

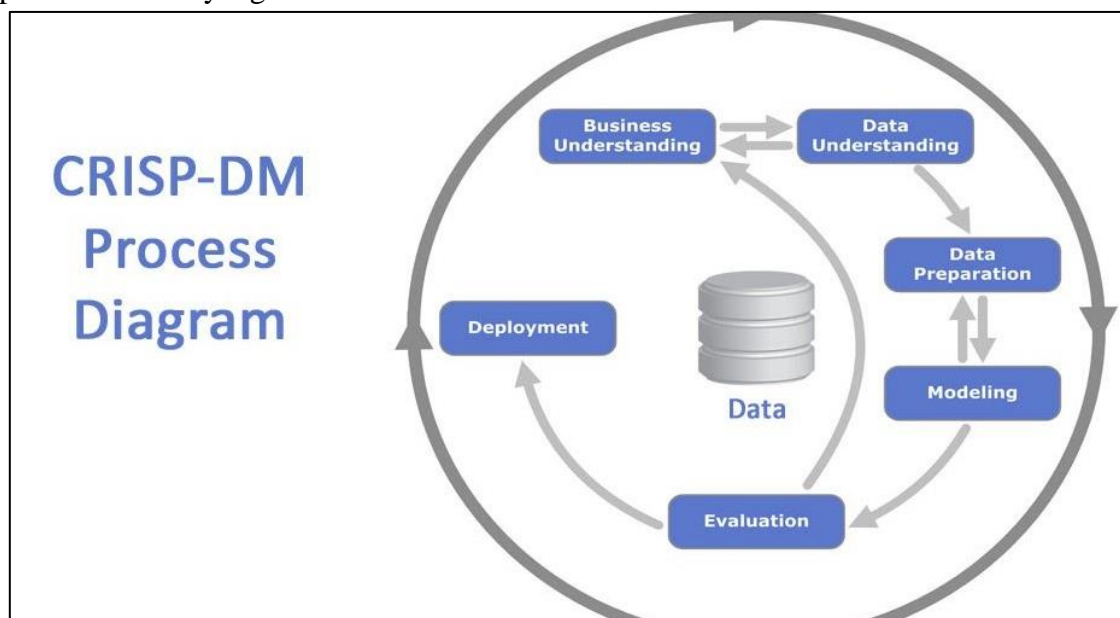
Aplikasi Data Science Dalam Mendeteksi Fraud Pada Industri Perbankan

[GitHub - reynaldi777/zenius-msib](https://github.com/reynaldi777/zenius-msib)

Penipuan merupakan salah satu tindakan kriminal yang merugikan berbagai pihak, dan juga merupakan tindakan yang melanggar hukum. Penipuan terus berkembang dengan berbagai macam motif dan media. Kasus penipuan juga merambat hingga ke dunia bisnis, dan ini dapat membawa kerugian yang cukup besar bagi pihak tertentu. Penipuan ini dilakukan dengan sengaja untuk kepentingan individu atau kelompok untuk tujuan tertentu. Tindakan penipuan ini dikenal dengan sebutan fraud.

Fraud adalah tindakan yang dengan sengaja dilakukan untuk memperoleh keuntungan yang tidak sah menggunakan berbagai cara dan manipulasi untuk mendapatkan keuntungan. Tindakan fraud semakin berkembang hingga hampir tidak terdeteksi, karena dilakukan dengan sangat rapi dan lebih struktur dengan bantuan teknologi. Oleh karena itu, sebuah perusahaan memerlukan bantuan teknologi untuk memecahkan masalah ini. Data science hadir sebagai solusi untuk mengatasi permasalahan ini. Melalui data science, fraud dapat terdeteksi sehingga kerugian yang dapat ditimbulkan pun dapat dicegah.

Dalam penggunaan data science, agar sistematis maka digunakan suatu metodologi praktis dalam prosesnya. Beberapa praktisi mengusulkan metode proses seperti workflow dengan pendekatan yang sederhana untuk memperbesar peluang keberhasilan dalam berbagai proyek. Metodologi tersebut adalah Cross-Industry Standard-Process for Data Mining (CRISP-DM). Metodologi ini terdiri dari serangkaian tahapan, mulai dari analisa masalah, hingga impact cari model yang disusun.



Model CRISP-DM

Contoh Use-Case Dari Data Science Pada Fraud Detection Menggunakan Decision Tree Model

Business Understanding

Perusahaan bernama FinTech baru saja meluncurkan produk terbarunya, yaitu sebuah Mastercard kepada pelanggan. Sebagai perusahaan yang saat ini sedang naik daun, Fintech ingin untuk memberikan pengalaman yang terbaik kepada pelanggan dengan mengurangi jumlah verifikasi, misalnya dengan menggunakan password sekali pakai (OTP). Salah satu pendekatan untuk proyek ini adalah dengan meningkatkan sistem deteksi penipuan (fraud) pada kartu credit perusahaan, khususnya untuk mendeteksi secara real time aktivitas transaksi yang mencurigakan. Dapatkah kita secara otomatis mendeteksi transaksi palsu secara real time berdasarkan aktivitas transaksi biasa?

Data Understanding and Preparation

Pertama tama penyimpangan tertentu dari kebiasaan belanja sebagai transaksi yang berpotensi penipuan akan ditetapkan. Dikarenakan beberapa transaksi menggunakan kartu kredit di beberapa jenis toko lebih mudah dikonversi ke dalam cash, maka itu akan lebih besar kemungkinannya terhadap penipuan daripada tempat lain. Oleh karena itu KKP (Kode Kategori Pedagang) akan dikelompokkan sesuai dengan keadaan aslinya (resiko ketersediaan untuk melakukan penipuan), sehingga akan ada lebih sedikit kode KKP untuk disortir satu per satu. Pengelompokan KKP ini juga dapat dibantu oleh ahli domain.

Dalam sejarahnya, baik bank digital maupun bank tradisional lainnya, distribusi data yaitu rasio transaksi palsu dan transaksi normal itu sangat tidak seimbang. Jumlahnya jauh lebih banyak transaksi yang normal daripada transaksi hasil penipuan. Oleh karena itu, teknik under sampling (under-sampling catatan transaksi normal dalam set pelatihan ke angka yang berarti) akan memungkinkan model untuk mempelajari kedua jenis transaksi ini.

Data Requirement

Data yang dibutuhkan adalah variable profil kartu yang akan mengungkapkan karakteristik perilaku penggunaan kartu. Untuk memprediksi secara real-time apakah suatu transaksi curang, diperlukan model untuk membandingkan apakah transaksi menyimpang dari kebiasaan belanja historis atau tidak. Data historis yang diperlukan: Lokasi geografis, Hari dalam sebulan, Jam buka, Jenis pedagang menurut KKP (Kode Kategori Pedagang).

Data Collection

Data yang dibutuhkan akan didapat dengan meminta data rahasia dari gudang data perusahaan kartu kredit. Data historis akan mencerminkan profil penggunaan kartu pelanggan.

Analytical Approach

Sebuah model pendekatan Decision Tree akan dikembangkan untuk mempelajari serangkaian pelatihan dari data set sejarah transaksi hasil penipuan dan transaksi sah. Penyimpangan statistik dari profil kartu kredit pelanggan akan bertindak sebagai variabel yang menunjukkan inkonsistensi dibandingkan dengan tren historis kartu kredit.

Modelling and Evaluation

Variabel yang digunakan adalah penyimpangan tertentu dari statistik rata-rata profil, seperti statistik regional, statistik jumlah transaksi harian, dan statistik transaksi.

Untuk mengatasi masalah pada nilai yang hilang, setiap pemisahan dalam Decision Tree, model akan mengidentifikasi bidang input yang paling mirip dan menetapkannya sebagai pengganti bidang yang hilang. Seperti disebutkan di atas, metode under sampling akan mengurangi kemungkinan False Negatives; di mana transaksi penipuan ditandai sebagai normal.

TRAINING AND TEST SET SIZES FOR THE SAMPLES

Samples		# of Records	
		Training Set	Test Set
Set-1F-To-1N	Normal	1362	583
	Fraud	657	316
Set-1F-To-4N	Normal	3404	1460
	Fraud	657	316
Set-1F-To-9N	Normal	6787	2942
	Fraud	657	316

Pembagian Dataset untuk pelatihan dan pengujian

Rasio penipuan terhadap transaksi normal yang digunakan untuk menentukan rasio optimal untuk set pelatihan adalah 1:1, 1:4, dan 1:9. 70% dari data--70% dari transaksi normal dan 70% dari transaksi penipuan--digunakan sebagai set pelatihan untuk model Pohon Keputusan; sedangkan 30% data digunakan sebagai set tes untuk evaluasi model. Di setiap sampel, jumlah transaksi penipuan tetap, karena catatan penipuannya rendah; jumlah transaksi normal berubah sesuai dengan rasio yang diberikan di atas.

Referensi:

- Bryan. T. (2020, February 21). Data Science Methodology (CRISP-DM) for credit card fraud detection using decision tree model. LinkedIn. Retrieved September 20, 2022, from <https://www.linkedin.com/pulse/data-science-methodology-credit-card-fraud-detection-using-bryan-tan/>
- Ragil, Y., & Flinsetyadiz. (2020, July 13). *Metodologi crisp-DM Beserta Contoh kasusnya*. Flin Setyadi. Retrieved September 20, 2022, from <https://flinsetyadi.com/metodologi-crisp-dm-beserta-contoh-kasusnya/>