

BAB I PENDAHULUAN

1.1. Latar Belakang

Kartu kredit adalah teknologi yang mempermudah berbisnis. Sistem ini melakukan transaksi dengan nasabah bank. Sistem yang digunakan oleh bank dalam kartu kredit adalah sistem peminjaman dimana pelanggan dapat membayar secara berkala terhadap barang yang mereka beli atau pembayaran hutang sesuai dengan persyaratan atau perjanjian yang ditentukan. Sistem ini dapat melakukan transaksi secara online, sehingga transaksi dapat dilakukan di banyak tempat. Tak hanya menggunakan sistem hutang yang konsumen dapat membayar secara berkala atau membuat ketentuan supaya memudahkan konsumen, bank juga menerapkan sistem hadiah terhadap nasabah bank supaya tetap mempertahankan nasabahnya.

Kartu kredit juga menggunakan sistem bayar langsung melalui rekening yang dimiliki oleh nasabah *bank*, sehingga uang yang digunakan untuk transaksi langsung menggunakan uang nasabah. Transaksi menggunakan kartu kredit dapat digunakan secara *online* atau *offline*. Transaksi online cukup memasukkan nomor dari kartu kredit kedalam input metode pembayaran di toko online atau *e-commerce* atau yang disebut komersial/ belanja online, sedangkan transaksi *offline* langsung menggunakan kartu kredit untuk pembayaran di tempat dengan mesin *Electronic Data Capture*(EDC) atau alat gesek kartu kredit untuk pembayaran non tunai [1].

Meskipun metode kartu kredit banyak keuntungan terhadap aktifitas jual-beli, akan tetapi sistem kartu kredit mempunyai masalah tersendiri. Dengan berkembangnya jaringan, masalah yang dihadapi oleh sistem ini adalah penipuan kartu kredit atau kartu kredit palsu untuk transaksi atau disebut *credit card fraud*(CCF) atau *fraud* dalam kartu kredit [1].

R.Goyal membuat catatan berdasarkan statistik tahun 2012 terdapat 6 negara yang tercatat mengalami *fraud* dalam kartu kredit yaitu Ukraina, Indonesia, Yugoslavia, Malaysia, Turki , USA dan negara lainya. Berdasarkan catatan juga Ukraina memiliki nilai *fraud* sebesar 19%, kemudian survey di Indonesia memiliki tingkat *fraud* sebesar 18,3%. Negara setelah Indonesia adalah Yugoslavia memiliki tingkat *fraud* dalam kartu kredit sebesar 17,8% kemudian diikuti oleh Malaysia sebesar 5,9 % dan Turki 9% kemudian USA dan negara lainya hanya memiliki kasus 1% dari *fraud* dalam kartu kredit [2].

R. Jhangiani menyatakan di dalam penelitiannya, berdasarkan laporan perkembangan *fraud* dalam komersial/berbisnis secara elektronik yang terjadi pada quarter pertama tahun 2018 dibandingkan dengan tahun 2016 melebihi perkembangan transaksi dengan 83%. Pada daftar *fraud* dalam komersial/berbisnis secara elektronik melaporkan perkembangan *fraud* pada akun online departement store meningkat dari 0.06% dari 2016 hingga 0.23% di tahun 2017. Memperlihatkan baha lebih dari 10% total kerugian dari fraud. Penambhan *fraud* dalam kartu kredit masih pada 0.1% untuk semua kartu transaksi, yang menyebabkan kerugian keuangan sebagai kasus *fraud* dalam transaksi. Akan

tetapi dengan bertambahnya penggunaan kartu kredit dalam *volume* dan nilai, nilai proporsi dari *fraud* tetap sama [3].

R.Goyal membagi kategori dari *fraud* dalam kartu kredit menjadi 2 kategori yaitu *offline* dan *online fraud*. *Offline fraud* berupa pencurian kartu kredit milik orang lain. Kemudian *online fraud* dapat berupa *telecommunication fraud*, *computer penetration*, *bankruptcy fraud*, *theft / fake fraud*, *application fraud*. Pertama adalah *Telecommunication fraud* adalah bentuk *fraud* dalam kartu kredit dengan menggunakan layanan telekomunikasi seperti telpon dan sms, korban yang diincar adalah konsumen, pebisnis dan penyedia komunikasi servis. Bentuk kedua adalah *computer intrusion* yaitu mengakses dan manipulasi data tanpa ada izin dari pemilik, penyusup bisa orang luar atau orang yang tau sistem keamanan. Bentuk ketiga adalah *bankruptcy fraud*, yaitu bentuk *fraud* paling kompleks karena tak hanya menyangkut masalah CC. Keempat adalah *Theft Fraud/ Counterfeit Fraud*, bentuk ini menyatakan bahwa kartu ini bukan pemilik yang asli, ketika ada laporan maka bank akan menginvestigasi, secara bersamaan pelaku *fraud* akan menggunakan secara *remote* atau mengontrol secara kejauhan ketika melakukan *fraud*. Kelima adalah *Applications Fraud*, yaitu ketika seseorang memberikan data palsu yang berarti *application fraud*. [4].

Fraud dalam kartu kredit adalah masalah *global* karena perkembangan yang cepat terhadap komersial online dan pembayaran online, *fraud* dalam kartu kredit tidak hanya merugikan pihak nasabah tapi juga merugikan pihak bank dan pihak penjual. *Fraud* dalam kartu kredit adalah penggunaan kartu secara ilegal atau biasanya terjadi transaksi dengan kebiasaan yang tidak semestinya. *Fraud* dalam

kartu kredit ada 4 kebiasaan yang menjadikan *fraud* dalam kartu kredit yaitu kartu kredit yang tercuri atau kartu yang hilang, *mail theft*, *fake card* dan penggunaan bukan pengguna yang asli [5].

Penanggulangan *fraud* dalam kartu kredit dapat dilakukan dengan 2 hal yaitu pencegahan dan deteksi. Pencegahan adalah menghindari berbagai macam serangan dari pemalsu dengan menggunakan beberapa layar dari penjagaan, kemudian jika Pencegahan terjadi kegagalan maka dilakukan deteksi supaya langsung memperingatkan terhadap transaksi palsu atau *fraud* dalam kartu kredit yang dilakukan.

Aplikasi untuk deteksi *fraud* dalam kartu kredit penting untuk mengurangi *fraud* kartu kredit. Penggunaan pembelajaran mesin untuk deteksi *fraud* dalam kartu kredit juga sedikit perkembangannya karena kurangnya data yang disediakan dikalangan umum, dikarenakan sensitifitas data yang dirahasiakan. Deteksi *fraud* dalam kartu kredit menggunakan metode *manual* tidak disarankan karena makan waktu dan rendah akurasi. Penggunaan *machine learning* atau pembelajaran mesin direkomendasikan karena mengatasi masalah deteksi secara *manual* [6].

Pembelajaran mesin adalah metode yang digunakan komputer untuk analisa data dalam bersekala tinggi secara otomatis dan termasuk cabang dari kecerdasan komputasional yang mempelajari pola untuk membuat keputusan. Pembelajaran mesin dapat dikategorikan dalam 2 hal yaitu *supervised* dan *unsupervised*, untuk deteksi kartu kredit digunakan *supervised data*, dalam *machine learning* juga ada 2 tipe data pertama adalah data kategori dan data angka. Data dalam bentuk dapat

diganti dengan angka supaya dapat mengatasi masalah categorical data yang tidak dapat dihitung numerik secara keseluruhan [5].

Sebelumnya telah dilakukan penelitian untuk deteksi *fraud* dalam kartu kredit, seperti Kuldeep Randhawa pada tahun 2018 telah menggunakan *ada boost* dengan *voting* hasil akurasi dari metode ini adalah 0,942. Kemudian Shiyang Xuan *et al.* di tahun yang sama menggunakan *random forest* untuk deteksi *fraud* dalam kartu kredit dengan B2C *dataset*. Dengan akurasi rata – rata sebesar 95% dan nilai F1 sebesar 87%. kekurangan metode yang digunakan oleh Kuldeep Randhawa yaitu nilai *precision* yang didapatkan tidak tinggi, kemudian untuk Shiyang Xuan masalah seperti *imbalace dataset* dan *big data* membuat kurang efektif untuk prediksi dan hanya untuk *dataset* kecil[4]

Pada deteksi *fraud* terdapat masalah yang dihadapi ketika melakukan proses pengujian data, tantangan yang pertama yaitu dihadapi yaitu pada tidak seimbang nya ukuran data, biasanya data pelaku *fraud* dalam kartu kredit rasionya dibawah 10%. Kedua adalah perbedaan pada misklasifikasi kepentingan, yaitu ketika proses deteksi mempunyai versi yang berbeda juga mempunyai makna yang berbeda atau salah prediksi ketika melakukan deteksi. Ketiga adalah *overlapping data* yaitu data transaksi memungkinkan dapat dideteksi sebagai transaksi *fraud* tapi sebenarnya *false positive*. Keempat adalah kekurangan adaptasi dalam perkembangan zaman atau metode yang kuno, yaitu harus sering mengembangkan algoritma untuk deteksi karena perubahan jaman. Kelima karena biaya deteksi terkadang kita juga harus memastikan biaya karena kondisi perangkat keras juga harus dipertimbangkan untuk memperoleh hasil yang

optimal. Keenam tidak ada standard matrik evaluasi khusus untuk deteksi *fraud* dalam kartu kredit, untuk membandingkan antara hasil dengan system. [6]

Dalam paper ini akan dilakukan klasifikasi atau deteksi *fraud* dalam kartu kredit dengan metode *lasso logistic ensemble*. Metode ini dipilih karena mampu menyelesaikan masalah tidak seimbang data dan memberikan akurasi prediksi yang tinggi. *Lasso* dan *logistic regression* adalah bentuk dari regresi linier. *logistic regression* adalah regresi linier untuk klasifikasi yang bagus untuk metode klasifikasi karena metode ini mempunyai akurasi yang tinggi dalam prediksi kelas, *lasso* adalah regresi linier dengan *L1 penalty*. *Ensemble method*, metode yang digunakan, jika data mengalami ketidakseimbangan dalam *dataset* yang digunakan [7].

Dalam penelitian ini dibangun sistem untuk deteksi *fraud* dalam kartu kredit. Data mempunyai jumlah total 284807 data dengan *positive frauds* 492 dari semua transaksi dengan menerapkan *lasso logistic ensemble* sebagai salah satu cara untuk deteksi *fraud* dalam kartu kredit.

1.2. Rumusan Masalah

Pada bab Rumusan Masalah ada tiga poin penting yaitu permasalahan, usulan metode dan pertanyaan penelitian.

1.2.1. Permasalahan

Banyaknya kasus *fraud* pada transaksi kartu kredit, diperlukan deteksi untuk identifikasi *fraud* dalam kartu kredit. Dalam deteksi *fraud* terdapat permasalahan *imbalanced dataset* dimana jumlah pengguna *fraud* kurang dari 5%. Oleh karena itu dibutuhkan suatu metode untuk mengatasi *imbalanced dataset*

ketika melakukan proses training data supaya meningkatkan tingkat akurasi prediksi *fraud*.

1.2.2. Metode Usulan

Penelitian ini menggunakan *lasso logistic regression ensemble method* untuk mengatasi imbalance *dataset* dalam deteksi *fraud* dalam kartu kredit serta mendapatkan tingkat akurasi yang tinggi dalam klasifikasi, sebelum melakukan pelatihan, dataset diseleksi fitur yang digunakan dengan *r-squared* supaya meningkatkan akurasi prediksi kelas.

1.2.3. Pertanyaan Penelitian

1. Bagaimana implementasi dataset yang tidak seimbang untuk deteksi menggunakan *lasso logistic regression ensemble* ?
2. Bagaimana tingkat akurasi dan ROC AUC dari deteksi *credit card fraud* dengan *lasso logistic regression ensemble method* ?

1.3. Tujuan dan Manfaat Penelitian

1.3.1. Tujuan

Penelitian ini bertujuan untuk

1. Mengetahui implementasi *lasso logistic regression ensemble method* untuk deteksi pada dataset yang tidak seimbang
2. Mengetahui tingkat akurasi dan nilai ROC AUC dari deteksi *fraud* dalam kartu kredit menggunakan *lasso logistic regression ensemble method*

1.3.2. Manfaat

Manfaat penelitian sebagai berikut:

1. Alat untuk membantu institusi atau perusahaan keuangan seperti perbankan dalam mencegah kerugian dalam transaksi yang bersifat *fraud*

1.4. Batasan Masalah

Adapun batasan masalah penelitian ini adalah:

1. Data yang digunakan adalah data *credit card fraud dataset* yang didapatkan dari kaggle : <https://www.kaggle.com/mlg-ulb/creditcardfraud>
2. Fitur yang digunakan untuk target prediksi adalah fitur *class*, yang mempunyai nilai fitur 0 dan 1.
3. Jumlah fitur untuk model prediksi sebanyak 30 fitur.

1.5. Sistematika Proposal

Sistematika laporan dibagi menjadi beberapa bab yaitu :

1. BAB I PENDAHULUAN : Bab ini membahas tentang latar belakang mengenai ruang lingkup masalah yang dibahas, solusi yang diusulkan untuk dapat menyelesaikan permasalahan, tujuan dan manfaat dengan diterapkannya solusi yang diusulkan, serta sistematika penulisan laporan.
2. BAB II KAJIAN PUSTAKA : Pada bab ini dibahas mengenai metode usulan yang dipilih, contoh penggunaan metode dan penelitian terkait
3. BAB III USULAN METODE : bab ini berisi tahapan penelitian metode yang diusulkan, arsitektur, skenario pengujian serta data yang digunakan
4. BAB IV HASIL DAN PEMBAHASAN : Bab ini menguraikan implementasi dari metode penelitian pada bab sebelumnya, scenario pengujian, proses hasil ujicoba sistem, dan analisis hasil ujicoba sistem.
5. BAB V KESIMPULAN : Bab ini berisi kesimpulan dari penelitian yang disesuaikan dengan pertanyaan penelitian dan saran terkait pengembangan aplikasi selanjutnya.
6. REFRENSI : Berisi tentang daftar jurnal-jurnal atau referensi yang dijadikan acuan dalam penelitian ini.