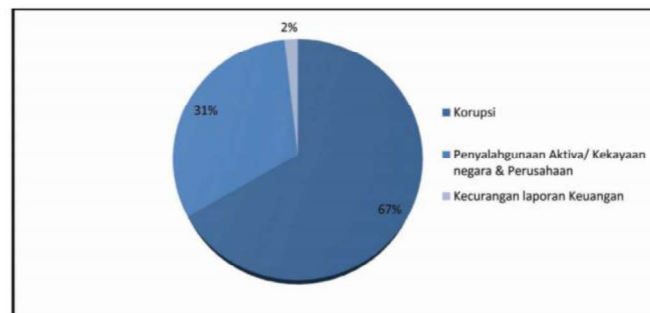


Bab I Pendahuluan

I.1 Latar Belakang

Pada perkembangan teknologi yang semakin modern, banyak orang yang ingin melakukan penipuan ataupun kecurangan untuk mendapatkan keuntungan yang merugikan banyak pihak (Utomo, 2018). Menurut *Association of Certified Fraud Examiners* (ACFE, 2020), “*Fraud is the use one’s occupation for personal enrichment through the deliberate misuse or application of the employing organization’s resources or assets*” yang artinya adalah “suatu tindakan untuk memperkaya diri melalui penyalahgunaan yang dilakukan secara sengaja atau penggunaan sumber daya organisasi atau aset-asetnya (ACFE, 2020). Pelaku-pelaku yang melakukan kejahatan *fraud* saat ini hanya terbatas pada golongan atas, tetapi sudah banyak pada lapisan pegawai bawah pada perusahaan. Hal ini tentunya menjadi satu hal yang perlu kita waspadai dan peduli terhadap sekeliling tempat kita bekerja. Pada penelitian yang telah dilakukan oleh *Association of Certified Fraud Examiners* (ACFE) Global menghasilkan bahwa pada setiap tahun rata-rata 5% dari pendapatan suatu organisasi akan menjadi korban *fraud* (Indonesia, Survey Fraud di Indonesia, 2016). Hasil *survey* yang dilakukan oleh ACFE tersebut dapat dilihat pada Gambar I.1.



Sumber: data diolah, 2016.

Gambar I.1 Fraud yang paling banyak di Indonesia
Sumber: (Indonesia, Survei Fraud Indonesia, 2016)

Berdasarkan hasil *survey* yang dilakukan oleh ACFE Indonesia pada tahun 2016, *fraud* yang paling banyak terjadi di Indonesia adalah korupsi (ACFE, 2020). Sebanyak 154 responden pada *survey fraud* Indonesia atau sebesar 67% memilih korupsi. Di Indonesia, berapa pun tingkat kerugian yang muncul dari *fraud*,

kerugian dari korupsi yang paling terbesar (Indonesia, Survei Fraud Indonesia, 2016).

Secara umum, *fraud* merupakan suatu tindakan penggunaan fasilitas telekomunikasi secara ilegal yang sengaja melakukannya dengan berbagai cara bentuk kecurangan, penipuan ataupun juga penggelapan oleh orang maupun perusahaan tertentu yang tujuannya adalah untuk menghindari biaya layanan atau pelacakan rekaman tagihan yang dilakukan secara ilegal (Sulisrudatin, 2018). Saat ini, banyak orang yang memanfaatkan SIMBox agar bisa memanipulasi nomor-nomor yang berasal dari luar negeri yang seharusnya menggunakan *roaming internasional*, tetapi dengan menggunakan SIMBox hitungannya menjadi panggilan lokal (Diskominfo, 2017). Pengguna alat ini dipastikan mendapatkan keuntungan yang besar. Karena panggilan dari luar negeri bisa dimanipulasi seolah-olah menjadi panggilan dalam negeri. Hal tersebut dilakukan untuk dapat menghindari tagihan komunikasi dan agar mendapatkan keuntungan serta dapat merugikan pihak-pihak yang bersangkutan. SIMBox merupakan sebuah perangkat yang tidak memiliki sertifikat (ilegal) (Susilo, 2006) (Puri, 2018).

Pelaku pengguna SIMBox banyak dilakukan oleh para *Fraudster* (sebutan untuk pelaku *fraud*) untuk mendapatkan keuntungan dengan memanfaatkan perbedaan tarif *incoming international* dengan tarif *domestic* terutama tarif retail *On-Net* (panggilan antar pelanggan dalam operator yang sama). Dengan mengalirnya *traffic incoming international* ke jalur SIMBox, maka pihak telekomunikasi yang seharusnya mendapatkan menjadi hanya mendapatkan *revenue* interkoneksi *domestic* yang berasal dari panggilan *On-Net domestic* dari SIM card yang ada pada perangkat SIMBox ke pelanggan penerima telepon. Karena SIMBox mengandalkan teknologi VoIP yang dimana *voice* dikonversi menjadi data dalam proses transmisinya, oleh karena itu akan menghasilkan suara yang lebih buruk dibandingkan dengan teknologi *voice* sebagaimana dengan jalur normal. Tentunya penurunan kualitas suara tersebut akan berdampak kepada komplain pelanggan dan penurunan kualitas dari pihak telekomunikasi bagi pelanggan. Dengan hilangnya *revenue* yang dirasakan oleh pihak telekomunikasi, hal ini akan berdampak pada turunnya pendapatan negara (AlBougha, 2016).

Menurut hasil monitoring perangkat Pos dan Informatika dari Dirjen SDPPI (Direktorat Jenderal Sumber Daya dan Perangkat Pos dan Informatika) Kominfo, SIMBox 2.26% sebagai alat dan perangkat telekomunikasi ilegal. Dampak dari SIMBox bagi perusahaan adalah kerugian pendapatan (*revenue loss*) dan *security compromise* (Diskominfo, 2017).

Kecurangan atau *fraud* saat ini sudah sangat meningkat, dengan meningkatnya kemajuan teknologi. Oleh karena itu, untuk mendeteksi data kecurangan SIMBox tersebut dapat menggunakan proses *data mining* (Rosario Taek, 2019). *Data mining* merupakan pengumpulan data, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam data berukuran besar (Nurchalifatun, 2015). *Data mining* terdapat beberapa algoritma diantaranya adalah algoritma *decision tree (C4.5)*, algoritma *K-Means*, algoritma *Support Vector Machines*, algoritma *Decision Tree*, algoritma *K-nearest Neighbors*, algoritma *Naïve Bayes*, algoritma *classification and regression trees*, dan algoritma *adaboost* (McLeod, Jr & Schell, 2007).

Dalam penelitian terdahulu banyak membahas mengenai implementasi beberapa algoritma. Berdasarkan penelitian yang dilakukan oleh (Salmu & Solichin, 2017) dalam memprediksi tingkat kelulusan mahasiswa menggunakan algoritma *Naïve Bayes*, diperoleh bahwa algoritma *Naïve Bayes* memperoleh akurasi sebesar 80.72%. Maka dapat disimpulkan, bahwa algoritma *Naïve Bayes* baik digunakan serta dijadikan acuan dalam memprediksi kelulusan mahasiswa.

Berdasarkan penelitian (Tempola, Muhammad, & Khairan, 2017) melakukan perbandingan klasifikasi antara algoritma *K-Nearest Neighbor* dan *Naïve Bayes* pada penentuan status gunung berapi dengan *K-Fold cross validation*. Pada hasil pengujian antara dua metode *machine learning* yang telah diterapkan pada sistem tersebut, diperoleh rata-rata akurasi sistem ketika menggunakan *K-Nearest Neighbor* sebesar 63,68%. Sedangkan ketika diterapkannya *Naïve Bayes classifier* dihasilkan rata-rata akurasi sistem sebesar 79,71% dan standar deviasi 3,55%. Dengan demikian ketika diterapkannya *Naïve Bayes classifier* akurasi sistem dalam melakukan klasifikasi lebih baik dibandingkan dengan algoritma *K-Nearest Neighbor*. Selain itu jarak akurasi setiap eksperimen dengan rata-rata akurasi lebih dekat ketika menggunakan algoritma *Naïve Bayes* dibandingkan dengan algoritma

K-Nearest Neighbor. Menurut penelitian (Nurhasan, Hikmah, & Utami, 2018) dalam melakukan perbandingan algoritma C4.5, *K-Nearest Neighbor* dan *Naïve Bayes* untuk klasifikasi menjadi penanggung jawab BSI *entrepreneur center* pada masing-masing kampus Universitas Bina Sarana Informatika. Berdasarkan hasil penelitian, algoritma *Naïve Bayes* memperoleh nilai akurasi yang paling tinggi dibandingkan algoritma C4.5 dan *K-Nearest Neighbor*, dengan nilai akurasinya sebesar 80%, metode C4.5 sebesar 73.33% dan metode K-NN sebesar 70%. Menurut penelitian (Azhar, 2019) dalam melakukan analisis kinerja algoritma *Naïve Bayes* dan *K-Nearest Neighbor* pada Sentimen Analisis dengan Pendekatan Lexicon di Media Twitter. Berdasarkan hasil penelitian, algoritma *Naïve Bayes* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma *K-Nearest Neighbor*. Uji coba yang dilakukan pada *Naïve Bayes* menghasilkan akurasi sebesar 80% sedangkan algoritma *K-Nearest Neighbor* menghasilkan akurasi sebesar 77%.

Sedangkan menurut penelitian yang dilakukan oleh (Pratiwi, 2016) dalam melakukan analisa perbandingan algoritma K-Means, *decision tree*, dan *Naïve Bayes* untuk sistem pengelompokan siswa otomatis. Berdasarkan hasil coba terhadap tiga algoritma menunjukkan bahwa *Naïve Bayes* mampu mengelompokkan siswa dengan tingkat akurasi 70.37% seperti guru mengelompokkan dan waktu yang dibutuhkan untuk mengelompokkan tercepat di antara K-Means dan *decision tree*.

Pada penelitian ini akan dibahas penggunaan klasifikasi *Naïve Bayes* dalam memprediksi data panggilan telepon. Pemilihan metode yang digunakan didasarkan oleh beberapa hal. Metode klasifikasi merupakan metode yang cocok digunakan dalam sistem prediksi, sedangkan algoritma *Naïve Bayes* merupakan implementasi yang sederhana dan waktu komputasi yang relatif cepat dalam proses pendeteksian SIM *Box*. Algoritma *Naïve Bayes* mampu mengestimasi parameter dengan jumlah data *training* yang kecil untuk melakukan klasifikasi dan mampu menangani nilai-nilai yang hilang selama perhitungan estimasi peluang.

Oleh karena itu berdasarkan penelitian yang telah dilakukan terdahulu didapatkan algoritma *Naïve Bayes* cukup baik dari algoritma lainnya, pada penelitian ini akan menggunakan algoritma *Naïve Bayes* untuk analisis dan deteksi

fraud pada data panggilan. Sehingga dapat menghasilkan data nomor panggilan yang terdeteksi sebagai *fraud* dan *not fraud* berdasarkan *behavior* pengguna panggilan telepon. Hasil dari penelitian ini diharapkan dapat membantu memprediksi nomor panggilan telepon *SIMBox fraud*.

I.2 Perumusan Masalah

Adapun rumusan masalah yang akan diselesaikan dalam penelitian ini berdasarkan dengan latar belakang yaitu bagaimana melakukan pendeteksian data panggilan yang menggunakan *SIM Box fraud* sehingga terindikasi sebagai kecurangan yang terjadi pada perusahaan telekomunikasi. Proses pendeteksian kecurangan ini dilakukan dengan menggunakan algoritma *Naïve Bayes* agar dapat melakukan identifikasi kecurangan pada data panggilan dengan cepat dan akurat.

I.3 Tujuan Penelitian

Melakukan pendeteksian secara cepat dan akurat pada data panggilan yang menggunakan *SIM Box* yang terindikasi sebagai kecurangan dengan menggunakan algoritma *Naïve Bayes* serta mengukur kinerja algoritma pada data panggilan yang melakukan kecurangan.

I.4 Manfaat Penelitian

Manfaat yang didapatkan dari penelitian ini, diantaranya:

1) Manfaat Keilmuan

Penelitian ini diharapkan memberikan informasi yang dapat berguna untuk mengatasi kecurangan secara cepat. Serta mampu memetakan perilaku panggilan telepon yang terdeteksi sebagai panggilan yang menggunakan *SIM Box*. Dengan melakukan pengolahan data panggilan dapat mengetahui seluruh aktifitas pengguna panggilan telepon yang terekam di dalam CDR. Selanjutnya adalah menganalisis *behavior* dari pengguna panggilan telepon berdasarkan parameter yang mempengaruhi proses pendeteksian kecurangan dengan menggunakan algoritma *Naïve Bayes*.

2) Manfaat Praktis

Penelitian ini diharapkan dapat memberikan manfaat bagi pihak PT XYZ agar dapat mendeteksi data nomor telepon yang menggunakan SIM *Box* dengan lebih cepat dan akurat sehingga dapat mempersingkat waktu. Sehingga perusahaan tidak kehilangan *revenue* yang berasal dari *incoming international*.

I.5 Batasan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan di atas, penelitian ini dibatasi mengingat rumusan masalah pada penelitian ini bersifat umum, diantaranya:

1. Data yang diambil adalah data panggilan bulan Juni, Juli dan Agustus tahun 2017.
2. Aplikasi yang digunakan dalam pengolahan data yaitu Excel, Pentaho Data Integration, MySQL dan Python.
3. Algoritma yang diimplementasikan adalah *Naïve Bayes*.
4. Mendeteksi nomor panggilan telepon *SIMBox fraud* dan *not fraud*.

I.6 Sistematika Penelitian

Sistematika penulisan penelitian ini terdiri dari bab yang berisi rincian penelitian, bab tersebut terdiri dari:

BAB I Pendahuluan

Bab ini berisi detail-detail penelitian yang terdiri dari latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan sistematika penulisan. Bab ini mengacu kepada bab-bab selanjutnya yang menjadi dasar mengenai penelitian.

BAB II Landasan Teori

Bab ini berisi definisi teori dari masing-masing cara dan tools yang digunakan dalam proses penelitian sehingga dengan adanya definisi dapat mempermudah sebagai acuan pelaksanaan penelitian.

BAB III Metodologi Penelitian

Bab ini berisi penjelasan metode secara konseptual yang disusun dari lingkungan, dasar ilmu, serta penelitiannya. Dan juga berisi penjelasan metode secara sistematis agar lebih mengetahui tahapan-tahapan yang dilakukan dalam bentuk swimlane.

BAB IV Analisis dan Pemodelan

Bab ini berisi mengenai data yang didapatkan, proses penelitian, hasil yang didapatkan sehingga dapat dianalisis.

BAB V Implementasi dan Pengujian

Bab ini berisi penjelasan dari proses penelitian dengan *data mining* yang dimanfaatkan dengan Algoritma *Naïve Bayes*.

BAB VI Kesimpulan dan Saran

Bab ini berisi kesimpulan yang dapat diambil dari hasil yang didapat pada penelitian sehingga dapat dijadikan sebagai bahan pertimbangan kedepannya, dan berisi saran agar sekiranya dapat menjadikan penelitian ini sebagai acuan untuk mengembangkannya di masa depan.