisiensi (hemat biaya, waktu, dan tenaga)
2. Pengambilan keputusan yang jelas
3. Hasil lebih akurat (lepas dari bias manusia)
4. Hasil lebih presisi (mampu mempertimbangkan banyak parameter sekaligus)

Walaupun demikian, model harus senantiasa dilatih menggunakan data terbaru dan dimonitor performanya agar memberikan hasil lebih maksimal.

Referensi

Tabagari, Salome.2015."Credit scoring by logistic regression".Diakses pada 18 September 2022
<https://core.ac.uk/download/pdf/79110695.pdf>

Wu, Songhao.2020."3 Best metrics to evaluate Regression Model?". Diakses pada 18 September 2022.
<https://towardsdatascience.com/what-are-the-best-metrics-to-evaluate-your-regression-model-418ca481755b>