Isnaeni Safitri

Topic 1 2 - Accelerated Machine Learning

Github Repository link: https://github.com/isnaenisft/Zenius-IsnaeniSafitri.git

---

Artikel tentang penggunaan Data Science dalam industri Banking

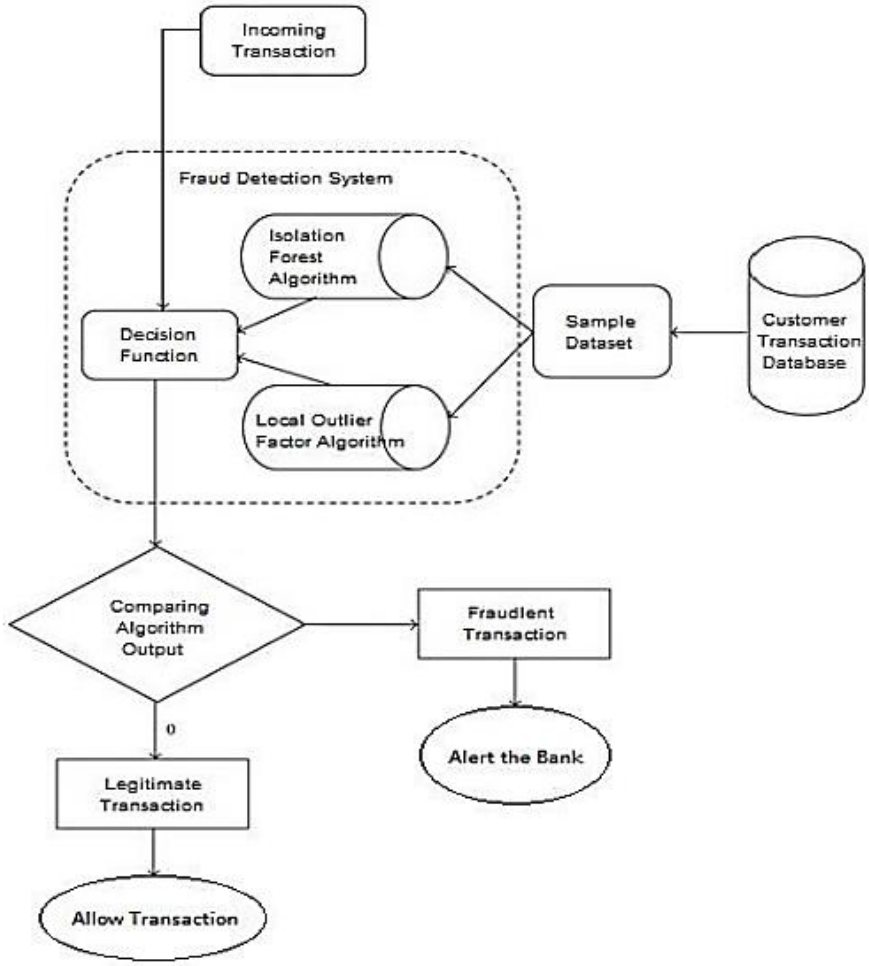
DETEKSI FRAUD PADA KARTU KREDIT

1. Pendahuluan

"Fraud" dalam transaksi kartu kredit adalah penggunaan akun yang tidak sah dan tidak diinginkan oleh orang lain selain pemilik akun tersebut. Tindakan pencegahan yang diperlukan dapat diambil untuk menghentikan penyalahgunaan ini dan perilaku fraud tersebut dapat dipelajari untuk meminimalkan dari kejadian serupa di masa mendatang. Dengan kata lain, fraud pada kartu kredit dapat didefinisikan sebagai kasus di mana seseorang menggunakan kartu kredit orang lain untuk alasan pribadi sementara pemilik dan otoritas penerbit kartu tidak menyadari bahwa kartu tersebut disalahgunakan. Deteksi fraud melibatkan pemantauan aktivitas pengguna untuk memperkirakan, memahami, atau menghindari perilaku yang tidak menyenangkan.

Masalah ini menuntut perhatian data scientist suatu perusahaan di mana solusi untuk masalah ini dapat diotomisasi. Masalah ini sangat menantang karena jumlah transaksi yang valid jauh lebih banyak dari transaksi fraud. Selain itu, pola transaksi sering kali mengubah sifat statistiknya seiring berjalannya waktu. Dalam kasus nyata, permintaan pembayaran dalam intensitas besar dapat discan oleh alat otomatis yang menentukan transaksi mana yang akan diotorisasi.

Algoritma machine learning digunakan untuk menganalisis semua transaksi resmi dan melaporkan transaksi yang mencurigakan. Laporan ini diselidiki, kemudian menghubungi pemegang kartu untuk mengkonfirmasi apakah transaksi asli atau fraud. Data scientist memberikan umpan balik ke sistem otomatis yang digunakan untuk melatih dan memperbarui algoritma untuk meningkatkan kinerja deteksi fraud. Alur proses deteksi fraud dalam kartu kredit adalah sebagai berikut.



Fraud Detection Flow (Maniraj, 2019)

2. Isi dan Pembahasan

Setelah mengetahui permasalahan bisnis yang akan diatasi, yaitu deteksi fraud pada penggunaan kartu kredit, selanjutnya adalah mengeksplor data yang dibutuhkan. Dataset yang digunakan dapat diperoleh dari kaggle. Di dalam dataset ini, ada 31 kolom di mana 28 dinamai V1-V28 untuk melindungi data sensitif. Kolom lain berisi waktu, jumlah, dan kelas. Waktu menunjukkan jeda waktu antara transaksi pertama dan yang berikutnya. Jumlah adalah jumlah uang yang ditransaksikan. Kelas 0 mewakili transaksi yang valid dan 1 mewakili fraud.

Tahapan selanjutnya adalah persiapan data menggunakan python melalui Jupyter Notebook atau Google Collab. Pada tahap ini akan dilakukan pengecekan terhadap data yang double atau data yang kosong. Setelah memeriksa dataset, plot histogram untuk setiap kolom. Hal ini dilakukan untuk mendapatkan representasi grafis dari dataset yang dapat digunakan untuk memverifikasi bahwa tidak ada nilai yang hilang dalam dataset. Dilakukan untuk memastikan bahwa tidak diperlukan imputasi nilai yang hilang dan

algoritma machine learning dapat memproses kumpulan data dengan lancar. Setelah melihat insight dari data yang kita punya dapat dilanjutkan untuk proses pembuatan model machine learning. Pendekatan algoritma machine learning yang dapat digunakan untuk memprediksi adanya fraud adalah outliers. Algoritma ini adalah bagian dari sklearn. Model ensemble dalam paket sklearn mencakup metode dan fungsi berbasis ensemble untuk klasifikasi, regresi, dan deteksi outlier. Data yang dimiliki cocok dengan model dan modul deteksi outliers berikut:

A. Local Outlier Factor

   Termasuk Unsupervised Outlier Detection, hal ini mengacu pada skor anomali setiap sampel, mengukur deviasi lokal dari sampel sehubungan dengan neighbours. Dengan membandingan nilai lokal sampel dengan their neighbours, dapat diidentifikasi sampel yang jauh lebih rendah dari neighbours. Nilainya cukup tidak baik dan dapat dianggap sebagai outliers.

B. Isolation Forest Algorithm

   Observasi dengan memilih secara acak nilai pemisah antara nilai maksimum dan minimum dari fitur yang ditetapkan. Dapat direpresentasikan sebagai pohon, jumlah pemisahan yang diperlukan untuk mengisolasi sampel setara dengan panjang jalur root node ke terminating node. Rata-rata panjang jalur ini memberikan keputasan yang kita gunakan.

3. Kesimpulan

Setelah pembuatan machine learning model, langkah selanjutnya adalah mengevaluasi model yang sudah dibuat. Model itu di testing pada catatatan transaksi yang berjalan selama 2 hari. Hasilnya model dapat mendeteksi tindakan fraud dengan akurasi lebih dari 96% dan presisi 33%. Persentase akurasi yang tinggi karena ketidakseimbangan data yang sangat besar antara jumlah transaksi yang valid dengan jumlah transaksi sesungguhnya. Karena model sudah dianggap bagus dalam menjawab business problem perusahaan, maka model dapat di deploy ke production server.

References:

Maniraj, S. P., Saini, A., & Ahmed, S. (2019). Credit Card Fraud Detection using Machine Learning and Data Science. *International Journal of Engineering Research & Technology (IJERT)*. Retrieved from credit-card-fraud-detection-using-machine-learning-IJERTV8IS09003120190913-7894-171thvf-with-cover-page-v2.pdf

Starter: Credit Card Fraud Detection 249b5984-4. (2019). Starter: Credit Card Fraud Detection 249b5984-4 | Kaggle

Credit Card Fraud Detection: Everything You Need to Know. Credit Card Fraud Detection: Everything You Need to Know (inscribe.ai)