

Volume 4 Nomor 3 September 2019

INFORMASI INTERAKTIF

JURNAL INFORMATIKA DAN TEKNOLOGI INFORMASI

PROGRAM STUDI TEKNIK INFORMATIKA – FAKULTAS TEKNIK -UNIVERSITAS JANABADRA

SISTEM PENDUKUNG KEPUTUSAN PENERIMA DANA DESA DENGAN MENGGUNAKAN METODE PROMETHEE DI KECAMATAN GODEAN KABUPATEN SLEMAN

Arif Budiman, Adi Prasetyo, M Hamzah

IMPLEMENTASI DETEKSI TEPI MENGGUNAKAN METODE *QUADRANT TREE CLASSIFIER* PADA PEMISAHAN OBJEK BERBASIS *DIGITAL IMAGE PROCESSING* (STUDI KASUS : OBJEK BENDERA NEGARA)

Azriel Christian Nurcahyo, Vera Wati, Dwindi Etika Profesi, Kusri

PEMBERDAYAAN KELOMPOK IBU-IBU RUMAH TANGGA MELALUI PENGEMBANGAN USAHA ANEKA KUE BERBAHAN TALAS DI KECAMATAN NGAGLIK KABUPATEN SLEMAN

Rini Raharti, Cungki Kusdarjito, Bambang Jatmiko, Puji Lestari, Siti Qomariyah

IMPLEMENTASI METODE K-NEAREST NEIGHBOR DAN REGRESI LINEAR DALAM PREDIKSI HARGA EMAS

Prabowo Budi Utomo, Ema Utami, Suwanto Raharjo

PREDIKSI KETERLAMBATAN PEMBAYARAN SPP SEKOLAH DENGAN *METODE K-NEAREST NEIGHBOR* (STUDI KASUS SMK AL-ISLAM SURAKARTA)

Robi Wariyanto Abdullah, Kusri, Emha Taufiq Luthfi

PREDIKSI CUSTOMER CHURN PERUSAHAAN TELEKOMUNIKASI MENGGUNAKAN *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR*

Kaharudin, Musthofa Galih Pradana, Kusri

PERENCANAAN RENCANA STRATEGIS SISTEM INFORMASI UNTUK TERCAPAINYA VISI MISI DARI RUMAH SAKIT

Siska Febriani, Bambang Soedijono, M. Rudyanto Arief

EVALUASI *USER INTERFACE* PADA APLIKASI E-COMMERCE (STUDI KASUS INFORMA & IKEA)

Sundari Ilkham, Kusri, M. Rudyanto Arief

PROGRESSIVE WEB APPS UNTUK REKAYASA HYBRID APPLICATION BERBASIS TEKNOLOGI *MEAN STACK*

Rusdy Agustaf, Bernard Renaldy Suteja

PENERAPAN METODE AHP DALAM PENENTUAN KRITERIA SISTEM PENDUKUNG KEPUTUSAN SELEKSI PENJUAL PADA KANTIN

Fandli Supandi, Kusri, Hanif Al Fatta



INFORMASI
INTERAKTIF

Vol. 4

No. 3

Hal. 131- 200

Yogyakarta
September
2019

ISSN
2527-5240

DEWAN EDITORIAL

- Penerbit** : Program Studi Teknik Informatika Fakultas Teknik Universitas Janabadra
- Ketua Penyunting
(Editor in Chief)** : Fatsyahrina Fitriastuti, S.Si., M.T. (Universitas Janabadra)
- Penyunting (Editor)** : 1. Selo, S.T., M.T., M.Sc., Ph.D. (Universitas Gajah Mada)
2. Dr. Kusriani, S.Kom., M.Kom. (Universitas Amikom Yogyakarta)
3. Jemmy Edwin B, S.Kom., M.Eng. (Universitas Janabadra)
4. Ryan Ari Setyawan, S.Kom., M.Eng. (Universitas Janabadra)
5. Yumarlin MZ, S.Kom., M.Pd., M.Kom. (Universitas Janabadra)
- Alamat Redaksi** : Program Studi Teknik Informatika Fakultas Teknik
Universitas Janabadra
Jl. Tentara Rakyat Mataram No. 55-57
Yogyakarta 55231
Telp./Fax : (0274) 543676
E-mail: informasi.interaktif@janabadra.ac.id
Website : <http://e-journal.janabadra.ac.id/>
- Frekuensi Terbit** : 3 kali setahun

JURNAL INFORMASI INTERAKTIF merupakan media komunikasi hasil penelitian, studi kasus, dan ulasan ilmiah bagi ilmuwan dan praktisi dibidang Teknik Informatika. Diterbitkan oleh Program Studi Teknik Informatika Fakultas Teknik Universitas Janabadra di Yogyakarta, tiga kali setahun pada bulan Januari, Mei dan September.

DAFTAR ISI

	<i>halaman</i>
Sistem Pendukung Keputusan Penerima Dana Desa dengan Menggunakan Metode Promethee di Kecamatan Godean Kabupaten Sleman Arif Budiman, Adi Prasetyo, M Hamzah	131-137
Implementasi Deteksi Tepi Menggunakan Metode <i>Quadrant Tree Classifier</i> pada Pemisahan Objek Berbasis <i>Digital Image Processing</i> (Studi Kasus : Objek Bendera Negara) Azriel Christian Nurcahyo, Vera Wati, Dwinda Etika Profesi, Kusrini	138-148
Pemberdayaan Kelompok Ibu-Ibu Rumah Tangga Melalui Pengembangan Usaha Aneka Kue Berbahan Talas di Kecamatan Ngaglik Kabupaten Sleman Rini Raharti, Cungki Kusdarjito, Bambang Jatmiko, Puji Lestari, Siti Qomariyah	149-154
Implementasi Metode K-Nearest Neighbor dan Regresi Linear Dalam Prediksi Harga Emas Prabowo Budi Utomo, Ema Utami, Suwanto Raharjo	155-159
Prediksi Keterlambatan Pembayaran SPP Sekolah Dengan Metode <i>K-Nearest Neighbor</i> (Studi Kasus Smk Al-Islam Surakarta) Robi Wariyanto Abdullah, Kusrini, Emha Taufiq Luthfii	160-164
Prediksi Customer Churn Perusahaan Telekomunikasi Menggunakan <i>Naïve Bayes</i> dan <i>K-Nearest Neighbor</i> Kaharudin, Musthofa Galih Pradana, Kusrini	165-171
Perencanaan Rencana Strategis Sistem Informasi Untuk Tercapainya Visi Misi dari Rumah Sakit Siska Febriani, Bambang Soedijono, M. Rudyanto Arief	172-178
Evaluasi <i>User Interface</i> pada Aplikasi E-Commerce (Studi Kasus Informa & IKEA) Sundari Ilkham, Kusrini, M. Rudyanto Arief	179-188
Progressive Web Apps Untuk Rekayasa Hybrid Application Berbasis Teknologi <i>Mean Stack</i> Rusdy Agustaf, Bernard Renaldy Suteja	189-194
Penerapan Metode Ahp Dalam Penentuan Kriteria Sistem Pendukung Keputusan Seleksi Penjual Pada Kantin Fandli Supandi, Kusrini, Hanif Al Fatta	195-200

PENGANTAR REDAKSI

Puji syukur kami panjatkan kehadirat Allah Tuhan Yang Maha Kuasa atas terbitnya JURNAL INFORMASI INTERAKTIF Volume 4, Nomor 3, Edisi September 2019. Pada edisi kali ini memuat 10 (sepuluh) tulisan hasil penelitian dalam bidang teknik informatika.

Harapan kami semoga naskah yang tersaji dalam JURNAL INFORMASI INTERAKTIF edisi September tahun 2019 dapat menambah pengetahuan dan wawasan di bidangnya masing-masing dan bagi penulis, jurnal ini diharapkan menjadi salah satu wadah untuk berbagi hasil-hasil penelitian yang telah dilakukan kepada seluruh akademisi maupun masyarakat pada umumnya.

Redaksi

PREDIKSI CUSTOMER CHURN PERUSAHAAN TELEKOMUNIKASI MENGUNAKAN NAÏVE BAYES DAN K-NEAREST NEIGHBOR

Kaharudin¹, Musthofa Galih Pradana², Kusri³

¹²³ Magister Teknik Informatika Universitas AMIKOM Yogyakarta

Email : ¹kahar.osvaldo@gmail.com, ²mgalihpradana@gmail.com, ³kusri@amikom.ac.id

ABSTRACT

For a company it is very vital. Customers are the key to running a business that is run. But in reality there are loyal customers and some are churned out. Churn is defined as the tendency of customers to stop doing business with a company. It is important for companies to be able to identify customers who have a tendency to be churn customers. Then a report is needed to be able to identify and make decisions for management. Prediction method using Naïve Bayes method produces an accuracy of 76% and K-Nearest Neighbor produces information with a K = 1 value of 73%, K = 3 which is 76% and K = 5 by 78% It can be concluded that the K-Nearest Neighbor Method with K = 3 has a better value. The results of customer predictions for a company can be used to take an example for the customer so they will not churn.

Keywords: Prediction, Customer, Churn, Naïve Bayes, Telecommunication, K-Nearest Neighbor.

1. PENDAHULUAN

Keberadaan pelanggan bagi sebuah perusahaan sangat vital. Pelanggan adalah konsumen yang akan menggunakan jasa atau produk yang ditawarkan dan dijual. Tanpa keberadaan pelanggan, sebuah perusahaan tidak akan mendapatkan keuntungan yang menjadi sebuah tujuan utama dibangunnya sebuah bisnis. Untuk mendapatkan pelanggan, perusahaan perlu melakukan promosi agar produk atau jasa yang ditawarkan dapat dikenal secara luas oleh masyarakat. Oleh karena itu banyak perusahaan berlomba-lomba dalam melakukan promosi, seperti menunjuk public figure sebagai brand ambassador, memberikan diskon besar dan masih banyak strategi yang lain dengan satu tujuan yakni mendapatkan pelanggan yang banyak.

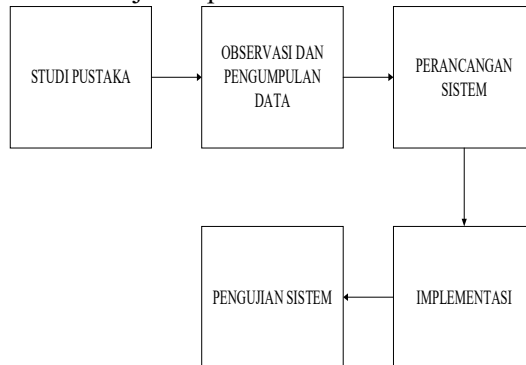
Setelah mendapatkan pelanggan sesuai yang diharapkan, masalah belum selesai, ada masalah lain yang mengintai perusahaan, yakni apakah pelanggan tersebut loyal atau malah Churn. Menurut Yu dkk., (2011) Customer Churn didefinisikan sebagai kecenderungan pelanggan untuk berhenti melakukan bisnis dengan sebuah perusahaan [1]. Hal ini telah menjadi isu penting yang merupakan salah satu tantangan utama oleh banyak perusahaan di era global ini dan harus dihadapinya. Berdasarkan transaksi yang ada dapat dilakukan analisa untuk mencari pola apakah pelanggan akan setia dengan perusahaan atau tidak.

Menurut Emmett C. Murphy dan Mark A. Murphy mendapatkan pelanggan baru jauh lebih sulit dibandingkan dengan mempertahankan pelanggan lama [2], serta biaya yang dikeluarkan perusahaan lima kali lipat lebih banyak dibandingkan dengan memuaskan dan mempertahankan pelanggan lama. Oleh karena itu perlu dilakukan prediksi tentang customer akan loyal atau tidak. Hal ini dapat bermanfaat bagi perusahaan dalam menentukan kebijakan apa yang akan diambil, seperti langkah promosi yang efektif. Proses prediksi yang dilakukan menggunakan 2 metode yaitu metode Naïve Bayes dan K-Nearest Neighbor. Kedua metode akan dibandingkan dan dicari hasil yang paling baik dan optimal dalam melakukan prediksi guna membantu perusahaan dalam menentukan kebijakan dan keputusan.

2. TINJAUAN PUSTAKA

Metode penelitian yang dilakukan ada 5 tahapan yakni : Studi Pustaka untuk mencari literatur dan referensi sebagai acuan melakukan penelitian baik dari buku, jurnal, prosiding dan yang lainnya. Selanjutnya observasi dan pengumpulan data yang dilakukan dengan mencari dataset yang akan digunakan dalam penelitian ini, berupa data transaksi pada perusahaan, data tersebut yang akan diolah dalam penelitian. Berikutnya tahap perancangan sistem dilakukan dengan merancang data flow diagram dan relasi tabel

sesuai dataset dan kebutuhan data yang ada. Tahap keempat adalah implementasi metode kedalam kode program sehingga memudahkan dalam menghasilkan analisa dan prediksi. Langkah yang terakhir adalah pengujian yang akan dilakukan dengan dataset yang ada dicoba kedalam sistem secara continue untuk mengetahui akurasi sistem. Gambar alur penelitian ditunjukkan pada Gambar 1.



Gambar 1 Alur Penelitian

2.1 Referensi

Penelitian terkait pernah dilakukan oleh Andri Wijaya dan Abba Suganda Girsang dengan judul *The Use Of Data Mining For Prediction Of Customer Loyalty*, dipublikasikan dalam *CommIT (Communication & Information Technology)*, May 31. Penelitian ini mengambil kesimpulan bahwa Algoritma Naïve Bayes menghasilkan akurasi sebesar 76 % [3]. Selanjutnya penelitian dengan judul *Evaluation of predictive ability of support vector machines and Naïve Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS* yang ditulis Binh Thai Pham, dkk pada *Journal of Geomatics* yang berkesimpulan secara keseluruhan, SVM dan NBT menunjukkan metode yang menjanjikan yang dapat digunakan untuk prediksi spasial tanah longsor di daerah rawan longsor [4].

Penelitian berikutnya yang ditulis oleh K.Kalaiselvi dan J. Sowmiya dengan judul *Predicting Instructors Performance in Higher Education Systems* yang dipublikasikan pada *European Alliance for Innovation (EAI)*. Penelitian ini menghasilkan kesimpulan Didapatkan hasil pengujian bahwa Algoritma J48DT mendapatkan hasil akurasi sebesar 84.8%, Naïve bayes 83.3% dan FA-Paired t-test 85.09% [5].

Penelitian rujukan ke empat adalah A Survey on Naive Bayes Based Prediction of Heart Disease Using Risk Factors dengan

penulis Sohana Saiyed, Nikita Bhatt and Amit P. Ganatra di *International Journal of Innovative and Emerging Research in Engineering*. Penelitian ini berkesimpulan bahwa Algoritma Naïve bayes mendapatkan nilai akurasi 86.38% [6].

2.2 Metode Naïve Bayes

Naïve Bayes merupakan sebuah model klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Naïve Bayes didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network [7]. Adapun rumus kedekatan pada Naïve Bayes ditunjukkan persamaan 1 [8] :

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

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
- $P(H)$: Probabilitas hipotesis H (prior probability)
- $P(X)$: Probabilitas dari X

2.3 Metode K-Nearest Neighbor

Algoritma K-Nearest Neighbor (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Data pembelajaran diproyeksikan k ruang berdimensi banyak, di mana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruang ini ditandai dengan kelas c, jika kelas c adalah klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut [9].

Langkah-langkah untuk menghitung metode Algoritma K-Nearest Neighbor [9]:

- a. Menentukan Parameter K (jumlah tetangga paling dekat).
- b. Menghitung kuadrat jarak Euclidian (queri instance) masing-masing objek terhadap data sampel yang diberikan.

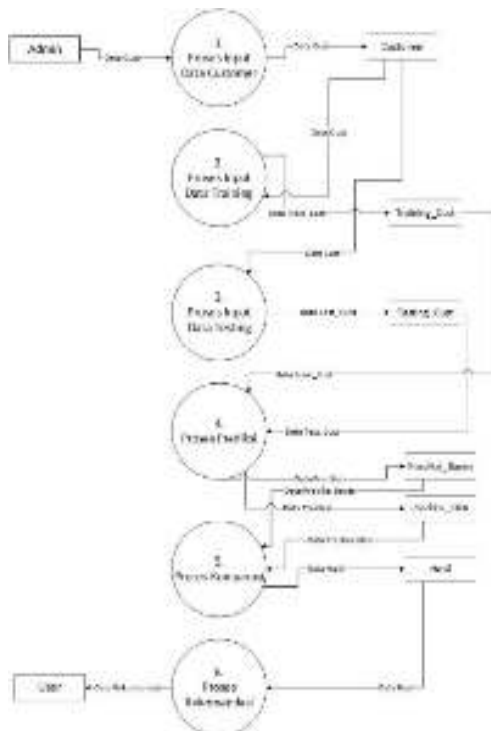
- c. Kemudian mengelompokkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak Euclidian terkecil.
- d. Mengumpulkan kategori Y (klasifikasi Nearest Neighbor)
- e. Dengan memakai kategori Nearest Neighbor yang paling mayoritas maka dapat diprediksi nilai query instance yang telah dihitung.

Untuk mendefinisikan jarak antara dua titik yaitu titik pada data training (x) & titik pada data testing (y) maka dipakai rumus Euclidean [9], seperti yang ditunjukkan pada persamaan (2) :

$$D(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (2)$$

2.4 Perancangan Data Flow Diagram

Data Flow Diagram (DFD) merupakan diagram yang digunakan untuk menggambarkan proses-proses yang terjadi pada sistem yang akan dikembangkan. Dengan model ini, data-data yang terlibat pada masing-masing proses dapat diidentifikasi. Rancangan DFD ditunjukkan oleh Gambar 2.

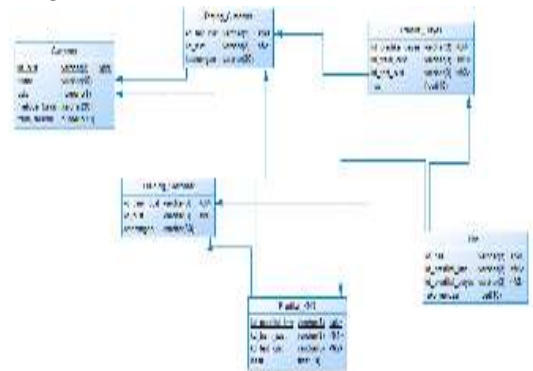


Gambar 2 Data Flow Diagram

2.5 Perancangan Database

Tabel penyimpanan data pada DFD selanjutnya perlu dilakukan design ke dalam

bentuk relasi tabel database yang ditunjukkan gambar 3.



Gambar 3 Relasi Tabel

2.6 Dataset Awal

Dataset awal yang digunakan dalam penelitian ini di dapatkan pada dataset google pada sebuah perusahaan telekomunikasi di Amerika Serikat. Data awal memiliki 20 variable dan ditunjukkan pada tabel 1.

Tabel 1 Dataset Awal

No	Nama Variable
1	CustomerID
2	Gender
3	SeniorCitizen
4	Partner
5	Dependents
6	Tenure
7	MultipleLines
8	InternetService
9	OnlineSecurity
10	OnlineBackup
11	DeviceProtection
12	TechSupport
13	StreamingTV
14	StreamingMovies
15	Contract
16	PaperlessBilling
17	PaymentMethod
18	MonthlyCharges
19	TotalCharges
20	Churn

Dari 20 data tersebut dipilih 9 variable yang memiliki korelasi yang kuat dan memiliki variasi yang tinggi untuk memprediksi customer churn. Variable terpilih ditunjukkan pada tabel 2.

Tabel 2 Variable Dataset

No	Nama Variable
1	CustomerID
2	Gender

3	SeniorCitizen
4	Tenure
5	InternetService
6	Contract
7	MonthlyCharges
8	TotalCharges
9	Churn

Dari variable dataset yang terpilih akan membantu dalam menentukan apakah customer termasuk customer churn ataupun tidak. Standar yang digunakan untuk customer dikatakan sebagai customer churn adalah customer tersebut memutuskan untuk tidak menggunakan layanan teleomukasi dari perusahaan tersebut.

2.7 Data Training

Data Training yang digunakan dalam penelitian ini ditunjukkan pada tabel 3.

Tabel 3 Data Training

customer ID	gender	tenure	Internet Service	Contract	Payment Method	Monthly Charges	Total Charges	Churn
1	Female	1	DSL	Month-to-month	Mailed check	2	1	No
2	Male	2	No	One year	Credit card	1	1	No
3	Male	3	DSL	Month-to-month	Bank transfer	3	1	Yes
4	Male	4	DSL	Month-to-month	Bank transfer	1	1	No
5	Female	5	No	One year	Mailed check	1	1	No

2.8 Penerapan Metode Naïve Bayes

Metode Naïve Bayes harus menggunakan data yang berupa kategorikal sehingga dataset yang digunakan terlebih dahulu harus dirubah yaitu pada atribut tenure, monthly charges dan total charges dimana data yang ada dikelompokkan sehingga menjadi beberapa kelompok, sampel data training pada metode Naïve Bayes ditunjukkan pada tabel 4.

Tabel 4 Sample Data Training Naïve Bayes

customer ID	gender	tenure	Internet Service	Contract	Payment Method	Monthly Charges	Total Charges	Churn
5948-UJZLF	Male	1	DSL	Month-to-month	Mailed check	2	1	No
5993-BQHEA	Male	2	No	One year	Credit card	1	1	No
3376-BMGFE	Female	1	Fiber optic	Month-to-month	Bank transfer	3	1	Yes
3318-ISQFQ	Female	2	No	Month-to-month	Bank transfer	1	1	No
2682-KEVRP	Female	3	No	One year	Mailed check	1	1	No

Adapun langkah-langkah untuk melakukan prediksi menggunakan algoritma Naïve Bayes adalah sebagai berikut :

1. Langkah pertama mendefinisikan Bobot Masing-Masing Atribut yang ditunjukkan pada tabel 5.

Tabel 5 Bobot Atribut

Atribut	Bobot
Gender	1
tenure	1
Internet Service	1
Contracts	1
Payment Method	1
Monthly Charges	1
Total Charges	1

2. Selanjutnya mendefinisikan bobot kedekatan atribut dengan gender, tenure, internet service, contract, payment method, monthly charge, dan total charge pada tabel 6, 7, 8, 9, 10, 11,12.

Tabel 6. Kedekatan nilai atribut gender

Nilai 1	Nilai 2	Kedekatan
Male	male	1
Female	female	1
Female	male	0,7
Male	female	0,7

Tabel 7. Sampel Kedekatan Nilai Atribut Tenure

Nilai 1	Nilai 2	Kedekatan
1	1	1
2	2	1
3	3	1
5	1	0,7
6	4	0,7
7	3	0,7

Tabel 8. Kedekatan Nilai Atribut Internet Service

Nilai 1	Nilai 2	Kedekatan
DSL	DSL	1
Fiber Optic	Fiber Optic	1
No	No	1
DSL	Fiber Optic	0,8
Fiber Optic	No	0,8

Tabel 9. Kedekatan Nilai Atribut Contract

Nilai 1	Nilai 2	Kedekatan
Month-to-month	Month-to-month	1

One year	One year	1
Two year	Two year	1
Two year	Month-to-month	0,8
One Year	Two year	0,8

Tabel 10. Kedekatan Nilai Atribut Payment Method

Nilai 1	Nilai 2	Kedekatan
Bank Transfer	Bank Transfer	1
Credit Card	Credit Card	1
Electronic check	Electronic check	1
Cash	Credit Card	0,8
Credit Card	Electronic check	0,8
Electronic check	Bank Transfer	0,8

Tabel 11. Sampel Kedekatan Nilai Atribut Monthly Charge

Nilai 1	Nilai 2	Kedekatan
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	1	0,6
7	3	0,6
8	9	0,6

Tabel 12. Sampel Kedekatan Nilai Atribut Total Charge

Nilai 1	Nilai 2	Kedekatan
1	1	1
2	2	1
3	3	1
4	4	1
5	5	1
6	1	0,6
7	3	0,6
8	9	0,6

3. Langkah ketiga adalah melakukan perhitungan data Testing dengan data Training yang ditunjukkan pada tabel 13.

Tabel 13. Sampel Data Testing

Customer ID	gender	tenure	Internet Service	Contract	Payment Method	Monthly Charges	Total Charges	Churn
6728-DKUCO	female	7	Fiber optic	One year	Electronic check	5	8	?

Untuk memprediksi apakah calon pembeli tersebut akan churn atau loya makal dilakukan perhitungan sebagai berikut :

a. Menghitung kedekatan data testing dengan data training no 1. Diketahui

a : Kedekatan nilai atribut Gender (male dengan male)

: 1

b : Bobot Atribut Gender

: 1

c : Kedekatan Nilai Tenure (7 dengan 1)

: 0,7

d : Bobot Nilai Tenure

: 1

e : Kedekatan nilai Atribut Internet Service (Fiber optic dengan DSL)

: 1

f : Bobot Nilai Atribut Internet Service

: 1

g : Kedekatan nilai Atribut Contract (Oneyear dengan Month to month)

: 0,6

h : Bobot Nilai Atribut Contract

: 1

i : Kedekatan nilai Atribut Payment Method (Electronic check dengan Mailed Check)

: 1

j : Bobot Nilai Atribut Payment Method

: 1

k : Kedekatan nilai Atribut Monthly Charge (5 dengan 2)

: 0,6

l : Bobot Nilai Atribut Monthly Charge

: 1

m : Kedekatan nilai Atribut Total Charge (8 dengan 1)

: 0,6

n : Bobot Nilai Atribut Total Charge

: 1

Dihitung :

Jarak =

$$\frac{(a \times b) + (c \times d) + (e \times f) + (g \times h) + (i \times j) + (k \times l) + (m \times n)}{b + d + f + h + j + l + n}$$

$$\text{Jarak} = \frac{(1 \times 1) + (0,7 \times 1) + (1 \times 1) + (0,6 \times 1) + (1 \times 1) + (0,6 \times 1) + (0,6 \times 1)}{1 + 1 + 1 + 1 + 1 + 1 + 1}$$

$$\text{Jarak} = \frac{5,5}{7}$$

$$\text{Jarak} = 0,786$$

Perhitungan diatas dilakukan terhadap keseluruhan data training yang ada sehingga, nilai jarak yang tertinggi merupakan hasil prediksi yang diperoleh.

2.9 Penerapan Metode K-Nearest Neighbor

Metode K-Nearest Neighbor harus berupa numeric sehingga data customer harus dirubah terlebih dahulu, yaitu atribut gender, internet service, contract, dan payment method. Sampel Data Training pada metode K-Nearest Neighbor ditunjukkan tabel 14, dan data testing pada tabel 15.

Tabel 14. Sampel Data Training K-Nearest Neighbor

Customer ID	gender	tenure	internet service	contract	payment method	monthly charge	total charge	class
00001	0	1	1	1	1	2000	5000	0
00002	1	2	1	2	2	3000	6000	0
00003	1	1	1	1	1	3000	3000	1
00004	1	2	1	2	2	4000	8000	0
00005	0	1	1	1	1	2000	2000	1

Tabel 15. Sampel Data Testing K-Nearest Neighbor

Customer ID	gender	tenure	internet service	contract	payment method	monthly charge	total charge	class
00006	0	2	1	1	1	3000	6000	0

Adapun langkah-langkah prediksi pada K-Nearest Neighbor adalah sebagai berikut ini :

1. Menentukan parameter k (jumlah tetangga paling dekat). Parameter k (jumlah tetangga paling dekat untuk melakukan memprediksi customer churn diatas adalah menggunakan nilai k=1, k=3, k=5.
2. Menghitung kuadrat jarak euclidean masing-masing objek terhadap data sample yang diberikan. Rumus yang digunakan dapat dilihat pada persamaan 3 :

$$D(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (3)$$

Sebagai contoh perhitungan diambil data testing yang pertama yaitu data testing lalu dikalikan terhadap seluruh data training yang ada:

$$D = \sqrt{(2 - 2)^2 + (1 - 26)^2 + (1 - 1)^2 + (1 - 1)^2 + (1 - 3)^2 + (29.85 - 44.45)^2 + (29.85 - 1183.8)^2}$$

$$= 1332229,603$$

Jadi jarak Euclidean Distance yang diperoleh adalah sebesar 1332229,603 Data testing tersebut dihitung Euclidean Distance nya keseluruhan data training yang ada, kemudian dirangking dengan mengambil 5 neighbor yang terdekat yaitu 5 nilai hasil perhitungan yang terkecil.

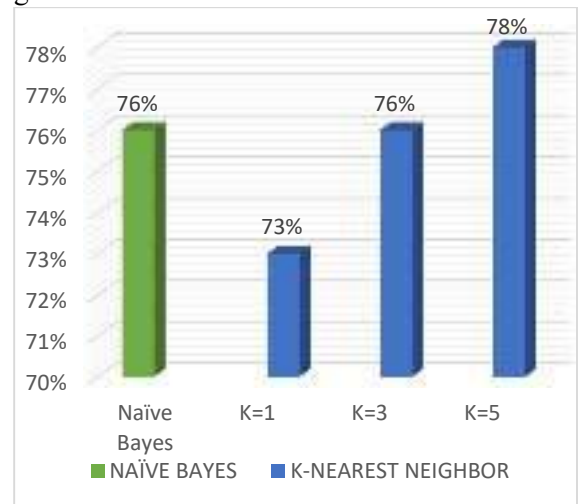
2.10 Komparasi Metode

Hasil dari pengujian prediksi customer churn dengan menggunakan Naïve Bayes dan K-Nearest Neighbor dapat dilihat pada tabel 16.

Tabel 16. Perbandingan Persentasi Nilai Akurasi Pengujian Algoritma

Algoritma			
Naïve Bayes	K-Nearest Neighbor		
	K=1	K=3	K=5
76 %	73 %	76 %	78 %

Untuk lebih melihat perbandingan akurasi dari hasil pengujian dapat melihat grafik pada gambar 4.



Gambar 4 Grafik Perbandingan Akurasi

3. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan dapat ditarik kesimpulan :

1. Atribut gender, tenure, internet service, contract, monthly charge, dan total charge dapat digunakan untuk melakukan prediksi customer churn.

2. Metode Naïve Bayes menghasilkan akurasi sebesar 76 %.
3. Untuk melakukan prediksi menggunakan Algoritma K-Nearest Neighbor untuk menghasilkan akurasi terbaik adalah K=5 yaitu memiliki tingkat akurasi 78%.
4. Metode yang menghasilkan akurasi yang lebih baik adalah metode K-Nearest Neighbor dengan nilai K=5.

Adapun saran dalam penelitian ini adalah sebagai berikut :

Peneliti selanjutnya dapat melakukan penambahan data training yang lebih banyak, peneliti selanjutnya juga dapat menggunakan data customer yang memiliki atribut yang lain seperti atribut total harga belanjaan, selain itu untuk melakukan prediksi customer churn juga dapat mencoba menggunakan metode yang lain seperti Decision Tree maupun Support Vector Machine.

DAFTAR PUSTAKA

- [1] X. Yu, S. Guo, J. Guo, and X. Huang, "An extended support vector machine forecasting framework for customer churn in e-commerce," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1425–1430, Mar. 2011.
- [2] E. C. Murphy and M. A. Murphy, *Leading On the Edge Of Chaos : The 10 Critical Elements for Success in Volatile Times*. USA: Prentice Hall Press, 2002.
- [3] Andri Wijaya, Abba Suganda Girsang, 2016, *The Use of Data Mining For Prediction Of Customer Loyalty*, *CommIT (Communication & Information Technology)*, May 31.
- [4] Binh Thai Pham, DieuTien Bui, Indra Prakash, M.B. Dholakia, 2016, *Evaluation of predictive ability of support vector machines and Naïve Bayes trees methods for spatial prediction of landslides in Uttarakhand state (India) using GIS*, *Journal of Geomatics Vol 10 No. 1*, Mar 29.
- [5] K.Kalaiselvi dan J. Sowmiya, 2018, *Predicting Instructors Performance in Higher Education Systems*, *European Alliance for Innovation (EAI)*, June 2018.
- [6] Sohana Saiyed, Nikita Bhatt, Amit P. Ganatra, 2016, *A Survey on Naive Bayes Based Prediction of Heart Disease Using Risk Factors*. *International Journal of Innovative and Emerging Research in Engineering*, Jan 2016.
- [7] Liao. 2007. *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application* . Singapore: World Scientific Publishing.
- [8] Budi Santosa. 2007. *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Graha Ilmu.
- [9] Siregar, A. M., & Puspabhuana, A., 2002. *Data Mining Pengolahan Data Menjadi Informasi dengan RapidMiner*. Sukoharjo: CV Kekata Group.