Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 6, Nomor 1, Oktober 2021
http://doi.org/10.33395/remik.v4i1.11196
e-ISSN : 2541-1330
p-ISSN : 2541-1332

Prediksi Customer Churn Pada Bisnis Retail Menggunakan Algoritma Naïve Bayes

[1]Firmansyah, [2] Agus Yulianto
Universitas Bina Sarana Informatika
Jakarta, Indonesia
Universitas Nusa Mandiri
Jakarta, Indonesia

firmansyah.fmh@bsi.ac.id, agus.aag@nusamandiri.ac.id

*Penulis Korespondensi
Diajukan : 25/10/2021
Diterima : 29/10/2021
Dipublikasi : 02/10/2021

ABSTRAK

Perusahaan retail seperti toko dan sejenisnya pada umumnya memiliki data customer yang menjadi anggota belanja dengan tujuan untuk menjalin hubungan yang saling menguntungkan antara customer dan toko, namun data customer seringkali tidak dikelola dan dianalisa dengan baik sehingga menyebabkan customer retail berhenti belanja di toko tersebut atau berpindah ke toko lain. Untuk dapat memprediksi kehilangan pelanggan, dibutuhkan model algoritma yang dapat memprediksi customer churn dengan akurat dan presisi. Untuk melakukan prediksi, maka dibuat model algoritma dengan naïve bayes dengan tujuan untuk memprediksi kemungkinan customer churn sehingga toko dapat melakukan tindakan preventif. Data customer diambil dengan klasifikasi yaitu jenis kelamin, grade point, kepemilikan kartu kredit, rentang usia, nilai rata-rata transaksi dan lama menjadi anggota. Data training dan testing dimasukkan ke dalam pemodelan naïve bayes dengan menghasilkan akurasi 80%. Naïve bayes terbukti dapat memprediksi kemungkinan customer churn berdasarkan data dan klasifikasi sehingga hasilnya dapat menjadi pendukung keputusan bisnis.

Kata Kunci: customer churn, data mining, machine learning, naïve bayes, prediction, retail

I. PENDAHULUAN

Retail adalah salah satu rangkaian aktivitas bisnis untuk menambah nilai guna barang dan jasa yang dijual kepada konsumen untuk konsumsi pribadi atau rumah tangga (Levy & Weitz, 2012). Dalam bisnis retail tentunya banyak jenisnya, seperti toko, mini market, super market hingga hyper market. Setiap produk yang dijual juga variatif mulai dari kebutuhan sehari-hari, kebutuhan rumah dan bisnis retail yang fokus pada kebutuhan tertentu. Peningkatan pelanggan merupakan salah satu faktor mendongkrak bisnis retail, sehingga perusahaan perlu mempertahankan customer untuk menjaga kelangsungan bisnis.

Pengelolaan customer bersumber dari data transaksi harian namun pengelolaan customer di perusahaan retail berbeda dengan perusahaan di segmen lainnya, pada perusahaan retail, customer perlu terdaftar sebagai anggota. Untuk terdaftar sebagai anggota klub belanja, customer dapat didaftarkan terlebih dahulu oleh toko untuk kemudian diberikan kartu anggota atau customer mendaftarkan diri di aplikasi yang disediakan oleh toko. Dengan menjadi anggota belanja, maka customer diuntungkan dengan beberapa penawaran seperti poin belanja, diskon dan promosi.

Customer yang loyal sering kali tidak lagi aktif melakukan belanja di toko langganannya disebabkan karena perusahaan retail tidak melakukan prediksi untuk mengantisipasi kehilangan pelanggan (customer churn). Data yang semakin besar dan kompleks menyulitkan dalam melakukan analisa dan prediksi terhadap data customer. Pemecahan masalahnya adalah dengan membuat model menggunakan naïve bayes untuk dapat memprediksi *customer churn*.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 6, Nomor 1, Oktober 2021 e-ISSN : 2541-1330
http://doi.org/10.33395/remik.v4i1.11196 p-ISSN : 2541-1332

Tujuan penelitian ini adalah untuk membuat model prediksi menggunakan naïve bayes dalam memprediksi customer churn sehingga perusahaan retail dapat memprediksi lebih dini potensi kehilangan pelanggan. Dengan prediksi customer churn juga perusahaan dapat meningkatkan efisiensi promosi dan mengurangi biaya yang terkait dengan customer churn (Liu & Zhuang, 2015)

II. STUDI LITERATUR

Penelitian Terdahulu

1. Penelitian Terkait

Penelitian menggunakan metode hibrida untuk memodelkan prediksi customer churn yaitu algoritma C5.0 dan algoritma Lolimot. Model digunakan untuk memprediksi customer churn di perusahaan telelomunikasi (Jamalian & Foukerdi, 2018). Membuat prediksi customer churn menggunakan machine learning dengan jumlah data berskala besar yang diambil dari Kaggle. Metode yang digunakan yaitu Natural Language Processing (NLP) (Jinde & Amit Savyanavar, 2020). Penelitian mengenai e-commerce, data yang diambil yaitu data transksi e-commerce berjumlah 626.275 dan 13 kolom data. Pemodelan menggunanakan multi layer perceptron (MLP) untuk memprediksi customer churn pada e-commerce (Pondel et al., 2021). Melakukan pemodelan menggunakan regresi logistic dan logit boost untuk memprediksi customer churn perusahaan telelomunikasi (Jain et al., 2020). Membuat pemodelan prediksi customer churn pada customer bank menggunakan algoritma Random Forest (Verma, 2020). Membuat prediksi menggunakan Artificial Neural Network untuk memprediksi customer churn, data yang digunakan sebagai data set yaitu data CRM (Customer Relationship Management) (Seyed et al., 2019). Dalam bidang perawatan kesehatan, penelitian memprediksi customer churn dengan model retrospective dengan neural network (Kwon et al., 2021).

2. Customer Churn

Customer churn atau kehilangan pelanggan merupakan istilah yang digunakan dalam bisnis untuk menyebut kehilangan pelanggan atau terputusnya hubungan antara customer dengan pemilik bisnis. Dalam industri teknologi informasi, customer churn mengacu kepada customer yang meninggalkan bisnis untuk berpindah kepada pesaing bisnis (Ahmed, 2019). Ada beberapa faktor yang mendorong customer meninggalkan bisnis atau produk tertentu (Zhao et al., 2021), yaitu :

1. Faktor harga
   Customer ingin membeli di toko dengan harga yang kompetitif dan biaya yang dikeluarkan sesuai dengan ekspektasi customer.
2. Faktor produk
   Faktor ini dapat dipengaruhi oleh kualitas seperti cacat pada produk yang dijual atau produk tidak sesuai dengan kebutuhan customer.
3. Faktor customer
   Yang mempengaruhi dalam faktor ini seperti tingkat konsumsi dan pendapatan dari customer dapat berdampak pada loyalitas.

3. Naïve Bayes

Naïve Bayes adalah salah satu metode dalam Data Mining, dimana data mining merupakan proses mengekstrak data dengan menggunakan metode dan algoritma tertentu untuk menghasilkan informasi yang lebih berguna sehingga dapat menjadi dasar dalam pengambilan keputusan (Firmansyah & Yulianto, 2021). Naïve Bayes merupakan teorema yang diperkenalkan oleh ilmuwan inggris, Thomas Bayes. Langkah-langkah dalam teorema bayes yaitu :

1. Menghitung jumlah kelas
2. Menghitung jumlah kelas per kelas
3. Menjumlahkan (mengkalikan) semua variable kelas
4. Bandingkan hasil per kelas
   Persamaan dari teorema bayes, yaitu :

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 6, Nomor 1, Oktober 2021
http://doi.org/10.33395/remik.v4i1.11196
e-ISSN : 2541-1330
p-ISSN : 2541-1332

$$P(C|X) = \frac{P(x|c)P(c)}{P(x)}$$

Dimana :
x : data untuk class yang belum diketahui
c : hipotesis data
P(c|x) : Probabilitas hipotesis berdasarkan kondisi (posteriory probability)
P(c) : Probabilitas hipotesis (prior probability)
P(x|c) : Probabilitas berdasakan kondisi pada hipotesis

Dalam teorema bayes, digunakan asumsi indepedensi yang tinggi (naif) bahwa masing-masing petunjuk saling independen, persamaannya yaitu :

$$P(c|X1, \dots, Xn) = P(C) \prod_{i=1}^{n} P(X1|C)$$

III. METODE

A. CRISP-DM (Cross-Industry Standard Process for Data Mining)

Metode yang digunakan untuk implementasi data mining menggunakan kerangka kerja CRISP-DM (Cross-Industry Standard Process for Data Mining) yaitu standar metode implementasi data mining untuk industri (Larose, 2006), tahapnya yaitu :

1. Business Understanding Phase
2. Data Understanding Phase
3. Data Preparation Phase
4. Modelling Phase
5. Evaluation Phase
6. Deployment Phase

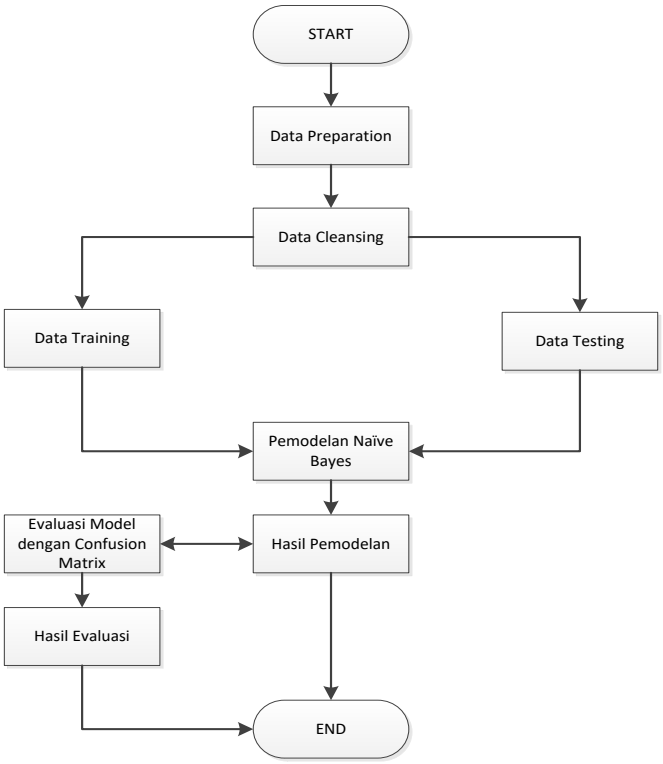
Gambar 1 CRISP-DM Process

B. Flowchart Pemodelan Data Mining

Flowchart atau Diagram alur digunakan untuk menggambarkan alur proses mulai dari dari preparation hingga hasil pemodelan. Diagram alur untuk menggambarkan proses pemodelan menggunakan naïve bayes ini seperti di bawah ini :

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 6, Nomor 1, Oktober 2021
http://doi.org/10.33395/remik.v4i1.11196
e-ISSN : 2541-1330
p-ISSN : 2541-1332



Gambar 2 Proses Pemodelan Naïve Bayes

IV. PEMBAHASAN DAN HASIL

Data yang digunakan untuk pemodelan bersumber dari data keanggotaan toko retail di 10 toko yang diambil secara acak (random sampling). Data keanggotaan merupakan data customer yang memiliki kartu belanja yang dapat digunakan untuk pengumpulan poin di toko. Total data awal yaitu berjumlah 258 baris dengan 8 atribut, kemudian dilakukan cleansing data sehingga data menjadi 123 baris dengan 6 atribut yaitu jenis kelamin, grade point, kepemilikan kartu kredit, rentang usia, nilai rata-rata transaksi, lama member dan class churn.

A. Data Training dan Data Testing

Data training berjumlah 123 baris dan 6 atribut dengan klasifikasi seperti di bawah ini :

Tabel 1. Klasifikasi variabel

Variable name	Class
Jenis kelamin	1) Laki-laki 2) Perempuan
Grade point	1) Low 2) Middle 3) High
Memiliki kartu kredit	1) 1 (Yes) 2) 0 (No)
Rentang usia	1) 0-20 2) 21-40 3) 41-60 4) >60
Nilai rata-rata transaksi	1) Low

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 6, Nomor 1, Oktober 2021
http://doi.org/10.33395/remik.v4i1.11196
e-ISSN : 2541-1330
p-ISSN : 2541-1332

	2) Middle 3) High
Lama member	1) <=3 2) <=6 3) >6
Class Churn	Yes/No

Dari data transaksi yang sudah diklasifikasikan, data training diambil secara acak sebesar 90% dan data testing sebesar 10%.

B. Algoritma Naïve Bayes

Dari data training kemudian dimasukkan ke dalam model algoritma naïve bayes sehingga menghasilkan model prediksi customer churn. Untuk menguji algoritma naïve bayes maka dimasukkan data testing ke dalam model prediksi, di bawah ini adalah langkah-langkah naïve bayes menggunakan 1 data testing dari 5 data testing sebagai sampel yaitu jika jenis kelamin=perempuan, point=low, memiliki kartu kredit=1, rentang usia=0-20, rata-rata transaksi=middle, lama member= <=6.

1. Menghitung probabilitas class/label
   P(Y=Yes)=44/118=0.372
   P(Y=No)=74/118=0.627

2. Menghitung jumlah kasus yang sama dengan class yang sama
   a. Menghitung probabilitas class jenis kelamin
      P(jenis kelamin=perempuan | Y=yes) = 28/46
      P(jenis kelamin=perempuan | Y=no) = 59/77
   b. Menghitung probabilitas class grade point
      P(grade point=low | Y=yes) = 15/46
      P(grade point =low | Y=no) = 22/77
   c. Menghitung probabilitas class memiliki kartu kredit
      P(memiliki kartu kredit=1 | Y=yes) = 19/46
      P(memiliki kartu kredit=1 | Y=no) = 27/77
   d. Menghitung probabilitas class rentang usia
      P(rentang usia=0-20 | Y=yes) = 7/46
      P(rentang usia=0-20 | Y=no) = 10/77
   e. Menghitung probabilitas class nilai rata-rata transaksi
      P(rata-rata transaksi=middle | Y=yes) = 13/46
      P(rata-rata transaksi=middle | Y=no) = 24/77
   f. Menghitung probabilitas class lama member
      P(lama member=<=6 | Y=yes) = 7/46
      P(lama member=<=6 | Y=no) = 29/77

3. Menjumlahkan semua variable Yes dan No
   P(jenis kelamin=perempuan | Y=yes) * P(grade point=low | Y=yes) *
   P(memiliki kartu kredit=1 | Y=yes) * P(rentang usia=0-20 | Y=yes) *
   P(rentang usia=0-20 | Y=yes) * P(rata-rata transaksi=middle | Y=yes) *
   P(lama member=<=6 | Y=yes).
   (28/46) * (15/46) * (19/46) * (7/46) * (13/46) * (7/46) =
   0,6086 * 0,3260 * 0,4130 * 0,1521 * 0,2826 * 0,1521 = 0,000535
   P(jenis kelamin=perempuan | Y=no) * P(grade point=low | Y=no) *
   P(memiliki kartu kredit=1 | Y=no) * P(rentang usia=0-20 | Y=no) *
   P(rentang usia=0-20 | Y=no) * P(rata-rata transaksi=middle | Y=no) *
   P(lama member=<=6 | Y=no)
   (59/77) * (22/77) * (27/77) * (10/77) * (24/77) * (29/77) =
   0,7662 * 0,2857 * 0,3506 * 0,1298 * 0,3116 * 0,3766 = 0.0011

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Riset dan E-Jurnal Manajemen Informatika Komputer
Volume 6, Nomor 1, Oktober 2021
http://doi.org/10.33395/remik.v4i1.11196
e-ISSN : 2541-1330
p-ISSN : 2541-1332

Nilai probabilitas tertinggi yaitu pada kelas No, sehingga kesimpulannya adalah customer akan masuk ke dalam kelas No atau tidak churn.

Akurasi data diukur dengan menggunakan confusion matrix dan rapid miner, hasil akurasi mencapai 80% dan precision mencapai 100%.

accuracy: 80.00%

	true No	true Yes	class precision
pred. No	3	1	75.00%
pred. Yes	0	1	100.00%
class recall	100.00%	50.00%	

Gambar 3 Accuracy

precision: 100.00% (positive class: Yes)

	true No	true Yes	class precision
pred. No	3	1	75.00%
pred. Yes	0	1	100.00%
class recall	100.00%	50.00%	

Gambar 4 Precision

Untuk grafik AUC (Area Under Curve) seperti di bawah ini :

Gambar 5 Grafik ROC

V. KESIMPULAN

Dari pemodelan naïve bayes untuk memprediksi customer churn memperoleh hasil akurasi mencapai 80% dan presisi sebesar 100% sehingga pemodelan ini dapat diterapkan untuk prediksi dengan sumber data yang sudah diklasifikasikan. Peningkatan jumlah data training tentunya akan mempengaruhi nilai akurasi dari pemodelan naïve bayes sehingga perlu dilakukan pemodelan hibrida menggunakan algoritma lain untuk meningkatkan akurasi seperti Bayesian network atau C5.0. Pemodelan naïve bayes juga hanya cocok untuk data yang sudah diklasifikasikan, untuk data numerical dapat menggunakan algoritma Bayesian network.

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

e-ISSN : 2541-1330
p-ISSN : 2541-1332

VI. REFERENSI

Ahmed, H. M. (2019). The Impact of Customer Churn Factors (CCF) on Customer's Loyalty: The Case of Telecommunication Service Providers in Egypt. *International Journal of Customer Relationship Marketing and Management*, *10*(1), 48–70. https://doi.org/10.4018/IJCRMM.2019010104

Firmansyah, & Yulianto, A. (2021). Market Basket Analysis for Books Sales Promotion using FP Growth Algorithm, Case Study : Gramedia Matraman Jakarta. *Journal of Informatics Anf Telecommunication Engineering*, *4*(January), 383–392.

Jain, H., Khunteta, A., & Srivastava, S. (2020). Churn Prediction in Telecommunication using Logistic Regression and Logit Boost. *Procedia Computer Science*, *167*, 101–112. https://doi.org/10.1016/j.procs.2020.03.187

Jamalian, E., & Foukerdi, R. (2018). A Hybrid Data Mining Method for Customer Churn Prediction. In *Technology & Applied Science Research* (Vol. 8, Issue 3). www.etasr.com

Jinde, V., & Amit Savyanavar, P. (2020). Customer Churn Prediction System using Machine Learning. *International Journal of Advanced Science and Technology*, *29*(5), 7957–7964.

Kwon, H., Kim, H. H., An, J., Lee, J. H., & Park, Y. R. (2021). Lifelog Data-based Prediction Model of Digital Health Care App Customer Churn: Retrospective Observational Study. *Journal of Medical Internet Research*, *23*(1). https://doi.org/10.2196/22184

Larose, D. T. (2006). *Data Mining Methods and Models*. Johns Wiley& Sons.

Levy, M., & Weitz, B. A. (2012). *Retailing Management Information Center*. New York: McGraw Hill Higher Education.

Liu, Y., & Zhuang, Y. (2015). Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data. *Journal of Computer and Communications*, *03*(06), 87–93.

Pondel, M., Wuczyński, M., Gryncewicz, W., Łysik, Ł., Hernes, M., Rot, A., & Kozina, A. (2021). Deep Learning for Customer Churn Prediction in E-Commerce Decision Support. *Business Information Systems*, 3–12. https://doi.org/10.52825/bis.v1i.42

Seyed, H., Iranmanesh, M., Hamid, M., Bastan, H., Shakouri, G., & Nasiri, M. M. (2019). Customer Churn Prediction Using Artificial Neural Network: An Analytical CRM Application. *Proceedings of the International Conference on Industrial Engineering and Operations Management*, *July 23-26*, 2214–2269.

Verma, P. (2020). Churn prediction for savings bank customers: A machine learning approach. *Journal of Statistics Applications and Probability*, *9*(3), 535–547. https://doi.org/10.18576/JSAP/090310

Zhao, M., Zeng, Q., Chang, M., Tong, Q., & Su, J. (2021). A Prediction Model of Customer Churn considering Customer Value: An Empirical Research of Telecom Industry in China. *Discrete Dynamics in Nature and Society*, *2021*. https://doi.org/10.1155/2021/7160527

This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.