

BAB I PENDAHULUAN

1.1 Latar Belakang

Fraud adalah penipuan yang disengaja dengan tujuan untuk mendapatkan keuntungan atau menyebabkan kerugian dengan trik implisit maupun eksplisit yang dilakukan oleh banyak pihak. Kemajuan dalam informasi dan teknologi secara tidak langsung meningkatkan kasus penipuan [1].

The Association of Certified Fraud Examiners (ACFE) mengklasifikasikan *fraud* ke dalam tiga bentuk berdasarkan perbuatan, yaitu penyimpangan atas aset (*asset misappropriation*), korupsi (*corruption*) dan kecurangan laporan keuangan (*fraudulent financial reporting*) [2].

Dalam penipuan penggunaan kartu kredit dibagi menjadi dua bagian yakni *inner card fraud* dalam arti adanya persetujuan antara pemegang kartu dan pihak bank untuk membuat identitas palsu dan melakukan penipuan dan *external card fraud* yang berarti adanya keterlibatan dalam pencurian kartu kredit [3].

Salah satu upaya untuk mendeteksi penipuan transaksi keuangan adalah menggunakan *machine learning*. Analisis *machine learning* dapat dilakukan pada data transaksi keuangan secara elektronik yang telah diambil. Selanjutnya, dapat mengidentifikasi karakteristik dan pola dari transaksi penipuan yang telah terjadi dan dapat digunakan untuk mendeteksi transaksi lainnya. Hal itu dapat membantu dalam mencegah terjadinya penipuan yang selanjutnya akan merugikan nasabah dan perbankan [1].

1.2 Batasan Masalah

Adapun batasan masalah yang digunakan pada penelitian ini yaitu :

1. Dataset yang digunakan untuk melakukan pengujian model adalah *dataset transaction.json*.
2. Model yang digunakan untuk mendeteksi *fraud* menggunakan algoritma Regresi Logistik dan *Random Forest*.

3. Tools yang digunakan adalah *Google Colaburatory* untuk *Data Preparation* dan *Google Data Studio* untuk penyajian data.
4. Menggunakan Teknik *SMOTE* untuk resampling data agar data seimbang.

1.3 Tujuan Penelitian

1. Mengetahui *feature important* atau variabel apa saja yang berpengaruh terhadap variabel dependen yaitu *Fraud*
2. Mengetahui nilai akurasi klasifikasi label fraud dengan algoritma Regresi Logistik dan *Random Forest*.
3. Memprediksi transaksi *fraud* setelah terjadinya transaksi dengan aplikasi berbasis website.

1.4 Manfaat Penelitian

Secara teoritis, hasil penelitian ini diharapkan dapat menjadi sumber referensi dan informasi untuk penelitian lebih lanjut terkait klasifikasi karakteristik variabel yang paling berpengaruh atau *feature important* sebuah data terutama yang terkait dengan transaksi.

1.5 Profil

Zenius merupakan platform belajar berbasis teknologi yang telah berdiri sejak tahun 2004 yang berfokus pada pemahaman konsep dan penalaran ilmiah dengan tujuan dapat memberikan *social impact* yang mampu mentransformasi masa depan bangsa. *Zenius* juga mulai mengembangkan produk yang ditujukan kepada para masyarakat umum dan pekerja, baik dalam sektor formal maupun informal, perusahaan dan perguruan tinggi, yakni *ZenPro*.

Melalui *ZenPro*, *Zenius* menjalin kemitraan dengan berbagai institusi, perusahaan dan juga para ahli di bidangnya. Salah satunya dengan Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi. *ZenPro* berupaya untuk mempersiapkan para lulusan sarjana Indonesia untuk mampu berpikir kritis, ilmiah dan terampil sebelum masuk ke dunia kerja melalui Program *Zenius Studi Independen Bersertifikat*.

Pada Program *Zenius* Studi Independen Bersertifikat terdapat 15 program salah satunya adalah *Data Warehousing, Analysis, and Visualization for Business Insights* atau lebih dikenal *Data Analyst*. Dilansir dari website kampus merdeka, Syarat keikutsertaan dalam kegiatan ini khususnya program *Data Analyst* adalah mahasiswa aktif dari perguruan tinggi Indonesia dengan jurusan: teknik informatika, ilmu komputer, statistika, aktuaria, matematika / matematika bisnis, sistem informasi. Mahasiswa minimal akan memasuki semester 6 pada semester genap 2022 dan tidak akan dinyatakan lulus hingga Juli 2022. Peserta yang telah mendaftar akan mengikuti seleksi administratif dan seleksi tertulis. Seleksi tertulis terkait materi bahasa Inggris, bahasa Indonesia, matematika, logika serta membuat esai singkat.

Pembelajaran pada program Studi Independen *Zenius* dibagi menjadi dua tahapan yaitu tahap *fundamental skills* dan tahap *learning path* masing-masing. Pembelajaran pada program ini dilaksanakan dengan dua metode yaitu sinkron yang dilaksanakan selama 5 bulan melalui platform *zoom* dan metode asinkron yang memungkinkan peserta untuk belajar mandiri menggunakan referensi materi yang 4 telah diberikan pada platform *Canvas*. Setiap peserta akan didampingi oleh satu mentor yang akan memonitor progres belajar peserta serta membantu *Subject Matter Expert* untuk memeriksa kelengkapan tugas. Selama program, pembelajaran *sinkronus* dilakukan sebanyak 3 kali dalam seminggu dengan durasi 3 jam 30 menit dan selanjutnya setiap peserta akan diberikan *assignment* untuk mengetahui pemahaman peserta akan materi yang telah diberikan. Setelah seluruh pemberian materi telah selesai, peserta dibagi menjadi beberapa kelompok dan diberikan *final project* mengenai pemodelan pada *python* serta visualisasi pada *Google Data Studio*. Peserta juga diharuskan mengerjakan *ZenCore* dengan minimal level pada akhir program adalah 60.

1.6 Landasan Teori

1.6.1 *Machine learning*

Machine Learning dapat diartikan sebagai algoritma atau program komputer yang dapat membuat sistem cerdas dengan mempelajari data data yang tersedia dimana algoritma atau program tidak didefinisikan secara terperinci [4].

Algoritma *machine learning* dibangun untuk menyelesaikan tiga kebutuhan, yaitu: prediksi, klasifikasi, dan *clustering*. Prediksi adalah kegiatan untuk menebak nilai luaran berdasarkan nilai masukan yang ditentukan. Klasifikasi adalah kegiatan membuat pengelompokan data-data berdasarkan kategori tertentu dan selanjutnya digunakan untuk mendeteksi kelompok suatu data. *Clustering* adalah kegiatan untuk mengelompokkan data berdasarkan tingkat kemiripan data-data tersebut [4].

1.6.2 Regresi logistik

Regresi logistik adalah analisis regresi untuk biner variabel. Dari sudut implementasi, logika regresi sederhana, mudah dimengerti dan disadari. Berjalan cepat dan biaya komputasi relatif rendah [5].

Klasifikasi Regresi logistik adalah salah satu pengklasifikasi linier dan sederhana serta mudah dibuat. Untuk mengatasi masalah ketidakseimbangan, fungsi kemungkinan Regresi logistik dimodifikasi sesuai dengan matriks biaya [6].

1.6.3 *Random forest*

Random Forest adalah pengklasifikasi *ensemble* yang terdiri dari sejumlah besar pohon keputusan, dan hasil akhir klasifikasi ditentukan oleh suara terbanyak dari semua pohon keputusan. Dapat menerapkan dua pendekatan untuk menangani kumpulan data yang tidak seimbang dengan menetapkan bobot kelas dan menetapkan ambang batas kemungkinan [6].

Klasifikasi *random forest* dilakukan dengan menggabungkan pohon dengan menjalankan eksperimen pada sampel yang disediakan. Saat menentukan kelas klasifikasi *random forest*, suara diambil dari beberapa pohon keputusan yang terbentuk, dan pemenang pohon ditentukan dengan suara terbanyak (majority vote).

Random forest mengambil dua parameter: jumlah pohon yang akan digunakan dan jumlah variabel bebas yang digunakan untuk mencapai nilai optimal dalam proses percabangan [7].

1.6.4 *Fraud*

Menurut *Association of Certified Fraud Examiners*(ACFE), kecurangan (*fraud*) sebagai tindakan penipuan atau kekeliruan yang dibuat oleh seseorang atau badan yang mengetahui bahwa kekeliruan tersebut dapat mengakibatkan beberapa manfaat yang tidak baik kepada individu atau pihak lain [2].

Fraud meliputi ketidakjujuran yang disengaja, kesalahan penyajian, manipulasi dan memberikan fakta yang dapat merugikan orang lain dan organisasi termasuk bank. *Fraud* juga meliputi pencurian, apropriasi, upaya untuk memperoleh sesuatu secara ilegal, dan kesalahan dalam membuat laporan keuangan termasuk aset dan kewajiban organisasi [8].

1.6.5 *Python*

Python adalah bahasa pemrograman tingkat tinggi, lintas platform, dan ditafsirkan yang berfokus pada keterbacaan kode. Menyediakan banyak library dengan kualitas tinggi dan dukungan untuk segala jenis komputasi ilmiah. Karakteristik ini menjadikan *Python* alat yang tepat untuk proyek penelitian dan industri di dengan penyelidikan yang complex [9].

1.6.6 *Google Colaboratory*

Google Colaboratory atau disingkat "*Colab*" adalah *notebook* *Jupyter cloud* yang digunakan secara luas untuk mengajarkan *machine learning* (pembelajaran mesin) dengan menulis teks penjelasan dan kode *Python* melalui browser. Pekerjaan ini memperkenalkan ekstensi *Colab* baru untuk mengajarkan desain sirkuit logika, bahasa *Verilog*, prosesor, dan arsitektur GPU. *Colab* memungkinkan kita untuk berbagi eksperimen yang dapat direproduksi di Web [10].

1.6.7 Google Data Studio

Google Data Studio adalah program visualisasi data baru yang dirancang sebagai alat yang mudah digunakan untuk mewakili kumpulan data kompleks dengan cara yang menarik dan jelas. Fungsi inti Data Studio adalah interpretasi visual bergaya dasbor dari media sosial dan analisis web seperti *Google AdWords* dan analisis *YouTube*; namun, dukungan alat seperti *MySQL* dan *Google Sheets* menunjukkan bahwa program ini dapat digunakan oleh peneliti untuk menginterpretasikan data mereka sendiri dalam format yang sama menarik dan ramah pengguna [11].

