

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi terkini membuat persaingan bisnis diberbagai sektor menjadi sangat ketat (Rahayu *et al.*, 2021). Dengan persaingan yang sangat ketat, *customer churn* dapat menjadi ancaman bagi perusahaan (De Caigny, Coussement and De Bock, 2018). *Customer churn* adalah hilangnya pelanggan yang berpindah dari satu sektor ke pesaing lain dalam waktu tertentu (Olaniyi *et al.*, 2020). Industri perbankan yang termasuk kedalam sektor *financial* tentu menghadapi tantangan mengenai permasalahan *churn* ini. *Customer* dapat berpindah ke bank yang berbeda karena alasan yang berfluktuasi, misalnya, layanan keuangan yang lebih baik dengan biaya lebih rendah, lokasi cabang bank, kualitas alat digital, suku bunga rendah, dan dll (Kaur and Kaur, 2020). *Customer churn* ini tentu saja meresahkan apabila tidak ditangani karena akan beresiko dalam penurunan *revenue* perusahaan (Herawati, Wibowo and Mukhlash, 2016). Jika *customer churn* terjadi pada perusahaan, maka perusahaan akan menghadapi kerugian yang besar dan menelan biaya yang besar untuk mencari *customer* yang baru (Santharam and Krishnan, 2018). Menurut Forbes, dibutuhkan lebih banyak biaya hingga lima kali lebih banyak untuk mendapatkan pelanggan baru daripada mempertahankan yang sudah ada (Wertz, 2018).

Kerugian akibat *customer churn* tidak boleh terjadi, seperti dari Hadist Rasulullah ﷺ yang berbunyi.

لَا ضَرَرَ وَلَا ضِرَارَ

“Tidak boleh membahayakan diri sendiri maupun membahayakan orang lain.”

Hadist tentang dilarang membahayakan (merugikan) diri sendiri maupun orang lain tersebut diriwayatkan oleh Imam Daruqutni dan Ibnu Abbas r.a dan Ubadah bin Shomith r.a. (HR. Ibnu Majah dan Daraqutni).

Dari dalil tersebut menjelaskan bahwa manusia diperintahkan untuk tidak melakukan sesuatu yang mengakibatkan mudharat (bahaya). Mencegah *customer churn* merupakan suatu merupakan upaya untuk menjauhkan perusahaan dari bahaya kerugian. Harvard Business School percaya bahwa dengan mengurangi tingkat churn pelanggan sebesar 5%,

perusahaan dapat meningkatkan keuntungan sebesar 25% hingga 95% (Plaksij, 2021). Dengan demikian, mencegah *customer churn* adalah sesuatu hal yang penting, terutama untuk bisnis seperti industri perbankan (Olaniyi *et al.*, 2020). Hal tersebut memiliki dampak yang signifikan tidak hanya meningkatkan keuntungan, tetapi juga meningkatkan daya saing (Amuda and Adeyemo, 2019). Data *customer bank* yang berjumlah besar dapat dimanfaatkan untuk memahami pola perilaku dan karakteristik dari *customer* yang akan *churn*. Membangun model prediksi dengan bantuan *machine learning* diperlukan untuk menemukan informasi dan pengetahuan dari data yang berjumlah besar tersebut sehingga menghasilkan model prediksi *customer churn*. Oleh karenanya, penting untuk membuat model prediksi *customer churn* untuk menyiapkan strategi pelanggan guna meningkatkan *customer retention dan revenue* pada bank.

Machine learning dan data mining adalah sub-bidang *artificial intelligence* yang banyak digunakan untuk membuat prediksi, termasuk memprediksi *customer churn* (Nalatissifa and Pardede, 2021). Sejumlah metode pembelajaran mesin dipraktikkan untuk memprediksi *customer churn* meliputi *decision trees*, *K-Nearest Neighbor (K-NN)*, *Naive Bayes*, *neural network*, *regression analysis-logistic regression*, *rule-based classification*, *first order inductive learning* dan *support vector machine* (Sivasankar and Vijaya, 2019). Penelitian terkait *customer churn* pernah dilakukan (Rahman and Kumar, 2020), memprediksi *customer churn* berbasis *machine learning*. Data yang digunakan merupakan data perbankan yang diperoleh dari situs Kaggle. Algoritma pengklasifikasi yang digunakan K-NN, SVM, *Decision Tree*, dan *Random Forest*. Hasil penelitian menggunakan model *Random Forest* setelah *oversampling* lebih baik dibandingkan dengan model lain dengan perolehan akurasi sebesar 85,18%. Pada penelitian sebelumnya evaluasi model berfokus pada hasil *accuracy*. Sedangkan pada kasus *customer churn* ini semakin banyak *customer* yang akan *churn* diprediksi benar akan *churn* jauh lebih penting. Oleh karena itu, pada penelitian skripsi ini penulis berniat untuk mengeksplorasi beberapa model *machine learning* serta teknik *resampling* yang dapat memprediksi *customer* yang akan *churn* dari bank yang berfokus pada *recall* sebagai hasil evaluasi performa model.

1.2 Perumusan Masalah

Diberikan *input X* berupa data *churn* nasabah bank. *Input X* = $[x_1, x_2, x_3, \dots, x_n]$ terdiri dari $n = 13$ *features* diantaranya adalah *row number*, *customer id*, *surname*, *credit score*,

geography, gender, age, tenure, balance, number of product, has credit card, is active member, dan estimated salary.

Diberikan *output* y berupa hasil prediksi kelas *customer churn*. *Output* $y = \in \{1,0\}$ bernilai 1 jika memenuhi kriteria *churn* dan 0 jika tidak.

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana membuat model klasifikasi $f(X) = y$? Model ini menerima input *features* X dan menghasilkan hasil prediksi kelas y ?
2. Bagaimana meningkatkan performa model klasifikasi algoritma *machine learning* dalam memprediksi kelas y ?
3. Bagaimana tanggapan Islam terhadap prediksi bank *customer churn* dengan *machine learning*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengimplementasikan berbagai algoritma *machine learning* pada prediksi bank *customer churn*
2. Mengimplementasikan teknik *resampling*
3. Meningkatkan performa dengan menerapkan *hyper parameter tuning*.
4. Memberikan kajian dari sudut pandang Islam terhadap prediksi bank *customer churn* dengan algoritma *machine learning*.
5. Mempermudah bank dalam memprediksi *customer* mereka yang berpotensi *churn* untuk membantu melakukan strategi pelanggan.

1.4 Research Gap

Pada penelitian (Husein and Harahap, 2021), memprediksi kemungkinan *churn* pelanggan pada sektor perbankan dengan *machine learning*. Data yang digunakan pada penelitian ini berupa kumpulan data *churn banking* yang bersumber dari Kaggle. Sebanyak lima model berbeda yaitu *Kneighbors Classifier, Logistic Regression, Linear SVC, Random Tree Classifier* dan *Random Forest Classifier*. Hasil dari pengujian diperoleh akurasi tertinggi sebesar 86% pada model *Random Forest Classifier*. Pendekatan *Anova* dan *Chi Square* merupakan pendekatan yang berpengaruh signifikan pada peningkatan hasil model prediksi.

Dalam Penelitian (Kumar, Tirupathaiiah and Reddy, 2019), yaitu memprediksi klien dari

perbankan dan industri dana menggunakan teknik *machine learning*. Dataset yang digunakan dalam pengujian adalah pertukaran nasabah bank. Penelitian ini menggunakan model algoritma *Decision Tree* dan *Neural Networks*. Hasil dari pengujian pada penelitian ini menghasilkan 86,52% untuk *Neural Network* dan 79,77% untuk *Decision Tree*.

Pada penelitian (Charandabi, 2020), memprediksi customer churn di industri perbankan. Data yang digunakan pada penelitian ini adalah data dari nasabah bank yang di peroleh dari situs Kaggle. Model yang digunakan pada penelitian ini adalah Naïve Bayes, K-NN, SVM, *Decision Tree*, *Random Forest*, dan ANN. Di antara enam model berbeda yang kembangkan, *Random Forest* dan ANN tampaknya lebih unggul dalam hal kinerja keseluruhan. Akurasi untuk *Random Forest* dan ANN pada data pengujian menghasilkan 86,4% dan 86,7%. Akurasi *Random Forest* menggunakan *balancing* data pada data pengujian *under-sampled* menghasilkan 79,8% dan *over-sampled* 81%. Akurasi ANN menggunakan *balancing* data pada data pengujian *under-sampled* menghasilkan 71,4% dan *over-sampled* 82,8%. Akurasi untuk *Random Forest* dan ANN tanpa adanya *outlier* pada data data pengujian menghasilkan 78,1% dan 81,3%.

1.5 Metode yang Diajukan

Tahap eksperimen dari penelitian ini terdiri dari beberapa tahapan. Tahap pertama dilakukan data *collection* terkait dengan dataset *Customer Churn*. Tahap selanjutnya adalah *preprocessing data* yang terdiri dari *remove unnecessary features*, *one-hot encoding*, dan *feature scaling*. Setelah selesai tahap *preprocessing* dilakukan *model development*.

Pada tahap *model development*, dimulai dengan *splitting* data menjadi data pengujian dan data pelatihan. Selanjutnya, untuk data pelatihan dilakukan 6 teknik *resampling*. Keenam teknik *resampling* tersebut yaitu *random under sampler*, *random under sampler*, SMOTE (*Synthetic Minority Over-sampling Technique*), ADASYN (*Adaptive Synthetic*), SMOTEEN (*Synthetic Minority Over-sampling Technique + Edited Nearest Neighbours*), dan SMOTETomek (*Synthetic Minority Over-sampling Technique + Tomek Links*). Setelah itu, data pengujian dilatih dengan menggunakan 12 model algoritma *machine learning* yang berbeda, yaitu: SVC (*Support Vector Classifier*), *Logistic Regression*, *KNeighbors Classifier*, *Decision Tree Classifier*, *Random Forest Classifier*, LGBM (*Light Gradient Boosted Machine*) *Classifier*, MLP (*Multilayer Perceptron*)

Classifier, Extra Trees Classifier, Gradient Boosting Classifier, XGB Classifier, dan AdaBoost Classifier.

Tahap selanjutnya, dilakukan evaluasi model terhadap model yang sudah dibangun dengan menggunakan *accuracy, precision, recall, f1-score*, dan AUC. Dalam penelitian ini hasil evaluasi model berfokus pada *recall*. Model dengan perolehan *recall* terbaik nantinya akan dilakukan *hyperparameter tuning* dengan menggunakan *gridsearchcv* untuk mendapatkan parameter terbaik agar menghasilkan hasil performa yang lebih baik lagi. Setelah pencarian parameter terbaik, kemudian model dievaluasi menggunakan nilai perolehan *recall*.

1.6 Hasil Evaluasi

Dari hasil eksperimen terdapat ketidakseimbangan kelas pada dataset. Oleh karena itu dilakukan beberapa teknik *resampling*. Pada eksperimen ini *recall* sebagai fokus utama untuk penilaian performa secara keseluruhan memperoleh kenaikan performa yang signifikan setelah menggunakan teknik *resampling*. Perolehan *recall* tertinggi diperoleh oleh algoritma LGBM Classifier disusul oleh XGB Classifier dan Gradient Boosting Classifier. Kenaikan performa *recall* dari ketiga algoritma tersebut diperoleh pada teknik *resampling cluster centroids*.

Pada hasil eksperimen penerapan *gridsearchcv* dengan $k=3$ pada teknik *cluster centroids* dan ketiga algoritma terbaik tersebut memperoleh peningkatan performa dari *recall*. Hasil dari eksperimen memperoleh *recall* sebesar 90,24% untuk XGB Classifier, 90,41% untuk LGBM Classifier, dan 89,21% untuk Gradient Boosting Classifier. Penerapan *gridsearchcv* dengan algoritma LGBM Classifier dengan teknik *resampling cluster centroid* merupakan model yang paling baik dengan perolehan *recall* sebesar 90,41%. Dari hasil eksperimen dengan perolehan *recall* sebesar 90,41% dari model tersebut sudah dapat diimplementasikan dalam memprediksi bank *customer churn*. Hanya saja model tersebut mempertahankan *precision*.

1.7 Kontribusi

Kontribusi dari penelitian skripsi ini adalah sebagai berikut:

1. Memberikan peningkatan performa dari berbagai model *machine learning* dengan menerapkan metode teknik *resampling* dalam memprediksi *customer churn*.

2. Membantu memprediksi *customer* yang berpotensi *churn* untuk mempermudah bank dalam melakukan strategi pelanggan.
3. Kontribusi dengan membagikan kode program dan laporan penelitian skripsi untuk pengembangan model selanjutnya.

1.8 Sistematika Penulisan

Sistematika penulisan penelitian skripsi ini terbagi dalam enam bagian. Bab I pendahuluan yang membahas latar belakang, tujuan, rumusan masalah, *research gap*, metode yang akan diajukan, hasil evaluasi, dan kontribusi. Bab 2 tinjauan pustaka yang membahas penelitian-penelitian sebelumnya terkait dengan topik penelitian. Bab 3 membahas metodologi penelitian yang terdiri dari desain eksperimen, data eksperimen, data *preprocessing*, dan *model development*. Bab 4 hasil dan pembahasan membahas evaluasi dan analisis dari penelitian skripsi. Bab 5 tinjauan Islam membahas pandangan Islam terkait dengan penelitian skripsi. Terakhir bab 6 berupa seluruh kesimpulan dan saran dari penelitian skripsi.