

Teknik *Weighting* untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi *Churn* Menggunakan *XGBoost*, *LightGBM*, dan *CatBoost*

Weighting Technique for Overcoming Class Imbalance in Churn Prediction Using XGBoost, LightGBM, and CatBoost

Wahyu Nugraha¹, Muhamad Syarif²

^{1,2}Sistem Informasi, Universitas Bina Sarana Informatika Kota Pontianak

E-mail: ¹wahyu.whn@bsi.ac.id, ²muhamad.mdx@bsi.ac.id

Abstrak

Churn merupakan kondisi dimana seseorang berpindah dari satu layanan ke layanan yang lain. *Churn* pelanggan menjadi masalah yang meningkat cukup signifikan dan menjadi tantangan utama yang harus dihadapi banyak perusahaan perbankan karena memiliki peran penting terhadap laba perusahaan. Oleh sebab itu, diperlukan cara untuk memprediksi perilaku *churn* tepat waktu agar bisa menerapkan retensi pelanggan. Namun, Permasalahan yang dihadapi oleh model prediksi *churn* adalah ketidakseimbangan kelas sehingga membuat model klasifikasi menghasilkan kinerja yang buruk. Solusi yang paling sering digunakan untuk mengatasi masalah ketidakseimbangan kelas terbagi menjadi tiga pendekatan yaitu pendekatan level data, level algoritma dan *ensemble*. Setiap pendekatan mengalami beberapa masalah yang sulit diprediksi ketika digunakan untuk menangani masalah ketidakseimbangan kelas. Pada penelitian ini, peneliti melakukan eksperimen menggunakan metode *ensemble* berbasis *boosting* untuk melakukan prediksi *churn* pelanggan dan mencoba meningkatkan kinerjanya pada *dataset* yang tidak seimbang dengan *parameter tuning* menggunakan *scale pos weight*. Algoritma klasifikasi yang digunakan yaitu *XGBoost* (*extreme gradient boosting*), *LightGBM* (*light gradient boosting machine*) dan *CatBoost*. Hasil eksperimen akan membandingkan kinerja dari ketiga algoritma berbasis *boosting* tersebut dengan menyesuaikan bobot parameternya sebanyak tiga kali. Dari hasil pengujian, model *CatBoost* memperoleh nilai *recall* tertinggi sebesar 0.79. Sedangkan untuk nilai *recall* terendah adalah model *CatBoost default* dengan nilai 0.47. Berdasarkan hasil eksperimen dapat disimpulkan bahwa model bekerja dengan cukup baik pada data yang tidak seimbang dengan memberikan mekanisme *hyperparameter scale pos weight* sehingga model dapat lebih fokus pada kelas minoritas yang sulit dideteksi.

Kata kunci: Prediksi *Churn*, *Imbalance Class*, Klasifikasi, *Boosting*

Abstract

Churn is a condition where a person moves from one service to another. Customer churn is a problem that has increased significantly and is a major challenge that must be faced by many banking companies because it has an important role in corporate profits. Therefore, it is necessary to predict churn on time in order to implement customer retention. However, the problem faced by the churn prediction model is class imbalance so that the classification model produces poor performance. The solutions that are most often used to solve the class imbalance problem are divided into three approaches, namely the data level approach, algorithm level and ensemble approach. Each approach suffers from some problems that are difficult to predict when used to deal with class imbalance problems. In this study, researchers conducted experiments using the boosting-based ensemble method to predict customer churn and tried to improve its performance on unbalanced datasets with tuning parameters using a pos weight scale. The classification algorithm used is XGBoost (extreme gradient boosting), LightGBM (light gradient boosting machine) and CatBoost. The experimental results will compare the performance of the three boosting-based algorithms by adjusting the parameter weights three times. The test results,

the CatBoost model obtained the highest recall value of 0.79. Meanwhile, the lowest recall value is the default CatBoost model with a value of 0.47. Based on the experimental results, it can be concluded that the model works quite well on unbalanced data by providing a scale post weight hyperparameter mechanism so that the model can focus more on minority classes that are difficult to detect.

Keywords: Churn Prediction, Imbalance Class, Classification, Boosting

1. PENDAHULUAN

Selama satu setengah dekade, bank memiliki jumlah nasabah yang kian hari terus mengalami peningkatan. Hal tersebut mendorong bank untuk tetap menjaga kualitas pelayanan yang mereka berikan [1]. Saat ini, industri perbankan mengalami perubahan yang besar dan kompleks, serta menghadapi banyak sekali tantangan [2]. Peran teknologi terkini membuat informasi lebih mudah diakses khususnya mengenai produk dan layanan dari setiap bank yang ada. Hal ini membuat orang mudah beralih diantara bank pesaing [3]. Bepindahnya loyalitas seseorang terhadap satu layanan ke layanan lain atau kompetitor lain disebut dengan fenomena ‘Churn’ [1][4]. Penyebab fenomena tersebut diantaranya ketersediaan teknologi terbaru, staf bank yang ramah terhadap pelanggan, suku bunga rendah, kedekatan lokasi geografis, serta beragam layanan yang ditawarkan [2]. Churn pelanggan telah menjadi masalah yang meningkat cukup signifikan dan menjadi tantangan utama yang harus dihadapi banyak perusahaan di seluruh dunia [5]. Churn memiliki efek yang luar biasa terhadap laba sebuah industri salah satunya perbankan. Oleh karena itu, penting untuk memprediksi churn dengan tepat agar bisa menerapkan retensi pelanggan tepat waktu [6].

Penelitian mengenai prediksi perilaku pelanggan bank khususnya mengenai perilaku churn telah banyak dilakukan sebelumnya [7]. Terdapat dua sudut pandang berbeda dari para peneliti mengenai penanganan masalah churn, dimana salah satunya berfokus pada peningkatan model prediksi churn yang lebih kompleks agar meningkatkan kinerja prediktif dari sebuah model [8]. Sedangkan pada sudut pandang yang lain peneliti ingin memahami apa penyebab yang mendorong churn pelanggan [8]. Churn pelanggan merupakan peristiwa yang jarang terjadi dibanyak industri tidak hanya dunia perbankan, ini menunjukkan bahwa jumlah churn secara signifikan lebih sedikit jumlah dibandingkan non-churner [3]. Walaupun jumlah churn pelanggan terlihat sedikit, kehilangan pelanggan menyebabkan berkurangnya laba yang cukup signifikan sehingga model prediksi untuk mendeteksi churn pelanggan menjadi sangat penting [9]. Model klasifikasi *machine learning* telah digunakan dalam membuat model prediksi churn diberbagai industri seperti telekomunikasi, perbankan dan *newspaper* [10]. Permasalahan yang dihadapi oleh model prediksi churn adalah ketidakseimbangan kelas dimana satu dari kelas terlalu dominan dibanding kelas yang lain [4]. *Imbalance class* merupakan kondisi dimana hampir semua *instance* (*majority class*) dimiliki oleh satu kelas, sedangkan sisanya dimiliki kelas lain (*minority class*) yang merupakan *impotent class* [1]. Model klasifikasi untuk mencari kinerja yang akurat terhadap data yang tidak seimbang cenderung mengklasifikasikan semua data ke dalam kelas mayoritas, sehingga membuat model klasifikasi menghasilkan kinerja yang buruk [11].

Solusi yang paling sering digunakan untuk mengatasi masalah ketidakseimbangan kelas terbagi menjadi tiga pendekatan [3]. Ketiga pendekatan tersebut telah digunakan oleh para peneliti pada dekade terakhir ini [12]. Pertama menggunakan pendekatan *data-level* yaitu melakukan *preprocessing* data pada *dataset* yang tidak seimbang [13]. Cara kedua menggunakan pendekatan teknik *algorithm-level* dengan memodifikasi algoritma *machine learning* untuk meningkatkan akurasi terhadap kelas minoritas. Pendekatan terakhir menggunakan teknik *ensemble*, yaitu menggunakan kedua teknik pendekatan level data dan algoritma di dalam kerangka (*cost-sensitive*) pembelajaran *ensemble*. Setiap pendekatan yang dikembangkan memiliki kelebihan dan kekurangan sehingga masing-masing pendekatan menunjukkan kemampuan serta keunggulannya terhadap masalah ketidakseimbangan kelas pada *domain* yang berbeda [3][14]. Salah satu metode atau teknik yang cukup populer dalam penanganan ketidakseimbangan kelas adalah metode

sampling yaitu *under-sampling* dan *over-sampling*. Selain itu, metode *ensemble* berbasis *bagging* dan *boosting* merupakan metode lain yang banyak digunakan untuk menangani masalah ketidakseimbangan kelas [15].

Metode-metode yang pernah digunakan mungkin menghadapi beberapa masalah yang sulit diprediksi ketika digunakan untuk menangani masalah ketidakseimbangan kelas [15]. Contohnya metode *under-sampling* yang memiliki potensi untuk menghilangkan beberapa data penting dalam proses pembelajaran model, selain itu, metode *over-sampling* yang kemungkinan dapat menyebabkan *overfitting*. Selanjutnya metode *ensemble* berbasis *bagging* dan *boosting* juga kemungkinan dapat menghilangkan beberapa data berguna karena menggunakan metode pengambilan sampel juga [16]. Walaupun pada penelitian sebelumnya menunjukkan bahwa *ensemble* belum memberikan performa yang baik. Namun, metode *ensemble* menunjukkan kinerja yang baik jika dikaitkan dengan ukuran statistik seperti AUC (*area under curve*) [3].

Pada penelitian ini, peneliti mencoba melakukan eksperimen menggunakan metode *ensemble* berbasis *boosting* untuk melakukan prediksi *churn* pelanggan dan mencoba meningkatkan kinerjanya pada dataset yang tidak seimbang dengan *parameter tuning* menggunakan *scale pos weight*. Metode klasifikasi menggunakan tiga algoritma populer yaitu *XGBoost* (*extreme gradient boosting*), *LightGBM* (*light gradient boosting machine*) dan *CatBoost*. Eksperimen ini akan memperlihatkan *XGBoost*, *LightGBM* dan *CatBoost* dalam menangani data yang tidak seimbang secara internal. Algoritma *boosting* secara umum menunjukkan hasil yang baik pada data yang tidak seimbang dengan memberikan kerja yang lebih fokus terhadap kelas minoritas. Pada algoritma *boosting* seperti *XGBoost*, *LightGBM* dan *CatBoost* tidak perlu dipasangkan dengan teknik manipulasi data karena dapat menyesuaikan skala bobotnya atau *scale-pos-weight* untuk mengatasi masalah data yang tidak seimbang [17]. Uji eksperimen ini akan membandingkan kinerja dari ketiga algoritma klasifikasi dengan menyesuaikan bobot parameternya sebanyak tiga kali. Pengujian sebanyak tiga kali menyebabkan kelas minoritas mendapatkan dampak tiga kali dan koreksi tiga kali kesalahan yang dilakukan terhadap kelas mayoritas. Hasil dari ketiga pengujian tersebut akan diukur menggunakan metrik evaluasi klasifikasi yaitu *confusion matrix* dimana penelitian ini tidak berfokus pada nilai *accuracy score*. Namun, berdasarkan permasalahan yang ada maka kami lebih berfokus pada nilai dari *recall score*. *Recall score* digunakan pada penelitian ini karena secara intuitif memiliki kemampuan pengklasifikasi untuk menemukan semua sampel positif atau minoritas sampel [18]. *Recall score* juga digunakan sebagai metrik evaluasi karena identifikasi *churn* merupakan masalah data yang tidak seimbang, sehingga kinerja model akan fokus pada kelas minoritas [17]. *Dataset* yang digunakan pada penelitian ini adalah *dataset churn modelling* yang sesuai dengan masalah penelitian yaitu ketidakseimbangan kelas. *Dataset* dapat diunduh pada <https://www.kaggle.com/datasets/shrutimechlearn/churn-modelling>. *Dataset* ini berisi rincian data pelanggan bank. Variabel target adalah variabel biner yang mencerminkan fakta apakah pelanggan menutup rekeningnya atau terus menjadi pelanggan. Hasil eksperimen menunjukkan bahwa algoritma *boosting* *CatBoost* mengalami peningkatan yang paling signifikan pada skor *recall* dari semua algoritma klasifikasi jika menggunakan parameter *scale pos weight*. *Scale pos weight* adalah parameter yang digunakan untuk memvariasikan rasio kelas positif dan negatif dengan tujuan untuk untuk mengontrol ketidakseimbangan kelas dengan lebih baik [19].

2. TINJAUAN STUDI

Prediksi *churn* pelanggan adalah proses mengidentifikasi pelanggan dengan kecenderungan tinggi untuk *churn*. prediksi *churn* dapat membantu perusahaan menemukan pelanggan yang berpotensi untuk *churn* sehingga bisa menawarkan insentif yang tepat kepada mereka demi upaya pencegahan terhadap *churn* pelanggan. Data yang berkualitas serta model pembelajaran prediktif yang baik adalah dua penentu utama model prediksi *churn* menjadi sukses. Penelitian mengenai *churn* prediksi telah dilakukan oleh beberapa peneliti. Diantaranya penelitian yang dilakukan oleh Gholamiangonabadi. Prediksi *churn* pelanggan di industri perbankan dilakukan dengan pendekatan dengan metode *neural network methods of multilayer perceptron*,

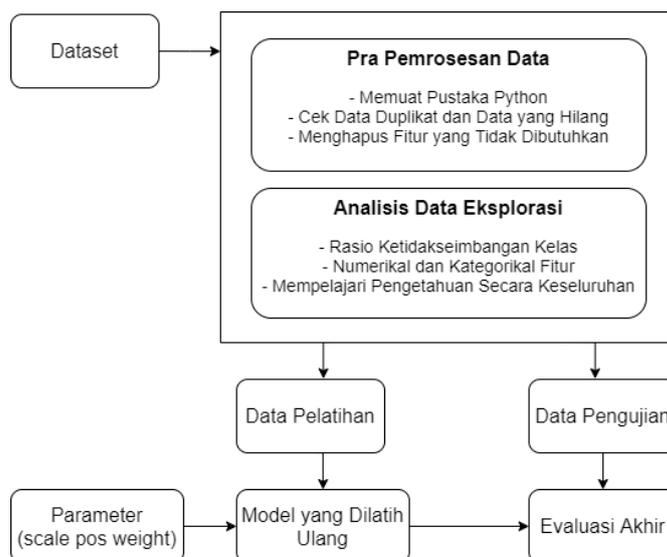
radial basis function, *support vector machine* dan *generalized regression*. Untuk meningkatkan akurasi dari prediksi digunakan *naïve bayes* sebagai *meta-classifier* [7]. Berdasarkan ciri umum *meta-classifier* disebut juga dengan *meta-nets* karena memiliki arsitektur *neural network* yang bekerja dengan mengambil semua nilai prediksi model klasifikasi [20]. Sehingga dapat disimpulkan bahwa *meta-classifier* merupakan parameter yang membantu memilih *classifier* agar sesuai dengan *output* dari masing-masing model. Hasilnya menunjukkan bahwa pendekatan ini telah menghasilkan peningkatan yang cukup signifikan dalam prediksi *churn* pelanggan.

Teknik *ensemble* beberapa dekade ini begitu menarik perhatian khususnya saat digunakan dalam prediksi *churn* pelanggan. *Ensemble* merupakan algoritma mempelajari data dengan menggunakan kombinasi dari beberapa algoritma atau model untuk mendapatkan output dengan akurasi yang lebih baik [21]. Cara melakukan *ensemble learning* salah satunya dengan proses *bagging* dan *boosting*. Hal tersebut terlihat dari hasil penelitian beberapa peneliti yang menunjukkan hasil yang lebih efektif daripada model tunggal. Misalnya penelitian yang dilakukan oleh Coussement [22] yang menerapkan CART *decision trees* dan GAM (*generalized additive model*) sebagai pembelajaran dasar untuk membangun hutan acak atau *random forests*. GAM *ensemble* dalam *churn* prediksi menunjukkan dampak yang menguntungkan dibandingkan dengan model tunggal. Selain itu, Penelitian oleh Lu [23] menggunakan *AdaBoost* untuk mengidentifikasi pelanggan telekomunikasi yang beresiko *churn*. Penelitiannya memperoleh kinerja klasifikasi yang lebih baik dari pada model logistik normal atau *non-ensemble*.

Sebagian besar data yang tersedia untuk prediksi *churn* seringkali tidak seimbang [4], yang berarti bahwa porsi jumlah *instance* pada kelas berbeda cukup signifikan. Distribusi kelas yang tidak seimbang menimbulkan tantangan pada algoritma *machine learning*. Oleh karena itu, pada penelitian ini peneliti mencoba untuk menggunakan model *ensemble* berbasis *boosting* dalam melakukan prediksi *churn* pelanggan. Selanjutnya peneliti mencoba untuk meningkatkan model *machine learning* menggunakan *parameter tuning* menggunakan *scale pos weight* untuk mengatasi ketidakseimbangan pada *dataset*.

3. METODE PENELITIAN

Pada penelitian ini optimasi model *machine learning* menggunakan parameter *scale pos weight* pada bahasa *python*. Gambar 1 menunjukkan tahapan eksperimen yang dilakukan yaitu dimulai dari *import dataset*, *data preprocessing*, *Exploratory Data Analysis (EDA)*, *model fitting* dengan parameter.



Gambar 1 Model Eksperimen *Churn Prediction*

Tahapan pertama melakukan pemilihan *dataset* yang tepat sesuai tujuan penelitian, sehingga *dataset churn* dengan kasus *imbalance class* dipilih. Selanjutnya pada tahap *preprocessing* dilakukan dengan memilih *library* yang dibutuhkan untuk proses *cleaning* data seperti mencari data yang duplikat dan *missing* serta menghapus beberapa variabel yang dianggap tidak diperlukan dalam membangun model prediksi. Setelah data siap maka dilakukan tahap analisis eksplorasi data yang bertujuan untuk melihat *imbalance ratio*, melihat variabel yang bersifat numerikal dan kategorikal dan diakhiri dengan mengambil seluruh informasi setiap *feature* atau variabel dari data.

Tahapan selanjutnya setelah data telah siap digunakan adalah membangun model. Data dibagi menjadi 70% data *training* dan 30% sebagai data *testing*. Data *training* harus lebih besar dari data *testing* agar pembelajaran mesin menjadi lebih baik. Kinerja model bisa berubah secara signifikan jika rasio *training* dan *testing* berbeda [24]. Rasio 70/30 untuk *training* dan *testing* dianggap sebagai rasio terbaik karena sebagian besar penelitian yang menggunakan *dataset churn* menggunakan rasio tersebut. Pada proses pembelajaran ketiga algoritma klasifikasi *XGBoost*, *LightGBM* dan *CatBoost* akan diuji bergantian sebanyak tiga kali berdasarkan *parameter tuning* menggunakan *scale pos weight*. Hasil dari setiap pengujian menggunakan metrik evaluasi *confusion matrix*. Matrik konfusi ini umumnya digunakan pada binari klasifikasi [25].

		Predicted Class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Gambar 2 *Confusion matrix for binary classification*

Confusion matrix berperan penting untuk mengetahui seberapa baik model yang telah dibangun. Dari tabel konfusi akan diperoleh *True Positif* (TP): Prediksi *class* positif yang diprediksi benar atau *real transaction* (data *churn* dan model memprediksi benar sebagai *churn*). *False Negatif* (FN): Prediksi *class* negatif, tetapi nilai sebenarnya positif atau prediksi salah (*type II error*). *False Positif* (FP): Prediksi *class* positif, tetapi nilai sebenarnya negatif atau prediksi salah (*type I error*). *True Negatif* (TN): Prediksi *class* negatif yang diprediksi benar atau nilai sebenarnya negatif (*non churn*). Semua nilai untuk mengukur kinerja berdasarkan matrik konfusi dapat menggunakan *accuracy*, *precision* dan *recall* pada persamaan berikut:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (1)$$

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

4. HASIL DAN PEMBAHASAN

Eksperimen untuk mencapai hasil dari penelitian ini terdiri dari beberapa tahap diantaranya *import data*, *data preprocessing*, *exploratory data analysis* (EDA), pengujian terhadap model berdasarkan data latih (menggunakan beberapa parameter) dan evaluasi hasil untuk menemukan hasil terbaik dari seluruh model klasifikasi berdasarkan tiap-tiap parameter pengujian.

4.1 Dataset

Dataset diperoleh dari *kaggle dataset repository* dengan format ekstensi CSV (*a comma separated values*). *Dataset* terdiri dari 14 variabel yang berisi rincian pelanggan bank. Variabel

target (*Exited*) adalah variabel biner yang menggambarkan apakah seorang pelanggan menutup rekeningnya atau terus menjadi pelanggan. Variabel prediktor terdiri dari *CreditScore*, *Geography*, *Gender*, *Age*, *Tenure*, *Balance*, *NumOfProducts*, *HasCrCard*, *IsActiveMember*, *EstimatedSalary*, dan target variabel *Exited*. Tabel 1 merupakan *overview* data primer *churn modelling* yang akan digunakan pada penelitian ini.

Tabel 1 *Sample Dataset Churn Modelling*

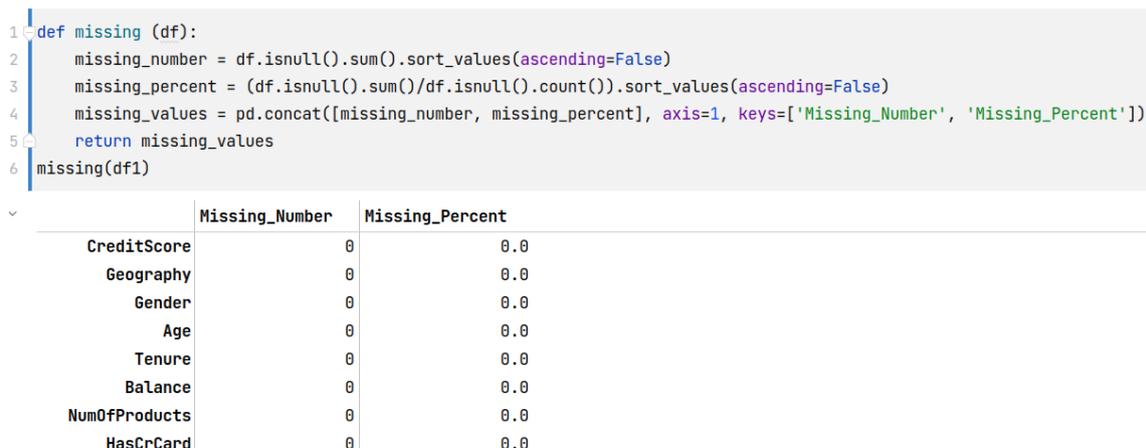
CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
619	France	Female	42	2	0	1	1	1	101348.88	1
608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
502	France	Female	42	8	159660.8	3	1	0	113931.57	1
699	France	Female	39	1	0	2	0	0	93826.63	0
850	Spain	Female	43	2	125510.82	1	1	1	79084.1	0

Penelitian ini memprediksi apakah pelanggan atau konsumen meninggalkan bank sehingga bank akan memberikan lebih banyak layanan dan memastikan untuk tidak meninggalkan bank. Fitur-fitur pada kumpulan data yang digunakan diantaranya:

1. *CreditScore*: angka antara 300-850 yang menggambarkan kelayakan pelanggan.
2. *Geography*: negara tempat pelanggan berasal.
3. *Tenure*: jumlah tahun nasabah di bank.
4. *Balance*: saldo pelanggan
5. *Numofproduct*: jumlah produk yang digunakan pelanggan.
6. *HasCrCard*: jumlah kartu kredit yang digunakan.
7. *Isactivemember*: binary flag untuk menunjukkan apakah klien aktif atau tidak dengan bank sebelumnya.
8. *Exited*: variabel target

4.2 Data Preprocessing

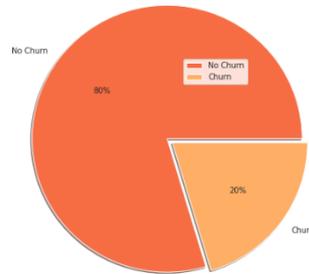
Tahap pertama eksperimen adalah import beberapa library yang dibutuhkan dalam pra-pemrosesan data. Eksperimen menggunakan bahasa *python* versi 3 dimana *library* yang digunakan diantaranya *numpy*, *pandas*, *matplotlib*, *seaborn*, *optuna*, *plotly* dan lainnya. Selanjutnya dengan *library pandas dataset* dengan format CSV dimuat. Selanjutnya dilakukan observasi data untuk menemukan duplikasi data serta mengecek apakah ada data yang hilang atau *missing value*. Gambar 3 menunjukkan bahwa pada dataset tidak terdapat data yang hilang atau *missing value*. Beberapa variabel yang dianggap tidak berguna dalam model prediksi akan dihilangkan seperti variabel *rownumber*, *customerid* dan *surname*.



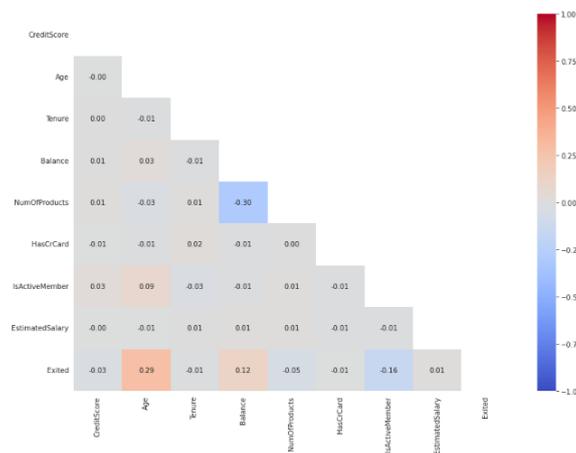
Gambar 3 Pengecekan *Missing Value*

4.3 Exploratory Data Analysis

Pada tahap ini akan dicari beberapa informasi penting seperti rasio dari *imbalance class*. Serta melihat hubungan atau korelasi antar atribut atau *feature* yang berkaitan erat dengan kondisi *churn*. Gambar 4 menunjukkan presentase *imbalance data* 20.37% mengalami *churn* dan 79.63% tidak *churn*. Sedangkan gambar 5 menggambarkan hubungan korelasi antar variabel dimana *Age* menjadi variabel yang memiliki korelasi tertinggi terhadap target dengan nilai 0,29 diikuti oleh *balance* sebesar 0,12.



Gambar 4 Imbalance Ratio



Gambar 5 Correlation heatmap Dataset

4.4 Model Fitting

Pengujian model klasifikasi pada eksperimen ini menggunakan tiga algoritma *boosting* yaitu *catboost*, *xgboost* dan *lightgbm* dengan komposisi 70% data *training* dan 30% sebagai data *testing*. Ketiga algoritma ini dipilih pada penelitian ini karena merupakan algoritma *boosting* yang cukup populer pada penelitian yang berkaitan dengan *machine learning*. Beberapa peneliti menggunakan algoritma ini di dalam penelitiannya [26][27]. Sehingga ketiga algoritma klasifikasi ini dirasa cukup menarik untuk dibahas. Pengujian dilakukan sebanyak tiga kali pada masing-masing algoritma berdasarkan *parameter tuning scale pos weight* dengan nilai parameter *default*, 3 dan 5.

4.4.1 CatBoost

Hasil klasifikasi dengan algoritma *catboost* yang dihasilkan dari *confusion matrix* dapat dilihat pada tabel 2, 3 dan 4.

Tabel 2 Confusion Matrix Catboost_default

True Label	0	2326	90
	1	307	277
		0	1
		Predicted Label	

Tabel 3 Confusion Matrix Catboost_adjusted_weight_3

True Label	0	2094	322
	1	168	416
		0	1
		Predicted Label	

Tabel 4 Confusion Matrix Catboost_adjusted_weight_5

True Label	0	1877	539
	1	123	461
		0	1
		Predicted Label	

4.4.2 XGBoost

Hasil klasifikasi dengan algoritma xgboost yang dihasilkan dari *confusion matrix* dapat dilihat pada tabel 5, 6 dan 7.

Tabel 5 Confusion Matrix XGBoost_Default

True Label	0	2299	117
	1	302	282
		0	1
		Predicted Label	

Tabel 6 Confusion Matrix XGBoost_adjusted_weight_3

True Label	0	2170	246
	1	234	350
		0	1
		Predicted Label	

Tabel 7 Confusion Matrix XGBoost_adjusted_weight_5

True Label	0	2081	335
	1	194	390
		0	1
		Predicted Label	

4.4.3 LightGBM

Hasil klasifikasi dengan algoritma *lightgbm* yang dihasilkan dari *confusion matrix* dapat dilihat pada tabel 8, 9 dan 10.

Tabel 8 Confusion Matrik LightGBM_default

True Label	0	2326	90
	1	292	292
		0	1
		Predicted Label	

Tabel 9 Confusion Matrix LightGBM_adjusted_weight_3

True Label	0	2104	312
	1	186	398
		0	1
		Predicted Label	

Tabel 10 Confusion Matrix LightGBM_adjusted_weight_5

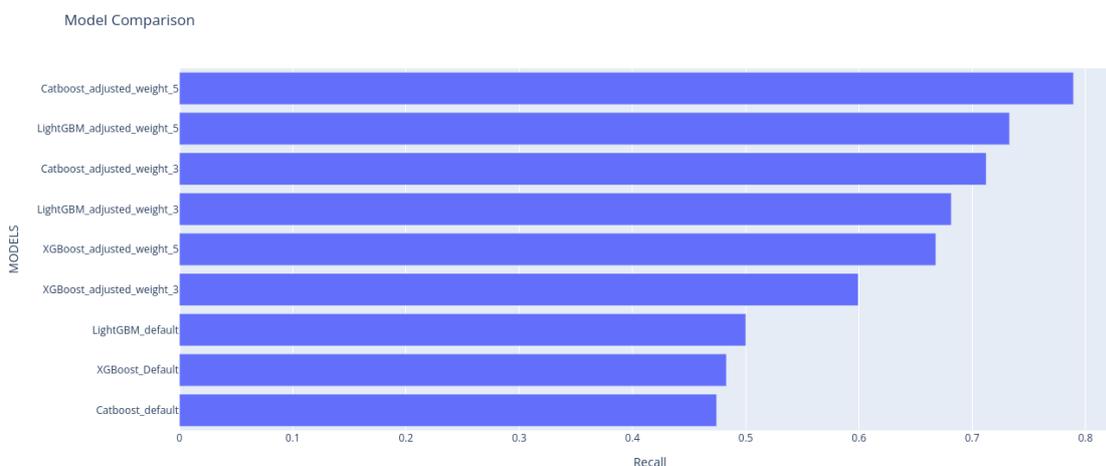
True Label	0	1965	451
	1	156	428
		0	1
		Predicted Label	

4.4.4 Hasil Keseluruhan dan Komparasi Model

Nilai yang dihasilkan dari pengujian terhadap keseluruhan model ditunjukkan pada tabel 11. Pada hasil tersebut terlihat bagaimana *CatBoost*, *XGBoost*, dan *LightGBM* menggunakan *hyperparameter scale po weight* menangani data yang tidak seimbang secara internal.

Tabel 11 Performa keseluruhan model terhadap data yang tidak seimbang

	Accuracy	Recall	Roc_Auc	Precision
<i>Catboost_default</i>	0.8677	0.4743	0.7185	0.7548
<i>Catboost_adjusted_weight_3</i>	0.8367	0.7123	0.7895	0.5637
<i>Catboost_adjusted_weight_5</i>	0.7793	0.7894	0.7831	0.461
<i>XGBoost_Default</i>	0.8603	0.4829	0.7172	0.7068
<i>XGBoost_adjusted_weight_3</i>	0.84	0.5993	0.7487	0.5872
<i>XGBoost_adjusted_weight_5</i>	0.8237	0.6678	0.7646	0.5379
<i>LightGBM_default</i>	0.8727	0.5	0.7314	0.7644
<i>LightGBM_adjusted_weight_3</i>	0.834	0.6815	0.7762	0.5606
<i>LightGBM_adjusted_weight_5</i>	0.7977	0.7329	0.7731	0.4869



Gambar 6 Komparasi Model Berdasarkan Nilai Recall

Pada gambar 6 menggambarkan bahwa dari model yang telah diujikan model *CatBoost* dengan *hyperparameter scale pos weight 5* memperoleh nilai *recall* tertinggi berdasarkan tabel 11 yaitu sebesar 0.79. Sedangkan untuk nilai *recall* terendah adalah *CatBoost default* dengan nilai 0,47. *Recall* adalah rasio $tp / (tp + fn)$ dengan perbandingan antara *True Positive* (TP) dengan banyaknya data yang sebenarnya positif. pada penelitian ini menggunakan *recall score* karena melihat kondisi yang berakibat lebih fatal jika model memprediksi customer yang tidak churn padahal churn. Jadi sebaiknya, model mempunyai nilai *recall* yang tinggi. Jika dilihat secara rata-rata maka penerapan *scale pos weight* cukup memberikan dampak yang signifikan terhadap performa model dengan data yang tidak seimbang. Peningkatan tersebut terlihat dari ketiga algoritma *boosting* yang digunakan.

5. KESIMPULAN DAN SARAN

Secara umum, algoritma *boosting* bekerja dengan baik bahkan pada data yang tidak seimbang. Model dibangun dari algoritma tersebut memberikan hasil yang baik pada data yang tidak seimbang dengan memberikan mekanisme untuk lebih fokus pada kelas minoritas dan menyesuaikan algoritma pelatihan. Karena dengan menyesuaikan bobot/*weight*, kelas minoritas mendapat dampak tiga kali lebih banyak dan koreksi tiga kali lebih banyak daripada kesalahan yang dilakukan pada kelas mayoritas. Penelitian ini juga mengamati bahwa ketika mengubah *scale pos weight* dari 3 menjadi 5, maka terjadi peningkatan yang signifikan pada skor *recall* khususnya pada algoritma *catboost* yang mengalami peningkatan paling signifikan. Penelitian ini dapat disimpulkan bahwa dengan mengubah *hyperparameter scale pos weight* kami mengamati perubahan pada nilai *recall*, akurasi, *roc-auc*, dan *precision*.

Peneliti menyarankan untuk menekankan bahwa pentingnya untuk mendapatkan pengetahuan yang lebih mendalam mengenai masalah yang dihadapi. Pada kasus ini *imbalance class* menjadi masalah yang sering dihadapi pada model prediksi *churn*. Namun, solusi yang diberikan harus berdasarkan permintaan pemangku kepentingan sehingga peneliti mampu membuat perubahan sesuai kebutuhan. Kelemahan pada metode yang digunakan peneliti terjadi ketika penggunaan nilai ekstrim pada *scale pos weight* dimana akan menyebabkan *overfit* pada minoritas kelas sehingga menghasilkan kinerja model yang buruk. Untuk itu dibutuhkan penelitian lebih lanjut untuk mengetahui penyebab agar bisa dibangun model prediksi yang jauh lebih baik.

DAFTAR PUSTAKA

- [1] M. A. H. Farquad, V. Ravi, and S. B. Raju, "Churn prediction using comprehensible support vector machine: An analytical CRM application," *Applied Soft Computing Journal*, vol. 19, pp. 31–40, 2014, doi: 10.1016/j.asoc.2014.01.031.
- [2] B. He, Y. Shi, Q. Wan, and X. Zhao, "Prediction of customer attrition of commercial banks based on SVM model," *Procedia Comput Sci*, vol. 31, pp. 423–430, 2014, doi: 10.1016/j.procs.2014.05.286.
- [3] B. Zhu, B. Baesens, and S. K. L. M. vanden Broucke, "An empirical comparison of techniques for the class imbalance problem in churn prediction," *Inf Sci (N Y)*, vol. 408, pp. 84–99, 2017, doi: 10.1016/j.ins.2017.04.015.
- [4] K. S. Karuppaiah and N. P. G. Palanisamy, "Enhanced Churn Prediction Using Stacked Heuristic Incorporated Ensemble Model," *Journal of Information Technology Research*, vol. 14, no. 2, pp. 174–186, 2021, doi: 10.1016/j.matpr.2020.12.893.
- [5] Y. Xie, X. Li, E. W. T. Ngai, and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Syst Appl*, vol. 36, no. 3 PART 1, pp. 5445–5449, 2009, doi: 10.1016/j.eswa.2008.06.121.
- [6] J. Xiao, G. Teng, C. He, and Z. Bing, "One-Step Classifier Ensemble Model for Customer Churn Prediction With Imbalance Class," *Advances in Intelligent Systems and Computing*, vol. 281, no. January, pp. 843–854, 2014, doi: 10.1007/978-3-642-55122-2.
- [7] D. Gholamiangonabadi, S. Nakhodchi, A. Jalalimanesh, and A. Shahi, "Customer churn prediction using a meta-classifier approach; A case study of Iranian banking industry," *Proceedings of the International Conference on Industrial Engineering and Operations Management*, vol. 2019, no. MAR, pp. 364–375, 2019.
- [8] A. de Caigny, K. Coussement, and K. W. de Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *Eur J Oper Res*, vol. 269, no. 2, pp. 760–772, 2018, doi: 10.1016/j.ejor.2018.02.009.
- [9] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Syst Appl*, vol. 38, no. 3, pp. 2354–2364, 2011, doi: 10.1016/j.eswa.2010.08.023.

- [10] T. Gattermann-Itschert and U. W. Thonemann, "How training on multiple time slices improves performance in churn prediction," *Eur J Oper Res*, vol. 295, no. 2, pp. 664–674, 2021, doi: 10.1016/j.ejor.2021.05.035.
- [11] N. Japkowicz and S. Stephen, "The Class Imbalance Problem: A Systematic Study," *Intelligent Data Analysis*, vol. 6, no. 5, pp. 429–449, 2002.
- [12] N. Moniz and H. Monteiro, "No Free Lunch in imbalanced learning," *Knowl Based Syst*, vol. 227, p. 107222, 2021, doi: 10.1016/j.knosys.2021.107222.
- [13] H. Ali, M. N. M. Salleh, R. Saedudin, K. Hussain, and M. F. Mushtaq, "Imbalance class problems in data mining: A review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1552–1563, 2019, doi: 10.11591/ijeecs.v14.i3.pp1552-1563.
- [14] A. Fernández, S. García, F. Herrera, and N. v. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, 2018, doi: 10.1613/jair.1.11192.
- [15] Z. Sun, Q. Song, X. Zhu, H. Sun, B. Xu, and Y. Zhou, "A novel ensemble method for classifying imbalanced data," *Pattern Recognit*, vol. 48, no. 5, pp. 1623–1637, 2015, doi: 10.1016/j.patcog.2014.11.014.
- [16] B. Zhu *et al.*, "Improving Resampling-based Ensemble in Churn Prediction Seppe vanden Broucke," *Proc Mach Learn Res*, vol. 74, no. September, pp. 79–91, 2017.
- [17] R. Zhong, R. L. Johnson, and Z. Chen, "Using Machine Learning Methods to Identify Coals from Drilling and Logging-While-Drilling LWD Data," *Asia Pacific Unconventional Resources Technology Conference, Brisbane, Australia, 18–19 November 2019*, pp. 970–994, 2020, doi: <https://doi.org/10.15530/AP-URTEC-2019-198288>.
- [18] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," *European Conference on Information Retrieval*, vol. 3408, pp. 345–359, 2005, Accessed: Jan. 22, 2023. [Online]. Available: https://doi.org/10.1007/978-3-540-31865-1_25
- [19] M. Syed, J. Marshall, A. Nigam, and N. v. Chawla, "Gender Prediction Through Synthetic Resampling of User Profiles Using SeqGANs," *International Conference on Computational Data and Social Networks*, vol. 11917, pp. 363–370, 2019.
- [20] M. Buscema, S. Terzi, and W. Tastle, "A new meta-classifier," in *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 2010. doi: 10.1109/NAFIPS.2010.5548298.
- [21] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 42, no. 4, pp. 463–484, 2012, doi: 10.1109/TSMCC.2011.2161285.
- [22] K. Coussemont and K. W. de Bock, "Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning," *J Bus Res*, vol. 66, no. 9, pp. 1629–1636, 2013, doi: 10.1016/j.jbusres.2012.12.008.
- [23] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting," *IEEE Trans Industr Inform*, vol. 10, no. 2, pp. 1659–1665, 2014, doi: 10.1109/TII.2012.2224355.
- [24] Q. H. Nguyen *et al.*, "Influence of data splitting on performance of machine learning models in prediction of shear strength of soil," *Math Probl Eng*, vol. 2021, 2021, doi: 10.1155/2021/4832864.
- [25] A. Luque, A. Carrasco, A. Martín, and A. de las Heras, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit*, vol. 91, pp. 216–231, 2019, doi: 10.1016/j.patcog.2019.02.023.
- [26] S. Demir and E. K. Sahin, "Predicting occurrence of liquefaction-induced lateral spreading using gradient boosting algorithms integrated with particle swarm optimization: PSO-

- XGBoost, PSO-LightGBM, and PSO-CatBoost,” *Acta Geotech*, Jan. 2023, doi: 10.1007/s11440-022-01777-1.
- [27] E. al Daoud, “Comparison between XGBoost, LightGBM and CatBoost Using a Home Credit Dataset,” *International Journal of Computer and Information Engineering*, vol. 13, no. 1, pp. 6–10, 2019, doi: doi.org/10.5281/zenodo.3607805.