

# 1 Pendahuluan

## 1.1 Latar Belakang

Persaingan antara perusahaan telekomunikasi seluler pada masa ini adalah bagaimana cara mempertahankan pengguna layanan yang masih aktif dalam menggunakan layanan dan menarik pengguna baru yang berpotensi menjadi pelanggan. Pelanggan menjadi salah satu faktor utama dalam kesuksesan yang dicapai dalam perusahaan telekomunikasi seluler [1]. Munculnya berbagai jenis penyedia jasa telekomunikasi dapat menjadi faktor perpindahan pelanggan menjadi masalah utama atau yang sering disebut *churn*.

*Churn* adalah salah satu aksi yang berkaitan dengan pelanggan, yaitu pelanggan beralih dari satu perusahaan ke perusahaan yang lain [2]. *Customer churn* merupakan masalah yang sangat diperhitungkan oleh perusahaan, terdapat proses yang dimiliki oleh perusahaan untuk menangani masalah ini yaitu *customer management*. Proses ini bertujuan untuk mempertahankan pelanggan yang menguntungkan. Perusahaan menyadari bahwa strategi bisnis yang berorientasi pelanggan akan sangat penting untuk mempertahankan keuntungan [1].

Salah satu strategi yang dibutuhkan dalam penanganan kasus *churn* adalah *Customer Relationship Management* (CRM). Strategi ini mendapatkan informasi langsung dari masing-masing pelanggan, data yang dimasukkan dalam proses registrasi. Strategi ini dapat dilakukan untuk mengidentifikasi pelanggan yang kemungkinan akan membatalkan kontrak mereka [3]. *Churn prediction* harus dilakukan untuk mengetahui kemungkinan pelanggan berpindah layanan. Banyak perusahaan yang menggunakan teknik data mining untuk analisis *churn*. Berbagai teknik data mining seperti *sequential patterns*, *genetic modelling*, *classification trees*, *neural networks*, and *SVM* telah dilakukan untuk mengeksplorasi *churn* [4].

Penerapan teknik *churn prediction* menggunakan metode data mining untuk melakukan klasifikasi, dan terdapat masalah yang timbul pada kasus ini yaitu *imbalance data*. Masalah ini disebabkan karena jumlah dari data

*non-churn* dibandingkan data *churn* tidak seimbang [5]. Dengan perbandingan data yang tidak seimbang maka akan mempengaruhi metode klasifikasi dari data mining untuk memprediksi kelas *churn*, akurasi tinggi sering dicapai dalam klasifikasi tanpa penanganan *imbalance data* karena hanya berfokus pada data mayoritas dan untuk data minoritas dianggap sebagai data langka atau data tidak disengaja [6].

Pada penelitian Tugas Akhir ini menggunakan metode *Logistic Regression* (LR) sebagai metode klasifikasi. *Logistic regression* merupakan model regresi biner statistik standar [2]. Metode *logistic regression* mempelajari hubungan antara variabel dependen dan beberapa variabel independen. Nama regresi logistik digunakan apabila variabel dependen hanya memiliki dua nilai, contohnya 0 dan 1 [7]. Kegunaan regresi logistik yang dapat melakukan klasifikasi dengan nilai dependen variabel 0 dan 1 akan sangat berguna bagi *churn prediction* yang memiliki nilai variabel dependen atau kelas *churn* 0 dan 1.

Salah satu metode yang dapat digunakan untuk menangani *imbalance data* adalah metode *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE merupakan salah satu metode *oversampling* untuk menangani masalah *imbalance data* dengan melakukan peningkatan pada data minor dengan cara duplikasi. Dalam hal sampling sintesis merupakan metode yang ampuh, telah terbukti pada setiap penelitian [8].

## 1.2 Perumusan Masalah

Berdasarkan paparan masalah pada latar belakang, maka permasalahan yang akan dibahas dan diteliti adalah:

1. Bagaimana cara penanganan terhadap *imbalanced data* dengan menggunakan metode *SMOTE*(*Synthetic Minority Over-sampling Technique*) ?
2. Bagaimana menerapkan *Logistic Regression* dalam menangani kasus *churn prediction*?
3. Bagaimana cara mengukur dan menganalisis performansi *Logistic Regression* dalam menangani kasus *churn prediction*?

### 1.3 Tujuan

Berdasarkan perumusan masalah diatas, maka tujuan dari penulisan Tugas Akhir ini adalah:

1. Menganalisis permasalahan *imbalanced data* dengan menggunakan teknik *SMOTE* (Synthetic Minority Over-sampling Technique)
2. Membangun sistem *churn prediction* dengan menerapkan metode *Logistic Regression* dan *SMOTE* (Synthetic Minority Over-sampling Technique)
3. Menganalisis performansi dari metode *Logistic Regression* dengan membandingkan dengan metode klasifikasi lain dalam bentuk *precision, recall, F1-measure*, dan *accuracy*.

### 1.4 Batasan Masalah

Batasan masalah yang digunakan dalam penulisan Tugas Akhir ini adalah:

1. Data berupa data .xlsx berisi data teks dengan format excel
2. Data yang digunakan pada analisis yaitu data perusahaan PT.Telekomunikasi Indonesia. Dengan data testing berjumlah 100580 record dan data training 99807 record.
3. Data yang digunakan berasal dari data pelanggan PT.Telekomunikasi Indonesia
4. Menganalisis hasil dari sistem yang dibuat menggunakan metode *Logistic Regression* saja dan sistem yang menggunakan metode *Logistic Regression* ditambah dengan dengan *SMOTE*(*Synthetic Minority Over-sampling technique*)
5. Penanganan data pada seleksi atribut hanya dilakukan terhadap data numerik dan data kategorikal diasumsikan sebagai data numerik.

### 1.5 Metodologi Penyelesaian Masalah

Metode yang digunakan untuk menyelesaikan masalah dalam Tugas Akhir ini adalah:

1. Studi literatur terdapat dua tahap, yaitu:

- a. Tahap pertama dilakukan pencarian penelitian terkait yang memiliki hubungan dengan Tugas Akhir ini, diantaranya *logistic regression, churn prediction, handling imbalance*, SMOTE, dan beberapa pengukuran performansi.
  - b. Tahap kedua dilakukan pendalaman materi mengenai hal-hal yang berkaitan dengan Tugas Akhir
2. *Feature selection* merupakan tahap yang dilakukan untuk mengetahui keterkaitan antar variabel/atribut didalam data. Sehingga model klasifikasi dapat didapat. Tahap ini hanya mengolah data numerik dan untuk data kategorikal diasumsikan sebagai data numerik.
  3. Mempelajari konsep dari SMOTE dan tahap-tahap yang terdapat didalamnya untuk penanganan *imbalance data*.
  4. Mempelajari konsep dari *logistic regression* dan tahap-tahap didalamnya untuk klasifikasi.
  5. Melakukan penggabungan *dataset* dari data *training clean* dan *testing clean*. Setelah tahap penggabungan dilakukan *split* data menjadi *data training* dan *data testing* untuk digunakan dalam pembangunan model klasifikasi dan pengukuran performansi model klasifikasi.
  6. Melakukan penanganan *imbalance data* menggunakan SMOTE pada *data training*.
  7. Membangun model klasifikasi menggunakan metode *logistic regression*.
  8. Menganalisis model klasifikasi menggunakan pengukuran yang dinyatakan dalam *precision, recall*, dan *F1-measure* dan *overall accuracy*.
  9. Menarik kesimpulan dari hasil penelitian dan menyusun Tugas Akhir

## **1.6 Sistematika Penulisan**

Untuk memahami lebih jelas Tugas Akhir ini, terdapat materi-materi yang dikelompokkan menjadi beberapa sub bab dengan sistematik penulisan sebagai berikut:

### **BAB 1 Pendahuluan**

Pada bab ini berisi latar belakang, perumusan masalah, tujuan, batasan masalah, metodologi penyelesaian masalah dan sistematika penulisan.

## **BAB 2 Tinjauan Pustaka**

Bab ini berisi penjelasan tentang penelitian terkait dan teori yang diambil dari kutipan buku maupun paper yang berkaitan dengan penyusunan laporan Tugas Akhir.

## **BAB 3 Perancangan Sistem**

Bab ini berisi penjelasan mengenai rancangan sistem *preprocessing data*, *imbalance data* dan klasifikasi dengan *logistic regression (LR)* yang meliputi flowchart pada masing-masing proses.

## **BAB 4 Pengujian dan Analisis**

Bab ini berisi penjelasan skenario pengujian yang dilakukan untuk mencapai tujuan penelitian yang ditulis pada Bab 1 Pendahuluan. Pembahasan hasil pengujian berdasarkan skenario pengujian yang dituliskan. Selain itu, pada bab ini juga dijelaskan analisis terhadap hasil pengujian tersebut. Hasil dari kegiatan analisis ini menjadi dasar pengambilan kesimpulan.

## **BAB 5 Kesimpulan dan Saran**

Bab ini berisi kesimpulan dan saran yang berkaitan dengan analisa sistem berdasarkan yang telah diuraikan pada bab-bab sebelumnya.