

BAB V PENUTUP

5.1 Kesimpulan

Customer churn analysis adalah metode yang digunakan oleh perusahaan untuk mengidentifikasi dan memprediksi pelanggan yang berpotensi untuk tidak kembali lagi bertransaksi di perusahaan. Hal tersebut merupakan masalah yang sangat penting karena pelanggan yang tidak loyal dapat mengurangi potensi pendapatan perusahaan dan meningkatkan biaya untuk mendapat pelanggan baru. Untuk melakukan *customer churn analysis*, dapat menggunakan komputer dan *software* Google Colab dengan metode langkah-langkah sebagai berikut:

- a. Akuisisi data : mengambil data mentah (*raw*).
- b. *Data preprocessing* : mengubah format data menjadi format yang mudah dianalisis komputer.
- c. *Feature engineering* : mengedit/menambah baru *feature-feature* yang berpotensi meningkatkan kinerja *modelling*.
- d. *Modelling* : memprediksi pelanggan *churn* atau tidak menggunakan algoritma tertentu.
- e. *Evaluation* : menghitung tingkat akurasi prediksi model.

Untuk meningkatkan hasil akurasi, peneliti melakukan *oversampling* dengan metode SMOTE dan *feature selection* dengan metode korelasi *spearman*. Dari metode *feature selection*, terdapat *feature-feature* terbaik yang digunakan untuk analisis model yaitu *shipments_fee*, *click_detail_items*, *total_transaction_label*, *total_amount_paid*, dan *search_count*.

Pada analisis ini, peneliti menggunakan tiga metode algoritma yaitu *Logistic Regression*, *Decision tree* dan *Random Forest Classifier* serta menggunakan *assessment accuracy*. Model yang dipakai memiliki dua metode yaitu model tanpa SMOTE dan *feature selection* dan model dengan SMOTE dan *feature selection*. Berdasarkan hasil penelitian, *decision tree* dan *random forest classifier* dengan SMOTE dan *feature selection* memiliki tingkat akurasi terbaik dengan nilai akurasi sebesar 75%. Namun, bila tanpa SMOTE dan *feature selection*, *random forest*

classifier memiliki akurasi yang lebih baik (71%) daripada *decision tree* (70%). Apabila dikaitkan dengan bisnis di mana terkadang perlu dilakukan analisis model yang cepat (dan perusahaan cenderung memiliki data yang sangat besar sehingga kecepatan dan ketepatan menjadi tujuan utama) maka penulis memilih model *random forest classifier* dengan *SMOTE* dan *feature selection* menjadi model algoritma terbaik untuk penelitian ini.

5.2 Saran

Beberapa saran yang diusulkan oleh peneliti di antaranya:

- a. Perlu adanya perhitungan secara manual terhadap tahapan-tahapan dan model algoritma dengan mengambil sampel pada *dataset* agar semakin meningkatkan akurasi hasil.
- b. Perlu melakukan *hyperparameter tuning* lebih lanjut terhadap model untuk meningkatkan akurasi model.
- c. Bila memungkinkan, dapat menggunakan perangkat keras yang lebih mumpuni atau layanan *notebook* berbayar seperti Colab Pro agar proses penelitian semakin cepat dan akurat.