

PREDIKSI LOYALITAS PELANGGAN INDIHOME DENGAN METODE *K-NEAREST NEIGHBOR*

Ivo Colanus Rally Drajana

Fakultas Ilmu Komputer Universitas Ichsan Gorontalo

Jl. Achmad Nadjamuddin No. 17 Telp. 0435-829975 Fax. 0435-829976 Kota Gorontalo

ivocolanusrally@gmail.com

Perusahaan informasi serta komunikasi penyedia jasa dan jaringan telekomunikasi lengkap di Indonesia adalah Telkom. Telkom merupakan perusahaan yang menyediakan jasa dalam bidang telekomunikasi terkemuka di Indonesia, dengan demikian Telkom wajib memberikan layanan terbaik kepada pelanggannya agar tetap loyal menggunakan produk dari Telkom, salah satunya Indihome. Salah satu permasalahan yang dihadapi yang bergerak dibidang penjualan jasa adalah mempertahankan loyalitas pelanggannya. Loyalitas pelanggan perlu diprediksi dengan akurat agar dapat membantu dalam proses pengambilan keputusan. Metode yang digunakan untuk memprediksi loyalitas pelanggan adalah metode K-Nearest Neighbor (K-NN). Tujuan dari algoritma KNN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data. Prediksi Loyalitas Pelanggan Indihome berdasarkan data Pelanggan algoritma K-NN telah selesai dilakukan. Pada eksperimen algoritma K-NN model terbaik dilihat berdasarkan nilai error terkecil yaitu 0.100, variabel periode 3, validation 10 dan nilai k 9.

Kata Kunci : Telkom, Loyalitas Pelanggan, K-Nearest Neighbor

I. PENDAHULUAN

Telkom merupakan perusahaan yang menyediakan jasa dalam bidang telekomunikasi terkemuka di Indonesia, dengan demikian Telkom wajib memberikan layanan terbaik kepada pelanggannya agar tetap loyal menggunakan produk dari Telkom, salah satunya Indihome. Loyalitas pelanggan adalah komitmen pelanggan selama bertahun-tahun untuk berlangganan kembali atau melakukan pembelian ulang produk atau jasa secara konsisten dimasa yang akan datang [1]. Perubahan gaya hidup masyarakat masa kini yang membutuhkan akses informasi yang beragam, mudah, cepat dan handal adalah peluang sekaligus tantangan bagi operator layanan telekomunikasi untuk memenuhi kebutuhan masyarakat. Laju perkembangan teknologi akses jaringan dan teknologi berbasis internet memungkinkan bagi operator layanan untuk dapat menyediakan layanan yang beragam (*multi service*) seperti layanan suara (telepon), data (internet) dan gambar (IP-TV) dalam satu jaringan akses dengan layanan *triple-play*. IndiHome merupakan layanan Triple Play dari Telkom yang terdiri dari Internet on Fiber atau High Speed Internet, Phone (Telepon Rumah), dan IPTV (UseTV Cable). Akan tetapi produk Indihome belum

banyak diketahui calon pelanggan apa itu Indihome, keunggulan dan fasilitas penggunaan Indihome.

Salah satu permasalahan yang dihadapi oleh perusahaan yang bergerak dibidang penjualan jasa adalah mempertahankan loyalitas pelanggannya. Dengan demikian loyalitas pelanggan perlu diprediksi dengan akurat agar dapat membantu dalam proses pengambilan keputusan terkait masalah tersebut. Dengan melakukan prediksi dapat membantu pihak Telkom untuk mengetahui apakah pelanggan Telkom loyal terhadap produk yang disediakan yaitu Indihome. Pelanggan yang dikatakan loyal dan tidak loyal (Telkom Gorontalo) yaitu berdasarkan umur pelanggan, perilaku bayar dan riwayat tunggakan. Metode yang digunakan untuk memprediksi loyalitas pelanggan adalah metode *K-nearest neighbor*. *K-Nearest Neighbor* adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari label class pada KNN. Tujuan dari algoritma KNN adalah mengklasifikasikan objek baru berdasarkan atribut dan training data [2]. KNN juga merupakan metode yang baik digunakan untuk melakukan suatu prediksi, berdasarkan penelitian sebelumnya dengan judul Model algoritma KNN untuk prediksi kelulusan mahasiswa memberikan hasil yang baik dengan tingkat akurasi 85,15% [3].

II. LANDASAN TEORI

A. Loyalitas Pelanggan

Loyalitas adalah wujud perilaku dari unit-unit pengambilan keputusan dalam melakukan pembelian secara *continue* terhadap barang ataupun jasa terhadap salah satu perusahaan yang dipilih. Loyalitas pelanggan juga bisa disebut komitmen pelanggan bertahun-tahun untuk berlangganan kembali atau membeli secara berulang produk atau jasa secara konsisten dimasa mendatang, meskipun pengaruh situasi pemasaran mempunyai potensi untuk menyebabkan perubahan perilaku [1].

B. Prediksi

Prediksi dapat kita artikan sebagai suatu kegiatan memperkirakan apa yang akan terjadi dimasa mendatang dengan memanfaatkan data-data lama dengan indikator tertentu dalam melakukan peramalan di masa mendatang. Ada beberapa permasalahan yang membutuhkan kegiatan

prediksi diantaranya, prediksi jumlah produksi, prediksi pendapatan, prediksi harga, prediksi tingkat kelulusan, prediksi pendapatan dan beberapa prediksi lainnya [4].

C. Data Mining

Data mining merupakan proses menggunakan teknik statistik, kecerdasan buatan, matematika, dan *machine learning* dalam mengekstraksi dan mengidentifikasi informasi yang dibutuhkan dan pengetahuan yang terkait dari beberapa *big database* [5].

Data mining juga merupakan proses menganalisa data dari perspektif yang berbeda-beda kemudian menyimpulkan menjadi informasi penting yang dapat dipakai untuk menambah keuntungan, meminimalisir biaya pengeluaran. Secara teknis, data mining disebut juga sebagai proses menemukan korelasi bahkan pola dari ratusan atau ribuan *field* dari sebuah relasional *big database* [6].

Beberapa teknik dimiliki data mining sesuai tugas yang bisa dilakukan, yaitu :

1. Deskripsi
Peneliti-peneliti biasanya mencoba menemukan cara dalam mendeskripsikan pola dan trend tersembunyi dalam data.
2. Estimasi
Estimasi bisa juga disebut dengan klasifikasi dengan mengekecualikan variabel tujuan yang lebih kearah numerik dari pada kategori.
3. Prediksi
Prediksi mirip dengan estimasi dan klasifikasi, namun prediksi hasilnya menunjukkan keadaan yang belum terjadi (terjadi dimasa mendatang).
4. Klasifikasi
Dalam mengklasifikasikan variabel tujuannya bersifat kategorik. Sebagai contoh, kita akan mengklasifikasikan penjualan dalam tiga kelas, yaitu penjualan tinggi, penjualan sedang, dan pejualan rendah.
5. Clustering
Clustering lebih kearah pengelompokan *record*, pengamatan, atau kasus didalam kelas memiliki kemiripan.
6. Asosiasi
Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

Sebagai rangkaian proses, data mining dapat dibagi menjadi beberapa tahapan. Tahap-tahap tersebut bersifat interaktif, pengguna terlibat langsung atau dengan perantaraan *knowledge base*.

Tahap-tahap data mining adalah sebagai berikut:

- a. Pembersihan data (*data cleaning*)
Pembersihan data adalah merupakan proses dalam menghilangkan-kan *noise* dan data yang tidak konsisten dan data tidak relevan.
- b. Integrasi data (*data integration*)
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru.
- c. Seleksi data (*data selection*)
Data yang ada dalam database sering kali tidak semuanya terpakai, oleh karena itu hanya data yang

cocok untuk dianalisis yang akan diambil dari *database*.

- d. Transformasi data (*data transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining.
- e. Proses *mining*
Merupakan suatu proses pertama saat metode diterapkan untuk mendapatkan pengetahuan berharga dan tersembunyi dari data.
- f. Evaluasi pola (*pattern evaluation*)
Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan.
- g. Presentasi pengetahuan (*knowledge presentation*)
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

D. K-Nearest Neighbor

Algoritma *K-Nearest Neighbor* merupakan sebuah metode klasifikasi terhadap objek berdasarkan data pembelajaran yang memiliki jarak paling dekat dengan objek tersebut. K-NN memakai klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji baru. K-NN mengambil konsep '*learning by analogy*', kemudian data *learning* dideskripsikan oleh atribut numerik *n*-dimensi. Setiap data *learning* merepresentasikan satu titik, dengan tanda *c* didalam ruang *n*-dimensi. Namun jika ada data *query* yang memiliki label tidak diketahui diinputkan, maka K-NN akan mencari *k* buah data *learning* yang memiliki jarak paling dekat dengan data *query* didalam ruang *n*-dimensi. Dari jarak data *query* dan data *learning* dihitung dengan mengukur jarak antara titik yang merepresentasikan data *query* ke semua titik yang merepresentasikan data *learning* dengan persamaan *Euclidean Distance*. Ketepatan algoritma K-NN sangat dipengaruhi oleh ada atau tidak fitur yang tidak relevan, ataupun jika bobot dari fitur tersebut tidak berbanding dengan relevansinya terhadap klasifikasi. *K* buah data *learning* yang terdekat akan melakukan *voting* dalam menentukan label mayoritas. Kemudian label dari data *query* akan ditentukan oleh label mayoritas, jika ada beberapa label mayoritas maka label data *query* dapat dipilih dengan cara acak di antara label-label mayoritas yang telah ada. [7].

Dalam mendefinisikan jarak diantara dua titik pada data training (*x*) dan titik pada data testing (*y*), maka digunakan persamaan *Euclidean*, seperti dibawah ini.

$$(x,y)=\sqrt{\sum(xk-yk)2nk-1} \dots\dots\dots(1)$$

Data training *x* dan titik data testing *y* akan diklasifikasi, dimana $x=x1,x2,\dots,xi$ dan $y=y1,y2,\dots,yi$ dan *I* merepresentasikan nilai atribut serta *n* merupakan dimensi atribut.

III. METODE PENELITIAN

Metode dalam penelitian ini adalah metode eksperimen dimana metode ini melakukan manipulasi satu atau lebih variabel dengan cara tertentu yang memengaruhi satu atau lebih variabel yang akan di ukur.

A. Sumber Data

Sumber data pada tahap ini adalah data primer, yaitu data yang diperoleh secara langsung dari sumbernya, kemudian untuk mengetahui permasalahan atau kendala digunakan teknik wawancara.

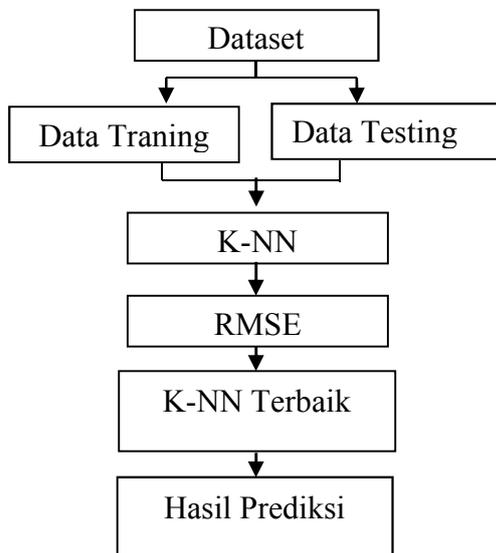
Pengumpulan data berupa pernyataan (statement) tentang keadaan, kegiatan tertentu, sifat, dan sejenisnya.

Metode pengumpulan data dalam penelitian ini yaitu:

1. Observasi
Pengamatan menggunakan semua indera (pendengaran, penciuman, penglihatan). Pencatatan hasil dapat dilakukan dengan bantuan alat rekam langsung mengenai loyalitas pelanggan.
2. Dokumen
Pengambilan data dokumen tertulis ataupun elektronik dari lembaga/institusi. Dokumen dibutuhkan untuk mendukung kelengkapan data yang berkaitan dengan objek penelitian tentang dokumen loyalitas pelanggan.
3. Wawancara
Dilakukan terhadap pihak-pihak yang terkait untuk mendapatkan informasi situasi atau kendala serta permasalahan yang dihadapi.

B. Metode Prediksi

Prediksi sama dengan ramalan atau perkiraan. Prediksi atau peramalan studi terhadap data historis dengan tujuan untuk hubungan kecenderungan dan pola sistematis. Model K-NN untuk memprediksi Loyalitas Pelanggan Indihome dapat ditunjukkan pada gambar dibawah ini.



Gambar 1. Model KNN untuk Prediksi Loyalitas Pelanggan

Dari gambar diatas model K-NN untuk memprediksi Loyalitas Pelanggan Indihome dapat dinarasikan berikut :

1. Dataset : Tahap pertama mengumpulkan dataset Loyalitas Pelanggan, setelah dataset terkumpul,

selanjutnya membagi dataset menjadi dua bagian data traning dan data testing.

2. Data traning merupakan data *learning*, dimana data akan di uji dengan data testing untuk mendapatkan hasil prediksi.
3. Data testing merupakan data yang akan diuji dengan data traning menggunakan metode K-NN.
4. RMSE digunakan untuk mengevaluasi kinerja hasil prediksi, untuk mengetahui model terbaik dari model K-NN.
5. Hasil prediksi merupakan *output* atau keluaran dari model K-NN, apakah dengan melihat model terbaik dapat digunakan untuk memprediksi data baru yang ingin diketahui hasilnya.

IV. HASIL DAN PEMBAHASAN

Dalam peneliti ini dilakukan beberapa tahap. Tahap pertama adalah tahap pengumpulan data. Data yang digunakan adalah data PT.Telkom Gorontalo. Data tersebut berupa data Pelanggan Indihome. Pelanggan bisa dikatakan loyal dan tidak loyal berdasarkan pada umur pelanggan, perilaku bayar di atas tanggal 20 dan riwayat tunggakan. Contoh dataset Pelanggan Indihome :

Tabel 1. Data Pelanggan PT. Telkom

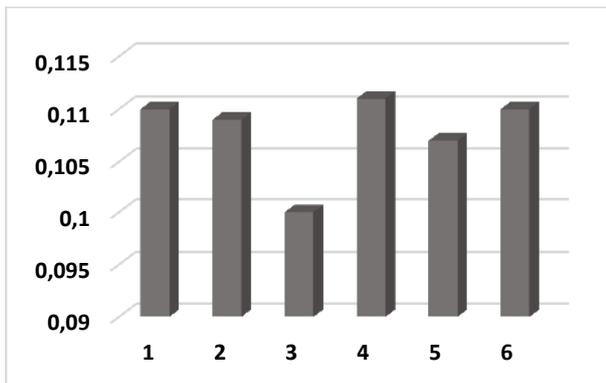
N O	UMUR BERLANGGAN	UMUR PELANGGAN	PEKERJAAN	PERILAKU BAYAR			Riwayat tunggakan (bulan)	Status
				< 20	20	> 20		
1	17	40	SOPIR	1	6	10	2	Tidak Loyal
2	5	24	MAHASISWA	4	0	1	2	Loyal
3	20	34	PNS	10	7	3	0	Loyal
4	14	40	PNS	10	3	1	0	Loyal
5	15	47	WIRASWASTA	5	4	6	2	Tidak Loyal
6	16	32	WIRASWASTA	6	5	5	0	Loyal
7	19	65	PENSIUNAN	9	5	5	2	Tidak Loyal
8	20	54	WIRASWASTA	7	9	4	2	Tidak Loyal
9	20	61	WIRASWASTA	6	5	9	0	Tidak Loyal
10	19	36	WIRASWASTA	9	4	6	0	Tidak Loyal
11	20	37	WIRASWASTA	6	6	8	2	Tidak Loyal
12	20	54	PEDAGANG	7	7	6	2	Tidak Loyal
13	20	41	SWASTA	9	5	6	0	Tidak Loyal
14	17	54	WIRASWASTA	7	5	5	2	Tidak Loyal
15	20	45	GURU	8	6	6	0	Tidak Loyal
...
105	21	47	GURU	5	9	7	2	Tidak Loyal

Penelitian ini bertujuan untuk menerapkan algoritma *K-Nearest Neighbor* (K-NN) pada data Loyalitas Pelanggan Indihome yang bertujuan untuk mencari model yang terbaik didalam memprediksi loyal tidaknya pelanggan.

Tabel 2. Nilai RMSE Loyalitas Pelanggan Indihome K-NN

Variabel	Cross Validation	K	RMSE
1	10	3	0.110
2	10	11	0.109
3	10	9	0.100
4	10	7	0.111
5	10	5	0.107
6	10	9	0.110

Pada Tabel 2. Diatas merupakan tabel hasil eksperimen Loyalitas Pelanggan Indihome dengan model menggunakan algoritma K-NN dari 1 sampai 6 variabel periode menggunakan *cross validation* 10 dan nilai *k* 1, 3, 5, 7, 9, 11, dilihat berdasarkan nilai RMSE terkecil. Hasil eksperimen model terbaik dilihat berdasarkan nilai *error* terkecil 0.100 dengan jumlah variabel periode 3 jumlah *validation* 10 dan nilai *k* yaitu 9.



Gambar 2. Grafik Model K-NN Loyalitas Pelanggan Indihome

Pada gambar 2 adalah grafik model terbaik untuk data Loyalitas Pelanggan Indihome, model terbaik dihasilkan oleh periode 3, nilai *validation* 10 dengan *k* 9, dengan nilai RMSE 0.100.

Prediksi Loyalitas Pelanggan Indihome berdasarkan data Pelanggan algoritma K-NN telah selesai dilakukan. Pada eksperimen algoritma K-NN model terbaik dilihat berdasarkan nilai *error* terkecil yaitu 0.100, variabel periode 3, *validation* 10 dan nilai *k* 9. Dalam membantu PT.Telkom dalam memprediksi Loyal tidaknya pelanggan agar lebih akurat, penerapan yang peneliti telah lakukan

yaitu algoritma K-NN dapat mempermudah dan lebih efisiensi dalam memprediksi Loyalitas Pelanggan. Manfaat dari penelitian ini bisa digunakan sebagai masukan dalam bertindak dimasa mendatang.

V. KESIMPULAN

Prediksi Loyalitas Pelanggan Indihome dengan algoritma *K-Nearest Neighbor* (K-NN) dengan telah selesai dilakukan. Melihat pada hasil eksperimen yang telah dilakukan, menunjukkan bahwa algoritma *k-Nearest Neighbor* (K-NN) mampu menghasilkan nilai *error* kecil. Disarankan untuk peneliti selanjutnya agar dapat melakukan uji coba dengan menambahkan *feature selection* dengan tujuan mendapatkan hasil *error* yang lebih kecil.

DAFTAR PUSTAKA

[1] Hatta. Iha. H, 2017. “*Feature*, Nilai, Kepuasan dan Loyalitas Pelanggan,” *Jurnal Riset Manajemen dan Bisnis*, ISSN : 2527-7502, Volume 2, Nomor 1, Universitas Pancasila, Februari, Jakarta.

[2] Bode. Andi, 2019. “Seleksi Fitur Untuk Prediksi Rating film Hollywood Menggunakan Model *K-Nearest Neighbor*,” *Jurnal JUPITER*, ISSN : 2356-4865, Volume 5, Nomor 1. Universitas Ichsan Gorontalo.

[3] Rohman. A, 2015. “Model Algoritma *K-Nearest Neighbor* (K-NN) Untuk Prediksi Kelulusan Mahasiswa,” *Jurnal NEO TEKNIKA*, ISSN : 2502-7662, Volume 1, Nomor 1, Universitas Pandanaran Semarang.

[4] Ayuningtias. L. P, Ifan. M, Jumadi, 2017. “Analisa Perbandingan Logic Fuzzy Metode Tsukamoto, Sugeno, dan Mamdani,” *Jurnal Teknik Informatika UIN Syarif Hidayatullah*, Bandung.

[5] Ridwan. M, Suyono. H, Sarosa. M, 2013. “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma *Naïve Bayes Classifier*,” *Jurnal EECCIS*, Volume 7, Nomor 1, Juni.

[6] Maburur. A. G, Lubis. R, 2012. “Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit,” *Jurnal Komputer Dan Informatika (KOMPUTA)*, Edisi 1, Volume 1, Universitas Komputer Indonesia, Bandung.

[7] Drajana. Ivo. C. R, 2018. “Prediksi Jumlah Produksi Coconut Oil Menggunakan *k-Nearest Neighbor* dan *Backward Elimination*,” *Jurnal Tecnoscienza*, ISSN : 2615-3319, Volume 3, Nomor 1, Oktober