

ARTIKEL *USE-CASE DATA SCIENCE*

PENGEMBANGAN MODEL *FRAUD DETECTION* BANK BKI

LATAR BELAKANG

Sektor perbankan memiliki fungsi yang sangat vital dalam pembangunan perekonomian dan stabilitas suatu negara sebagai penghimpun dana masyarakat melalui bank. Bank merupakan badan usaha di bidang keuangan yang menarik dan mengeluarkan uang dalam masyarakat, terutama memberikan kredit dan jasa dalam proses transaksi dan peredaran uang. Berdasarkan aktivitas usahanya di Indonesia, bank dibagi menjadi dua yaitu bank umum dan bank perkreditan rakyat. Bank umum merupakan bank yang melaksanakan kegiatan usaha baik secara konvensional maupun syariah. Bank umum adalah bank yang umum digunakan untuk menabung. Sesuai operasionalnya, bank umum yang beroperasi konvensional menjalankan kegiatan usaha dengan penetapan harga mengikuti ketentuan dari bank sentral yaitu Bank Indonesia.¹ Dengan segala fungsinya, sebagai sektor vital tentunya keamanan dari ancaman yang dapat merugikan seperti penipuan (*fraud*) harus menjadi perhatian.

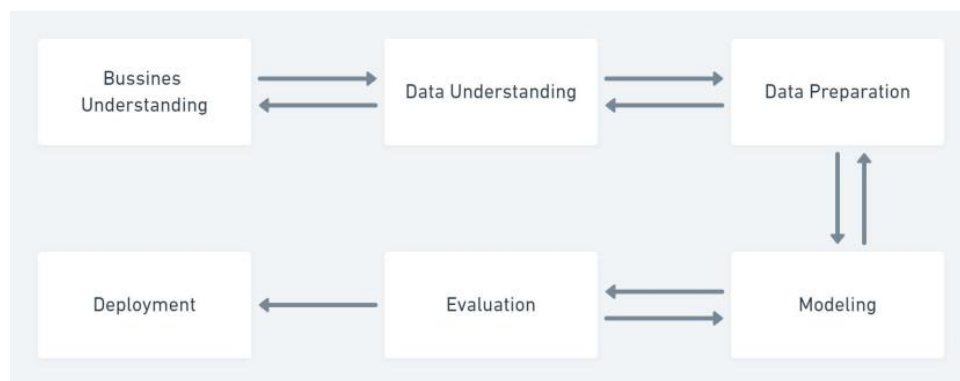
Berdasarkan Survei Fraud Indonesia tahun 2019, pihak yang paling banyak dirugikan akibat adanya fraud adalah industri keuangan dan perbankan sebanyak 41.4 %. *Fraud* atau kecurangan merupakan tindakan yang dilakukan secara sengaja oleh oknum baik internal maupun eksternal suatu perusahaan atau lembaga yang merugikan pihak lain untuk memperoleh keuntungan. Dalam industri perbankan, tindakan yang termasuk dalam *fraud* meliputi kecurangan, penipuan, penggelapan aset, pembocoran informasi, tindak pidana perbankan dan tindak lain yang dapat dipersamakan dengan *fraud*.² Tindakan fraud diklasifikasikan menjadi 3 bidang penyimpangan mendasar yaitu bidang operasi, bidang kredit, serta bidang teknologi dan informasi. *Fraud* bidang operasi mencakup kecurangan aktivitas pendanaan, dan pencairan ilegal deposito nasabah. *Fraud* bidang kredit mencakup pemberian kredit fiktif, agunan, fiktif oleh nasabah, penilaian jaminan yang terlalu tinggi, perubahan bunga ilegal sehingga terjadi peningkatan biaya dana dan kelebihan dari dana dipindahkan ke rekening pegawai bank, serta rekayasa

pemberian kredit. *Fraud* bidang teknologi dan informasi mencakup pencurian data nasabah, dan *skimming*. Sebagai bentuk jaminan keamanan oleh Bank, berbagai upaya inovasi untuk meningkatkan keamanan transaksi selalu dilakukan dan mengikuti peraturan strategi anti-fraud yang telah diatur dalam POJK 39/2019 yang berlaku bagi bank umum dengan memuat 4 (empat) pilar yakni: pencegahan; deteksi; investigasi, pelaporan, dan sanksi; dan pemantauan, evaluasi, dan tindak lanjut.³ **Sejalan dengan pilar kedua yaitu deteksi, dapat dikembangkan strategi anti-fraud dengan memanfaatkan teknologi yaitu sebuah mesin yang mampu mendeteksi potensi tindak kecurangan melalui anomali dari kebiasaan transaksi.**

Dengan kemajuan teknologi, sebuah mesin pendeteksi kecurangan (*fraud detection*) dapat dibangun menggunakan pembelajaran mesin (*machine learning*) oleh seorang Ilmuwan Data (*Data Scientist*).⁴ Namun, hal ini tidak serta merta mengakibatkan mesin tersebut dengan mudah dibangun. Pemahaman terhadap ilmu data (*Data Science*) sangat diperlukan dan model standar dalam pengembangan mesin juga menjadi hal yang krusial. Salah satu model standar pengembangan mesin pendeteksi kecurangan yang dapat diterapkan adalah Cross Industry Standar Process for Data Mining (CRISP-DM). **Berdasarkan permasalahan tersebut, diperlukan pembahasan mendalam mengenai pengembangan mesin pendeteksi kecurangan (*fraud detection*) menggunakan metode CRISP-DM.** Diharapkan nantinya pembahasan ini dapat membantu para pengembang untuk membangun mesin fraud detection pada sektor perbankan.

MODEL PROSES CRISP-DM

Cross-Industry Standard Process for Data Mining adalah model proses data mining (datamining framework) yang diprakarsai oleh 5 perusahaan yaitu **Integral Solution Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA**. Framework ini kemudian dikembangkan kembali oleh ratusan perusahaan dan organisasi sebagai metodologi terbuka (open source) bagi data mining. **Model proses CRIPS-DM** memberikan gambaran tentang siklus hidup pengembangan proyek data mining melalui 6 tahapan yaitu *Bussiness Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*.²



BUSSINESS UNDERSTANDING

Terdapat permasalahan dengan kondisi Anda adalah seorang Data Scientist di Bank BKI. **Anda diminta untuk membuat model machine learning yang mampu memprediksi suatu transaksi apakah fraud atau bukan fraud.** Berdasarkan riwayat transaksi nasabah, diperoleh data-data yang merupakan transaksi fraud dan tidak fraud yang bisa anda gunakan untuk membuat model machine learning. Variabel input (prediktor) berupa 26 variabel yang berkaitan dengan transaksi, sedangkan variabel output (target) merupakan kondisi apakah transaksi tersebut fraud atau tidak.

Berdasarkan kondisi tersebut maka permasalahan yang dapat diketahui adalah terdapat aktivitas fraud pada bank tersebut dan memerlukan model

machine learning untuk mempercepat pendeteksian fraud tersebut dengan memanfaatkan data riwayat transaksi.

DATA UNDERSTANDING

Untuk menyelesaikan permasalahan tersebut digunakan data yang dapat diakses pada <https://www.kaggle.com/competitions/fraud-detection-shift-mega/data>. Selanjutnya untuk memahi data, dilakukan langkah Explanatory Data Analysis (EDA).



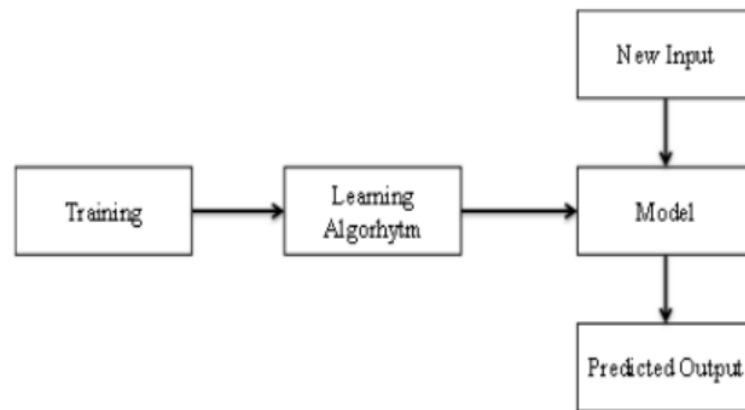
Berdasarkan data set tersebut dibagi atas 2 data yaitu data predict sebagai data yang harus diprediksi apakah terdapat fraud atau tidak dan fraud dataset sebagai data latih model machine learning nanti. Total terdapat 28 variabel pada data latih dan 27 varibael pada data prediksi. Pada data latih tidak ditemukan null value dan telah terlabel.

DATA PREPARATION

Pada tahap ini dapat dilakukan berbagai cara untuk membuat data lebih sederhana, bersih, dan mudah diproses. Tahap yang umum dilakukan pertama adalah data cleaning yang meliputi penanganan data missing value dan data dengan korelasi rendah. Selanjutnya, dapat dilakukan data labeling untuk setiap data yang terindikasi atau berpotensi melakukan tindak fraud. Setelah itu maka data siap untuk lanjut ke tahap berikutnya.

MODELING

Selanjutnya proses modeling dengan menentukan algoritma machine learning yang akan digunakan. Skema dari proses modeling akan terlihat sebagai berikut



Beberapa algoritma yang akan digunakan adalah K-Nearest Neighbors (KNN), Random Forest, Decision Treem Support Vector Machine (SVM), dan Logistic Regression. Pemilihan beberapa algoritma ditujukan untuk menemukan model terbaik nantinya dalam proses fraud detection.

EVALUATION

Pada tahap ini, model akan diuji dan dievaluasi performanya. Untuk setiap algoritma menghasilkan akurasi sebagai berikut

No	Method	Train score	Test score
1	K-Nearest Neighbors (KNN)	0.763671	0.738223
2	Random Forest	0.994999	0.745437
3	Decision Tree	0.994999	0.639633
4	Support Vector Machine (SVM)	0.740743	0.738113
5	Logistic Regression	0.740743	0.738113

Berdasarkan tabel tersebut diketahui algoritma terbaik dalam deteksi fraud pada kasus ini adalah Random Forest. Hasil deteksi fraud ini akan terformat sebagai berikut

Id	flag_transaksi_fraud
6678	0
437	1
1696	1
8868	1
7014	0

DEPLOYMENT

Tahap terakhir model akan dibentuk menjadi sebuah servis berupa API menggunakan Fast API, kemudian akan diberikan kepada team pengembang aplikasi untuk dapat digunakan sebagai mesin pendeteksi fraud.

DAFTAR PUSTAKA

1. Agita Natalia. Mengenal Sektor Perbankan dan Jenis-jenisnya Lebih Dalam. ajaib.co.id. Published January 3, 2020. Accessed September 18, 2022. <https://ajaib.co.id/mengenal-sektor-perbankan-dan-jenis-jenisnya-lebih-dalam/>
2. Salleh MFM, Suryanto T. Fraud Detection on Banking Industry in South Sumatera: A Study on the Role of Internal Auditors'. *International Journal of Shari'ah and Corporate Governance Research*. 2019;2(2). doi:10.46281/ijscgr.v2i2.399