

## PREDIKSI KELANCARAN PIUTANG PELANGGAN PADA PT. CITRA INA FEEDMILL DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBORS

Auki Akbar <sup>1\*)</sup>, Riza Adrianti Supono <sup>2)</sup>

<sup>1</sup>Magister Manajemen Sistem Informasi, Universitas Gunadarma

E-mail: [aukiakbar123@gmail.com](mailto:aukiakbar123@gmail.com)

<sup>2</sup>Magister Manajemen Sistem Informasi, Universitas Gunadarma

E-mail: [adrianti@staff.gunadarma.ac.id](mailto:adrianti@staff.gunadarma.ac.id)

**Abstract :** *Giving credit to customers is a solution that is often done by business actors today, such as companies. By offering credit solutions, companies can attract potential customers and make it easier for customers to make payments. However, being able to provide credit to customers can cause losses for the company when customers are unable to pay their periodic credits. For this reason, in this study an analysis was carried out to categorize customer payment capabilities using the Naïve Bayes Algorithm and the K-Nearest Neighbors Algorithm. The results of the study, obtained the best data mining model to predict the age classification of customer receivables using the KNN algorithm which is optimized with a feature selection algorithm with an accuracy of 99%.*

**Keywords:** *algorithm, data mining, KNN, Naïve Bayes, customer receivables*

### 1. PENDAHULUAN

Kelancaran piutang pelanggan merupakan salah satu elemen penting agar neraca keuangan suatu perusahaan stabil. Pemberian piutang kepada pelanggan, dalam hal ini pemberian kredit, juga sangat berguna bagi pelanggan. Hal itu dikarenakan, akan meringankan pelanggan dalam membeli produk perusahaan. Selain itu perusahaan juga mendapatkan keuntungan lain dengan menerapkan solusi pemberian kredit yaitu menarik minat pelanggan.

Pemberian kredit yang dilakukan perusahaan pun biasanya dilakukan dengan banyak pertimbangan dikarenakan pengetahuan mengenai kondisi pelanggan yang melakukan kredit kepada perusahaan sedikit, sehingga resiko kerugian karena memberikan kredit kepada pelanggan pun tetap ada.

Permasalahan pemberian kredit kepada pelanggan juga dialami oleh PT. Citra Ina Feedmill (PT.Citrafeed). PT. Citrafeed adalah suatu perusahaan yang memproduksi pakan ternak dan hasil produksinya didistribusikan ke perusahaan pemasok bahan makanan atau peternakan di seluruh Indonesia. Sehingga, perusahaan ataupun peternakan tersebut merupakan pelanggan dari PT. Citrafeed.

Masalah pemberian kredit tak luput dari permasalahan dari PT. Citrafeed. Walaupun setiap pelanggan yang terdaftar merupakan pelanggan yang telah di survei oleh karyawan perusahaan. Tetapi kondisi umur piutang yang terlampau lama, terkadang masih bisa terlihat di laporan umur piutang.

Dengan menggunakan data mining, diharapkan permasalahan tersebut dapat diselesaikan. Beberapa penelitian sebelumnya menjadi acuan untuk dikembangkannya penelitian ini, seperti

penelitian mengenai analisa mengenai data koperasi pada PUSKOPDIT Jakarta pada tahun 2018. Penelitian ini, menggunakan Algoritma Naïve Bayes yang dioptimasi dengan Algoritma Particle Swarm Optimizer (Mardiana, 2018). Penelitian berikutnya mengenai analisa akurasi model yang dibangun dengan menggunakan algoritma Naïve Bayes dan K-Nearest Neighbors dan kombinasi kedua algoritma tersebut untuk memprediksi kredit nasabah Bank Danamon (Gultom & Simanjuntak, 2020) dan penelitian mengenai prediksi tingkat kelancaran kredit bank menggunakan algoritma naïve bayes berbasis algoritma forward selection (Hasan, 2017)

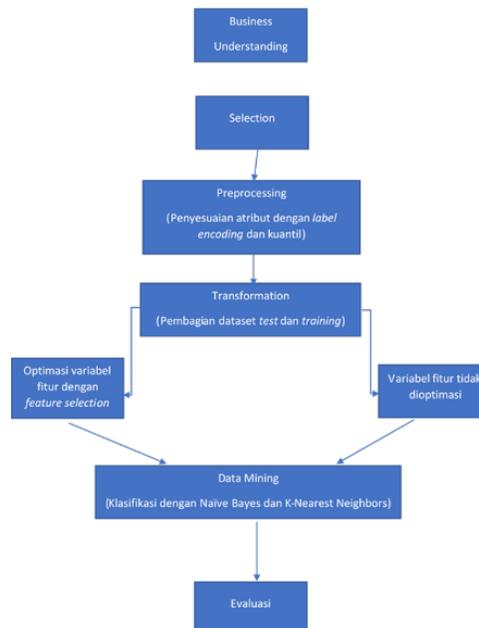
Laporan umur piutang pada PT.Citrafeed, mengklasifikasikan umur piutang menjadi periode *di bawah 30 hari, di bawah 60 hari, di bawah 90 hari, di bawah 120 hari, dan lebih dari 120 hari*. Tujuan penulis pada penelitian ini, yaitu membuat model data mining untuk memprediksi umur piutang sesuai klasifikasi umur piutang yang ada pada laporan tersebut. Kemudian mengevaluasi model yang telah dibuat menggunakan kedua algoritma tersebut dan mengkomparasikan algoritma tersebut ketika variabel fitur dioptimasi terlebih dahulu menggunakan algoritma *feature selection*.

## 2. METODE PENELITIAN

Pada bagian ini akan dijelaskan hal-hal yang mendasari penelitian dan bagaimana penelitian ini dilakukan.

### *Knowledge Discovery in Database*

Metode penelitian yang digunakan dalam penelitian ini mengikuti kaidah dasar dari *Knowledge Discovery in Database* atau *data mining* yang meliputi proses *selection*, data preprocessing, data transformation, *data mining*, dan *evaluation* (Fayyad, 1997). Berikut ini, adalah tahap-tahap dalam penelitian ini yang digambarkan dalam bentuk diagram.



Gambar 1 Kerangka Kerja Penelitian

### **Dataset**

Dataset yang digunakan adalah kumpulan data yang telah diolah berdasarkan laporan umur piutang. Normalisasi data digunakan juga dalam membentuk dataset sebelum diolah menggunakan algoritma *data mining*. Beberapa hal yang dilakukan untuk menormalisasi data seperti menghilangkan atribut nama karyawan, mengubah kategori karyawan menjadi bilangan biner menggunakan metode *one hot encoding*, menghilangkan atribut nama sales dan menggantinya dengan nomor 1 hingga 16, menghilangkan atribut tanggal faktur dan menggantinya dengan atribut bertipe numerik yang berisi selisih tanggal faktur dan tanggal laporan diambil, mengganti besaran nominal kredit pelanggan dan mengelompokkannya ke dalam kategori 0 hingga 3 menggunakan metode kuantil, dan menyimpan variabel-variabel target ke dalam 1 atribut saja.

### **Naïve Bayes**

*Naïve Bayes* adalah algoritma klasifikasi sederhana yang menggunakan aturan Bayes dengan asumsi yang kuat bahwa setiap atribut pada data dikondisikan independen berdasarkan kelas yang ditentukan. *K-Nearest Neighbors*. Algoritma *Naïve Bayes* sangat sering digunakan dalam pembentukan model data mining dikarenakan mudah digunakan dan akurasi yang tinggi. Tetapi disamping itu, kelemahan dari algoritma ini dikarenakan sifat algoritma ini yang menganggap bahwa setiap atribut independen, menyebabkan ketika korelasi antara atribut sangat rendah, maka menyebabkan algoritma ini akan kurang bekerja dengan baik. Secara umum, persamaan yang digunakan oleh algoritma ini merupakan persamaan *posterior*, yang diturunkan dari aturan *bayes* (Berrar, 2018). Persamaan posterior ditunjukkan oleh persamaan 1.

$$P(y|x) = P(y)P(x|y)/P(x) \quad (1)$$

Pada penelitian ini, Algoritma *Naïve Bayes* digunakan sebagai estimator untuk membangun model *data mining*.

### **K-Nearest Neighbors**

*K-Nearest Neighbors* adalah algoritma klasifikasi yang mengelompokkan data berdasarkan jarak dari kategori yang ada dalam ada dalam algoritma tersebut (Kramer, 2013). Kelebihan dari algoritma ini adalah algoritma ini bersifat *nonparametrik* artinya tidak mengasumsikan apapun mengenai distribusi *instance* di dalam dataset sehingga hasil prediksi yang dihasilkan sangat fleksibel. Proses pengolahan dataset yang terjadi menggunakan algoritma KNN, dapat dijelaskan sebagai berikut :

1. Menentukan parameter k (jumlah tetangga paling dekat)
2. Menghitung jarak Euclidean objek terhadap data training yang diberikan
3. Mengurutkan hasil yang didapatkan dari langkah nomor 2, dari nilai yang paling kecil hingga nilai yang paling besar
4. Membuat kategori baru dari kumpulan nilai di langkah 3, kemudian diklasifikasikan menjadi nilai *K* kelompok
5. Dengan menggunakan kategori *K-Nearest Neighbors* yang paling mayoritas maka dapat diprediksikan kategori objek.

Persamaan Jarak Euclidean dapat dilihat pada persamaan 2.

$$d(p,q) = d(q,p) = \sqrt{\sum_{f=1}^n (q_i - p_i)^2} \quad (2)$$

Pada penelitian ini, Algoritma *K-Nearest Neighbors* juga digunakan untuk membangun model *data mining*.

**Metode Elbow**

Metode *Elbow* digunakan untuk mencari jumlah klasifikasi *K* terbaik pada algoritma yang membutuhkan nilai tersebut, seperti Algoritma *K-Means*, *K-Nearest Neighbors*, dll. Metode *Elbow* dibuat dengan memetakan nilai *Root Mean Square Error* dari model data mining yang dibuat dari nilai-nilai *K* tertentu ke dalam grafik. Hasil nilai *K* paling optimal adalah, titik yang membentuk siku pertama, sebelum posisi titik terhadap sumbu *y* naik drastis yang menandakan tingkat *error* dari model yang telah dibuat meningkat (Mabayoje et al., 2019).

**Feature Selection**

*Feature Selection* pada penelitian ini digunakan untuk menyeleksi variabel fitur yang benar-benar berpengaruh. Variabel fitur tersebut akan diurutkan dari hasil perhitungan dengan distribusi *chi square* ( $\chi^2$ ) secara *descending*. Nantinya hasil variabel fitur yang diseleksi menggunakan metode ini, akan dikomparasikan model data miningnya dengan algoritma yang tidak menggunakan metode *feature selection*.

**Penghitungan Evaluasi**

Setelah model data mining selesai dibangun, langkah akhir yang dilakukan adalah mengevaluasi model data mining tersebut. Hal itu bertujuan untuk mengetahui model data mining yang dikembangkan sudah cukup baik atau tidak. Kriteria yang digunakan dalam penelitian ini adalah dengan menghitung nilai *accuracy*, *precision*, *recall* dan *f1 score*.

Nilai *accuracy* adalah rasio ketepatan hasil prediksi terhadap seluruh data target (Agarwal, 2014). Sedangkan *precision* adalah rasio ketepatan hasil prediksi positif terhadap jumlah data hasil prediksi positif. Lalu nilai *recall* adalah rasio ketepatan hasil prediksi positif terhadap data aktual. Dan *f1-score* adalah rasio yang menentukan kekuatan *precision* dan *recall*. Jika nilai *f1 score* mendekati nilai 1, artinya rasio *precision* dan *recall* sudah cukup baik (Buckland & Gey, 1994).

Nilai-nilai tersebut bisa lebih mudah dihitung, ketika hasil prediksi dipetakan ke dalam *confusion matrix*. Berikut ini adalah penggambaran dari *confusion matrix*.

Tabel 1 Confusion Matrix

		Nilai Aktual	
		True	False
Nilai Prediksi	True	True Positive	False Positive
	False	False Negative	True Negative

Tabel dari *confusion matrix* terdiri atas nilai *true positive*, *false positive*, *false negative*, dan *true negative*. Nilai-nilai tersebut dapat dijelaskan sebagai berikut.

1. *True Positive*, menjelaskan bahwa nilai yang diprediksi dalam kategori ini merupakan nilai positif dan nilai aktual juga positif.

2. *True Negative*, menjelaskan bahwa nilai yang diprediksi dalam kategori ini merupakan nilai negatif dan nilai aktual juga negatif
3. *False Positive*, menjelaskan bahwa nilai yang diprediksi dalam kategori ini positif sedangkan nilai aktualnya negatif.
4. *False Negative*, menjelaskan bahwa nilai yang diprediksi dalam kategori ini negative sedangkan nilai aktualnya positif.

Dan dari nilai-nilai ini, rasio accuracy, precision, recall dan f1-score dapat dihitung sebagai berikut.

$$1. \text{ Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$2. \text{ Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$3. \text{ Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$4. \text{ f1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

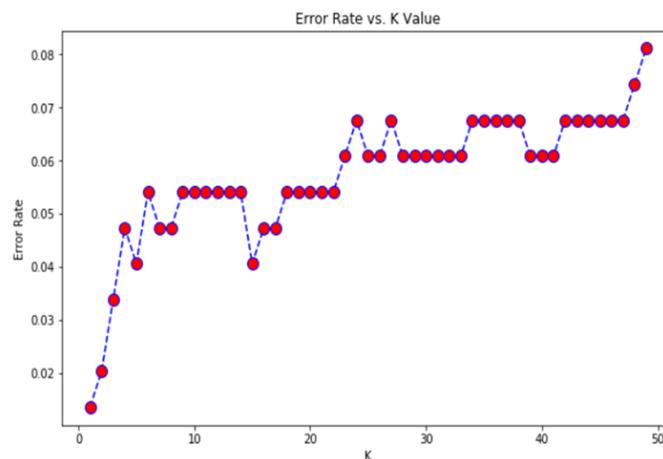
### 3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, yang akan dibahas adalah menentukan model data mining terbaik dalam memprediksi data umur piutang pelanggan. Sebagai langkah awal, yang dilakukan adalah mengumpulkan dataset terlebih dahulu. Contoh dataset yang akan digunakan dapat dilihat pada Gambar 2.

SINAR MEKAR OKI FARM	Perseoran	Purnomo	14-Jan-19	46935000	46935000	0	0	0	0	0
SINAR MEKAR OKI FARM	Perseoran	Purnomo	21-Jan-19	28495000	28495000	0	0	0	0	0
SINAR MEKAR OKI FARM	Perseoran	Purnomo	21-Jan-19	16800000	16800000	0	0	0	0	0
SINAR MEKAR OKI FARM	Perseoran	Purnomo	28-Jan-19	54805000	54805000	0	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	11-Dec-18	46800000	0	46800000	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	13-Dec-18	49600000	0	49600000	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	14-Dec-18	46800000	0	46800000	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	15-Dec-18	46800000	0	46800000	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	18-Dec-18	46800000	46800000	0	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	18-Dec-18	47790000	47790000	0	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	21-Dec-18	49700000	49700000	0	0	0	0	0
PSB FARM (Poncol Sejaht	Perseoran	Purnomo	24-Dec-18	49700000	49700000	0	0	0	0	0

Gambar 2 Contoh Dataset yang Akan Digunakan

Selanjutnya, pemodelan *data mining* yang dibuat menggunakan algoritma K-Nearest Neighbors memerlukan nilai K sebagai jumlah klasifikasi untuk algoritma tersebut. Pencarian nilai K optimal dilakukan dengan menggunakan Metode *Elbow*. Pemetaan nilai K menggunakan Metode Elbow dapat dilihat pada Gambar 3.

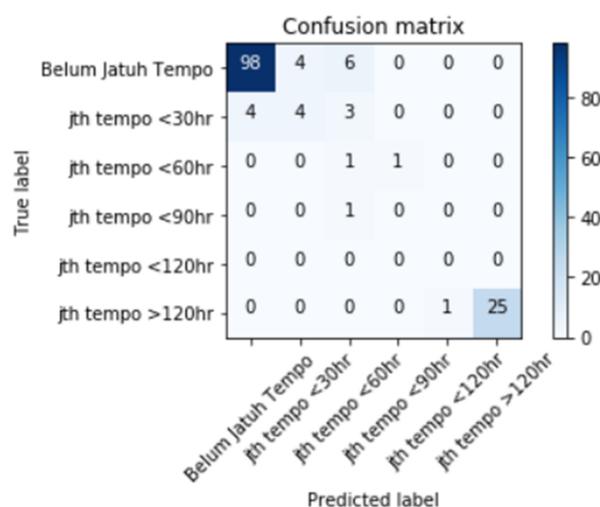


Gambar 3 Metode Elbow

Terlihat pada Gambar 3, bahwa titik terendah berada di sekitar nilai  $K = 2$ , sehingga bisa diasumsikan bahwa nilai  $K$  paling optimal yang digunakan untuk membuat model data mining menggunakan K-Nearest Neighbors adalah dengan nilai  $K$  adalah 2.

Pada proses ini, hasil pemodelan data mining menggunakan *Naïve Bayes* dan *KNN* maupun variabel fitur dioptimasi terlebih dahulu menggunakan Algoritma *Feature Selection* baru dibangun model data mining menggunakan kedua algoritma tersebut telah selesai dilakukan. Sehingga hal yang akan fokus dibahas pada bagian ini adalah mengevaluasi model data mining.

Langkah pertama yang dilakukan pada proses evaluasi adalah menentukan confusion matrix untuk memetakan hasil prediksi dari model data mining dan menghitung nilai *accuracy*, *precision*, *recall* dan *f1-score*. Pada penulisan ini, contoh yang akan diberikan untuk penggambaran confusion matrix dan perhitungan nilai evaluasi menggunakan model data mining yang dibangun menggunakan algoritma *Naïve Bayes*. Confusion Matrix yang dihasilkan dari model data mining tersebut dapat dilihat pada gambar 4.



Gambar 4 Confusion Matrix pada Model Naive Bayes

Terlihat bahwa dari 148 data *testing* yang diprediksi pada model *Naïve Bayes*, bahwa Kategori “belum jatuh tempo” berhasil tepat diprediksi sebesar 98 data. Lalu kategori “jatuh

## *Template Jurnal Edunomika*

tempo < 30hari” berhasil tepat diprediksi sebesar 4 data. Lalu kategori “jatuh tempo < 60 hari” berhasil tepat diprediksi sebesar 1 data. Lalu kateogri “jatuh tempo < 90 hari” tidak ada data yang tepat di prediksi, dikarenakan model data mining ini memprediksi 1 data pada kategori “jatuh tempo < 60 hari” yang seharusnya diprediksi untuk kategori “jatuh tempo < 90 hari”. Untuk kategori “jatuh tempo < 120 hari” tidak ada juga hasil yang diprediksi dan untuk kategori “jatuh tempo >120 hari” data yang diprediksi tepat sebanyak 25 data. Sehingga data yang tepat di prediksi sebanyak 126 data dari 148 data *testing*.

Untuk perhitungan nilai-nilai evaluasi telah bisa dihitung dari confusion matrix tersebut. Berikut ini adalah hasil perhitungannya.

$$Accuracy = (98 + 4 + 1 + 0 + 0 + 25) / 148 = 0,864864865$$

Untuk kategori belum jatuh tempo, nilai precision, recall dan f1-score adalah sebagai berikut.

$$Precision = 98 / (98 + 4 + 0 + 0 + 0 + 0) = 0,960784314$$

$$Recall = 98 / (98 + 4 + 6 + 0 + 0 + 0) = 0,907407407$$

$$f1-score = 2 \times 0,9607 \times 0,9074 / (0,9607 + 0,9074) = 0,93$$

Untuk kategori jatuh tempo < 30, nilai precision, recall dan f1-score adalah sebagai berikut

$$Precision = 4 / (4 + 4 + 0 + 0 + 0 + 0) = 0,5$$

$$Recall = 4 / (4 + 4 + 3) = 0,36$$

$$f1-score = 2 \times 0,5 \times 0,36 / (0,5 + 0,36) = 0,42$$

Untuk kategori jatuh tempo < 60, nilai precision, recall dan f1-score adalah sebagai berikut.

$$Precision = 1 / (6 + 3 + 1 + 1 + 0 + 0) = 0,09$$

$$Recall = 1 / (0 + 0 + 1 + 1 + 0 + 0) = 0,5$$

$$f1-score = 2 \times 0,09 \times 0,5 / (0,09 + 0,5) = 0,15$$

Untuk kategori jatuh tempo < 90, nilai precision, recall dan f1-score adalah sebagai berikut.

$$Precision = 0 / (0 + 0 + 1 + 0 + 0 + 0) = 0$$

$$Recall = 0 / (0 + 0 + 1 + 0 + 0 + 0) = 0$$

$$f1-score = 0$$

Untuk kategori jatuh tempo < 120, nilai precision, recall dan f1-score adalah sebagai berikut.

$$Precision = 0 / (0 + 0 + 0 + 0 + 0 + 1) = 0$$

$$Recall = 0$$

$$f1-score = 0$$

Untuk kategori jatuh tempo > 120, nilai precision, recall dan f1-score adalah sebagai berikut.

$$Precision = 25 / (0 + 0 + 0 + 0 + 0 + 25) = 1$$

$$Recall = 25 / (0 + 0 + 0 + 0 + 1 + 25) = 0,96$$

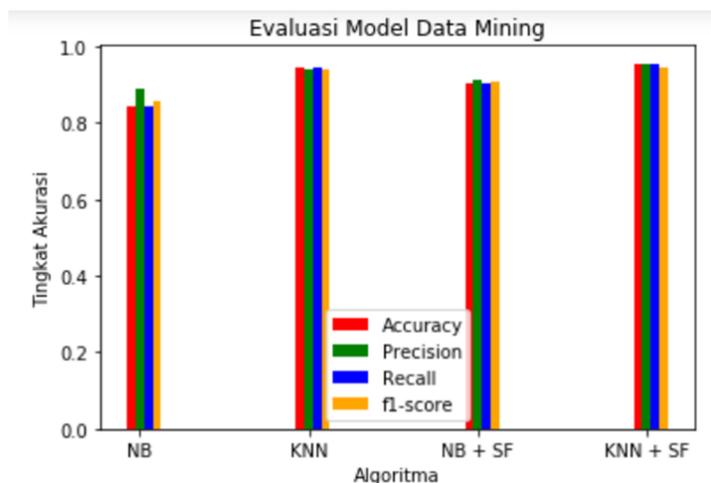
$$f1-score = 2 \times 1 \times 0,96 / (1 + 0,96) = 0,98$$

Sehingga nilai-nilai tersebut dapat dipetakan seperti yang terlihat pada Gambar 5.

	precision	recall	f1-score	support
Belum Jatuh Tempo	0.96	0.98	0.97	113
jth tempo <30hr	0.80	0.62	0.70	13
jth tempo <90hr	0.67	1.00	0.80	2
jth tempo <120hr	0.00	0.00	0.00	1
jth tempo >120hr	1.00	1.00	1.00	19
accuracy			0.95	148
macro avg	0.68	0.72	0.69	148
weighted avg	0.94	0.95	0.94	148

Gambar 5 Perhitungan nilai Evaluasi

Secara keseluruhan, untuk 4 model yang dibuat dapat dipetakan nilainya dalam grafik pada Gambar 6.



Gambar 6 Evaluasi Model Data Mining

Terlihat pada Gambar 6, berdasarkan skala 0 hingga 1, jika model data mining yang dibangun menggunakan Algoritma *Naïve Bayes* saja menghasilkan nilai *accuracy*, *precision*, *recall* dan *f1-score* sebesar 0,86, 0,89, 0,85 dan 0,86. Untuk model data mining yang dibangun menggunakan Algoritma *K-Nearest Neighbors* saja menghasilkan nilai *accuracy*, *precision*, *recall* dan *f1-score* sebesar 0,95, 0,94, 0,95 dan 0,94. Untuk model data mining yang dibangun dengan menggunakan Algoritma Feature Selection untuk memilih variabel fitur terbaik, kemudian menggunakan model *Naïve Bayes* untuk membangun model data miningnya, nilai *accuracy*, *precision*, *recall* dan *f1-score* yang dihasilkan sebesar 0,89, 0,91, 0,91 dan 0,91. Dan model yang dihasilkan dengan mengkombinasikan Algoritma *K-Nearest Neighbors* dengan Algoritma *Selection Features*, nilai *accuracy*, *precision*, *recall* dan *f1-score* yang dihasilkan sebesar 0,99, 0,96, 0,95 dan 0,95.

#### 4. KESIMPULAN

Penelitian ini bertujuan untuk memprediksi hasil klasifikasi umur piutang yang ada di PT. Citrafeed. Dikarenakan pemilihan pelanggan baru pada perusahaan tersebut masih dipilih oleh karyawan pemasaran, sehingga penelitian bisa dijadikan sebagai evaluasi terhadap laporan piutang pelanggan.

Model data mining yang dibangun pada penelitian ini, menggunakan Algoritma *Naïve Bayes* dan *K-Nearest Neighbors* serta menggunakan Algoritma Feature Selection untuk

menyeleksi fitur sebelum dibangun menjadi model data mining. Hasil yang didapatkan ketika membandingkan 4 model yang dibangun, bahwa model data mining yang dibangun menggunakan Algoritma K-Nearest Neighbors merupakan algoritma yang lebih baik digunakan untuk menyelesaikan permasalahan klasifikasi piutang pelanggan pada penelitian ini. Hal itu didasarkan oleh nilai accuracy, rata-rata precision, rata-rata recall, dan f1-score yang lebih tinggi dibandingkan algoritma Naïve Bayes. Ketika variabel fitur dioptimasi lagi dengan feature selection, rasio-rasio evaluasi menjadi lebih tinggi lagi. Sehingga algoritma feature selection sangat berguna untuk membuat model data mining yang lebih baik.

Beberapa saran yang penulis bisa berikan, agar bisa menyempurnakan penelitian ini dan menjadi bahan penelitian selanjutnya, diantaranya:

1. Pengembangan model data mining yang dilakukan masih dibuat dengan menggunakan 1 variabel target. Laporan umur piutang aslinya memisahkan klasifikasi umur piutang ke dalam beberapa kolom. Jika model data mining dapat dikembangkan dengan menggunakan banyak variabel target, mungkin model data mining yang dibangun lebih baik lagi.
2. Pengembangan model data mining menggunakan algoritma selain Naïve Bayes dan K-Nearest Neighbors mungkin saja merupakan model data mining yang lebih baik, sehingga penelitian menggunakan algoritma lainnya dalam memprediksi data piutang pelanggan sangat dianjurkan.
3. Pembagian dataset yang lebih beragam akan memperoleh hasil akurasi yang lebih beragam, Sehingga didapatkan model data mining yang lebih baik.
4. PT. Citrafeed dapat mengimplementasikan hasil penelitian ini ke dalam sistem informasi yang perusahaan tersebut gunakan, sehingga menjadi fitur untuk menentukan pelanggan baru yang lebih baik.

#### **DAFTAR PUSTAKA**

- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. In *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>
- Berrar, D. (2018). Bayes' theorem and naive bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(2018), 403–412. <https://doi.org/10.1016/B978-0-12-809633-8.20473-1>
- Buckland, M., & Gey, F. (1994). The relationship between Recall and Precision. *Journal of the American Society for Information Science*, 45(1), 12–19. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1<12::AID-ASIS>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<12::AID-ASIS>3.0.CO;2-L)
- Fayyad, U. (1997). Data mining and knowledge discovery in databases: Implications for scientific databases. *Scientific and Statistical Database Management - Proceedings of the International Working Conference*, 2–11. <https://doi.org/10.1109/ssdm.1997.621141>
- Gultom, F., & Simanjuntak, T. (2020). Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Dengan Menggunakan Algoritma Naïve Bayes Dan K-Nearest Neighbor. *METHOMIKA: Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi*, 4(1), 98–102. <https://doi.org/10.46880/jmika.v4i2.158>

- Hasan, M. (2017). Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma Naïve Bayes Berbasis Forward Selection. *ILKOM Jurnal Ilmiah*, 9(3), 317–324. <https://doi.org/10.33096/ilkom.v9i3.163.317-324>
- Kramer, O. (2013). Dimensionality Reduction with Unsupervised Nearest Neighbors. *Intelligent Systems Reference Library*, 51, 13–23. <https://doi.org/10.1007/978-3-642-38652-7>
- Mabayoje, M. A., Balogun, A. O., Jibril, H. A., Atoyebi, J. O., Mojeed, H. A., & Adeyemo, V. E. (2019). Parameter tuning in KNN for software defect prediction: an empirical analysis. *Jurnal Teknologi Dan Sistem Komputer*, 7(4), 121–126. <https://doi.org/10.14710/jtsiskom.7.4.2019.121-126>
- Mardiana, T. (2018). Optimasi Naïve Bayes Dengan Particle Swarm Optimization Dan Stratified Untuk Prediksi Kredit Macet Pada Koperasi. *Jurnal Riset Informatika*, 1(1), 43–50. <https://doi.org/10.34288/jri.v1i1.13>