

Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada “Lombok Vape On”

Yahya^{1*}, Winda Puspita Hidayanti²

^{1,2}Program Studi Sistem Informasi, Universitas Hamzanwadi
^{*}ayhay7078@gmail.com

Abstrak

Perusahaan yang dibangun dengan tujuan untuk menghasilkan barang atau jasa yang ditujukan untuk memenuhi kebutuhan konsumen juga untuk memperoleh keuntungan bagi perusahaan tersebut. Perusahaan sudah tentu memiliki tujuan yang sama yaitu mencapai target yang sudah direncanakan dan mampu menghasilkan pemasukan yang selalu meningkat disetiap tahunnya. Dan salah satu kegiatan usaha yang harus dilakukan agar perusahaan tetap berjalan dan berkembang adalah penjualan. Keputusan yang diambil pemegang tanggung jawab perusahaan akan mempengaruhi perusahaan dimasa depan. Salah satu keputusan yang harus ditentukan yaitu produk yang akan diproduksi dan dijual untuk periode selanjutnya. Dilihat dari era globalisasi saat ini bahwa usaha yang dapat mengikuti zaman yaitu produk rokok elektrik (Vape). Produk tersebut benar-benar diminati oleh kalangan anak muda hususnya para pria. Salah satu instansi yang mengambil keputusan untuk menggeluti bisnis tersebut yaitu Lombok Vape On. “Lombok Vape On” memiliki konsumen yang cukup banyak sehingga pendapatan yang didapat tentu banyak, akan tetapi pendapatannya tidak menentu dan tidak dapat diprediksi dengan baik setiap bulannya. Tujuan dari penelitian ini untuk mengklasifikasikan hasil penjualan dari Lombok Vape on untuk mengetahui efektifitas penjualan perbulannya agar toko tersebut dapat mengetahui pasang surut pemasukan yang didapat. Apabila instansi tersebut mengetahui tingkat penjualannya maka instansi tersebut dengan mudah mengubah strategi penjualan agar stabil dan meningkat. Untuk itu peneliti menerapkan metode K-Nearest Neighbor untuk mengklasifikasin masalah tersebut,dan diharapkan metode ini mampu mengatasi masalah pada “Lombok Vape On”.

Kata Kunci : *K-NN, Vape, Lombok Vape On*

Abstract

Companies that are built with the aim of producing goods or services that are intended to meet the needs of consumers as well as to obtain profits for the company. The company certainly has the same goal which is to achieve the planned target and be able to generate revenue that is always increasing every year. And one of the business activities that must be done so that the company continues to run and develop is sales. The decision taken by the holder of corporate responsibility will affect the company in the future. One of the decisions that must be determined is the product to be produced and sold for the next period. Judging from the current era of globalization, businesses that can keep up with the times are electric cigarette products (vape). These products are really in demand by young people especially men. One of the agencies that made the decision to run the business was Lombok Vape On. Lombok Vape On has quite a lot of consumers so that the income earned is certainly a lot, but the income is uncertain and cannot be well predicted every month. The purpose of this study is to classify the sales results from Lombok Vape on to determine the effectiveness of monthly sales so that the store can know the ups and downs of income obtained. If the agency knows the level of sales, the agency can easily change the sales strategy to be stable and increase. For this reason researchers applied the K-Nearest Neighbor method to classify the problem, and it is hoped that this method is able to overcome the problem in Lombok Vape On.

Keywords: *K-NN, Vape, Lombok Vape On*

1. Pendahuluan

Penjualan merupakan salah satu unsur penting dalam suatu bisnis di bidang pemasaran, setiap bisnis memiliki tujuan yang umum disetiap langkah dan prosesnya, tujuan yang dimaksud mampu mencapai target yang sudah direncanakan dan mampu menghasilkan pemasukan yang selalu meningkat disetiap tahunnya, selain itu bisnis yang baik memiliki tujuan jangka panjang supaya masa yang akan datang dapat mengalami perubahan ke arah yang lebih baik. Kegiatan perusahaan yang bisa menghasilkan keuntungan adalah penjualan, dan penjualan merupakan salah satu faktor terpenting dalam bisnis. Dengan adanya perkembangan yang semakin pesat di era globalisasi ini, pebisnis berlomba-lomba untuk mendapatkan banyak keuntungan dengan berbagai macam kegiatan penjualan dimana semakin banyak barang yang terjual maka semakin banyak pula keuntungan yang dicapai. Salah satu bisnis yang fenomena dikalangan masyarakat saat ini yaitu bisnis Vape (Rokok Elektrik) dimana belakangan ini kita tahu bahwa rokok elektrik sudah mulai diminati oleh masyarakat banyak terutama pada perkembangan gaya hidup dikalangan pria, namun selain pria, saat ini wanita remaja juga menjadi salah satu peminat dari rokok elektrik tersebut. Salah satu toko di Nusa Tenggara Barat yang memilih untuk mendirikan bisnis rokok elektrik yaitu "Lombok Vape on", toko tersebut menghasilkan penjualan yang cukup besar setiap tahun, tetapi hasil penjualan yang

dihasilkan tidak stabil dan tidak mampu diprediksi kenaikan dan penurunan penjualannya.

Dari penelitian yang dilakukan, berusaha untuk mengklasifikasikan hasil penjualan dari "Lombok Vape On" untuk mengetahui efektifitas penjualan. Data yang digunakan adalah data hasil penjualan tahun 2019 dan mengklasifikasikan hasil penjualan pada "Lombok Vape On" berdasarkan jumlah item atau jenis barang yang mampu terjual.

Dengan melakukan klasifikasi pada toko tersebut dapat mengetahui pasang surut pemasukan yang didapat dari toko "Lombok Vape On". Untuk mengetahui efektifitas dari toko tersebut, dilakuakn dengan menggunakan salah satu metode dalam data mining yang merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database yang besar, sedangkan metode yang digunakan yaitu metode K-Nearest Neighbor.

Kelebihan menggunakan metode K-Nearest Neighbor diperkirakan dapat memberikan hasil prosentase yang cukup bagus sehingga prosentase yang dihasilkan mampu membuat toko "Lombok Vape On" mengetahui pasang surut dari hasil penjualan. Apabila toko mengetahui peningkatan atau penurunan dari hasil penjualannya maka toko dapat segera mengambil tindakan untuk mengubah dan meningkatkan strategi yang jauh lebih baik lagi.

Berdasarkan uraian di atas, maka penelitian yang dilakukan adalah Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) Pada “Lombok Vape On”.

2. Tinjauan Pustaka

2.1. Penelitian terkait

Dalam penulisan skripsi ini, penulis terinspirasi menggunakan referensi dari penelitian-penelitian sebelumnya yang berkaitan dengan pembahasan skripsi yang dibuat. Penelitian ini terkait penggunaan data mining dengan menggunakan algoritma K-Nearest Neighbor. Berikut ini penelitian terdahulu yang berkaitan dengan proposal skripsi antara lain:

- Penelitian yang dilakukan (Resti Hutami, 2016), algoritma ini digunakan untuk melakukan prediksi data penjualan furniture pada CV. Octo Agung Jepara. Hasil penelitian menunjukkan metode yang diusulkan berhasil diimplementasikan untuk menyelesaikan kasus prediksi penjualan dengan tingkat error sebesar 6 persen dan akurasi 94 persen^[1].
- Penelitian yang dilakukan oleh Nobertus Krisandi. Dkk dengan judul “Algoritma K-Nearest Neighbor dalam Klasifikasi Data Hasil Produksi Kelapa Sawit pada PT. Minamas Kecamatan Parindu”. Data yang digunakan adalah data hasil produksi kelapa sawit (Tonase) dari 50 kelompok tani pada periode Juli-Desember 2011 pada PT. Minasa Kabupaten Sanggau. Nilai k yang

digunakan adalah k=1, k=3, k=5 dan k=7. Hasil penelitian menunjukkan bahwa, hasil produksi yang dominan adalah dengan nilai k=7 sebesar yang memiliki nilai accuracy sebesar 34%. Hal tersebut juga mengindikasikan bahwa K-Nearest Neighbor (KNN) dipengaruhi oleh jumlah klastering data^[2].

- Penelitian yang dilakukan oleh Ricky Imanuel Ndaumanu, Kusriani, M. Rudyanto Arief yang berjudul Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa Dengan Metode K-Nearest Neighbor. Pada penelitian ini, dikemukakan bahwa dari jumlah pendaftaran mahasiswa baru ini, banyak juga mahasiswa yang mengundurkan diri setiap tahunnya yang disebabkan berbagai masalah. Oleh karena adanya mahasiswa yang mengundurkan diri, penulis melakukan analisis pengunduran diri mahasiswa menggunakan algoritma K-Nearest Neighbor. Penelitian ini bertujuan untuk memprediksi seberapa tingkat akurasi pengunduran diri mahasiswa STIKOM UYELINDO Kupang^[3].

2.2. Landasan Teori

1. Pengertian Data Mining

Data *mining* merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data. Data *mining* adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari

pengolahan data dengan metode data mining ini dapat digunakan untuk mengambil keputusan di masa depan. Salah satu teknik yang dibuat dalam data *mining* adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan. Kebutuhan untuk prediksi juga dapat memanfaatkan teknik ini. Dalam data *mining*, pengelompokan data juga bisa dilakukan. Tujuannya adalah agar kita dapat mengetahui pola universal data-data yang ada^[1].

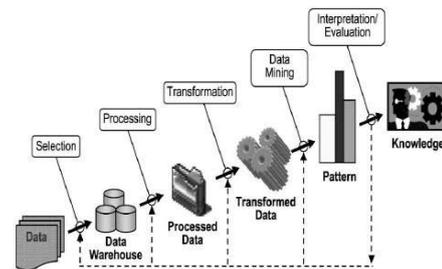
2. Pengertian Penerapan

Menurut Kamus Besar Bahasa Indonesia (KBBI), pengertian penerapan adalah perbuatan menerapkan, sedangkan menurut beberapa ahli, penerapan adalah suatu perbuatan mempraktekkan suatu teori, metode, dan hal lain untuk mencapai tujuan tertentu dan untuk suatu kepentingan yang diinginkan oleh suatu kelompok atau golongan yang telah terencana dan tersusun sebelumnya.

3. Tahapan Data Mining

Menurut Fayyad, 1996 dalam jurnal Yuli Sudriani yang berjudul "Tren Analisa Data Berskala Besar Terkait Penelitian Ekologi" menyatakan tahapan yang dilakukan pada proses data *mining* diawali dari seleksi data dari data sumber ke data target, tahap *preprocessing* untuk memperbaiki kualitas data, transformasi, data *mining* serta tahap interpretasi dan evaluasi yang menghasilkan

output berupa pengetahuan baru yang diharapkan memberikan kontribusi yang lebih baik secara detail dijelaskan sebagai berikut :



Gambar 1. Tahapan data mining

a. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional. Data hasil seleksi akan digunakan untuk proses data *mining*, dan disimpan dalam suatu berkas dan terpisah dari basis data operasional.

b. Cleaning Data

Proses *cleaning* dilakukan dengan menghilangkan noise, membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*).

c. Transformasi Data

Tahap ini merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data *mining*.

d. Data mining

Data mining merupakan proses mencari pola atau informasi yang sangat menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

e. Interpretation/Evaluation

Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan

bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

4. Klasifikasi

Klasifikasi adalah satu bentuk analisis data yang *menghasilkan* model untuk mendeskripsikan kelas data yang penting. Klasifikasi memprediksi kategori ke dalam label class. Klasifikasi merupakan proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau class data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui^[2].

Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *Decision/Classification Trees, Bayesian Classifiers/Naïve Bayes Classifiers, Neural Networks, Analisa Statistik, Algoritma Genetika, Rough Sets, K-Nearest Neighbor, Metode Rule Based, Memory Based Reasoning, dan Support Vector Machines (SVM)*. Proses klasifikasi didasarkan pada empat komponen (Gorunescu, 2011) ^[3]:

a. Kelas

Variabel dependen yang berupa kategorikal yang merepresentasikan label yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, *customer loyalty*, jenis gempa. *J Pikel* 1(1) : 65 -76 (2013).

b. Predictor Variabel independen

Predictor Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.

c. Training Datasets

Training dataset satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.

d. Testing dataset

Testing dataset berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi.

5. Algoritma K-NN (K-Nearest Neighbor)

K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma *supervised* dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasarkan atribut dan sampel latih.

Pengklasifikasian tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik uji, akan ditemukan sejumlah K objek (titik training) yang paling dekat dengan titik uji. Klasifikasi menggunakan *voting* terbanyak di antara klasifikasi dari K objek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *Euclidian*^[4].

Langkah-langkah untuk menghitung algoritma K-Nearest Neighbor:

- Menentukan parameter K (Jumlah tetangga paling dekat).

- Menghitung kuadrat jarak Euclid (queri instance) masing-masing objek terhadap data sampel yang diberikan.
- Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak Euclid terkecil.
- Mengumpulkan kategori Y (Klasifikasi Nearest Neighbor)

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data training), diantaranya euclidean distance dan manhattan distance (city block distance), yang paling sering digunakan adalah euclidean distance (Bramer,2007), yaitu^[5]:

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

Dimana $a = a_1, a_2, \dots, a_n$, dan $b = b_1, b_2, \dots, b_n$ mewakili n nilai atribut dari dua record. Untuk atribut dengan nilai kategori, pengukuran dengan euclidean distance tidak cocok. Sebagai penggantinya, digunakan fungsi sebagai berikut (Larose, 2006): $\text{different}(a_i, b_i) = \begin{cases} 0 & \text{jika } a_i = b_i \\ 1 & \text{lainnya} \end{cases}$. Dimana a_i dan b_i adalah nilai kategori. Jika nilai atribut antara dua record yang dibandingkan sama maka nilai jaraknya 0, artinya mirip, sebaliknya, jika berbeda maka nilai kedekatannya 1, artinya tidak mirip sama sekali. Misalkan atribut warna dengan nilai merah dan merah, maka nilai kedekatannya 0, jika merah dan biru maka nilai kedekatannya 1.

Untuk menghitung kemiripan kasus, digunakan rumus (Kusrini, 2009):

$$\text{Similarity}(p, q) = \frac{n \sum_{t=1}^n f(p_i, q_i) \times w_i}{w_i}$$

Keterangan :

P= Kasus baru

q = Kasus yang ada dalam penyimpanan

n = Jumlah atribut dalam tiap kasus

i = Atribut individu antara 1 sampai dengan n

f = Fungsi similarity atribut i antara kasus p dan kasus q

w= Bobot yang diberikan pada atribut ke-i

6. Definisi Dataset

Untuk istilah mainframe IBM untuk suatu file, lihat Kumpulan data (mainframe IBM). Untuk perangkat antarmuka telekomunikasi, lihat Modem. Kumpulan data (atau dataset) adalah kumpulan data. Dalam kasus data tabular, satu set data sesuai dengan satu atau lebih tabel database, di mana setiap kolom tabel mewakili variabel tertentu, dan setiap baris sesuai dengan catatan tertentu dari set data yang dimaksud. Kumpulan data mencantumkan nilai untuk setiap variabel, seperti tinggi dan berat objek, untuk setiap anggota kumpulan data. Setiap nilai dikenal sebagai datum. Kumpulan data juga dapat terdiri dari kumpulan dokumen atau file.

7. RapidMiner

RapidMiner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (*Graphical User Interface*) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat *open source* dan dibuat dengan menggunakan program Java di

bawah lisensi *GNU PublicLicence* dan *RapidMiner* dapat dijalankan di sistem operasi manapun. Dengan menggunakan *RapidMiner*, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. *RapidMiner* dikhususkan untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model *Bayesian*, *Modelling*, *Tree Induction*, *Neural Network* dan lain-lain. Banyak metode yang disediakan oleh *RapidMiner* mulai dari klasifikasi, *klustering*, asosiasi dan lain-lain.

3. Metode Penelitian

3.1. Teknik Pengumpulan Data

Untuk memperoleh data yang digunakan dalam penelitian ini, ada beberapa metode yang digunakan antara lain :

- Observasi
Teknik observasi yang dilakukan adalah mengumpulkan data dengan melakukan observasi atau pengamatan langsung ke Lombok Vape On untuk melakukan penelitian dan mencari informasi-informasi yang dibutuhkan dalam menyelesaikan penelitian ini.
- Interview (Wawancara)
Melakukan komunikasi langsung dengan teknik tanya jawab atau wawancara dengan pemilik toko Lombok Vape On.
- Studi pustaka
Suatu metode pengumpulan data dengan cara membaca dan mempelajari buku-buku yang berhubungan dengan penelitian ini,

serta referensi yang terkait dengan topik yang akan dibahas dalam penyusunan skripsi ini, serta mengunjungi (Browsing) situs-situs internet yang berhubungan dengan skripsi.

3.2. Teknik Pemrosesan Data

1. Eksperimen *Algoritma K-Nearest Neighbor*
Data berikut adalah datasets penjualan pada "Lombok Vape On". Dimana terdapat 9 *record* untuk data training yaitu no 1-9 dan 1 data testing yaitu no 10.

Tabel 1. Data Seleksi

No	Jenis	Keuntungan	Pendapatan	Jumlah	Kategori
1	Wrap baterai	251.2	538.2	287	Mencapai target
2	Aegis boost	5.060.000	42.500.000	104	Mencapai target
3	Coil aegis Boost 04	1.505.000	5.465.000	132	Mencapai target
4	Coil gear 300mah	3.860.000	9.340.000	137	Mencapai target
5	Gold cotton kecil	1.720.000	3.330.000	209	Mencapai target
6	One fire	6.940.000	9.980.000	304	Mencapai target
7	Smok coil 40 rpm	129	5.065.000	123	Mencapai target
8	Smok rpm coil 0 3	1.756.000	5.910.000	134	Mencapai target
9	Vooopoo vinci coil	1.841.000	7.756.000	169	Mencapai target
10	Kawat koil	1.260.000	2.820.000	312	?

Langkah-langkah dari *algoritma K-Nearest Neighbor* :

- Penentuan parameter $k=5$ (jumlah tetangga paling dekat). Disini dalam penentuan parameter $k=5$
- Menghitung *Eucliden distance*)

Tabel 2. Hasil *Euclidean distance*

No	Jenis	Profit	Pend.	Qty	Rem	ED
1	Wrap baterai	251.2	538.2	287	OK	2494852.4
2	Aegis boost	5.060.000	42.500.000	104	OK	39861540
3	Coil aegis Boost 04	1.505.000	5.465.000	132	OK	2656322.7
4	Coil gear 300mah	3.860.000	9.340.000	137	OK	7019287.7
5	Gold cotton kecil	1.720.000	3.330.000	209	OK	686804.2
6	One fire	6.940.000	9.980.000	304	OK	9139365.4
7	Smok coil 40 rpm	129	5.065.000	123	OK	2513799.1
8	Smok rpm coil 0 3	1.756.000	5.910.000	134	OK	3129555.2
9	Vooopoo vinci coil	1.841.000	7.756.000	169	OK	4970076.2

- Mengurutkan objek-objek tersebut kedalam kelompok yang mempunyai jarak *Euclidian* terkecil. Untuk mengurutkannya kita hanya perlu membuat urutan dari data yang mempunyai jarak terkecil ke data terbesar.

Tabel 3 Urutan Jarak *Euclidian* terkecil

No	Jenis	Distance	Rank
1	Aegis boost	39861540.36	1
2	One fire	9139365.405	2
3	Coil gear 300mah	7019287.715	3
4	Voopoo vinci coil	4970076.128	4
5	Smok rpm coil 0 3	3129555.245	5
6	Coil aegis Boost 04	2656322.652	6
7	Smok coil 40 rpm	2213799.124	7
8	Wrap baterai	2494852.437	8
9	Gold cotton kecil	686804.2011	9

- Mengumpulkan kategori Y (*Klasifikasi Nearest Neighbor*). Pada tahap ini hanya mengambil data sesuai dengan jumlah k yang kita tentukan adalah k=5. Jadi, kita memilih 5 data yang memiliki *Euclid* terkecil, dan hasilnya sebagai berikut :

Tabel 4. Urutan kategori Y

No	Jenis	Distance	Rangking
1	Aegis boost	39861540.36	1
2	One fire	9139365.405	2
3	Coil gear 300mah	7019287.715	3
4	Voopoo vinci coil	4970076.158	4
5	Smok rpm coil 0 3	3129555.245	5

- Dengan menggunakan kategori mayoritas, maka didapat hasil klasifikasi dan pengolahan seperti tabel di bawah:

Tabel 5. Urutan kategori mayoritas

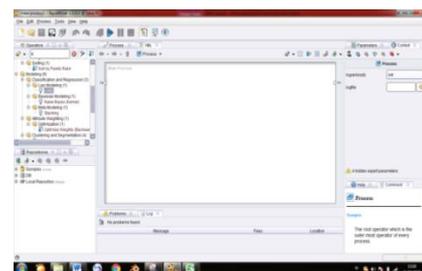
No	Jenis	Distance	Rangking	Kategori
1	Aegis boost	39861540.36	1	Mencapai target
2	One fire	9139365.405	2	Mencapai target
3	Coil gear 300mah	7019287.715	3	Mencapai target
4	Voopoo vinci coil	4970076.158	4	Mencapai target
5	Smok rpm coil 0 3	3129555.245	5	Mencapai target

Data yang memiliki *Euclid* terkecil adalah data ke 2, 6, 4, 9, 8 dari data yang sudah diuji tersebut terdapat semua data mencapai target. Sehingga mayoritas data mampu mencapai target. Jadi hasil klasifikasi dan pengolahan untuk kawat coil adalah penyandang mencapai target.

2. Pengujian Algoritma

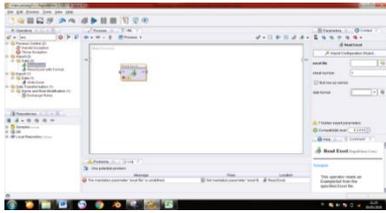
Diperlukan beberapa tahapan untuk mengolah data yaitu:

- Untuk menganalisa, dibutuhkan data training yang akan diolah dan dimasukkan kedalam *Rapidminer* dan menggunakan berbagai macam format seperti csv, xls, mdb, dan lain-lain. Adapun format data yang digunakan oleh penulis yaitu xls.
- Buka aplikasi *rapidminer*, kemudian akan tampil halaman awal, selanjutnya klik File dan pilih New Proses.



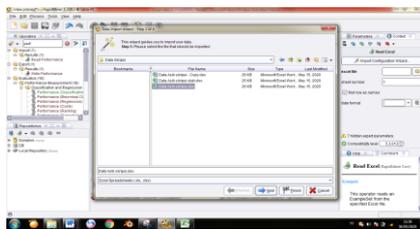
Gambar 2. Tampilan kerja *Rapidminer*

- Memasukkan data yang akan diolah dan sesuaikan format data yang dimiliki oleh peneliti, dan disini peneliti menggunakan format xls kemudian klik pada pencarian atau [filter] di *Rapidminer* seperti pada gambar di bawah ini, Kemudian pilih Import Wizard.



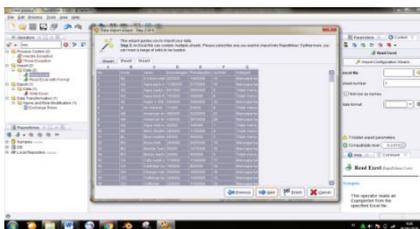
Gambar 3. Pemilihan format data

d. Tampilan data untuk import wizard diperlukan 5 langkah dalam tahapan ini, langkah ke-1 tentukan nama file yang berisi data training yang akan dimasukkan dan selanjutnya klik next pada gambar di bawah ini.



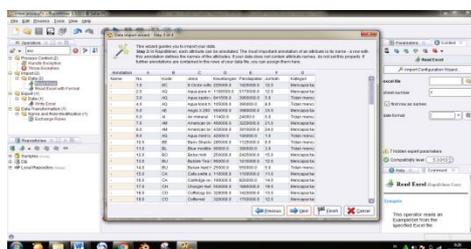
Gambar 4. Pemilihan Data

e. Tampilan data untuk menentukan tempat data tersimpan, selanjutnya klik next pada gambar di bawah ini:



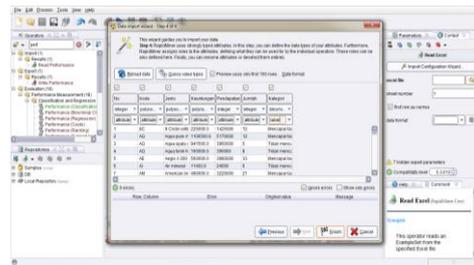
Gambar 5 Tampilan sheet data

f. Tampilan untuk melihat data yang kosong, atau data yang tidak dapat diolah. Selanjutnya klik next pada gambar di bawah ini:



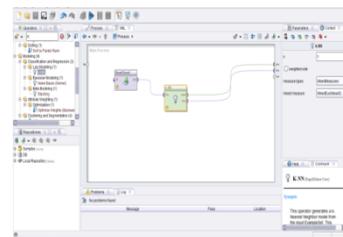
Gambar 6. Tampilan pengecekan data

g. Tahapan ini dilakukan untuk menentukan label pada data, dan apabila ada atribut yang tidak diperlukan, hilangkan centang paling atas sesuaikan urutan dari atribut yang tidak diperlukan tersebut. Kemudian pilih finish untuk masuk ke tahap selanjutnya, seperti gambar di bawah ini:



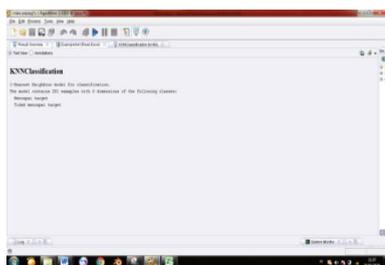
Gambar 7. Pemilihan tabel atribut dan label

h. Tahapan ini dilakukan untuk menambahkan model algoritma yang dibutuhkan, pilih *Tab Operator- Modelling-Classification and Regression-Lazy Modelling- k-NN*. Apabila sudah terpilih algoritma tersebut selanjutnya Geser *k-NN* ke arah *Main Process* dan hubungkan seperti gambar di bawah ini :



Gambar 8. Relasi data training dengan model

i. Tampilan untuk melihat hasil dari pengolahan menggunakan algoritma *K-Nearest Neighbor*, pilih menu *Process-Run* maka akan tampil hasil seperti berikut ini :



Gambar 9. Relasi data training dengan model

4. Hasil dan Pembahasan

1. Cros validation

Pengujian ini menggunakan 281 data, pengujian ini juga dilakukan untuk mengetahui pengaruh nilai k terhadap tingkat akurasi data, semakin tinggi jumlah k yang digunakan maka akan semakin bagus hasil dari pengujian. Nilai k yang terbaik pada algoritma ini tergantung pada jenis data yang digunakan. Nilai k yang bagus dapat dipilih dengan optimasi parameter, contohnya dengan menggunakan *cross validation*. Dalam kasus khusus, klasifikasi *K-Nearest Neighbor* dapat diprediksi berdasarkan training data yang paling dekat (dengan kata lain $k=1$). Nilai k (ketetanggaan paling dekat) yang digunakan dalam penelitian ini adalah 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 dengan *cross validation* 10.

accuracy: 86.48% +/- 5.00% (nilai: 86.48%)			
	true Mencapai target	true Tidak mencapai target	class precision
pred. Mencapai target	121	17	87.58%
pred. Tidak mencapai target	21	122	85.21%
class recall	85.21%	87.77%	

Algoritma *K-Nearest Neighbor* akan melakukan training terhadap data yang telah dipisahkan oleh *Cross Validation* menjadi dua tempat yaitu satu tempat untuk data training dan satu tempat untuk data testing. Training terdiri dari metode *K-Nearest Neighbor* yang digunakan dan testing terdiri dari *apply model* dan *performance*.

2. Confusion Matrix

Jumlah *true positif* (TP) adalah 121 *record* diklasifikasikan sebagai *class* yang mencapai target dan *false negatif* (FN) sebanyak 21 *record* diklasifikasikan sebagai *class* yang mencapai target namun pada kenyataannya adalah berada pada *class* yang tidak mencapai target. Berikutnya 122 *record false positif* (TN) diklasifikasikan sebagai *class* yang tidak mencapai target, dan 17 *record false positif* (FP) diklasifikasikan sebagai *class* yang tidak mencapai target namun ternyata diklasifikasikan sebagai *class* yang mencapai target. Berdasarkan Tabel 8 menunjukkan bahwa tingkat akurasi menggunakan algoritma *K-Nearest Neighbor* dengan *K-Fold Validation* 6 adalah sebesar 86.48%, dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *Ppv* dan *Npv* pada persamaan di bawah ini.

3. Kurva ROC



Gambar di atas adalah grafik ROC dengan nilai AUC (*Area Under Curve*) dengan hasil *accuracy* dengan *K-Fold Validation* 6 sebesar 86.48% dengan nilai akurasi *Excellent Classification*.

5. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa data penjualan yang

telah diolah dengan menerapkan teknik data mining menggunakan algoritma *K-Nearest Neighbor*. Dari data penjualan dapat diketahui bahwa rata-rata jumlah satu jenis barang penjualan 10 pcs dapat disebut sebagai kategori mampu mencapai target dan apabila kurang dari 10 pcs yang terjual maka akan disebut dengan kategori tidak mampu mencapai target. Dari kategori tersebut dapat diketahui hasil pengolahan jumlah barang yang paling banyak terjual oleh toko bisnis Lombok Vape On. Pengujian ini dilakukan dengan cara menghitung sesuai dengan langkah-langkah *algoritma K-NN (K-Nearest Neighbor)* untuk menghasilkan model sehingga masuk dalam kategori *excellent classification*, pengolahan dilakukan dengan menggunakan 9 *Validation*, dari pengolahan tersebut terdapat *K-Fold Validation 6* menjadi percobaan yang memiliki akurasi tertinggi. Dari percobaan yang telah dilakukan dapat diketahui hasil nilai *K-Fold Validation 6* dengan hasil akurasi diperoleh sebesar 86.48% dan *AUC* sebesar 0.874. Oleh karena itu metode *algoritma K-Nearest Neighbor* merupakan salah satu algoritma klasifikasi yang baik digunakan dan mampu mengklasifikasikan jumlah barang yang dapat terjual paling banyak. Adapaun jenis barang yang dikategorikan terjual paling banyak oleh toko bisnis tersebut yaitu jenis barang Aegis boost, One fire, Coil gear 300mah, Voofoo vinci coil, Smok rpm coil 0 3. Sehingga dalam menganalisa data metode ini cukup akurat untuk menganalisis data penjualan yang dikategorikan jenis barang paling dominan.

6. Daftar Pustaka

- [1] R. A. Pangestu, S. Rudiarto, and D. Fitriana, "Aplikasi Web Berbasis Algoritma K-Nearest Neighbour Untuk Menentukan Klasifikasi Barang Studi Kasus : Perum Peruri," vol. 2, no. 1, 2018.
- [2] B. Sawit, S. Bss, and M. Metode, "Data Mining Untuk Memprediksi Hasil Produksi Buah Sawit Pada Pt Bumi Sawit Sukses (Bss) Menggunakan Metode K-Nearest Neighbor", pp. 198–207, 2019.
- [3] H. B. Suhartini1, "Klasifikasi Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Kelayakan Bantuan Rehabilitasi Rumah Tidak Layak Huni Pada Desa Lenek Duren Kecamatan Aikmel Kabupaten Lombok Timur Suhartini1,Hariman," vol. 2, no. 2, pp. 79–85, 2019.
- [4] Yahya, "Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decision Tree(C4.5)," vol. 1, no. 1, pp. 56–63, 2018.
- [5] M. Rivki and A. M. Bachtiar, "Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia," *J. Sist. Inf.*, vol. 13, no. 1, p. 31, 2017.
- [6] Y. A. Setianto, K. Kusri, and H. Henderi, "Penerapan Algoritma K-Nearest Neighbour Dalam Menentukan Pembinaan Koperasi Kabupaten Kotawaringin Timur," *Creat. Inf. Technol. J.*, vol. 5, no. 3, p. 232, 2019.
- [7] U. B. Rahayu, U. Islam, N. Sunan, and G. Djati, "Penerapan Algoritma K-Nearest Neighbor Dan Algoritma Simple Multi Attribute Rating Technique Untuk Menentukan Strategi Penjualan Pada Pt . Inti (Persero)," 2016.
- [8] B. SAWIT, S. BSS, AND M. METODE, "Data Mining Untuk Memprediksi Hasil Produksi Buah Sawit Pada Pt Bumi Sawit Sukses (Bss) Menggunakan Metode K-Nearest Neighbor," PP. 198–207, 2019.