

1.PENDAHULUAN

1.1 Latar Belakang

Customer memiliki potensi untuk berhenti menggunakan produk atau jasa apabila merasa tidak puas. Kepuasan pelanggan ini dapat disebabkan oleh banyak faktor, mulai dari perubahan selera, perpindahan domisili, ataupun kurangnya *customer relationship management* sebuah perusahaan (Prabadevi *et al.*, 2023). Pelanggan yang berhenti menggunakan suatu jasa atau produk dan berpindah ke kompetitor lain biasa disebut dengan *customer churn*. Kehilangan customer tentunya berakibat pada berkurangnya pendapatan perusahaan tersebut. Meskipun perusahaan dapat mencari *customer* baru untuk menggantinya yang telah *churn*, namun menurut jurnal marketing mendapatkan *customer* baru membutuhkan biaya lebih besar bila dibandingkan biaya untuk mempertahankan *customer* (Wagh *et al.*, 2023).

Customer churn terjadi di semua perusahaan dan menjadi sebuah permasalahan di kebanyakan perusahaan (Kassem *et al.*, 2020), tidak terkecuali pada bidang usaha data center (Miller, 2023). PT X sendiri adalah perusahaan yang bergerak dalam bidang *data center*. Perusahaan menyediakan tempat dan fasilitas untuk *customer* menipkan server mereka. Sistem pembelian adalah pelanggan yang ingin menggunakan produk dari perusahaan ini harus mengambil kontrak yang berlaku 1 tahun. Salah satu tantangan yang sedang dihadapi oleh PT X adalah *customer churn*. Meskipun setiap tahun ada *customer* baru, namun jumlah pelanggan yang berhenti menggunakan juga terhitung banyak dan cukup mengurangi pendapatan dari perusahaan.

Customer churn dapat diminimalisir dengan adanya *customer analysis*. Menurut ahli, *customer analysis* atau analisa pelanggan adalah kegiatan yang dilakukan dengan data target pasar untuk memahami pasar (Moscato *et al.*, 2019). Analisa pelanggan dapat menghasilkan berbagai informasi yang bisa digunakan untuk pengambilan keputusan. Salah satu informasi tersebut adalah prediksi *customer churn* yang berkaitan langsung dengan kelayakan pelanggan (Madushanka *et al.*, 2015). Prediksi ini dapat menjadi alat untuk bersaing di era dimana

perusahaan berlomba - lomba untuk mempertahankan *customer*. Hasil dari prediksi *customer churn* dapat digunakan untuk mengetahui siapa saja *customer* yang memiliki kemungkinan untuk berpindah (Li. J, 2023). Dengan adanya informasi ini, perusahaan dapat menetapkan strategi yang tepat untuk memperkecil kemungkinan *churn* dan mengatur prioritas penanganannya.

Penelitian mengenai prediksi *customer churn* telah banyak dilakukan di bidang usaha yang beragam. Ada yang di bidang telekomunikasi, penyedia streaming film, penyedia streaming lagu, toko retail, bank, dan masih banyak lagi. Metode yang digunakan untuk melakukan prediksi juga beragam, mulai dari *Support Vector Machine*, *ANN*, *KNN*, *Decision Tree*, *Logistic Regression*, dan lain - lain. Di setiap penelitian tersebut tiap karakter data yang diteliti berbeda sehingga prediksinya harus menyesuaikan dengan jenis data dan jumlah data perusahaan. Keberagaman tersebut menyebabkan implementasi dari prediksi *churn* ini harus disesuaikan lagi dan dicoba lagi dengan data masing - masing.

Salah satu penelitian dilakukan oleh Firmansyah dan Agus terhadap *customer churn* perusahaan retail dengan menggunakan Naive Bayes. Perusahaan retail yang dimaksud adalah supermarket, *hypermarket*, dan *minimarket* yang memiliki kartu *membership*. Penelitian dilakukan dengan 258 data dengan 6 atribut yaitu jenis kelamin, grade point, kepemilikan kartu kredit, rentang usia, nilai rata-rata transaksi, lama member, dan *class churn*. Data kemudian dibagi 2 untuk diuji dan di latih. Hasil dari pengujian menggunakan *confusion matrix* adalah akurasi mencapai 80% dan presisi mencapai 100%. (Yulianto & Firmansyah, 2021)

Metode Support Vector Machines digunakan untuk memprediksi *customer churn* pada perusahaan di bidang telekomunikasi. Data yang dianalisis ada 5000 data dan atribut yang digunakan ialah 3G, total consumption, calling fees, local sms fees, international sms fees, international calling fees, local sms count, international MOU, total MOU, on net MOU, dan churn. Setelah dilakukan penelitian, hasilnya dibandingkan dengan metode - metode lain seperti c4.5, KNN, dan lain - lain. Hasilnya menunjukkan bahwa akurasi SVM terbaik yaitu 98,7% (Rodan et al., 2014).

Penelitian *customer churn* menggunakan metode Naive Bayes pernah dilakukan terhadap data *customer* toko yang berbasis pada e - commerce. *Customer* dibagi menjadi 10 segmen menggunakan RFM yaitu Champions, loyal customers, potential loyalist, recent customers, promising, customer needing attention, about to sleep, at risk, can't lose them, dan

hibernating. atribut yang diperhitungkan ada 4 yaitu customer id, invoice id, invoice, dan total. 32000 data dibagi menjadi 2 yaitu data testing dan data training dengan berbagai ratio. Hasil prediksi berbagai ratio memiliki nilai terbaik yaitu akurasi 97,27%, precision 100% dan recall sebesar 96,98%. Dari penelitian ini dapat dilihat bahwa metode Naive Bayes berhasil menghasilkan nilai yang cukup tinggi. (Novendri, R., 2021)

Meskipun telah banyak bidang usaha yang diteliti, namun prediksi churn pada pelanggan data center belum banyak dilakukan. Melihat performa dari metode Naive Bayes dan Support Vector Machine yang menghasilkan output yang cukup tinggi, maka diputuskan untuk menerapkan metode tersebut pada penelitian ini. Selain itu dengan data perusahaan yang kecil, dua metode ini dianggap paling cocok untuk memprediksi *customer churn*. Dalam penelitian ini akan dilakukan juga variable selection yang berguna untuk menambahkan akurasi.

Hasil dari kedua model ini kemudian dibandingkan untuk mengetahui metode mana yang lebih cocok pada PT X. Penelitian dilakukan untuk memprediksi *customer* mana yang memiliki potensi untuk beralih di PT X. Informasi yang didapat kemudian akan digunakan oleh *Customer Relationship Management* untuk meningkatkan hubungan dengan *customer* sehingga pendapatan perusahaan tidak berkurang.

1.2 Perumusan Masalah

1. Apa saja atribut yang mempengaruhi *customer churn* pada PT X?
2. Metode mana yang lebih cocok untuk memprediksi *customer churn* PT X, Naive Bayes atau Support Vector Machine?

1.3 Tujuan

Memprediksi *customer churn* perusahaan X dengan metode Naive Bayes dan Support Vector Machine dan membandingkan hasilnya untuk mengetahui metode mana yang lebih sesuai untuk memprediksi *customer churn* pada PT X sehingga dapat digunakan untuk mempertahankan pelanggan.

1.4 Ruang Lingkup

Ruang lingkup dibatasi pada :

- a. Input dari penelitian adalah Data PT X pada tahun 2017 - 2023
- b. Terdapat 2 data yang diberikan yaitu Data Customer dan Hasil Survey Customer
- c. Atribut yang digunakan dalam penelitian :
 - i. Data Customer : Customer Id, Jenis Customer, Lokasi, Tahun Join, Tahun Keluar, Total Tahun, Total Belanja, Jumlah Fasilitas, Jumlah Komplain, dan Label.
 - ii. Hasil Survey Customer : Id Survey, Id Customer, Tahun Survey, Data Center(Kualitas dan Keamanan Data Center, Peralatan Penunjang Keamanan Yang Dipasang, Kualitas dan Kerapian Instalasi, Kesesuaian Harga Dengan Yang Diterima), Technical Support (Cara Komunikasi Dengan Pelanggan, Penyelesaian Masalah, Kerapian, Perilaku, Ketepatan Waktu Pemenuhan Janji), Account Manager (Cara Komunikasi Dengan Pelanggan, Pemberian Solusi, Product Knowledge, Perilaku, Kerapian)
- d. Label Churn diberikan oleh perusahaan
- e. Proses Preprocessing yang akan dilakukan :
 - i. Data Cleaning
 - ii. Attribute selection

Atribut akan diseleksi lagi menggunakan bantuan fungsi correlation bila diperlukan. Variable selection ini yang digunakan untuk menentukan atribut yang mempengaruhi *customer* churn pada PT X.
- f. Data dibagi menjadi 2 yaitu data training dan data testing dengan pembagian yaitu 70% & 30%, 60% & 40%, dan 80% & 20%
- g. Pembuatan model dengan metode Naive Bayes dan Support Vector Machine yang akan digunakan untuk memprediksi *customer* churn perusahaan X
- h. Bahasa pemrograman yang digunakan untuk membuat model adalah Python
- i. Library yang digunakan untuk pembuatan model adalah scikit-learn
- j. Tipe Algoritma Naive Bayes yang digunakan adalah Gaussian Naive Bayes dan Bernoulli Naive Bayes
- k. Hasil dari tes menggunakan model adalah *customer* akan dilabeli menjadi 2 yaitu churn dan not churn dilengkapi dengan persentase prediksi dan penjelasannya

- l. Hasil tes kedua metode tersebut diuji menggunakan *confusion matrix* untuk tes performance
- m. Hasil nilai uji *confusion matrix* SVM dan Naive Bayes akan dibandingkan untuk melihat model mana yang lebih cocok untuk prediksi *customer churn* di PT X

1.5 Metodologi Penelitian

Langkah-langkah dalam pengerjaan Skripsi:

1. Studi Literatur
 - 1.1 Teori *Customer Churn*
 - 1.2 Teori Prediksi *Customer Churn* sebagai Usaha Mempertahankan Pelanggan
 - 1.3 Teori *Naive Bayes*
 - 1.4 Teori *Support Vector Machine*
2. Pengumpulan Data
 - 2.1 Pengumpulan data perusahaan
 - 2.2 Melakukan data *preprocessing*
3. Pembuatan Model Prediksi
 - 3.1 Membuat model dengan metode Naive bayes
 - 3.2 Membuat model dengan metode SVM
4. Pengujian dan Analisis
 - 4.1 Melakukan pengujian data dengan *confusion matrix*
 - 4.2 Melakukan survei terhadap Perusahaan
 - 4.3 Menganalisis hasil pengujian

5. Pengambilan Kesimpulan

5.1 Membuat kesimpulan dari hasil penelitian yang ada

6. Pembuatan Laporan

6.1 Membuat laporan dari hasil penelitian

1.6 Sistematika Penulisan

Penulisan laporan skripsi dibagi menjadi beberapa bab yaitu :

BAB 1 : PENDAHULUAN

Bab 1 membahas mengenai latar belakang masalah yang diangkat, rumusan masalah, tujuan, ruang lingkup penelitian, metodologi penelitian, dan sistematika penulisan

BAB 2 : LANDASAN TEORI

Bab 2 membahas mengenai teori - teori yang menjelaskan atribut penelitian dan metode yang digunakan untuk meneliti. Selain itu, membahas mengenai penelitian - penelitian sebelumnya yang berkaitan dengan penelitian yang dilakukan

BAB 3 : ANALISIS DAN DESAIN SISTEM

Bab 3 membahas mengenai analisis masalah yang ada, lalu bagaimana dapat diselesaikan dengan analisis solusi, lalu dilanjutkan dengan analisis dataset yang akan digunakan untuk prediksi, dan desain sistem, serta desain UI.

BAB 4 : IMPLEMENTASI SISTEM

Bab 4 membahas mengenai perangkat lunak yang digunakan dalam penelitian, implementasi model machine learning, dan website implement

BAB 5 : PENGUJIAN SISTEM

Bab 5 membahas mengenai perangkat keras dan perangkat lunak yang digunakan, feature selection, pengujian parameter SVM, pengujian parameter Naive Bayes, perbandingan evaluasi model, pengujian website sederhana

BAB 6 : KESIMPULAN DAN SARAN

Bab 6 membahas mengenai kesimpulan dan saran