

Komparasi Algoritme *K-Nearest Neighbors* Dan *Support Vector Machines* Dalam Prediksi Layanan Produk *ICONNET*

I Wayan Ady Purnawibawa^{1*}, I Nyoman Purnama², dan I Nyoman Yudi Anggara Wijaya³

Program Studi Teknik Informatika, STMIK Primakara

Jl. Tukad Badung No.135 Renon, Denpasar, Indonesia

*Corresponding Author: adypurnawibawa@gmail.com

Abstract

Prediction of customer needs is needed to help ICON+ management provide maximum and quality ICONNET services to customers, and is useful for company management in planning related products offered, as well as providing input to management regarding products that are in great demand by customers. This study implements and compares the K-Nearest Neighbor (K-NN) and Support Vector Machine (SVM) algorithms in predicting the service Iconnet products that are most in demand by customers, so as to facilitate ICON+ management in planning customer service provision, and preparing market strategies in the future. A total of 3,206 datasets (consisting of 2,565 training data and 641 testing data) ICONNET service enthusiasts for 1 year that have been cleaned, were tested on both algorithms, based on parameters Bandwidth, Request_Date, Status, customer address, and service fee. Algorithm performance accuracy was tested using Cross Validation, Confusion Matrix and ROC curve methods. The results of the accuracy test show that the performance of the K-NN algorithm is more accurate than the SVM algorithm in various test categories.

Keywords: *Data Mining; Customer interest; Performance Accuracy; Cross Validation; Confusion Matrix*

Abstrak

Prediksi kebutuhan pelanggan diperlukan untuk membantu manajemen ICON+ menyediakan layanan ICONNET secara maksimal dan berkualitas kepada pelanggan, serta berguna bagi jajaran manajemen perusahaan dalam melakukan perencanaan terkait produk yang ditawarkan, serta memberi masukan pada pihak manajemen mengenai produk yang banyak diminati oleh pelanggan. Penelitian ini mengimplementasikan dan membandingkan algoritme *K-Nearest Neighbor (K-NN)* dan *Support Vector Machine (SVM)* dalam memprediksi layanan produk *Iconnet* yang paling diminati oleh pelanggan, sehingga dapat mempermudah manajemen ICON+ dalam perencanaan penyediaan layanan pelanggan, dan penyusunan strategi pasar di masa mendatang. Sebanyak 3.206 dataset (terdiri atas 2.565 data *training* dan 641 data *testing*) peminat layanan ICONNET selama 1 tahun yang telah dibersihkan, diuji pada kedua algoritme tersebut, berdasarkan parameter *Bandwith, Request_Date, Status, alamat pelanggan, dan biaya layanan*. Akurasi kinerja algoritme diuji menggunakan metode *Cross Validation, Confusion Matrix* serta *ROC curve*. Hasil uji akurasi menunjukkan kinerja algoritme *K-NN* lebih akurat dari algoritme *SVM* pada berbagai kategori pengujian.

Kata kunci: *Data Mining; Minat pelanggan; Akurasi Kinerja; Cross Validation; Confusion Matrix*

1. Pendahuluan

Memahami kebutuhan pelanggan atau konsumen adalah salah satu strategi yang sangat penting dari sekian banyak strategi pemasaran. Setiap produk yang ditawarkan kepada pelanggan, fungsinya tidak akan terwujud dengan baik hingga kebutuhan konsumen dapat ditangani dengan baik. Dengan memahami kebutuhan pelanggan, perusahaan dapat menyediakan produk atau layanan yang dapat membantu pelanggan, sehingga perusahaan akan

dengan mudah menghasilkan penjualan. Memahami kebutuhan pelanggan juga dapat membantu perusahaan membuat rencana pemasaran [1].

Layanan akses jaringan telekomunikasi merupakan salah satu bentuk layanan jasa yang berkembang pesat saat ini, seiring dengan perkembangan Teknologi Informasi. Suatu jenis layanan yang sering digunakan berkaitan dengan akses jaringan telekomunikasi yaitu teknologi WiFi. Teknologi WiFi memungkinkan pengguna untuk mengakses Internet dan mentransfer data dari berbagai perusahaan penyedia layanan internet ke gedung dan rumah yang telah memiliki akses WiFi.

Salah satu penyedia layanan WiFi rumahan adalah PT. Indonesia *Comnets Plus* (ICON+) dengan produk layanan ICONNET, yang merupakan anak perusahaan PT PLN (Persero). Sebagai operator layanan WiFi, ICONNET menyediakan berbagai paket layanan seperti ICONNET 10 Mbps, 20 Mbps, 50 Mbps, dan 100 Mbps. Kualitas pelayanan menjadi salah satu faktor yang menentukan kepuasan pelanggan. Prediksi kebutuhan pelanggan diperlukan untuk membantu manajemen ICON+ menyediakan layanan ICONNET secara maksimal kepada pelanggan, juga berguna bagi jajaran manajemen dalam melakukan suatu perencanaan terkait produk yang ditawarkan, serta memberi masukan pada pihak manajemen mengenai produk yang banyak diminati oleh calon pelanggan.

Teknologi data mining merupakan salah satu teknologi terkini yang memanfaatkan data historis atau data operasional organisasi sebagai pijakan dalam menentukan perencanaan arah bisnis perusahaan di masa mendatang. Data mining mengekstraksi data untuk menemukan pola (*pattern recognition*) yang penting dalam tumpukan data dalam database menjadi suatu pengetahuan (*knowledge discovery*) [2]. Beragam fungsi pada data mining diantaranya: fungsi deskripsi, fungsi prediksi, fungsi estimasi, fungsi *clustering*, fungsi klasifikasi, serta fungsi asosiasi [3].

Dalam hal prediksi bisnis, terdapat beberapa algoritme data mining yang dapat digunakan, seperti algoritme *K-Nearest Neighbors* dan *Support Vector Machines*. Penelitian sebelumnya telah menggunakan metode K-NN memprediksi pelayanan produk Indihome mempergunakan oleh Haspriyanti dan Prasetyaningrum [4]. Pada penelitian tersebut, akurasi algoritme *K-Nearest Neighbors* untuk memprediksi layanan produk Indihome sebesar 99%. Dalam bidang kesehatan, K-NN telah diuji untuk memprediksi penyakit [5, 6] dengan akurasi mencapai 85%. Dalam bidang pendidikan, algoritme K-NN telah diuji untuk prediksi prestasi siswa [7] dan prediksi kelulusan siswa [8], serta prediksi bidang bisnis lainnya [9].

Selain penelitian tersebut, ada juga penelitian mengenai implementasi *data mining* untuk memprediksi dengan menganalisis sentimen terhadap layanan Indihome berdasarkan *Twitter* dengan metode *Support Vector Machine* (SVM) yang dilakukan oleh Tineges, Triayudi, dan Sholihati [10]. Pada penelitian tersebut, akurasi dari algoritme *Support Vector Machine* (SVM) untuk analisis sentimen terhadap layanan Indihome berdasarkan *Twitter* sebesar 87%. Algoritme SVM juga telah digunakan dalam prediksi bidang bisnis pariwisata [11, 12] dan bidang bisnis lainnya [13, 14].

Tujuan penelitian ini mengimplementasikan dan membandingkan algoritme *K-Nearest Neighbor* dan *Support Vector Machine* dalam memprediksi Layanan Produk *Iconnet* yang paling diminati oleh pelanggan, sehingga dapat mempermudah manajemen ICON+ dalam perencanaan penyediaan layanan pelanggan dan penyusunan strategi pasar di masa mendatang.

2. Tinjauan Pustaka

Terdapat beberapa topik penelitian sebelumnya yang berkaitan dengan prediksi layanan konsumen, sebagai berikut:

Penggunaan Data Mining untuk Prediksi Minat *Customer* Penjualan Handphone dengan Algoritme *Apriori*, telah diteliti oleh Wahyuni, Sulistianingsih, Hariyanto, dan Lumbanbatu [15]. Penelitian tersebut menganalisis data penjualan *handphone* yang bersumber dari *database* sistem informasi transaksi penjualan. Uji data menggunakan aplikasi data mining *weka* dalam menemukan hubungan pola penjualan *handphone* antar item. Proses pengolahan data dimulai *praprosesing* dengan memilih variabel data, kemudian menemukan nilai *spot* dari tiap item set *handphone* dan kombinasi antara jenis *handphone*. Berdasarkan hasil pencarian nilai *spot* dan kombinasi antara jenis *handphone*, ditemukan nilai *confidence* dalam tiap kombinasi. Kombinasi yang memenuhi nilai minimum *spot* dan minimum *confidence* akan menjadi sebuah aturan asosiasi. Aturan asosiasi yang dihasilkan menjadi informasi jenis *handphone* yang paling banyak

terjual selama 1 bulan sampai 2 tahun. Hasilnya, di peroleh *knowlwge* jenis *handphone* yang paling diminati dan hubungan antara jenis atau tipe *hanphone* tersebut. *Knowledge* tersebut dapat dijadikan dasar menentukan persediaan setiap jenis *handphone*. Algoritme Apriori juga telah digunakan oleh Longgana, Irvan, dan Wilarto [16] dalam memprediksi penjualan produk telekomunikasi yang paling diminati oleh *customer*.

Algoritme C.45 telah digunakan oleh Susana, Khofidoh, dan Al Afghani [17] dalam memprediksi Peminat Kuota Internet. Hasil penelitian mereka menunjukkan bahwa semua perusahaan penyedia jasa layanan termasuk Telkomsel, Indosat, dan axis diminati oleh konsumen. Dalam menganalisis data menggunakan algoritme C.45 berbasis *tools Rapidminer*, yang pertama bahwa hasil pohon keputusan dalam menentukan minat konsumen dilihat dari harga, sedangkan tingkat akurasi metode C.45 pada penelitian tersebut mencapai 94,67%.

Nofriansyah, Erwansyah, dan Ramadhan [18] menguji penerapan algoritme *Naive Bayes* untuk mengetahui minat beli pelanggan terhadap kartu internet XL. Pada penelitian tersebut, algoritme *Naive Bayes* memprediksi minat pembeli produk kartu XL di masa mendatang berdasarkan pengalaman para pembeli di masa lalu melalui *database* penjualan. Hasil penelitian dapat digunakan untuk mendukung manajemen perusahaan meningkatkan strategi pemasaran.

Arwa Ulayya Haspriyanti dan Putri Taqwa Prasetyaningrum [4] pada penelitian “Penerapan Data Mining Untuk Prediksi Layanan Indihome Menggunakan Metode *K-Nearest Neighbor*”. Pada penelitian ini, pengubahan data primer yang telah diuji menunjukkan bahwa nilai akurasi data sebesar 99,99% dari implementasi data mining yang menghasilkan 1 jenis produk layanan terlaris dari 3 produk layanan. Pada penelitian ini bisa dijadikan sebagai acuan utama dalam menggunakan metode *K-Nearest Neighbor* dengan pemilihan atribut terbaik agar mendapatkan performa terbaik prediksi *K-Nearest Neighbor*

Rian Tingeges, Agung Triayudi dan Ira Diana Sholihati [19] pada penelitian “Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi *Support Vector Machine (SVM)*”. Pada penelitian ini, penerapan SVM dalam memprediksi sentimen yang ada pada cuitan Twitter mengenai layanan Indihome didapatkan hasil dengan pengujian menggunakan *Confussion Matrix*, didapat akurasi sebesar 87% dengan *precision* 86%. Penulis menggunakan penelitian ini untuk referensi efektivitas tingkat akurasi dari metode *Support Vector Machine* dengan metode pengujian *confussion matrix*.

Perbedaan penelitian ini dengan penelitian [15]-[18] adalah penggunaan metode yang berbeda dalam memprediksi layanan konsumen. Namun demikian, penelitian ini memiliki kesamaan dengan penelitian [4] dan penelitian [19], yaitu menggunakan metode *K-Nearest Neighbor* seperti pada penelitian [4] dan metode SVM seperti pada penelitian [19]. Pebedaan terletak pada objek penelitian dan parameter-parameter analisis yang digunakan.

3. Metodologi

3.1 Metode *K-Nearest Neighbor*

Algoritme *K-Nearest Neighbors* merupakan metode yang menggunakan algoritme *supervised learning*. Algoritme *K-Nearest Neighbor* yaitu suatu metode guna melaksanakan klasifikasi kepada objek didasarkan atas data pengajaran dengan jarak terdekat terhadap objeknya itu [20].

Algoritme *K-Nearest Neighbor* dengan pencarian jarak menggunakan rumus *Euclidian* disajikan berikut:

- 1) Menentukan Nilai K (tetangga) terdekat
- 2) Mempersiapkan data training berupa nilai kriteria suatu data baru yang belum diketahui statusnya.
- 3) Menghitung jarak setiap sampel data traning terhadap data yang akan diuji (data uji) berdasarkan persamaan:

$$d(x, y) = \sqrt{\sum_{i=1}^n (xi - yi)^2} \dots\dots\dots (1)$$

Dengan:

- d = jarak
- xi = sampel data

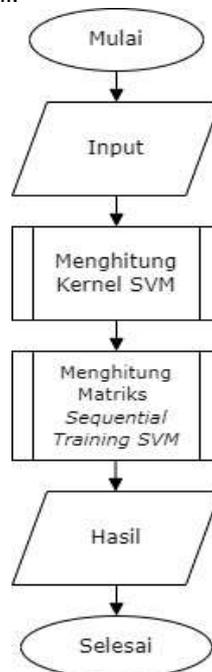
y_i = data uji
 i = variabel data
 n = dimensi data

- 4) Menetapkan status data uji berdasarkan nilai rata-rata K buah sampel data training terdekat.

Proses algoritme *K-Nearest Neighbor* dimulai dari menentukan *dataset*. Setelah *dataset* tersedia, berikutnya mencari parameter K (jumlah tetangga paling dekat). Parameter K pada *testing* ditentukan berdasarkan nilai K optimum pada saat *training*. Selanjutnya melakukan perhitungan metrik jarak diantaranya (*Euclidean distance*), *Manhattan Distance*, *Chebyshev Distance (Maksimal Distance)* selanjutnya setiap objek kepada data sampelnya yang diberi. Kemudian melaksanakan pengurutan berbagai objeknya itu kedalam kelompok berdasarkan bobot *uniform* dan jarak (*distance*). Dengan menggunakan kategori mayoritas, maka didapatkan hasil perhitungan dari algoritme *K-Nearest Neighbor*.

3.2 Metode Support Vector Machine (SVM)

Dalam perhitungan menggunakan algoritme *Support Vector Machines*, ada beberapa langkah-langkah yang dilakukan. Langkah awal dalam proses pengklasifikasian *Support Vector Machines* adalah menginput *dataset* yang akan diolah, lalu dilakukan pemilahan data *train* dan data *test* secara acak. Kemudian mendefinisikan data *train* yang ada untuk dilakukan perhitungan menggunakan beberapa kernel diantaranya kernel linear, RBF, polynomial, dan sigmoid untuk perbandingannya yang selanjutnya diubah menjadi bentuk matriks. Kemudian membuat model dari data *train* yang diproses menggunakan metode algoritme *Support Vector Machine* dan mengklasifikasikan data didasarkan atas model yang telah ditentukan. Kemudian melaksanakan perhitungan nilainya *sequential training* yakni melaksanakan perhitungan kepada testing *Support Vector Machine*, dengan langkah pertama yaitu menghitung nilai biasanya, lalu menghitung nilai $f(x_i, x_{test})$ yakni dengan mengamati nilai paling besar dalam kelas negatif maupun positif melalui penggunaan fungsi kernel yang dipakai. Sesudah nilai sebelumnya diperoleh, tahapan berikutnya yakni menghitung nilai $f(x)_{test}$. Tahapan proses *Support Vector Machine* bisa diamati dalam gambar ini.



Gambar 1. Flowchart Algoritme Support Vector Machines

3.3 Data dan Parameter

Rancangan penelitian terdapat mengenai pembahasan penelitian secara garis besar yang dilakukan oleh penulis dalam menyelesaikan penelitian. Penulis menggunakan dataset

peminat layanan ICONNET satu tahun terakhir yaitu pada Mei 2020 – Mei 2021 yang berasal dari PT. Indonesia Comnets Plus (ICON+) SBU Regional Bali Nusra yang berjumlah 4.141 data.

Analisis dan seleksi data dilakukan dengan menelusuri apakah didalam sebuah data masih terdapat data yang kosong / *null* dalam dataset tersebut. Hasilnya data yang semula berjumlah 4141 data menjadi 3206 data setelah dilakukan seleksi data (*data selection*), seperti Gambar berikut:

	BANDWIDTH	REQUEST_DATE	STATUS	NAMA_PROV	NAMA_KAB	NAMA_KEC	NAMA_KEL	BIAYA_LAYANAN
2	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
3	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
4	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
5	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
6	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
7	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
8	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
9	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
10	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
11	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
12	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
13	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
14	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
15	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
16	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
17	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
18	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
19	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
20	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
21	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
22	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
23	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
24	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
25	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
26	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
27	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
28	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
29	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
30	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
31	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
32	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000
33	10 Mbps	Mei 2020	PA	NUSA TENGGARA BARAT	KOTA MATARAM	AMPENAN	PEJERUK	134000

Gambar 2. Data Peminat ICONNET dengan Atribut yang Telah Disesuaikan

Data dan atribut pada Gambar 2 sudah sesuai dengan atribut yang nantinya akan digunakan dalam melakukan pengolahan datanya. diantaranya:

- Bandwith*, mewakili layanan yang digunakan,
- Request_Date*, mewakili periode waktu berminat berlangganan,
- Status*, mewakili status pelanggan dalam hal ini sudah melakukan pembayaran awal (PA) atau Pendaftar (Register),
- Nama_Provinsi*, *Nama_Kabupaten*, *Nama_Kecamatan*, dan *Nama_Kelurahan* mewakili tempat pelanggan berada,
- Biaya_Layanan* mewakili harga layanan ICONNET.

Dengan demikian data sudah sesuai dan siap untuk dilakukan implementasi *data mining* menggunakan metode *K-Nearest Neighbors* dan *Support Vector Machines* untuk mengetahui perbandingan dari suatu prediksi layanan ICONNET.

3.4 Teknik Pengujian Metode

Pengujian tingkat akurasi metode mempergunakan metode *Cross Validation*, *Confusion Matrix* serta *ROC curve*.

1) *Cross Validation* adalah metode dalam melakukan perkiraan kesalahan prediksi untuk evaluasi kinerja model. Dalam *cross validation* dikenal sebagai estimasi rotasi, dengan membagi data menjadi himpunan bagian K dengan ukuran yang hampir sama, model dalam klasifikasi di *train* dan diuji sebanyak k. Setiap pengulangan salah satu himpunan bagian akan digunakan sebagai data pengujian dan sub kelompok data K lainnya berfungsi sebagai data *train*. Tata cara *K-fold cross validation* sebagai berikut [12]:

- Total instance* dibagi menjadi N bagian,
- Fold ke-1* adalah ketika bagian ke 1 menjadi data uji (*testing data*) dan sisanya menjadi data latihan (*training data*). Selanjutnya, menghitung akurasi suatu hasil pengukuran dengan angka atau data yang sebenarnya berdasarkan porsi data tersebut. Perhitungan akurasi tersebut menggunakan persamaan sebagai berikut :

$$\text{Akurasi} = \frac{\sum \text{jumlah data diprediksi benar}}{\sum \text{jumlah prediksi yang dilakukan}} \times 100 \dots\dots\dots(2)$$

- c) *Fold* ke-2 yaitu saat bagiannya ke-2 menjadi data uji (*testing data*) dan sisanya menjadi data latih (*training data*). Selanjutnya, menghitung akurasi berdasarkan porsi data tersebut, dan seterusnya sampai meraih *fold* ke-k. Hitung rerata akurasi dari k sebuah akurasi diatas. Rata-rata akurasi ini menjadi akurasi final.
- 2) *Confusion matrix* merupakan salah satu evaluasi model yang berupa tabel guna mengungkapkan klasifikasi total data pengujian benar serta total data pengujian yang salah. Penilaian model ini, dilandaskan kepada uji untuk memprediksi objek dengan benar maupun salah. Urutan ujinya tersebut disajikan dalam bentuk tabel yang bernama *confusion matrix*. Setiap sel tabelnya berisikan angka yang menampilkan berapa banyak keluaran yang ada dari kelas yang diamatinya dalam melakukan prediksi. Berikut contoh *confusion matrix* ditampilkan pada Tabel 1 [21].

Tabel 1: *Confusion Matrix*

Klasifikasi		Kelas Sebenarnya	
		1	0
Kelas Prediksi	1	TP	FN
	0	FP	TN

Keterangan:

TP (*True Positive*): Interpretasi: hasil memprediksi positif dan itu benar,

TN (*True Negative*): Interpretasi: hasil memprediksi negative dan itu benar,

FP (*False Positive*): Interpretasi: hasil memprediksi positif dan itu salah,

FN (*False Negative*): Interpretasi: hasil memprediksi negative dan itu salah.

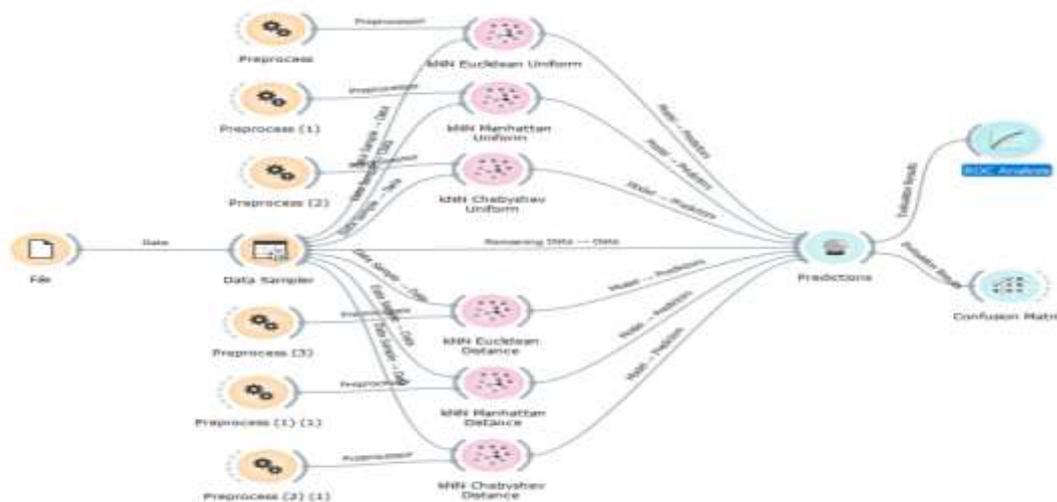
- 3) *ROC (Receiver Operating Characteristics) curve* merupakan teknik dalam menunjukan teknik klasifikasi mengacu pada performa. ROC dapat memvisualisasikan *confusion matrix*. ROC merupakan sebuah grafik dua dimensi dengan *false positive* yang merupakan garis horizontal dan *true positive* yang merupakan garis vertikal. Kurva inilah yang menunjukkan tingkat akurasi dari suatu model [22].

Penyajian kepada evaluasi AUC untuk mengklasifikasi model prediksi yang diciptakan dapat dikelompokan sebagai berikut [21]:

- >0.90 – 1 = Luar biasa
- >0.8 – 0.9 = Sangat baik
- >0.7 – 0.8 = Baik
- >0.6 – 0.7 = Cukup baik
- 0.5 – 0.6 = Tidak baik

4) Hasil dan Pembahasan

4.1 Pengolahan *Data Mining* Menggunakan Algoritme *K-Nearest Neighbors*.



Gambar 3. Tampilan Pengolahan Data Mining Algoritme K-NN pada *Orange Data Mining*

Pada penelitian ini, *dataset* akan dibagikan jadi data *training* serta data *testing*. Rasio atau perbandingan data yang dipakai dalam menentukan data *training* serta data *testing* adalah didasarkan pada Prinsip Pareto yang dikenal dengan aturan 80/20 [23]. Maka ditentukanlah data *training* 80% dengan data berjumlah 2.565 data, sedangkan data *testing* 20% dengan data berjumlah 641. Gambar 3 merupakan tampilan implementasi pengolahan *K-Nearest Neighbor* pada *Orange Data Mining*

Berikut merupakan hasil dari evaluasi dari 6 model *K-Nearest Neighbor*.

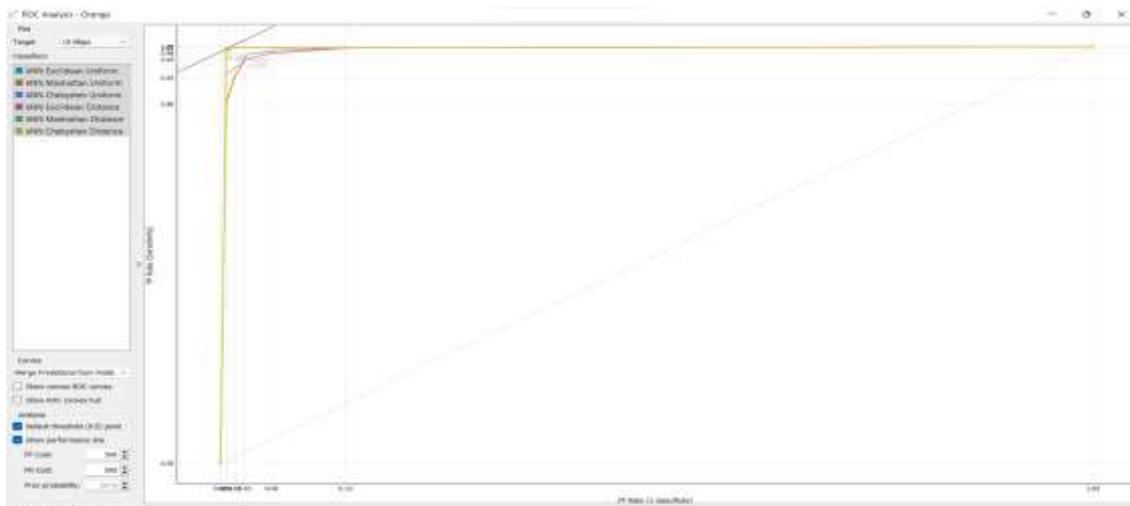
Model	AUC	CA	F1	Precision	Recall
kNN Euclidean Uniform	0.990	0.958	0.957	0.959	0.958
kNN Manhattan Uniform	0.988	0.953	0.952	0.954	0.953
kNN Chebyshev Uniform	0.994	0.969	0.969	0.970	0.969
kNN Euclidean Distance	0.994	0.986	0.986	0.986	0.986
kNN Manhattan Distance	0.992	0.980	0.980	0.980	0.980
kNN Chebyshev Distance	0.996	0.992	0.992	0.992	0.992

Gambar 4. Hasil Evaluasi Nilai KNN

Dari hasil evaluasi 6 model *K-Nearest Neighbor*, diketahui dari 641 *data testing*, didapat bahwa KNN dengan metrik *Chebyshev* berdasarkan bobot *distance* merupakan algoritme yang paling unggul dibandingkan dengan metrik KNN lainnya. *Classification accuracy* yang didapat adalah 99,2%. Berikut tampilan dari *confusion matrix* KNN *Chebyshev Distance* pada Gambar 5.

		Predicted				Σ
		10 Mbps	20 Mbps	50 Mbps	100 Mbps	
Actual	10 Mbps	317	3	0	0	320
	20 Mbps	2	286	0	0	288
	50 Mbps	0	0	25	0	25
	100 Mbps	0	0	0	8	8
Σ		319	289	25	8	641

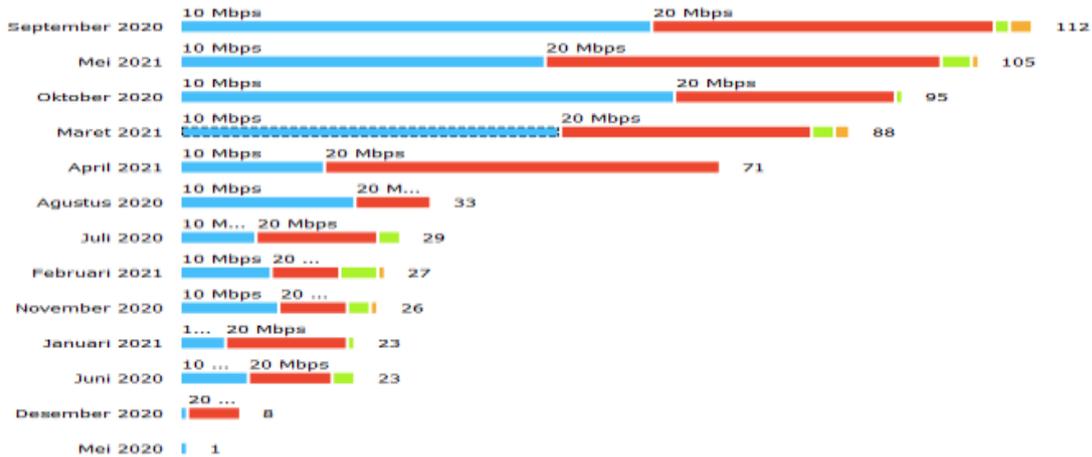
Gambar 5 Hasil *Confusion Matrix* KNN *Chebyshev Distance*



Gambar 6. Hasil ROC Analysis KNN

Selain itu grafik ROC menginformasikan AUC yang dihasilkan algoritme *K-Nearest Neighbor* sebesar 0,996 atau 99,6% sehingga termasuk dalam kelompok “*Excellent Classification*”. Berikut tampilan hasil *ROC Curve* pada gambar 6.

Dari tingkat akurasi dan *ROC Curve* yang didapat dari *KNN Chebyshev Distance* tersebut, adapun pola data prediksinya divisualisiskan dalam Gambar 7.

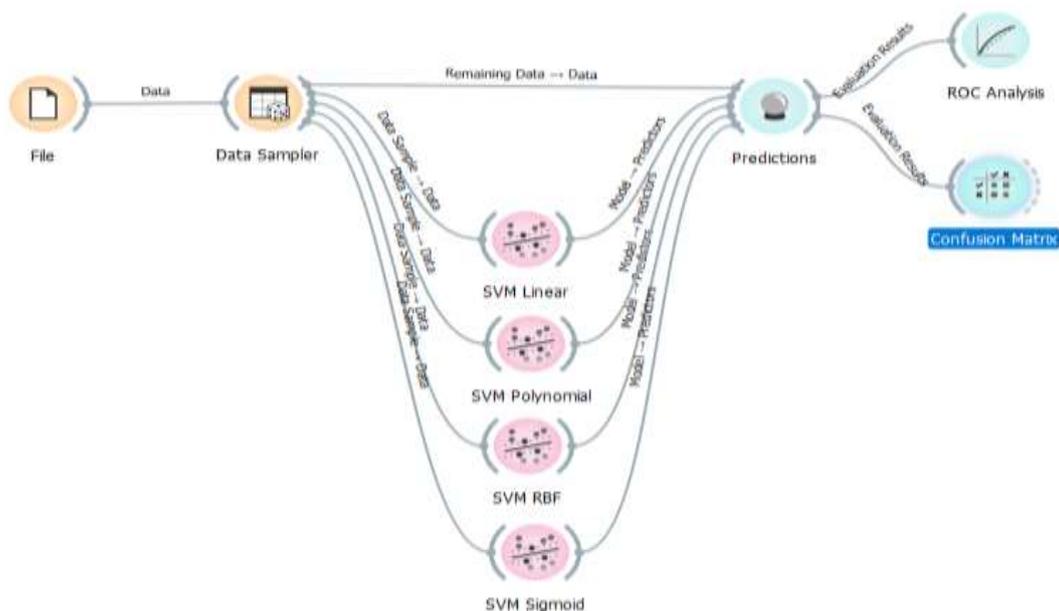


Gambar 7. Hasil Prediksi *KNN Chebyshev Distance* Produk ICONNET Terlaris

Pada Gambar 7, prediksi penjualan produk ICONNET terlaris terdapat pada bulan September, berdasarkan data dari rentang waktu Mei 2020 hingga Mei 2021. Produk bandwidth yang paling banyak diminati oleh pelanggan adalah produk ICONNET 10 Mbps.

4.2 Proses Pengolahan *Data Mining* Menggunakan Algoritme *Support Vector Machines*.

Dalam proses pengolahan *data mining* menggunakan SVM, penulis menggunakan metode perhitungan SVM yang sudah disediakan oleh Orange Data Mining yaitu diantaranya SVM Linear, SVM Polynomial, SVM RBF (*Radial Basis Function*) dan SVM Sigmoid. Hasilnya terlihat pada Gambar 8.



Gambar 8. Tampilan Pengolahan *Data Mining* SVM

Gambar 9 merupakan hasil dari evaluasi dari 4 model *Support Vector Machine*.

Model	AUC	CA	F1	Precision	Recall
kNN Euclidean Uniform	0.990	0.958	0.957	0.959	0.958
kNN Manhattan Uniform	0.988	0.953	0.952	0.954	0.953
kNN Chebyshev Uniform	0.994	0.969	0.969	0.970	0.969
kNN Euclidean Distance	0.994	0.986	0.986	0.986	0.986
kNN Manhattan Distance	0.992	0.980	0.980	0.980	0.980
kNN Chebyshev Distance	0.996	0.992	0.992	0.992	0.992

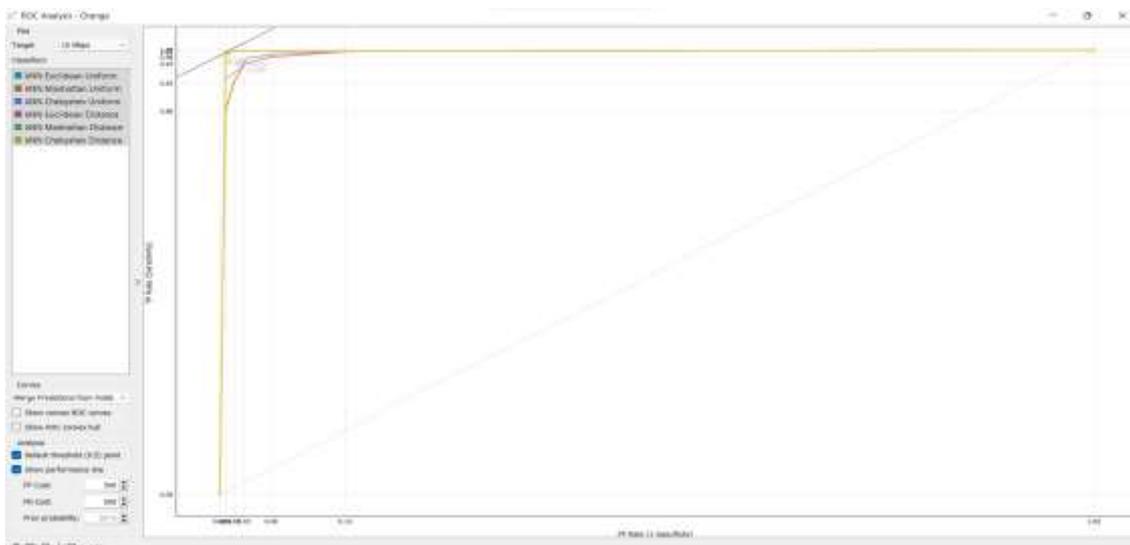
Gambar 9. Hasil Evaluasi Nilai *Support Vector Machine*

Dari hasil di atas, SVM dengan kernel Linear lebih unggul. Tingkat akurasi yang dihasilkan oleh algoritme SVM Linear adalah mencapai 94,4% dengan *Precision* 94.8%. Lalu nilai *Recall* didapat 94,4% dan nilai *F-1 Score* 94,4%. Berikut tampilan dari *cross validation* SVM Linear pada gambar 10.

	Predicted				Σ
	10 Mbps	20 Mbps	50 Mbps	100 Mbps	
10 Mbps	287	33	0	0	320
20 Mbps	3	285	0	0	288
50 Mbps	0	0	25	0	25
100 Mbps	0	0	0	8	8
Σ	290	318	25	8	641

Gambar 10. Hasil *Confusion Matrix* SVM Linear

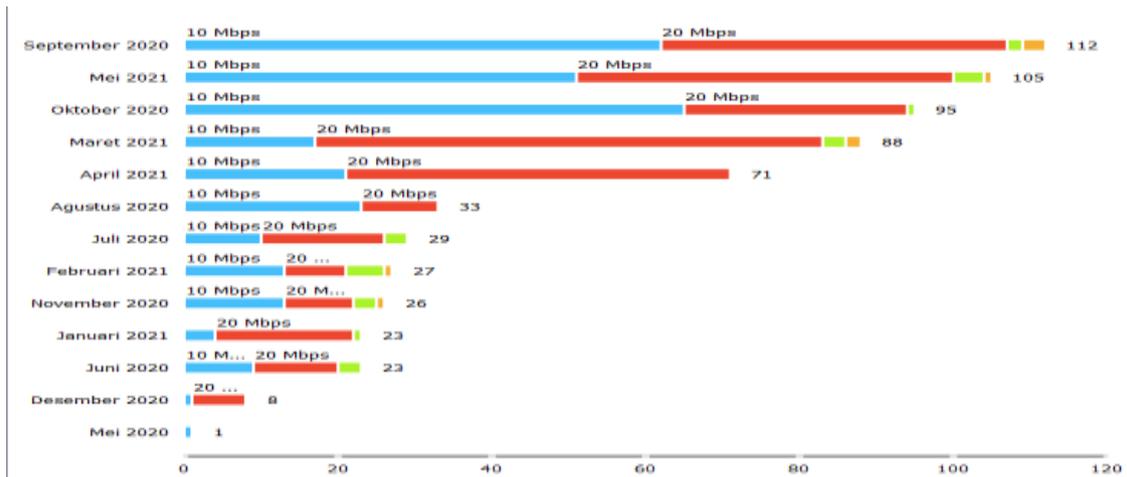
Dilanjutkan dengan nilai ROC Curve/AUC (*Area Under the Curve*) yang tertinggi didapatkan oleh SVM Linear yaitu 0,973. Berikut tampilan *ROC Curve Support Vector Machine*.



Gambar 10 Hasil *ROC Analysis* SVM

Dari tingkat akurasi dan ROC Analysis yang didapat dari SVM Linear tersebut adapun pola data prediksinya divisualisikan dalam Gambar 11.

Pada Gambar 11. diatas diketahui prediksi berdasarkan algoritme SVM Linear penjualan produk ICONNET terlaris terdapat pada bulan September berdasarkan data dari rentang waktu Mei 2020 hingga Mei 2021. Produk bandwidth yang paling banyak diminati oleh pelanggan adalah produk ICONNET 10 Mbps



Gambar 11. Hasil Prediksi SVM Linear Produk ICONNET Terlaris

4.3 Hasil Evaluasi Perbandingan Algoritme *K-Nearest Neighbors* dan *Support Vector Machine* menggunakan *Cross Validation*

Hasil dari evaluasi *cross validation* menggunakan model algoritme *K-Nearest Neighbor Chebyshev Distance* dan *Support Vector Machine Linear* berdasarkan 3026 data peminat layanan ICONNET yang akan dilakukan pengujian, maka diperoleh hasil dari *accuracy*, *precision*, *recall*, dan *f-1 score* bisa dilihat pada Tabel 2 berikut.

Tabel 2. Hasil Evaluasi *Cross Validation*

Algoritme	Accuracy	Precision	Recall	F1-Score	AUC / ROC Curve
KNN Chebyshev Distance	98,4%	98,4%	98,4%	98,4%	99,5%
SVM Linear	96,5%	96,5%	96,5%	96,5%	99,6%

Hasil klasifikasi pada algoritme KNN *Chebyshev Distance* dan SVM Linear bisa dilihat pada Tabel 1, bahwa nilai akurasi dari *K-Nearest Neighbor Chebyshev Distance* paling unggul dibanding *SVM Linear* yaitu **98,4 %**. Berdasarkan Tabel 1 diatas, diperlihatkan juga perbandingan dari dua model AUC. Terlihat bahwa nilai AUC yang paling tinggi adalah metode *Support Vector Machine Linear* yaitu **99,6%** dengan kategori *Excellent Classification* (sangat baik). Berbeda tipis dengan KNN Chebyshev Distance yaitu sebesar **99,5%** (*Excellent Classification*).

4.4 Evaluasi Hasil Perbandingan Prediksi Algoritme *K-Nearest Neighbor* dan *Support Vector Machine*

Dibawah ini adalah Tabel 3 dari perbandingan hasil klasifikasi prediksi Produk Layanan ICONNET mempergunakan metode *K-Nearest Neighbor* serta *Support Vector Machine*. Klasifikasi prediksi tersebut sudah dilakukan pembagian data training maupun data testing. Menggunakan perbandingan data training 80% serta data testing 20%.

Tabel 3. Hasil Evaluasi Perbandingan KNN dan SVM

Algoritme	Accuracy	Precision	Recall	F1-Score	AUC / ROC Curve
KNN Chebyshev Distance	99,2%	99,2%	99,2%	99,2%	99,6%
SVM Linear	94,4%	94,8%	94,4%	94,4%	97,3%

4.5 Pembahasan

Berdasarkan hasil evaluasi perbandingan algoritme K-NN dan SVM, didapat bahwa nilai akurasi tertinggi diperoleh oleh algoritme *K-Nearest Neighbor Chebyshev Distance* dengan nilai **99,2%** dilanjutkan oleh SVM Linear **94,4%**. Sedangkan untuk nilai *Precision* tertinggi diperoleh oleh algoritme *K-Nearest Neighbor Chebyshev Distance* **99,2%** dilanjutkan oleh SVM Linear **94,8%**. Lalu untuk nilai *Recall* tertinggi didapatkan algoritme *K-Nearest Neighbor Chebyshev Distance* yaitu sebesar **99,2%** dilanjutkan oleh dan SVM Linear **94,4%**. Selain itu, itu *F-1 Score* tertinggi diperoleh *K-Nearest Neighbor Chebyshev Distance* dengan **99,2%** dilanjutkan dengan SVM Linear **99,2%**. Selanjutnya untuk nilai AUC / *ROC Curve* tertinggi diperoleh oleh *K-Nearest Neighbor Chebyshev Distance* dengan nilai **99,6%** dilanjutkan dengan SVM Linear dengan perolehan nilai **97,3%**. Hasil pengujian akurasi tersebut menunjukkan kinerja algoritme *K-Nearest Neighbor Chebyshev Distance* lebih tinggi pada seluruh kriteria dibandingkan algoritme SVM Linear, baik dari sisi nilai **accuracy, precision, recall, f-1 score** dan nilai **ROC Curve**. Dengan demikian, dapat dikatakan bahwa kinerja algoritme *K-Nearest Neighbor* lebih baik dari algoritme SVM. Temuan ini sama dengan temuan Naufal, Adiwijaya, dan Astuti [24] yang menemukan kinerja akurasi algoritme K-NN yang lebih tinggi dari akurasi algoritme SVM dalam kasus deteksi penyakit kanker pada manusia, namun berbeda dengan temuan Budianto, Ariyuna, dan Maryono [25] yang menemukan kinerja akurasi algoritme K-NN yang lebih rendah dari akurasi algoritme SVM dalam kasus pengenalan karakter plat kendaraan bermotor

5. Simpulan

Prediksi layanan produk ICONNET diperoleh hasil bahwa kinerja *K-Nearest Neighbor* dengan metrik *Chebyshev Distance* lebih unggul dari *SVM Linear*. Didapatkan bahwa dari data *training* sebanyak 2565 data dan data *testing* 641 data. Dengan menggunakan *K-Nearest Neighbor Chebyshev Distance* memiliki nilai *accuracy* 99,2% dan *precision* 99,2% sedangkan SVM Linear memiliki nilai akurasi 94,4% dan *precision* 94,8%. Hasil dari implementasi *K-Nearest Neighbor Chebyshev Distance* cukup baik, dilihat dari nilai *Accuracy, Precision, Recall*, dan *F1-Score* yang mendapatkan nilai diatas 90%. Lalu nilai AUC yang didapat 99,6% yang dikategorikan kelompok "*Excellent Classification*" (sangat baik). Berdasarkan prediksi *KNN Chebyshev Distance* maupun SVM Linear, produk ICONNET terlaris terdapat pada bulan September dengan produk ICONNET 10 Mbps yang paling banyak diminati.

Kontribusi riset ini bisa digunakan oleh pihak manajemen PT. Indonesia Comnets Plus (ICON+) untuk memprediksi dalam melakukan pengambilan keputusan mengenai produk ICONNET kedepannya dan bisa memberikan masukan untuk dijadikan sebagai bahan pertimbangan pengetahuan mengenai implementasi *data mining* pada layanan ICONNET menggunakan metode *K-Nearest Neighbors* yang sangat direkomendasikan untuk digunakan dikarenakan lebih unggul dari SVM Linear.

Daftar Referensi

- [1] S. Nuryudayatun, "Membangun Kepuasan Pelanggan dengan Melakukan Customer Understanding". *Jurnal Bisnis, Manajemen, dan Akuntansi*, vol. 1, no. 2, pp. 17-24, 2013.
- [2] D. T. Larose, *Discovering knowledge in data: an introduction to data mining*. Hoboken, N.J: Wiley-Interscience, 2005.
- [3] D. T. Larose, *Data Mining Methods and Models: Larose/Data Mining Methods and Models*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005.
- [4] A. U. Haspriyanti and P. T. Prasetyaningrum, "Penerapan Data Mining Untuk Prediksi Layanan Produk Indihome Menggunakan Metode K-NearstNeighbor," *Journal of Information System and Artificial Intelligence*, vol. 1, no. 2, pp. 57–64, 2021.
- [5] T. Praningki and Budi, "Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN". *Creative Information Technology Journal*, vol. 4, no. 2, pp. 83-93, 2018.
- [6] H.A. Wowiling and J. Johan, "Aplikasi Diagnosis Penyakit Kulit Pada Kucing Menggunakan K-Nearest Neighbor (K-NN)". *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi (JMApTeKsi)*, vol. 2, no. 1, pp. 56-61, 2020.
- [7] Y. Umaidah and P. Purwantoro, "Penerapan algoritme K-Nearest Neighbor (K-NN) Dengan Pencarian Optimal untuk Prediksi Prestasi Siswa". *Journal of Information System, Informatics and Computing*, vol. 3, no. 2, pp. 1-8, 2019.

- [8] A.S. Mananta and G.A. Sandag, "Prediksi Kelulusan Mahasiswa Dalam Memilih Program Magister Menggunakan Algoritme K-NN". *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 10, no. 2, pp. 90-96, 2021.
- [9] H.S. Perdana, R. Ruliah, and S. Fathimah, "Penentuan Keberhasilan Inseminasi Buatan Pada Sapi Menggunakan K-Nearest Neighbor". *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 7, no. 1, pp. 11-20, 20128.
- [10] R. Tineges, A. Triayudi, and I. D. Sholihati, "Analisis Sentimen Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM)," *mib*, vol. 4, no. 3, pp. 650-658, 2020, doi: 10.30865/mib.v4i3.2181.
- [11] A. Darmawan, N. Kustian, and W. Rahayu, "Implementasi data mining menggunakan model svm untuk prediksi kepuasan pengunjung taman tabebuya". *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 2, no. 3, pp. 299-307, 2018.
- [12] N.P.N. Hendayanti, I.K.P. Suniantara, and M. Nurhidayati, "Penerapan Support Vector Regression (SVR) Dalam Memprediksi Jumlah Kunjungan Wisatawan Domestik Ke Bali". *Jurnal Varian*, vol. 3, no. 1, pp. 43-50, 2019.
- [13] Y.V. Wijaya, A. Erfina, and C. Warman, "Analisis Sentimen Seputar UU ITE Menggunakan Algoritma Support Vector Machine". *Progresif: Jurnal Ilmiah Komputer*, vol. 17, no. 2, pp. 1-14, 2021.
- [14] Y. Ramdhani and A. Mubarak, "Analisis Time Series Prediksi Penutupan Harga Saham Antm. Jk Dengan Algoritme SVM Model Regresi". *Jurnal Responsif: Riset Sains dan Informatika*, vol. 1, no. 1, pp. 77-82, 2019.
- [15] S. Wahyuni, I. Sulistianingsih, E. Hariyanto, and O.C.V. Lumbanbatu, "Data Mining Prediksi Minat Customer Penjualan Handphone Dengan Algoritme Apriori". *JURNAL UNITEK*, vol. 14, no. 2, pp. 10-19, 2021.
- [16] P.F. Longgana, I. Irvan, and A.H. Wilarto, "Penentuan Minat Konsumen Terhadap Produk Menggunakan Algoritme Apriori Pada PT. Telkom Indonesia". *Jurnal Teknoinfo*, vol. 16, no. 2, pp. 340-353, 2022.
- [17] H. Susana, S. Khofidoh, and A. Al Afghani, "Implementasi Data Mining Metode C. 45 untuk Memprediksi Peminat Kuota Internet pada Masa Covid 19. *MEANS (Media Informasi Analisa dan Sistem)*, vol. 6, no. 2, pp. 155-157, 2021.
- [18] D. Nofriansyah, K. Erwansyah, and M. Ramadhan, "Penerapan Data Mining dengan Algoritme Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi). *Jurnal Saintikom*, vol. 15, no. 2, pp. 81-92, 2016
- [19] A. Sarino, "Upaya Menciptakan Kepuasan Pelanggan Dengan Pengelolaan Service Quality (Servqual)," *Jurnal MANAJERIAL*, vol. 9, no. 2, pp. 19-24, 2010, doi: 10.17509/manajerial.v9i2.1798.
- [20] T. Thi Bi Dan, S. Widya Sihwi, and R. Anggrainingsih, "Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret," *ITSmart*, vol. 4, no. 2, pp. 84-91, 2016, doi: 10.20961/its.v4i2.1770.
- [21] J. Demšar et al., "Orange: Data Mining Toolbox in Python," *Journal of Machine Learning Research*, vol. 14, no. 35, pp. 2349-2353, 2013.
- [22] Defiyanti, "Analisis dan Prediksi Kinerja Mahasiswa Menggunakan Teknik Data Mining," *Syntax J. Inf.*, vol. 2, no. 1, pp. 1-8, 2016, doi: 10.35706/syji.v2i01.192.
- [23] I. Loelianto, Moh. S. S. Thayf, and H. Angriani, "Implementasi Teori Naive Bayes Dalam Klasifikasi Calon Mahasiswa Baru Stmik Kharisma Makassar," *SINTECH Journal*, vol. 3, no. 2, pp. 110-117, 2020, doi: 10.31598/sintechjournal.v3i2.651
- [24] S.A. Naufal, A. Adiwijaya, and W. Astuti, "Analisis Perbandingan Klasifikasi Support Vector Machine (SVM) dan K-Nearest Neighbors (KNN) untuk Deteksi Kanker dengan Data Microarray. *JURIKOM (Jurnal Riset Komputer)*, vol. 7, no. 1, pp. 162-168, 2020.
- [25] A. Budianto, R. Ariyuana, and D. Maryono, "Perbandingan K-Nearest Neighbor (KNN) Dan Support Vector Machine (SVM) Dalam Pengenalan Karakter Plat Kendaraan Bermotor. *Jurnal Ilmiah Pendidikan Teknik dan Kejuruan*, vol. 11, no. 1, pp. 27-35, 2018.