

Penerapan Metode *Machine Learning* untuk Prediksi Nasabah Potensial menggunakan Algoritma Klasifikasi Naïve Bayes

Devi Fitrihanah¹, Saruni Dwiasnati^{2,*}, Hanny Hikmayanti H³, Kiki Ahmad Baihaqi⁴

1,2Fakultas Ilmu Komputer, Teknik Informatika, Universitas Mercu Buana

3,4 Fakultas Teknik dan Ilmu Komputer, Teknik Informatika Universitas Buana Perjuangan Karawang

Article Info

Article history:

Received March 20, 2021

Revised June 4, 2021

Accepted June 13, 2021

Keywords:

Potential Customers

Machine Vision

Machine Learning

Classification

Naïve Bayes

ABSTRACT

Customers are people who trust the management of their money in a bank or other financial service party to be used in banking business operations, thereby expecting a return in the form of money for their savings. To reach information to increase company profits, a method is needed to be able to provide knowledge in supporting the data that the company has. The model can be obtained by using predictive data processing of customer data that is categorized as potential or not potential. Data processing can be done using Machine Learning, namely classification techniques. This technique will produce a churn prediction model for determining the category of customers who fall into the Potential or Not Potential category and find out what accuracy value will be generated by applying the classification technique using the Naïve Bayes Algorithm. The parameters used in this study are Gender, Age, Marital Status, Dependent, Occupation, Region, Information. The data used are 150 data from customers who have participated in the savings program to find out whether the customer is in the Potential or Non-Potential category. The accuracy results generated using this data are 86.17% of the tools used by Rapidminer.

Copyright © 2021 Universitas Indraprasta PGRI.

All rights reserved.

Corresponding Author:

Saruni Dwiasnati,

Department of Informatic,

Universitas Mercu Buana, Jakarta

Jl. Meruya Selatan, Kebon Jeruk, Jakarta Barat

Email: saruni.dwiasnati@mercubuana.ac.id

1. PENDAHULUAN

Jasa keuangan merupakan suatu produk dan jasa keuangan yang disediakan oleh penyedia jasa keuangan, seperti Bank, Layanan Pembayaran, Bank Daearah, Perusahaan Asuransi, Koperasi, atau bahkan penyedia E-Money. Jasa Keuangan di bagi menjadi 2 bagian, yaitu Jasa Keuangan Perbankan yaitu suatu badan usaha yang berada di bawah pengawasan dan pengaturan OJK yang menghimpun dana dari masyarakat dalam bentuk simpanan serta menyalurkan dana tersebut kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dan Jasa Keuangan non Perbankan yaitu Lembaga-lembaga yang melakukan kegiatan keuangan selain bank sebagai contoh asuransi, dana pensiun, pasar modal.

Munculnya data mining didasari dari dengan adanya masalah data explosion atau tsunami data yang dialami oleh banyak bidang industri yang telah mengumpulkan data sekian tahun lamanya seperti data pembelian, data penjualan, data nasabah, data transaksi, dan data-data lainnya yang dapat menghasilkan sebuah informasi yang berguna unntuk masa depan. Adanya banyak data dalam perusahaan inilah yang disebut dengan data explosion atau ledakan data. Penggunaan aplikasi data mining di Indonesia lebih banyak dipakai dalam perbankan, industri, dan jasa [1]. Namun dalam kenyataannya tidak semua nasabah masuk dalam kategori Potensial untuk di tawari sebuah produk yang terdapat pada perusahaan asuransi tersebut. Kategori ini dipengaruhi oleh beberapa parameter yang dapat menopang hal tersebut, antara lain Jenis Kelamin, Umur, Status Perkawinan, Tanggungan, Pekerjaan, Wilayah, dan Informasi, maupun parameter tersebut tidak dapat

menjadi parameter baku dalam penentuan kategori tersebut. Sehingga tidak sedikit agent salah dalam memberikan informasi produk yang dibutuhkan oleh nasabah yang mengakibatkan tidak ada pertumbuhan laba karena produk-produk perusahaan asuransi tersebut hanya beberapa produk yang laku terjual padahal produk-produk tersebut bagus untuk di beli oleh nasabah yang lebih berpotensi. Maka metode Machine Learning ini sangat di butuhkan untuk menentukan *accuracy* dalam memberikan informasi untuk penentuan nasabah potensial.

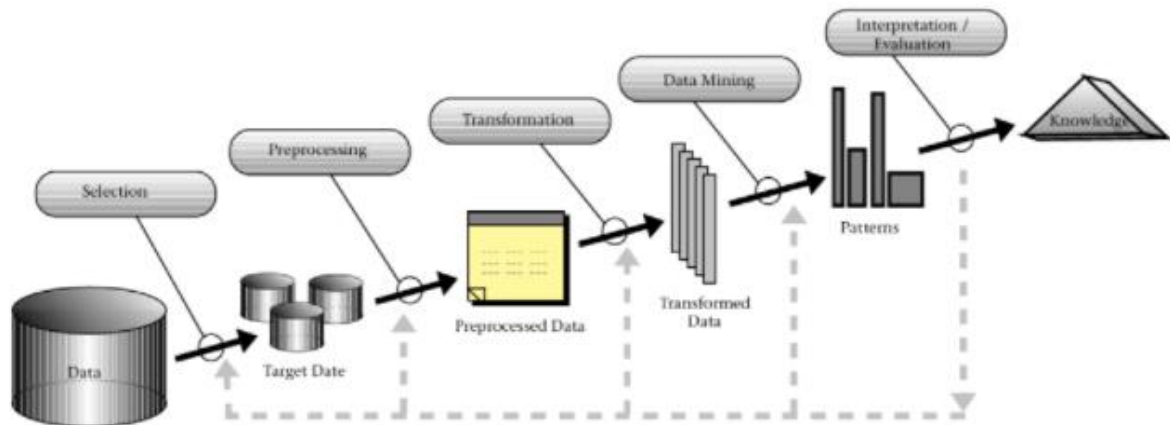
Dalam upaya pemberian informasi tentang prediksi nasabah potensial pada perusahaan asuransi. Para agent dan divisi yang memberikan data kepada agent pada perusahaan asuransi belum memiliki suatu model yang dapat membantu dalam melakukan prediksi nasabah tersebut masuk dalam kategori Potensial atau Tidak Potensial. Informasi tersebut diperoleh dengan menggunakan proses pengolahan data yang dimiliki oleh agent perusahaan asuransi tersebut terhadap data nasabah.

Untuk menentukan nasabah yang dianggap potensial atau tidak potensial ini sebuah Bank memerlukan sebuah system agar tepat sasaran. Salah satu yang dapat diterapkan adalah dengan menerapkan metode Data Mining. Dengan menggunakan teknik ini proses menentukan nasabah potensial atau tidak potensial dapat dibantu dengan penggunaan aplikasi komputer yang dibuat berdasarkan algoritma yang cocok dengan data yang dimiliki untuk menghasilkan nilai *accuracy*. Dengan strategi marketing yang terarah diharapkan biaya yang dikeluarkan semakin kecil sedangkan laba yang didapat semakin besar agar tetap dapat bertahan dengan bisnis yang di jalankan. *Model Machine Learning* (ML) yang terintegrasi dengan algoritma machine vision telah digunakan sudah banyak di implementasikan di berbagai bidang, seperti bidang pertanian, kesehatan, jasa keuangan. Namun dalam penelitian ini model machine learning di implementasikan dalam bidang penjualan perbankan, hal ini di lakukan untuk menentukan peningkatan jumlah nasabah potensial yang masih bias di lakukan untuk menambah laba pada perusahaan tersebut. Pada penelitian yang di lakukan oleh [2] menyebutkan bahwa klasifikasi masyarakat miskin berdasarkan data penduduk miskin yang diperoleh dari Kecamatan Tibawa dengan menggunakan teknik *data mining*. Atribut yang digunakan dalam melakukan klasifikasi masyarakat miskin adalah Umur, Pendidikan, Pekerjaan, Penghasilan, Tanggungan, Status (Kawin atau Belum Kawin). Metode yang akan digunakan adalah metode *Naïve Bayes Classifier*, yang merupakan salah satu teknik pengklasifikasian dalam data mining. Berdasarkan penelitian yang dilakukan dihasilkan kesimpulan bahwa sistem klasifikasi masyarakat miskin di wilayah pemerintahan Kecamatan Tibawa Kabupaten Gorontalo menghasilkan pengujian *confusion matrix* dengan teknik *split* validasi, penggunaan metode klasifikasi *naïve bayes* terhadap dataset yang telah diambil pada objek penelitian diperoleh tingkat akurasi sebesar 73% atau termasuk dalam kategori *Good*. Sementara nilai *Precision* sebesar 92% dan *Recall* sebesar 86%. Studi kasus lainnya menyebutkan bahwa dengan menggunakan data mining dapat menganalisa nasabah yang berpotensi [3] untuk mengikuti tabungan deposito berjangka, hasil pengolahan data dibandingkan dan menunjukkan bahwa Algoritma *Naïve Bayes* dengan *Particle Swarm Optimization* memiliki *accuracy* tertinggi sebesar 97,04%.

Tujuan yang ingin dicapai dari penelitian ini adalah untuk mengetahui nilai *classifier* dari algoritma yang digunakan untuk menghasilkan nilai terbaik dalam hal *accuracy*, *precision*, dan *recall*. Penelitian ini diharapkan dapat berguna bagi para pemberi informasi data untuk di tindak lanjuti oleh para agen pada perusahaan asuransi tersebut dalam menentukan kategori nasabah tersebut masuk dalam kategori Potensial atau Tidak Potensial untuk di hubungi oleh para agen yang di miliki oleh perusahaan asuransi tersebut.

2. METODE

Penelitian ini menerapkan algoritma *machine vision* untuk mengumpulkan data dan memilih data nasabah tersebut masuk dalam kategori nasabah potensial atau tidak potensial pada tahapan yang berbeda dengan menggunakan beberapa data yang di dapatkan hasilnya. Teknik menganalisis data dalam penerapan data mining ini menggunakan proses tahapan *Knowledge Discovery in Databases* (KDD) yang terdiri dari *Database*, *Data Selection*, *Data Cleaning*, *Data transformation*, *Data mining*, dan *interpretation*. Dalam KDD terdapat enam fase atau tahapan proses. Adapun tahapan proses data mining dengan menggunakan metode KDD digambarkan pada gambar 1 berikut ini:



Gambar 1. Tahapan Metode Penelitian Menggunakan KDD

Untuk memperbesar keberhasilan, penelitian ini akan dilakukan dengan metode KDD dengan tahapan penelitian meliputi:

1. *Database*
Data yang digunakan dalam penelitian ini adalah data yang di ambil dari kumpulan data set yang berjumlah 5.000 data nasabah yang di ambil kisaran pada Januari 2017 sampai dengan Juni 2019.
2. *Data Selection*
Langkah selanjutnya setelah data set di dapatkan maka buatlah parameter-parameter yang dibutuhkan untuk menghasilkan model algoritma yang dibutuhkan.
3. *Data Cleaning*
Setelah data dipisahkan sesuai dengan parameter yang dibutuhkan maka langkah selanjutnya ada membersihkan data-data yang mengalami salah penulisan, kurangnya ejaan pada tulisan tersebut, data tidak bermakna.
4. *Data transformation*
Melakukan proses klasifikasi untuk menghasilkan kelompok-kelompok data yang sejenis
5. *Data mining*
Membuat model yang sesuai dan menarik dengan data dan algoritma yang digunakan. Pada tahap ini menggunakan data yang sudah di import dari excel ke tools yang digunakan dan di beri label untuk menentukan apa yang ingin di ketahui hasil nilai nya.
6. *Interpretation.*
Implementasikan model yang sudah di rancang pada tools yang digunakan dan pembuatan laporan.

2.1. Data Mining

Data Mining adalah proses penggalian data dari tumpukan *database* yang berukuran besar yang digunakan untuk menemukan *knowledge* berupa informasi penting dan bermanfaat [4]. *Data mining* merupakan sebuah proses untuk menemukan dan memberikan gambaran berupa pola struktural dalam data sehingga dapat digunakan sebagai alat bantu dalam membantu memberikan penjelasan terhadap sekumpulan data, selain itu juga membuat prediksi dari data tersebut. Bidang yang termasuk dalam data mining adalah bagian dari *Knowledge Discovery Data* (KDD) yaitu proses untuk meringkas sebuah informasi yang bermanfaat yang sebelumnya belum diketahui dan tidak terlihat. *Data mining* juga digunakan untuk mengembangkan model yang dapat digunakan untuk memahami fenomena dari analisis data dan prediksi atau perkiraan.

Knowledge Discovery Data terdiri dari 5 tahapan dalam prosesnya [5]:

1. Langkah pertama adalah pemilihan data atau *data selection*. Pada proses ini data secara keseluruhan perlu diseleksi yang akan digunakan dalam proses data mining. File tersebut disimpan dalam sebuah berkas terpisah dari data sebelumnya.

2. Pembersihan data dilakukan dengan membuang data yang tidak konsisten, membuang data yang berulang atau sama, memperbaiki data yang kurang lengkap.
3. Transformasi pada data, data perlu ditransformasikan agar sesuai dengan algoritma data mining yang digunakan.
4. Data Mining, dilakukan pemrosesan data sesuai dengan algoritma yang di pilih.
5. Evaluasi, hasil pengolahan dari data mining kemudian di sajikan kedalam laporan atau bentuk yang mudah dimengerti oleh orang lain.

2.2. Klasifikasi

Proses penemuan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [4]. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/classification trees*, *Bayesian classifiers/ Naïve Bayes classifiers*, *Neural networks*, Analisa Statistik, Algoritma Genetika, *Rough sets*, *k-nearest neighbor*, *Metode Rule Based*, *Memory based reasoning*, dan *Support vector machines (SVM)* [6].

2.3. Algoritma Naïve Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [7]. Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam *Data Mining*. Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

2.4. Data Nasabah

Data yang digunakan dalam penelitian saat ini yang berhasil terkumpul terdiri dari data nasabah tahun 2017 sampai dengan tahun 2019 dengan berbagai kriteria yang sesuai dengan output yang dapat dijadikan sampel dalam penelitian yang akan dijadikan sebagai validasi model, sebanyak 150 record dan data set yang berhasil dikumpulkan sekitar 5.000 data yang masih belum dapat digunakan. Agar data set dapat digunakan mesti di sesuaikan dengan parameter yang di butuhkan, dengan data yang terbaik tanpa ada missing value dan sesuai dengan standar dari data yang dapat digunakan oleh algoritma tersebut.

2.5. Rapidminer

Rapidminer adalah salah satu software untuk pengolahan *Machine Learning*, *data mining*, *Text Mining*, dan *Predictive Analytics*. Pekerjaan yang dilakukan oleh *rapidminer text mining* adalah berkisar dengan analisis teks, mengekstrak pola-pola dari dataset yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan database. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program *Java* di bawah lisensi *GNU Public Licence* dan *Rapid Miner* dapat dijalankan di sistem operasi manapun [8]. *Rapid Miner* dikhususkan untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model *Bayesian Modelling*, *Tree Induction*, *Neural Network* dan lain-lain. Banyak metode yang disediakan oleh *Rapid Miner* mulai dari klasifikasi, *clustering*, asosiasi dan lainnya. Jika tidak ada

sebuah divisi pemasaran, data tersebut di ambil dari tahun 2017 hingga 2019. Dari tahap pemilihan data ini didapatkan data sebanyak 5.000 dari data mentah yang di dapatkan dari dua tahun tersebut dari beberapa departement. Parameter yang digunakan untuk data nasabah yaitu Jenis Kelamin, Umur, Tanggungan, Pekerjaan, Wilayah, Waktu, dan label yang digunakan pada penelitian itu adalah Data Potensial.

3.2. Data Pre-processing (Pembersihan Data)

Pada penelitian ini tahap pembersihan data dilakukan karena melihat data yang tersedia pada database di perusahaan tersebut masih sangat minim yang dibutuhkan untuk pengolahan data untuk penentuan nilai accuracy nya. Tahap pembersihan ini berfungsi untuk menghilangkan data duplikasi, kesalahan, dan validation rules pada database dari sebuah data mentah tersebut.

3.3. Transformation Data

Tahap ini merupakan proses untuk melakukan perubahan bentuk tabel terhadap data yang telah dipilih, sehingga data tersebut sesuai untuk proses *data mining*. Tabel yang akan digunakan pada penelitian ini berupa tabel dengan Kecamatan, Intensitas Curah Hujan, Debit Air, Luas Wilayah, Lamanya Hujan, Kepadatan Penduduk, dan label yang digunakan pada penelitian itu adalah Target.

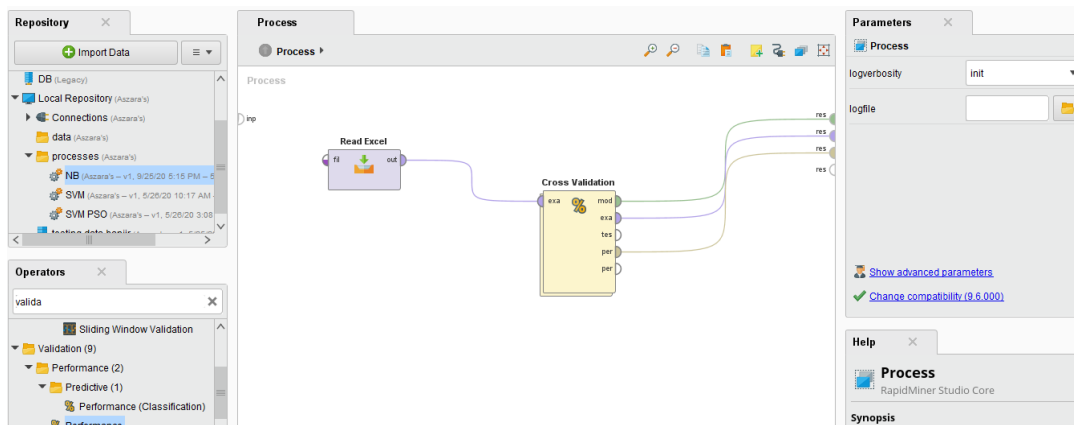
Tabel 1. Tabel Data Training

Jenis Kelamin	Umur	Tanggungan	Pekerjaan	Wilayah	Keterangan Waktu	Target
L	Konsolidasi	1	PNS	Tengah	Pagi	Potensial
L	Konsolidasi	2	Profesional	Tengah	Pagi	Potensial
L	Konsolidasi	2	Profesional	Tengah	Siang	Potensial
P	Konsolidasi	2	Profesional	Tengah	Siang	Potensial
P	Konsolidasi	2	Profesional	Tengah	Siang	Potensial
L	Konsolidasi	2	Wiraswasta	Tengah	Siang	Potensial

3.4. Data Mining

Proses ini bertujuan untuk mendapatkan pola-pola dan informasi yang tersembunyi di dalam basis data yang telah melewati tahap transformation. Pada tahap ini dibuat tiga form utama yang digunakan untuk penelitian yaitu form k-Fold Cross validation, uji akurasi, dan prediksi Naïve Bayes. Metode K-Fold Cross Validation dan Uji akurasi digunakan untuk mengetahui nilai k-Optimal yang dapat dihasilkan. Setelah mendapatkan nilai k-Optimal, maka selanjutnya menggunakan nilai k tersebut untuk melakukan prediksi nasabah yang masuk dalam kategori potensial ada berapa dan yang masuk dalam kategori tidak potensial ada berapa.

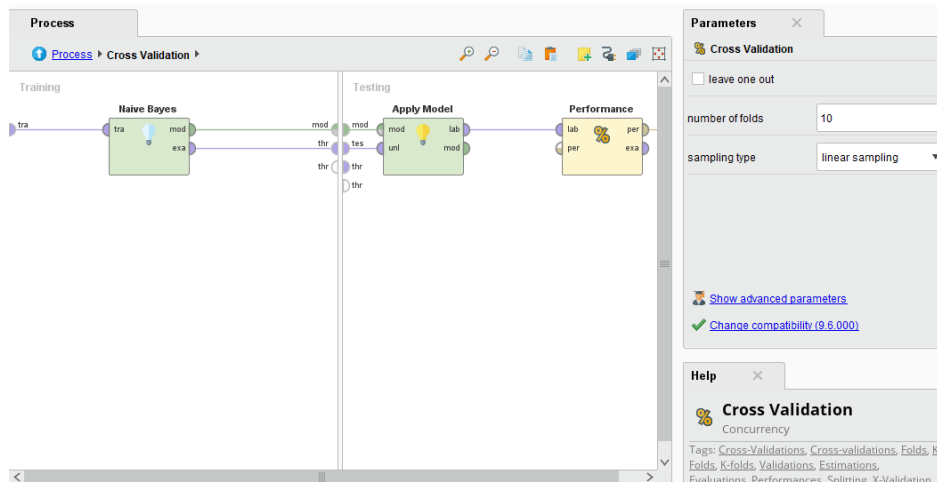
3.4.1. Hasil Klasifikasi Dengan Naïve Bayes:



Gambar 3. Desain Proses algoritma Naïve Bayes

Pemodelan dilakukan dengan memasukkan data training yang berjumlah 150 record yang di input pada Read Excel dan dihubungkan dengan Cross Validation. Setelah itu double klik cross validation sampai muncul jendela baru yang dari design yang diinginkan.

3.4.2. Hasil Desain Cross Validation



Gambar 4. Desain Cross Validation

Dalam jendela Training tersebut dimasukkan Naïve Bayes, dan untuk jendela Testing dimasukkan data Apply model dan Performace dengan mencoba number of folds nya sebanyak K-10, karena di rasa menggunakan K-Fold nya sampai 10 hasil yang dihasilkan cukup bagus. Setelah proses dijalankan akan mendapatkan hasil yang bisa dilihat pada menu result.

3.4.3. Nilai Accuracy menggunakan algoritma Naïve Bayes

Table View Plot View

accuracy: 86.17% +/- 9.94% (micro average: 86.09%)

	true Tidak Potensial	true Potensial	class precision
pred. Tidak Potensial	69	18	79.31%
pred. Potensial	3	61	95.31%
class recall	95.83%	77.22%	

Gambar 5. Result Hasil Accuracy

Tabel hasil pengolahan data menunjukkan bahwa accuracy yang diperoleh sebesar 86.17%. Hasil analisa antara data yang di testing dengan data training di Rapid Miner dapat dilihat pada Gambar 5. Untuk menghitung akurasi sebagai berikut:

- Jumlah data yang diuji : 150
- Jumlah data yang di prediksi benar “Potensial” : 61
- Jumlah data yang di prediksi benar “Tidak Potensial” : 18
- Jumlah data yang di prediksi salah “Potensial” : 3
- Jumlah data yang di prediksi salah “Tidak Potensial” : 69

Dari perhitungan tersebut dapat disimpulkan bahwa klasifikasi dengan menggunakan algoritma Naive bayes untuk menentukan nasabah potensial menghasilkan tingkat akurasi sebesar 86.17% dan tingkat error 13.83%.

3.5. Evaluation

Pada tahap ini dilakukan pencarian k-Optimal pada algoritma Naïve Bayes untuk prediksi Nasabah yang potensial dan nasabah yang tidak potensial berdasarkan dengan parameter yang digunakan. Setelah

mendapatkan nilai k-Optimal maka dapat digunakan pada algoritma Naïve Bayes untuk memprediksi nasabah tersebut masuk ke dalam kategori potensial atau tidak potensial.

4. PENUTUP

Beberapa kesimpulan yang didapat dari penelitian ini adalah pemilihan algoritma yang digunakan tergantung perlu atau tidaknya melihat semua data set sebelum melakukan proses pada setiap *instancenya*, apakah menggunakan algoritma yang memproses setiap *instance* secara langsung, atau algoritma yang membaca semua *training instance*. Dari perhitungan yang telah dilakukan maka klasifikasi dengan menggunakan algoritma *Naive bayes* untuk menentukan nasabah potensial menghasilkan tingkat akurasi sebesar 86.17% dan tingkat error 13.83%. Hal ini berguna untuk menentukan kategori nasabah masuk dalam kategori potensial atau tidak potensial untuk di hubungi oleh agen perusahaan asuransi tersebut.

UCAPAN TERIMA KASIH

Peneliti ingin mengucapkan terima kasih kepada Universitas Mercu Buana yang sudah mendukung di berbagai sektor untuk menyelesaikan penelitian ini, dan Universitas Buana Perbangsa Karawang untuk mensupport penelitian ini dapat terlaksana dengan baik.

DAFTAR PUSTAKA

- [1] A. Fadli, "Konsep Data Mining," *Konsep Data Mining*, pp. 1–9, 2003.
- [2] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018.
- [3] C. Agustina, "Analisa Nasabah Potensial Tabungan Deposito Berjangka Menggunakan Teknik Klasifikasi Data Mining," *J. Teknol. Inf. dan Terap.*, vol. 5, no. 2, pp. 105–112, 2018.
- [4] S. Wahyuni, "Implementasi Rapidminer Dalam Menganalisa Data Mahasiswa Drop Out," *J. Abdi Ilmu*, vol. 10, no. 2, pp. 1899–1902, 2018.
- [5] E. T. L. Kusriani, *Algoritma data mining*. 2009.
- [6] A. Jananto, "Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa," *Dinamik*, vol. 18, no. 1, p. 242727, 2013.
- [7] P. Purnama and C. Supriyanto, "Deteksi Penyakit Diabetes Type II Dengan Naive Bayes Berbasis Particle Swarm Optimization," *J. Teknol. Inf.*, vol. 9, 2013.
- [8] S. Haryati, A. Sudarsono, and E. Suryana, "implementasi data mining untuk memprediksi masa studi mahasiswa menggunakan algoritma c4. 5 (studi kasus: universitas dehasen bengkulu)," *J. Media Infotama*, vol. 11, no. 2, 2015.
- [9] M. Bramer, *Principles of data mining*, vol. 180. Springer, 2007.
- [10] A. V. Pakpahan and H. Irawan, "Penerapan Algoritma Naive Bayes Untuk Menentukan Nasabah Potensial pada AJB BUMIPUTERA 1912," *J. Komput. Bisnis*, vol. 10, no. 2, 2017.