

BAB I PENDAHULUAN

1.1 Latar Belakang

Persaingan bisnis di zaman sekarang sangatlah ketat. Segala hal akan dicoba suatu perusahaan bisnis agar dapat meraih untung lebih banyak agar tetap berada dalam persaingan. Salah satu permasalahan penting yang dihadapi perusahaan-perusahaan besar adalah *Customer Churn*. *Customer Churn* adalah berakhirnya hubungan antara pelanggan dengan suatu penyedia layanan jasa yang terjadi karena suatu hal yang dialami oleh pelanggan.

Perusahaan-perusahaan besar saling bersaing dalam mempertahankan pelanggan agar bisnis mereka dapat terus berkembang. Perusahaan dapat mencegah seorang pelanggan yang akan churn dengan prediksi churn dan pada akhirnya membantu perusahaan tersebut agar tidak kehilangan pendapatan [6]. Tetapi, terdapat beberapa kendala pada prediksi churn, salah satunya yaitu *imbalanced data*. Oleh karena itu, diperlukan model prediksi yang baik agar prediksi churn mendapatkan hasil yang akurat dan tahan pada masa yang kompetitif seperti sekarang [3].

Algoritma klasifikasi dapat berfungsi secara optimal ketika *dataset* yang digunakan dalam proses klasifikasinya adalah *dataset* yang persebaran kelasnya merata. Dalam proses prediksi churn, *dataset* yang digunakan adalah *dataset* yang persebarannya tidak merata. *Imbalanced dataset* terjadi ketika persebaran kelas dalam suatu *dataset* tidak seimbang. Hal ini berdampak negatif terhadap hasil prediksi suatu perusahaan karena dapat mengakibatkan hilangnya pendapatan.

Undersampling adalah salah satu teknik data mining yang dapat digunakan untuk menangani permasalahan *imbalanced dataset*. *Undersampling* akan memilih beberapa data pada kelas mayoritas dan menghapusnya dari *dataset* agar *dataset* menjadi seimbang. Teknik *undersampling* memiliki kelebihan pada waktu komputasi dan penggunaan memori [8] jika jumlah data pada *dataset* yang digunakan sangatlah banyak.

Underbagging adalah metode yang menggabungkan *undersampling* dengan *bagging*. Cara kerjanya adalah dengan membuat beberapa gugus data untuk dijadikan data latih untuk membuat model *classifier*, kemudian dilakukan *voting* terhadap hasil klasifikasi dari tiap model.

Support Vector Machine (SVM) adalah salah satu algoritma untuk klasifikasi yang mendefinisikan sebuah *hyperplane* untuk mengklasifikasikan suatu *dataset*. *Support Vector Machine* termasuk ke dalam metode *supervised learning* yaitu *classifier* yang memerlukan data latih. Pada algoritma SVM, *hyperplane* yang digunakan adalah *hyperplane* yang memiliki jarak terbesar antar kelas yang ada pada *dataset* tersebut. Pada pengujian yang dilakukan oleh Farquad, Ravia, dan Raju, *classifier* SVM dapat menghasilkan nilai F_1 -measure sebesar 84.03% dalam memprediksi churn pada data Bank di wilayah Latin Amerika [19]. Pada [20] juga ditunjukkan bahwa *classifier* SVM dapat menghasilkan akurasi yang lebih tinggi dibandingkan dengan *classifier* ANN, *Decision Tree*, *Logistic Regression*, dan *Naïve bayesian classifier* yaitu sebesar 90.88%.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana proses penanganan *imbalanced data*?
2. Bagaimana pengaruh penanganan *imbalanced data* proses prediksi churn?
3. Bagaimana menganalisis hasil dari pengujian sistem?

Batasan masalah dari tugas akhir ini adalah data yang digunakan adalah data pelanggan internet SPEEDY wilayah Telkom regional 7 yang dibagi menjadi 2 bagian yaitu, data latih dan data uji.

1.3 Tujuan

Tujuan yang dicapai pada penelitian ini adalah:

1. Mengimplementasikan metode *underbagging* untuk menangani permasalahan *imbalanced data* pada *dataset* churn yang digunakan.
2. Mengimplementasi metode *Support Vector Machine* pada *dataset* churn.
3. Menganalisis hasil model klasifikasi yang digunakan untuk prediksi churn.

1.4 Metodologi Penelitian

Metodologi penelitian yang akan dilakukan adalah:

1. Studi Literatur

Mempelajari dan memahami referensi yang berhubungan dengan tugas akhir yang dilakukan termasuk tentang *Undersampling*, *Bagging*, *Support Vector Machine*, *Confusion Matrix*, dan prediksi churn. Referensi yang digunakan berasal dari jurnal, buku, dan sumber lain yang ada di internet.

2. Analisis dan Perancangan Sistem

Merancang alur sistem prediksi churn menggunakan *underbagging* dan SVM. Dimulai dari *preprocessing* data sebelum digunakan, hingga pengukuran kinerja sistem menggunakan *confusion matrix*.

3. Implementasi Sistem

Membangun sistem yang sesuai dengan rancangan sistem. Setelah sistem terimplementasi dengan baik, akan didapatkan hasil yang dapat diukur menggunakan *confusion matrix*.

4. Analisis Hasil Implementasi Sistem

Menganalisis kinerja dari model SVM dan teknik *handling imbalanced data underbagging* dengan menggunakan *confusion matrix* sebagai pengukuran kinerja sistem.

5. Pembuatan Laporan Tugas Akhir

Membuat laporan tugas akhir berdasarkan hasil implementasi sistem dan analisis dari implementasi sistem yang dibuat. Laporan tugas akhir ini terdiri atas lima bab, yaitu pendahuluan, tinjauan pustaka, perancangan sistem, hasil dan analisis, dan kesimpulan dan saran.