

Prediksi Kelalaian Pinjaman Bank Menggunakan *Random Forest* dan *Adaptive Boosting*

<http://dx.doi.org/10.28932/jutisi.v6i1.2313>

Joseph Sanjaya^{#1}, Erick Renata^{#2}, Vincent Elbert Budiman^{#3}, Francis Anderson^{#4}, Mewati Ayub^{#5}

[#]Jurusan Magister Ilmu Komputer, Universitas Kristen Maranatha
Jl. Surya Sumantri No.65, Sukawarna

¹mi1879011@student.it.maranatha.edu

²mi1879010@student.it.maranatha.edu

³mi1879007@student.it.maranatha.edu

⁴mi1879015@student.it.maranatha.edu

⁵mewati.ayub@it.maranatha.edu

Abstract — A loan is one of the most important products on the bank, which used for main revenue. All bank tries to find the most effective business strategy to persuade a customer to use the loan, but loan default has a negative effect after the application is approved. Loan default causes loss on the bank, therefore it is mandatory to calculate in order to decrease the risk of the loan default. This study uses the random forest and adaptive boosting machine learning methods to get the prediction and decision. The random forest uses a voting method from many decision trees and adaptive boosting can support to increase accuracy, stability and handle an underfit or overfit problem. The experimental results show that Adaptive Boosted Random Forest outperformed normal random forest and Deep learning Neural Network (DNN) in recall rate evaluation metrics with small trade-offs in the accuracy.

Keywords— Adaptive Boosting; Bank; Loan Default; Machine learning; Random Forest;

I. PENDAHULUAN

Pinjaman adalah produk yang penting pada bank, dimana pinjaman merupakan prioritas utama pemasukan utama pada bank, tetapi seringkali pinjaman berdampak negatif di saat peminjam tidak bisa melunaskan pinjamannya atau disebut kredit macet. Peristiwa kredit macet adalah peristiwa negatif pada bank yaitu kerugian. Terlalu banyak kredit macet pun dapat menjadi suatu penilaian bank, dimana, menurut peraturan PBI Nomor 6/10/PBI/2004 tanggal 12 April 2004 mengenai *Non-Perform Loan (NPL)*, suatu bank diharuskan memiliki *NPL* dibawah 5%. Ada banyak faktor yang dapat memengaruhi macetnya kredit, maka dalam penelitian ini akan dilakukan pembelajaran menggunakan pembelajaran mesin. Dengan menggunakan pembelajaran mesin, diharapkan fitur dan faktor yang memengaruhi suatu

pinjaman menjadi kredit macet dapat terlihat. Metode yang akan digunakan adalah metode *random forest*, yaitu metode *voting* pada banyak pohon pengambilan keputusan. Menurut Cuiqing Jiang, *random forest* merupakan metode yang baik untuk memprediksi *loan default* dikarenakan memiliki akurasi yang tinggi dan terhindar dari *overfitting*, selain itu, digunakan metode *adaboost* juga, untuk meningkatkan akurasi, stabilitas, dan masalah *overfit* dan *underfit* [9].

Berikut ini merupakan 3 hal yang menjadi fokus permasalahan:

- Masalah kredit macet merupakan fokus utama di dalam penelitian ini, penelitian diharapkan dapat menghasilkan prediksi *loan default customer* sehingga dapat mengurangi kemungkinan terjadinya *huge loss* pada perbankan.
- Performa dari *model* prediksi *loan default* yang akan digunakan pada penelitian ini.
- Keadaan *bank* yang diteliti berdasarkan *risk assessment NPL (Non-Perform Loan)*.

Berdasarkan fokus permasalahan yang dijabarkan diatas, berikut tujuan dari penelitian yang dilakukan:

- Merancang sistem prediksi *loan default customer* sehingga dapat mengurangi kemungkinan terjadinya *huge loss* pada perbankan. Diharapkan penelitian dapat menghasilkan *model* yang memiliki akurasi dan *recall rate* yang tinggi.
- Mendapatkan hasil *benchmarking* pada performa *model* yang diajukan.
- Menganalisis keadaan *bank* yang diteliti berdasarkan *risk assessment NPL*.

Penelitian ini diharapkan memiliki beberapa manfaat, yakni:

- mengenali cara pemodelan algoritma *Random forest* dan *Adaboost* untuk Prediksi,

- b. mengurangi resiko terjadinya *loan default* pada perbankan yang dapat menyebabkan *huge loss* pada bank,
- c. memberikan prediksi *negative behavior customer* sehingga data dapat dipakai untuk membuat keputusan dengan tingkat *risk* rendah,
- d. membuat proses *loan auditing* lebih cepat, karena sudah menggunakan *machine learning* (tidak *manual*).

II. KAJIAN LITERATUR

A. Machine learning

Machine learning adalah algoritma atau *model* statistik yang berfungsi untuk melakukan tugas spesifik tanpa suatu instruksi dengan mengandalkan "*pattern*". Algoritma *machine learning* akan membuat suatu *model* matematis berdasarkan data, yang biasa disebut "*data training*", yang digunakan untuk membuat suatu prediksi atau melakukan keputusan tanpa harus secara eksplisit diprogramkan untuk melakukan hal tersebut [4].

Machine learning dibagi menjadi dua, yakni *unsupervised learning* dan *supervised learning*. *Supervised learning* merupakan pembelajaran menggunakan data yang mengandung *input* serta *output* yang diinginkan. Klasifikasi dan Regresi merupakan bentuk dari *supervised learning*. Algoritma Klasifikasi berguna untuk memisahkan dan mengelompokkan data untuk melakukan prediksi sebagaimana akan digunakan pada *Loan Prediction*.

Unsupervised learning merupakan pembelajaran menggunakan data yang hanya memiliki *input* tetapi tidak mempunyai *output* spesifik. Fungsi utama dari *unsupervised learning* adalah menemukan fitur atau *pattern* dan akan mengkategorikan hasilnya sebagai prediksi [12].

B. Data Analytics (DA)

Data Analytics (DA) adalah proses memeriksa kumpulan data (*data set*) untuk menarik kesimpulan tentang informasi yang dikandungnya. DA digunakan dalam industri komersial untuk memungkinkan organisasi membuat keputusan bisnis yang lebih tepat dan oleh para ilmuwan atau peneliti untuk memverifikasi atau menyangkal *model*, teori, dan atau hipotesis ilmiah.

Inisiatif *Data Analytics* dapat membantu pebisnis meningkatkan pendapatan, meningkatkan efisiensi operasional, mengoptimalkan kampanye pemasaran dan upaya layanan pelanggan, merespons lebih cepat tren pasar yang sedang berkembang dan mendapatkan keunggulan kompetitif atas pesaing - semuanya dengan tujuan akhir untuk meningkatkan kinerja bisnis. Bergantung pada aplikasi tertentu, data yang dianalisis dapat terdiri dari catatan historis atau informasi baru yang telah diproses untuk penggunaan analitik waktu-nyata. Selain itu, dapat berasal dari campuran sistem *internal* dan sumber data eksternal [7].

C. Data Preprocessing

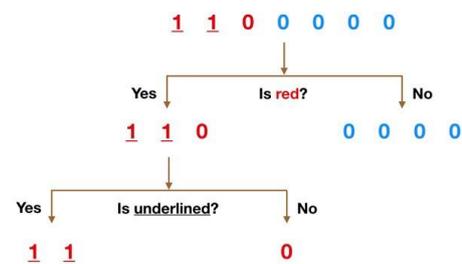
Data preprocessing adalah langkah yang sering diabaikan tetapi merupakan langkah penting dalam proses penambangan data. Pengumpulan data biasanya merupakan proses yang dikendalikan secara longgar, sehingga berada di luar jangkauan nilai, misalnya kombinasi data yang tidak mungkin, nilai yang hilang, dan lain-lain. Menganalisis data yang belum disaring dapat membuat masalah di hasil akhir karena hasil akhir tidak sesuai dengan ekspektasi pada pengolahan data. Dengan demikian, representasi dan kualitas data adalah yang pertama dan terpenting sebelum menjalankan analisis.

Persiapan data dapat mengambil waktu yang cukup besar dalam pemrosesan. *Data Preprocessing* meliputi persiapan data, ditambah dengan integrasi, pembersihan, normalisasi dan transformasi data; dan tugas pengurangan data. Hasil yang diharapkan setelah rangkaian tugas pemrosesan data yang andal adalah *dataset* final, yang bisa dianggap benar dan bermanfaat untuk algoritma penambangan data lebih lanjut [8].

D. Random Forest Algorithm

Random Forest Algorithm merupakan salah satu algoritma *machine learning* yang digunakan sebagai klasifikasi. Klasifikasi merupakan bagian penting dari *machine learning* yang bertugas untuk mengelompokkan data bergantung pada kecenderungannya.

Decision Tree merupakan bagian dari *Random forest*, yaitu merupakan pohon keputusan untuk membedakan suatu data. Sebagai contoh, diberikan data yaitu dua angka 1 dan lima angka 0, dan setiap angka memiliki warnanya masing masing, lalu data ingin dikelompokkan, dan fitur yang diambil dari data tersebut adalah warna dan garis bawah, maka pengerjaanya seperti gambar 1 [9].



Gambar 1. Cara Kerja *Decision Tree*

Random forest algorithm merupakan kumpulan dari *decision tree* yang beroperasi menjadi suatu gabungan fungsional. Setiap *decision tree* memiliki kesimpulan prediksi klasifikasi dan hasil prediksi akan digabungkan. *Random forest algorithm* memiliki hasil yang lebih baik dibandingkan dengan *model* individual lainnya dikarenakan *Random forest algorithm* menggunakan *decision tree* yang tidak memiliki korelasi. Kesalahan prediksi dalam satu *decision tree* dapat ditutupi dengan kebenaran yang

didapatkan dari *decision tree* lainnya asalkan arah pembuatan *decision tree* benar.

Ada beberapa ketentuan agar *random forest algorithm* baik digunakan yaitu:

- Decision tree* harus dibuat berdasarkan sinyal kecenderungan suatu data, tidak disarankan untuk menduga kecenderungan suatu data.
- Prediksi yang dilakukan oleh setiap *decision tree* harus memiliki korelasi yang rendah dengan *decision tree* lainnya.

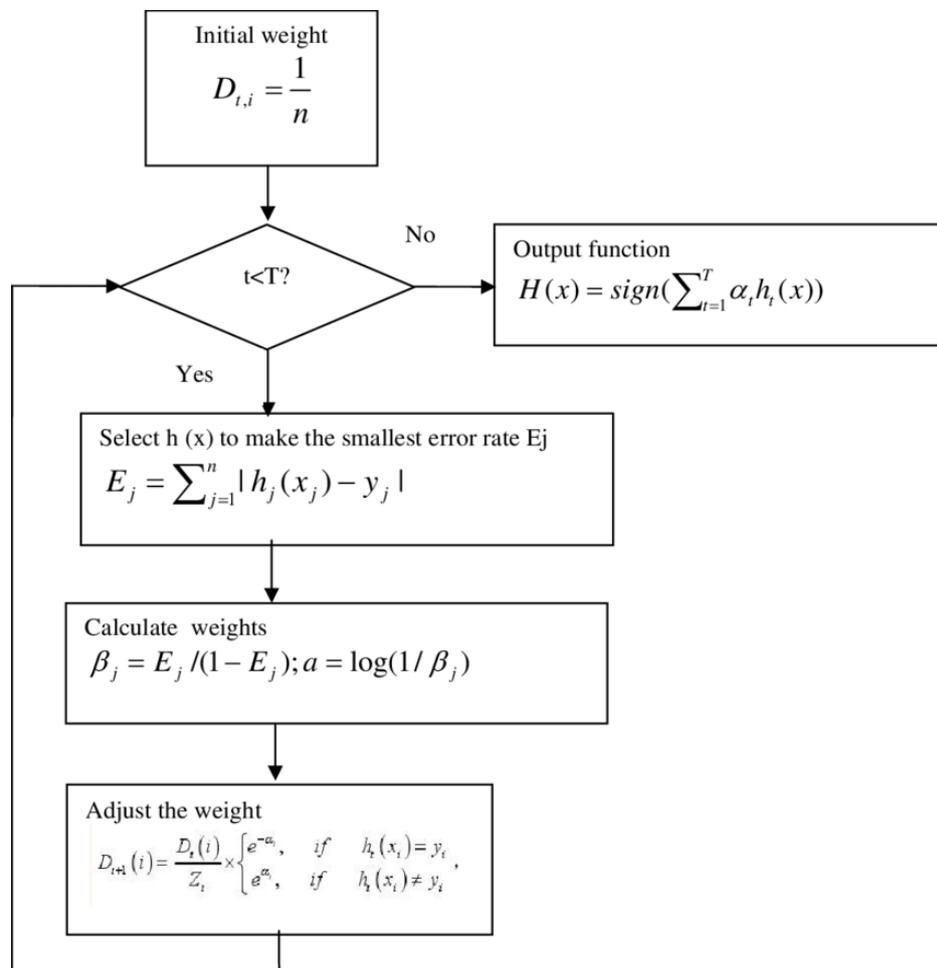
Random forest algorithm dipilih untuk melakukan prediksi *loan* yang baik atau tidak dikarenakan beberapa hal yakni:

- Random forest algorithm* berjalan efisien pada data yang jumlahnya banyak, dan memiliki akurasi yang tinggi dibandingkan algoritma lainnya

- Dapat berjalan dengan baik dengan kelas yang populasinya tidak seimbang [4].

E. Adaboost Algorithm

Algoritma *Adaboost* adalah salah satu dari *boosting classification algorithm* yang dapat meningkatkan kelompok klasifikasi "lemah" menjadi klasifikasi "kuat". Langkah pertama pada algoritma ini biasanya menggunakan algoritma klasifikasi dasar. Kemudian menyesuaikan bobot sampel sesuai dengan hasil klasifikasi dasar, yang membuat sampel yang diklasifikasi lebih diperhatikan. Selanjutnya sampel yang disesuaikan digunakan untuk melatih *base learner* berikutnya. Setelah iterasi, pembobotan ditambahkan ke *base learner* untuk membentuk klasifikasi akhir. Cara kerja algoritma ini dapat dilihat pada gambar 2 [8].



Gambar 2. Cara Kerja *Adaptive Boosting*

F. Confusion matrix

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode

klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya [5].

Berdasarkan jumlah keluaran kelasnya, sistem klasifikasi dapat dibagi menjadi 4 (empat) jenis yaitu klasifikasi *binary*, *multi-class*, *multi-label* dan *hierarchical* [12]. Pada klasifikasi *binary*, data masukan dikelompokkan ke dalam salah satu dari dua kelas. Jenis klasifikasi ini merupakan bentuk klasifikasi yang paling sederhana dan banyak digunakan. Contoh penggunaannya antara lain dalam sistem yang melakukan deteksi orang atau bukan, sistem deteksi kendaraan atau bukan, dan sistem deteksi pergerakan atau bukan.

G. Risk assessment

Risk assessment adalah metode yang sistematis untuk menentukan apakah suatu kegiatan/aset mempunyai risiko yang dapat diterima. *Risk assessment* sangat penting karena membantu menciptakan kesadaran tentang bahaya dan risiko yang didapatkan dari aset yang dimiliki. Hal ini bertujuan untuk mengurangi kemungkinan bahaya dengan menambahkan langkah-langkah pengendalian yang diperlukan dan tindakan pencegahan.

Penilaian risiko juga memprioritaskan bahaya dan membantu menentukan apakah tindakan pengendalian yang ada memadai, hal ini dilakukan untuk memeringkatkan risiko dan meyakinkan pihak manajemen bahwa terdapat risiko-risiko yang menjadi prioritas untuk dikelola secara efektif [13].

H. Non-Perform Loan (NPL)

Non-Perform Loan (NPL) adalah rasio yang menggambarkan tingkat persentase tertentu antara total kredit bermasalah dengan total kredit yang diberikan atau dengan kata lain kredit yang kolektabilitasnya masuk dalam kategori kolektabilitas tidak lancar. Dengan rumus yang diberikan sebagai berikut [13]:

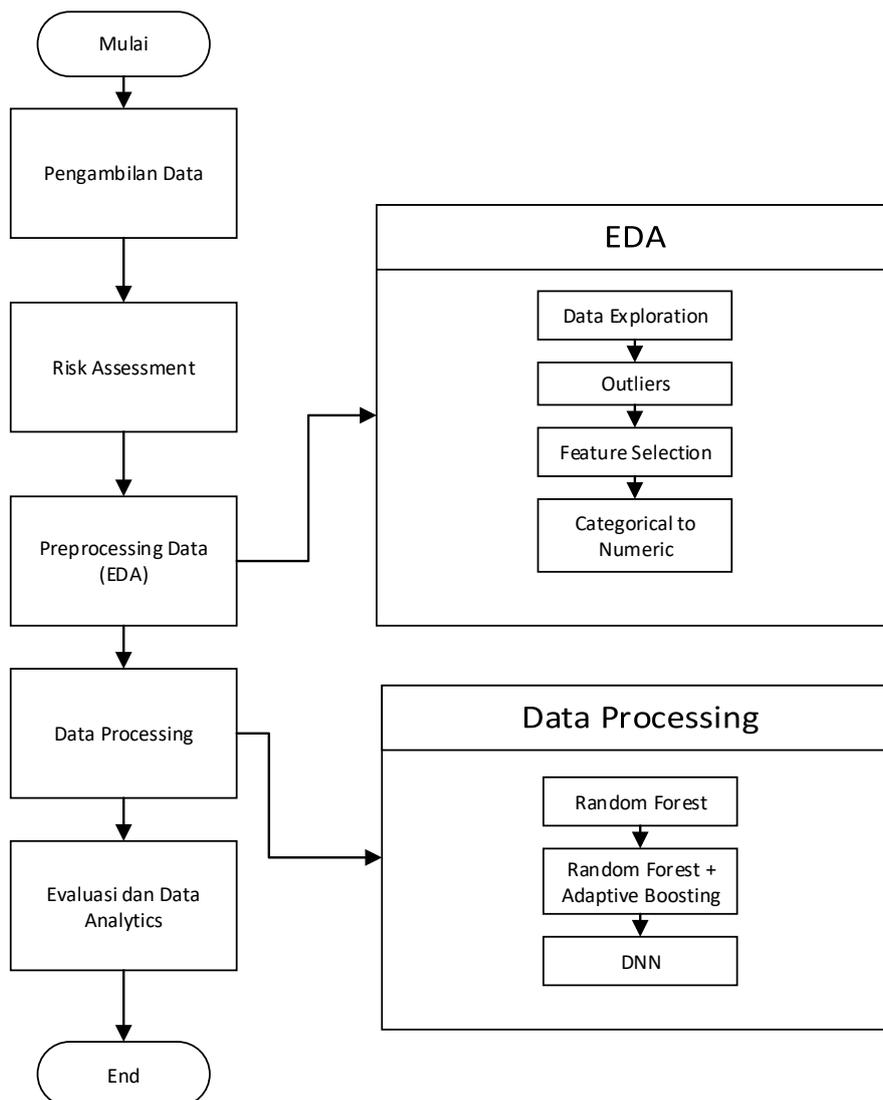
$$NPL\ Net = \frac{Bad\ Loan}{Loan\ Amount} \times 100\% \quad (1)$$

III. PERANCANGAN

A. Research Design

Perancangan penelitian ini akan menggunakan proses seperti pada gambar 3, yaitu:

- Pengambilan data yang berhubungan dengan *loan* pada jangka waktu 5 tahun.
- Menganalisis keadaan *loan* pada bank yang bersangkutan menggunakan *risk assessment*.
- Preprocessing* data atau *EDA* sehingga didapatkan fitur yang dapat menghasilkan *model* yang efektif.
- Data processing* menggunakan 3 metode yang akan dibandingkan yaitu *Random forest*, *Random forest* dan *Adaptive Boosting*, dan *DNN*.
- Evaluasi *model* yang telah dibuat, lalu bandingkan performa ketiga *model*.



Gambar 3. Research Design

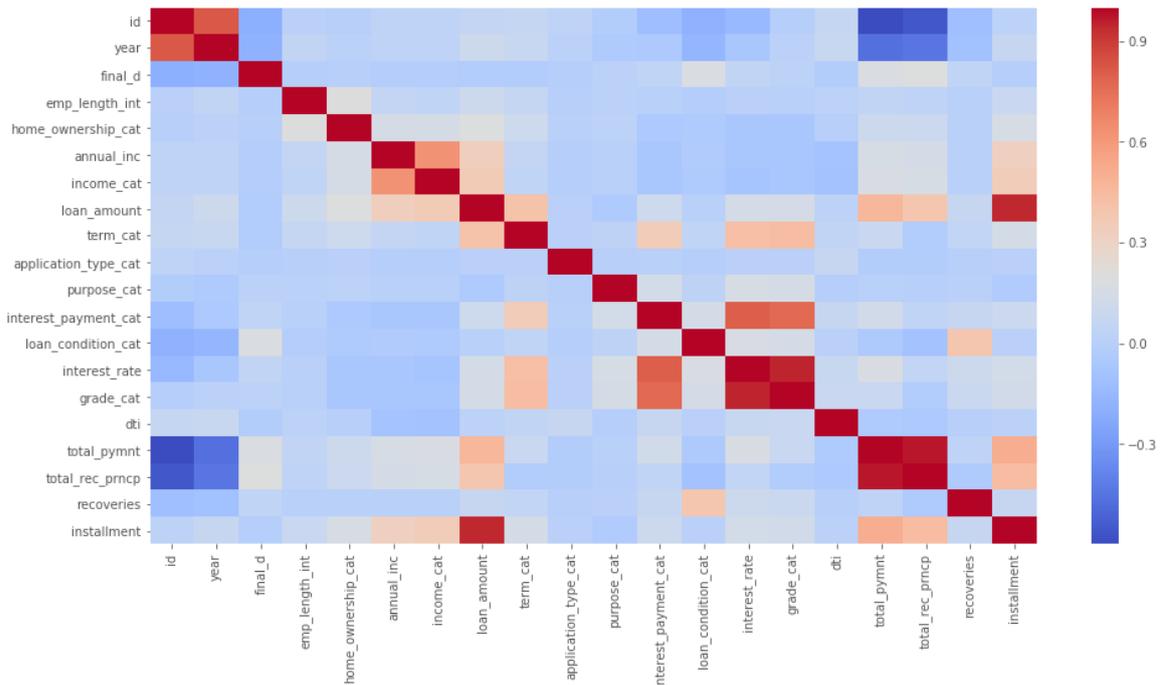
B. Sumber Data

Sumber data diperoleh dari salah satu bank di Indonesia. Data tersebut merupakan data peminjaman (*loan*). Data-data tersebut antara lain:

- a. Data Kondisi Peminjaman
Data ini merupakan data yang akan dipakai sebagai output dari prediksi. Data terdiri dari status baik atau buruknya sebuah pinjaman.
- b. Data Kepemilikan Rumah Kustomer
Data ini merupakan salah satu fitur *input* yang akan dipakai. Data berupa kepemilikan rumah yaitu *RENT*, *OWN*, *MORTGAGE*, *OTHER*, *NONE*, dan *ANY*.
- c. Data Tujuan Peminjaman
Data ini merupakan salah satu fitur *input* yang akan

dipakai. Data berupa tujuan dilakukannya peminjaman.

- d. Data *Grade* Peminjaman
Data ini merupakan salah satu fitur *input* yang akan dipakai. Data berupa klasifikasi kelas dari pinjaman.
- e. Data Kategori Pendapatan
Data ini merupakan salah satu fitur *input* yang akan dipakai. Data berupa klasifikasi kelas dari pendapatan customer yang melakukan pinjaman.
- f. Data lain-lain.
Data ini merupakan input fitur-fitur lain yang dapat dilihat pada gambar 4.



Gambar 4. Heatmap Correlation

C. Perancangan NPL Risk assessment

Berikut ini adalah perancangan *NPL Risk assessment* yang akan dipakai sebagai hasil *risk assessment* penelitian ini. Langkah-langkah yang digunakan:

- Mengambil dan menghitung total pengeluaran bank terhadap *loan* selama 5 tahun.
- Menghitung nilai *NPL* setiap tahunnya.
- Membandingkan nilai *NPL*, jika diatas 5% maka bank dinyatakan tidak sehat sehingga risiko tinggi.

Risk Assessment ini dilakukan agar memastikan kualitas data bank yang dipakai merupakan data yang baik, karena data yang buruk tidak akan menghasilkan model yang baik.

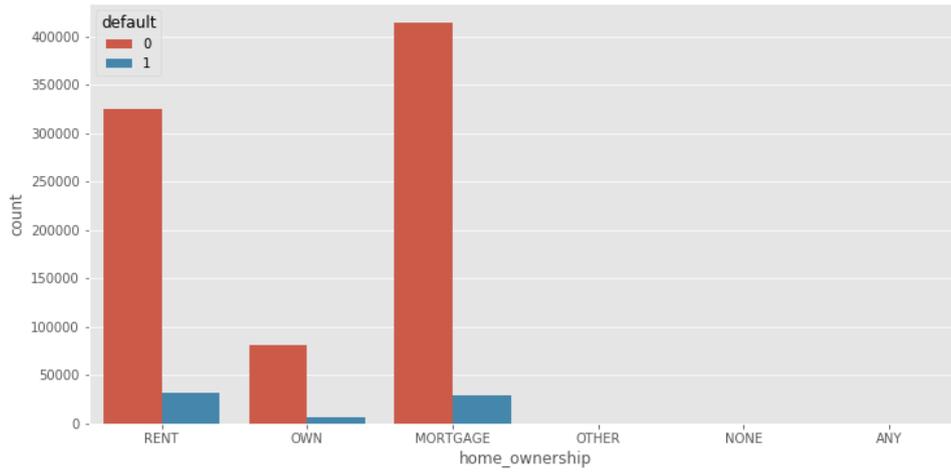
D. Pre-Processing Data

Berikut ini adalah langkah-langkah yang dilakukan untuk tahap *Preprocessing* data :

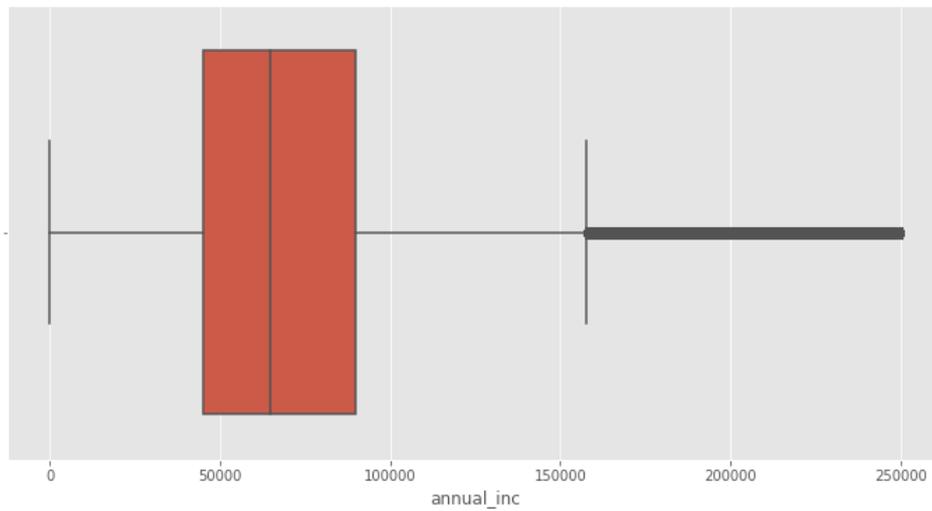
- Melihat korelasi data menggunakan *pearson* dan *heatmap*, dapat dilihat pada gambar 4. Korelasi berkisar dari 0 hingga 1. Nilai yang mendekati nol (warna biru) berarti tidak ada tren linier antara kedua variabel. Semakin dekat ke nilai 1 (warna merah), semakin berkorelasi positif.
- Membuat fitur kondisi pinjaman menjadi nilai 1 dan 0, Jika kondisi pinjaman baik maka nilai 0 dan sebaliknya.
- Membuang fitur yang dianggap tidak penting untuk pemrosesan data.
- Mengubah fitur lamanya bekerja, menggunakan *Min*

max scaler.

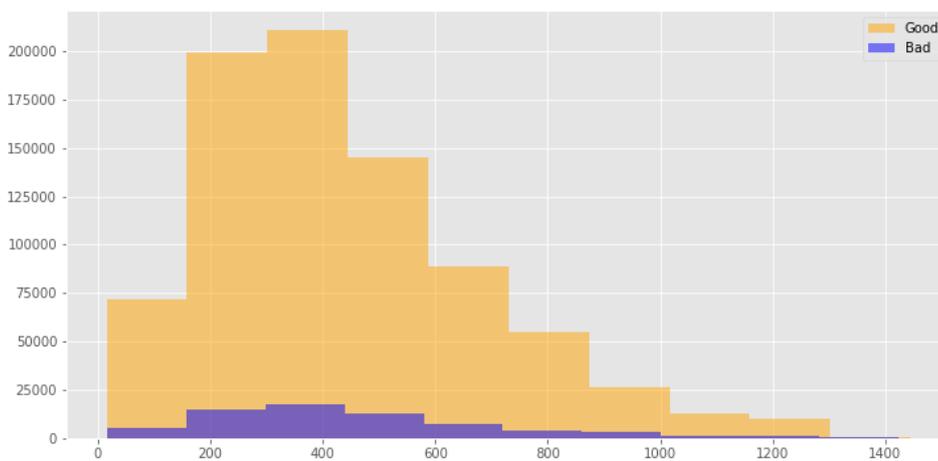
- Melihat hubungan fitur kepemilikan rumah dengan kondisi pinjaman menggunakan *countplot*, hasil dapat dilihat pada gambar 5. Jika dilihat untuk kepemilikan “*OTHER, NONE, dan ANY*” tidak berdampak pada data kondisi pinjaman yang ditandakan oleh nilai 0 dan 1 karena jumlahnya sangat sedikit atau bahkan tidak ada, sehingga fitur ini dapat dibuang.
- Melihat hubungan antara fitur lain dengan kondisi pinjaman.
- Menggunakan *boxplot* untuk mendeteksi adanya *Outlier*, *boxplot* ini ditampilkan pada gambar 6. Jika dilihat terdapat cukup banyak data *outlier* sehingga data harus dipangkas agar tidak berdampak pada akurasi model.
- Membandingkan besaran *installment* dengan keadaan pinjaman dengan menggunakan *histogram plot*, hasilnya dapat dilihat pada gambar 7.
- Setelah diketahui *properties* dari setiap data, maka dilakukan pemangkasan data yang dianggap *outlier*, perubahan jenis data agar menghemat *memory*, dan perubahan data klasifikasi menggunakan metode *one-hot encode*.
- Pembagian data *test* dan data *train*, pada penelitian ini adalah 70%: 30%.



Gambar 5. Hubungan Kepemilikan Rumah dengan Kondisi Pinjaman



Gambar 6. Penggunaan Boxplot untuk Mencari *Outlier*

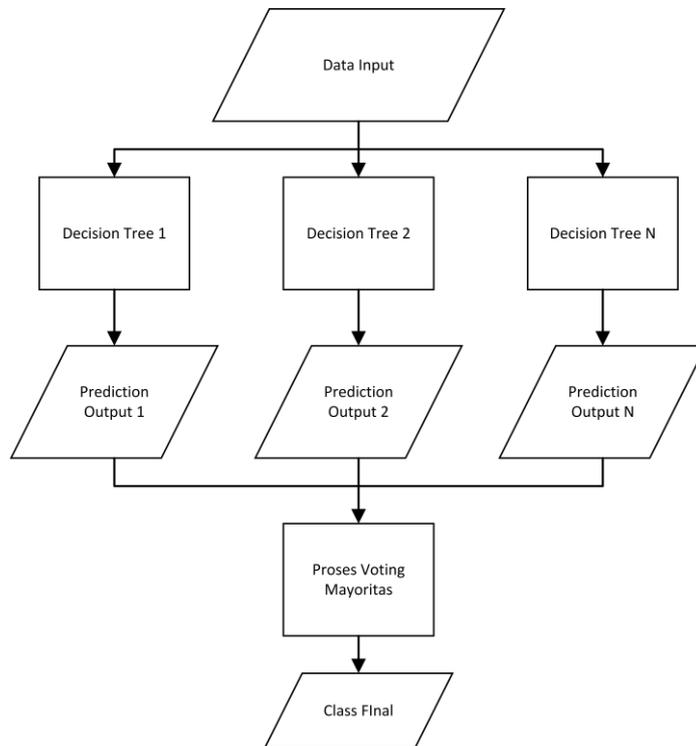


Gambar 7. Hubungan *Installment* Rumah dengan Kondisi Pinjaman

E. Perancangan Model Random Forest (Baseline)

Berikut ini adalah perancangan model Random forest yang akan dipakai sebagai base learner untuk Adaboost algorithm. Model dapat dilihat pada flowchart gambar 8.

Pada model yang dirancang Random forest ditetapkan untuk memiliki 100 cabang Decision Tree, gini criterion, dan “best” splitter.



Gambar 8. Flowchart Model Random Forest

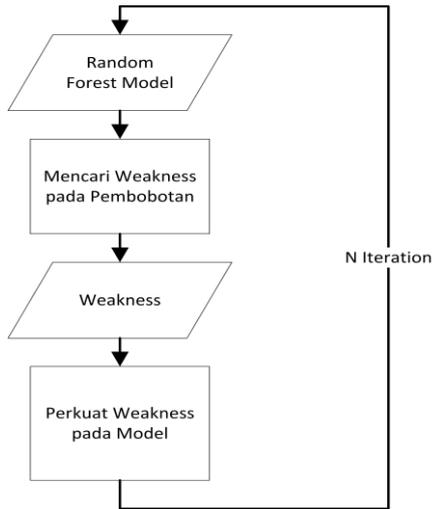
F. Perancangan Model Adaboost Algorithm

Berikut ini adalah perancangan model Adaboost menggunakan random forest algorithm yang sebelumnya dibuat sebagai base learnernya.

TABEL I
TREND NON-PERFORM LOAN (NPL) BANK DALAM 5 TAHUN TERAKHIR

Tahun	Total (Rp)	Uang Masuk (Bad Loan) (Rp)	Uang Masuk (Good Loan) (Rp)	NPL	Kenaikan
2011	261.683.825	25.209.111	258.419.151	0.10	0.01
2012	718.411.025	74.555.875	698.316.879	0.10	-0.03
2013	1.981.989.025	154.381.839	1.720.877.961	0.08	-0.04
2014	3.503.830.175	127.871.407	2.099.869.695	0.04	-0.03
2015	6.417.580.175	31.130.055	1.295.165.135	0.00	0.00

Flowchart untuk model Adaboost dapat dilihat pada gambar 9. Pada model ini ditetapkan untuk melakukan 100 kali iterasi



Gambar 9. Flowchart Model Adaboost

G. Perancangan Model DNN

Berikut ini adalah perancangan model DNN yang akan dipakai sebagai pembanding pada saat benchmarking model yang ada pada penelitian ini. Hal ini dilakukan untuk membandingkan performa metode machine learning yang digunakan dengan metode deep learning. Model DNN digunakan sebagai pembanding karena model DNN merupakan metode klasifikasi deep learning yang cukup baik dan mudah digunakan. Berikut parameter-parameter yang digunakan:

- Batch Size: 10000
- Epoch: 1000
- Shuffle: True
- Features: [emp_length, ann_inc, loan_amt, int_rate, dti, mortgage, own, rent, high, low, medium, short, long, credit, debt, low_int]
- Classes: 2

IV. HASIL EKSPERIMEN DAN EVALUASI

A. Hasil Analisis Risk Assessment NPL

TABEL II
TOTAL DANA YANG SUDAH DIKELUARKAN SELAMA 5 TAHUN

2011	2012	2013	2014	2015
Rp. 261.683.825	Rp. 718.411.025	Rp. 1.981.989.225	Rp. 3.503.830.175	Rp. 6.417.580.175

Pada tabel I ditampilkan data tren Non-Perform Loan (NPL) bank dalam 5 tahun terakhir. Berdasarkan tabel II, dari tahun 2011 hingga 2015 total peminjaman bank yang bersangkutan selalu mengalami kenaikan. Jika dilihat dari

tabel I, maka perhitungan NPL nilainya <5%.

Menurut Peraturan Bank Indonesia Nomor 6/10/PBI/2004 tanggal 12 April 2004 tentang Sistem Penilaian Tingkat Kesehatan Bank Umum, semakin tinggi nilai NPL (diatas 5%) maka bank tersebut tidak sehat. Semakin tinggi nilai NPL maka semakin tinggi risiko kredit yang akan ditanggung pihak bank, sesuai dengan Peraturan Bank Indonesia [10].

NPL yang tinggi menyebabkan menurunnya laba yang akan diterima oleh bank. Penurunan laba mengakibatkan dividen yang dibagikan juga semakin berkurang sehingga pertumbuhan tingkat return saham bank akan mengalami penurunan.

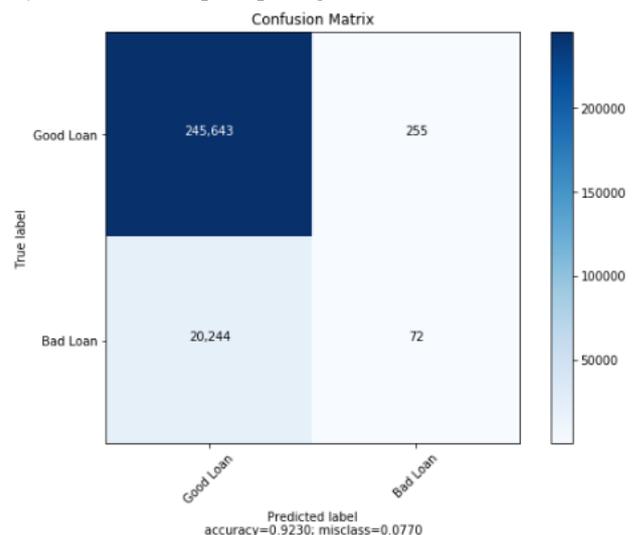
Sesuai dengan data pada tabel I, nilai NPL bank masih dalam batas toleransi yang diatur dalam undang-undang. Dengan ini, kita dapat menyimpulkan bahwa data ini cukup baik untuk digunakan. Tetapi data yang dipakai mempunyai masalah yaitu penyebaran data yang tidak merata.

B. Hasil Percobaan pada Model Random Forest

TABEL III
HASIL MODEL RANDOM FOREST

N-Tree	Accuracy	Recall Rate
10	42.3432%	0.1612%
30	63.13421%	0.2754%
50	84.6431%	0.2912%
70	90.0213%	0.3123%
100	92.2998%	0.3545%

Pada tabel III, ditampilkan accuracy maksimal pada jumlah Decision Tree dari 10 sampai dengan 100. Jika diperhatikan akurasi dari model sangat tinggi tetapi recall rate sangat rendah. Hal ini mungkin dikarenakan distribusi data yang tidak seimbang. Dari model ini juga didapatkan Confusion matrix seperti pada gambar 10.



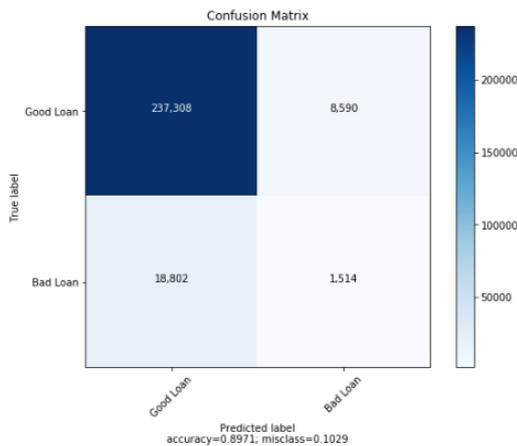
Gambar 10. Confusion matrix Model Random Forest

C. Hasil Percobaan pada Model Adaboost

TABEL IV
HASIL MODEL ADABOOST

N-Tree	N-Iteration	Accuracy	Recall Rate
100	10	39.5543%	0.0413%
100	30	43.54213%	2.1237%
100	50	62.431%	3.3522%
100	70	82.1103%	5.2331%
100	100	89.7105%	7.4522%

Pada tabel IV, terlihat bahwa dengan menggunakan model ini recall rate dapat ditingkatkan dengan tradeoff accuracy yang tidak terlalu besar. Dengan model ini didapatkan juga confusion matrix yang dapat dilihat pada gambar 11.



Gambar 11. Confusion matrix Model Adaboost

Jika diperhatikan pada gambar 10 dan gambar 11, terdapat error type I (false positive) dan error type II (false negative) yang jumlahnya cukup banyak. Hal ini terjadi dikarenakan sebaran data yang tidak seimbang antara good dan bad loan.

D. Hasil Percobaan pada Model DNN

TABEL V
HASIL MODEL DNN

Global Step	Loss	Exec. Time (sec)
100	0.6911358	12.449
2800	0.58773583	12.833
4100	0.5654046	12.681
4900	0.86097443	14.436

Pada tabel V, terlihat bahwa dengan menggunakan model DNN ini didapatkan loss terkecil dengan execution time yang cukup baik pada global step 4100. Selanjutnya akan

digunakan parameter global step 4100 untuk pengukuran accuracy dan recall rate-nya. Dengan cara ini didapatkan:

- Accuracy: 92.368543%
- Recall rate: 0.0000213%

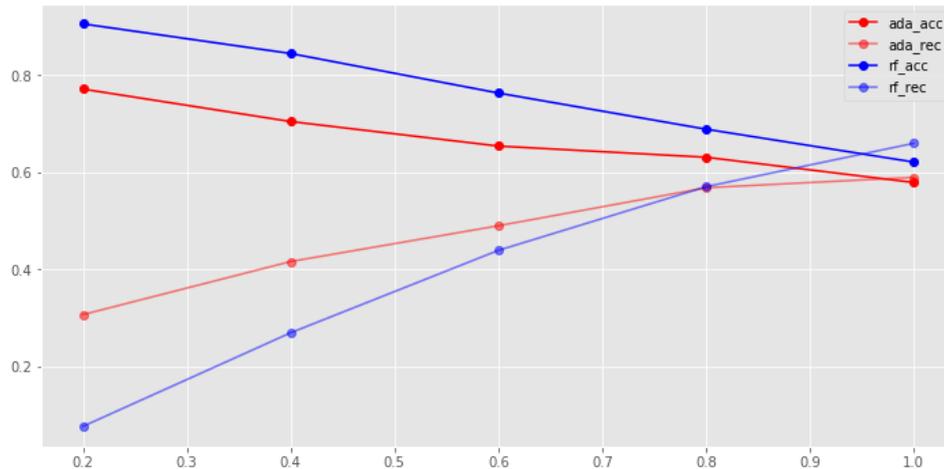
Tanpa ada perubahan data sampel dapat disimpulkan bahwa model DNN memiliki akurasi tertinggi. Model Random forest dengan Adaboost memiliki recall rate tertinggi. Tetapi nilai recall rate pada DNN masih sangat rendah, hal ini mungkin dikarenakan distribusi data yang tidak seimbang pada fitur kondisi peminjaman yaitu "Good Loan" dan "Bad Loan". Oleh karena itu, selanjutnya akan dilakukan analisis jika dilakukan undersampling pada jumlah data mayoritas.

E. Hasil Percobaan pada Model setelah dilakukan Undersampling

TABEL VI
HASIL MODEL SETELAH DILAKUKAN UNDERSAMPLING

Ratio	Model	Accuracy	Recall Rate
0.2	RF	90.628%	7.70312%
	RF + Adaboost	77.195%	30.725%
	DNN	82.431%	0.0352%
0.4	RF	84.5241%	26.9836%
	RF + Adaboost	70.50681%	41.6667%
	DNN	65.3541%	2.8414%
0.6	RF	76.3772%	43.96535%
	RF + Adaboost	65.4514%	49.03032%
	DNN	64.4312%	4.71872%
0.8	RF	68.9265%	57.08309%
	RF + Adaboost	63.1454%	56.86158%
	DNN	46.2341%	12.54814%
1.0	RF	62.1552%	66.02678%
	RF + Adaboost	57.9481%	59.0126%
	DNN	45.6324%	21.5657%

Pada tabel VI, terlihat mulai muncul model yang cukup baik. Model yang baik memiliki accuracy dan recall rate yang tinggi. Jika dilihat model DNN tidak beroperasi cukup bagus ketika dilakukan undersampling terhadap data. Model Random Forest muncul sebagai model yang paling bagus ketika dilakukan undersampling. Grafik perbandingan performa antara Random Forest dan Random Forest dan Adaboost dapat dilihat pada gambar 12.



Gambar 12. Perbandingan antara Model RF dan RF + Adaboost

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Pada perbandingan ketiga metode, *DNN* unggul pada akurasi dari prediksi modelnya. *Adaptive Boosted Random forest* menunjukkan akurasi terendah dari ketiga metode yang diuji tetapi menunjukkan *recall rate* yang tertinggi dibandingkan yang lain. *Undersampling* data dapat meningkatkan akurasi dan *recall rate* dari *random forest* pada kasus data ini.

Sesuai dengan tren data dalam 5 tahun terakhir, bank yang diteliti tidak terindikasi melanggar peraturan PBI Nomor 6/10/PBI/2004 tanggal 12 April 2004 mengenai *NPL* diatas 5%. *NPL* Bank masih dalam batas toleransi yang diatur undang-undang.

B. Saran

Sebaiknya *model* digunakan dalam proses indikasi *risk loan default* secara *real* dilapangan sehingga dapat dibandingkan hasil yang didapat dibandingkan cara konvensional.

UCAPAN TERIMA KASIH

Terima kasih diucapkan sebesar – besarnya kepada para pengajar dan staf Fakultas Teknologi Informasi, Program Studi Magister Ilmu Komputer Universitas Kristen Maranatha Bandung atas dukungan dan bantuannya dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] B. Chang, R. Yang, C. Guo, S. Ge, and L. Li, "A new application of optimized random forest algorithms in intelligent fault location

of rudders," *IEEE Access*, vol. 7, pp. 94276–94283, 2019, doi: 10.1109/ACCESS.2019.2926109.

- [2] L. Zhu, D. Qiu, D. Ergu, C. Ying, and K. Liu, "A study on predicting loan default based on the random forest algorithm," *Procedia Computer Science*, vol. 162, pp. 503–513, 2019, doi: 10.1016/j.procs.2019.12.017.
- [3] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.
- [4] J. R. Koza, F. H. Bennett, D. Andre, and M. A. Keane, *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*, Artificial Intelligence in Design '96, Dordrecht: Nederland, Springer, 1996, pp. 151–170.
- [5] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab*, Yogyakarta, Andi, 2012.
- [6] S. Garcia, J. Luengo, F. Herrera, *Data Preprocessing in Data Mining*, seri Intelligent Systems Reference Library, Switzerland, Springer, 2014.
- [7] F. Provost and T. Fawcett, *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*, 1. ed., 2. release. Beijing: O'Reilly, 2013.
- [8] F. Wang, Z. Li, F. He, R. Wang, W. Yu, and F. Nie, "feature learning viewpoint of adaboost and a new algorithm," *IEEE Access*, vol. 7, pp. 149890–149899, 2019, doi: 10.1109/ACCESS.2019.2947359.
- [9] C. Jiang, Z. Wang, R. Wang, and Y. Ding, "Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending," *Ann Oper Res*, vol. 266, no. 1–2, pp. 511–529, Jul. 2018, doi: 10.1007/s10479-017-2668-z.
- [10] H. Darmawi, *Manajemen Risiko*, Jakarta : Bumi Aksara, 2010.
- [11] P. K. Ozili, "Non-performing loans and financial development: new evidence," *The Journal of Risk Finance*, vol. 20, no. 1, pp. 59–81, Jan. 2019, doi: 10.1108/JRF-07-2017-0112.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
- [13] T. Aven, "Risk assessment and risk management: review of recent advances on their foundation," *European Journal of Operational Research*, vol. 253, no. 1, pp. 1–13, Aug. 2016, doi: 10.1016/j.ejor.2015.12.023.