

Klasifikasi Penjualan Jus Dengan Menggunakan Algoritma K-Nearest Neighbor (K-NN) Untuk Penerapan Konsep Up-Selling

Studi Kasus pada PT. Rafid Kintan Gemilang

Ismi Anisa Sopiyantri, Tacbir Hendro Pudjiantoro, Irma Santikarama

Jurusan Informatika, Fakultas Sains dan Informatika

Universitas Jenderal Achmad Yani

Jl. Terusan Sudirman, Cimahi

ismianisas@gmail.com

Abstrak— Data mining merupakan suatu proses mengolah data untuk mengekstraksi data guna mendapatkan pola tertentu atau informasi lebih berarti dari sejumlah data besar yang ada didalam database. Salah satu bagian dari data mining adalah klasifikasi untuk mengelompokkan objek menjadi kelas tertentu menggunakan Algoritma K-Nearest Neighbor (K-NN) yang melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat. Penelitian ini akan membahas persoalan proses klasifikasi data penjualan jus menggunakan algoritma k-nearest neighbor untuk penerapan konsep up-selling. Pada proses penjualan jus menghasilkan data besar namun analisis pemanfaatan data penjualan jus yang ada di perusahaan belum sampai pada bentuk pengetahuan terhadap pengelompokan produk potensial paling laku, laku, dan kurang laku untuk dimunculkan secara langsung dan cepat berdasarkan data sebelumnya yang sudah ada sebagai acuan pembuatan rekomendasi produk dalam meningkatkan penjualan. Dengan menggunakan Algoritma K-Nearest Neighbor (K-NN) dan menerapkan konsep up-selling, akan menghasilkan tools yang dapat mengklasifikasi produk paling laku, laku dan kurang laku hingga mendapatkan kombinasi produk untuk rekomendasi perusahaan terhadap pengambilan keputusan.

Kata kunci— Data Mining; Klasifikasi; Algoritma K-NN; Up-Selling

I. PENDAHULUAN

Penggunaan data mining telah banyak digunakan oleh beberapa perusahaan dalam pengolahan data untuk menghasilkan suatu informasi yang bermanfaat, salah satunya dibidang makanan dan minuman. Data mining adalah proses analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkan yang belum diketahui sebelumnya yang dipahami dan berguna bagi pemilik data tersebut. Data mining memiliki beberapa bagian yang diantaranya *associan rule*, *prediction* [1], *classification*, *clustering* [2] dan *regression*. Secara garis besar data mining dapat dikelompokkan menjadi 2 yaitu *descriptive mining* dan *predictive* [3]. *Descriptive mining* adalah suatu proses untuk menemukan karakteristik penting dari data dalam suatu *database*, sedangkan *predictive mining* adalah proses untuk menemukan pola data dengan menggunakan beberapa *variable* lain dimasa depan [4]. Salah

satu algoritma yang digunakan dalam klasifikasi adalah algoritma K-Nearest Neighbor (K-NN) [5]. Klasifikasi merupakan salah satu teknik data mining yang termasuk kedalam *predictive*, klasifikasi bertujuan untuk mengelompokkan objek menjadi kelas tertentu berdasarkan nilai atribut yang berkaitan dengan objek yang diamati tersebut [6].

Algoritma *K-Nearest Neighbor* adalah klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. *K-Nearest Neighbor* akan mencari *k* buah data *learning* yang jaraknya paling dekat dengan data *query* dalam ruang *n*-dimensi. Jarak antara data *query* dengan data *learning* dihitung dengan cara mengukur jarak antara titik yang merepresentasikan data *query* dengan semua titik yang merepresentasikan data *learning* dengan rumus *Euclidean Distance* [5]. Keuntungan menerapkan algoritma *K-Nearest Neighbor* adalah proses pelatihan berjalan lebih cepat dan lebih fleksibel karena didasarkan pada kedekatan jarak data latih yang ada [7]. Beberapa penelitian terdahulu telah banyak menggunakan algoritma *K-Nearest Neighbor* salah satunya adalah digunakan untuk klasifikasi perbandingan tingkat akurasi bahan baku minuman *wine* dengan tingkat akurasi sebesar 68,75% dan sebesar 72,97% [8], menentukan tingkat kematangan susu kefir dengan hasil pengujian nilai akurasi yang dihasilkan yaitu 80% dengan nilai *k=5* untuk proses *sorting* [9], dan untuk penjualan terdapat klasifikasi standar kualitas makanan berdasarkan *grade* terbaik untuk dipasarkan ke pelanggan dengan nilai *k* yang digunakan adalah 1 (*nearest*), 3 dan 5 yang dipilih berdasarkan uji coba setelah tahap *training* [10].

Klasifikasi penjualan dilakukan dari data penjualan yang terdapat pada *database* perusahaan. Proses klasifikasi menghasilkan kelas – kelas tertentu, dari data uji dengan jarak terdekat yang nantinya hasil tersebut akan dijadikan penerapan dalam konsep *Up-Selling* [11]. *Up-Selling* adalah strategi bagi perusahaan untuk meningkatkan keuntungan dengan menentukan persepsi, pengetahuan dan kebijakan. Pelaksanaan teknik *up-selling* seharusnya tidak diketahui oleh pelanggan karena memungkinkan mengurangi rasa nyaman dari pelanggan tersebut [12]. Penelitian sebelumnya yang menggunakan strategi *up-selling* salah satunya adalah

rekomendasi untuk menawarkan pencarian komputer untuk calon pembeli dengan *filter* tipe komputer yang mempunyai nilai jual lebih tinggi [13], membuat sistem penjualan dengan konsep *up-selling* untuk menawarkan produk yang cocok untuk dikombinasi dengan nilai *confidence* 100% sehingga mendapatkan target penjualan yang luas dan *omzet* meningkat [14], dan menawarkan berbagai macam produk yang lebih baik dibandingkan yang akan dibeli. *Up-selling* disini berperan untuk meningkatkan loyalitas pelanggan dan meningkatkan penjualan dengan mengarahkan pelanggan untuk membeli produk yang lebih mahal [15].

Salah satu perusahaan yang dijadikan sebagai studi kasus pada penelitian ini adalah PT. Rafid Kintan Gemilang yang bergerak dibidang makanan dan minuman terutama pada produksi pengolahan buah – buahan dan sayuran menjadi produk minuman yang siap saji. Dalam proses penjualan jus terdapat data besar dengan jumlah ± 39.127 record data yang tersimpan secara digital dalam format *file excel* (.xlsx) selama lima tahun pada periode 2015 – 2019. Dari data besar penjualan jus yang di hasilkan, membuat data tersebut dapat dimanfaatkan kembali untuk keberlangsungan perusahaan itu sendiri.

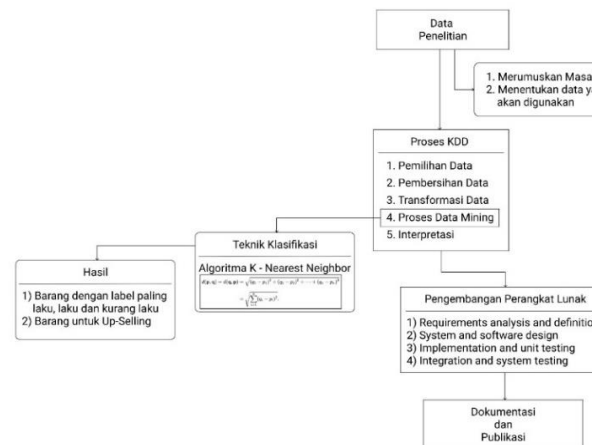
Pemanfaatan data penjualan jus dilakukan dengan cara menganalisis data untuk kepentingan *strategy marketing* dalam meningkatkan penjualan. Akan tetapi dalam proses memanfaatkan data tidak selamanya berjalan sesuai dengan harapan seperti analisis yang dilakukan oleh perusahaan khususnya bagian *marketing*. Permasalahan yang terjadi adalah analisis data yang dilakukan oleh bagian *marketing* hasilnya belum berbentuk pengetahuan. Pengetahuan yang dimaksud merupakan pengetahuan untuk mengelompokkan satu data baru secara langsung dan memunculkannya secara cepat berdasarkan data sebelumnya yang sudah ada.

Dari uraian tersebut penting untuk dilakukan penelitian dalam melakukan analisis pengolahan data menggunakan teknik data mining khususnya mengklasifikasikan produk paling laku, laku dan kurang laku untuk dijadikan rekomendasi produk guna meningkatkan penjualan dan loyalitas pelanggan dalam memberikan manfaat lebih untuk konsumen dari setiap produk yang dibeli. Klasifikasi tersebut menggunakan algoritma *k-nearest neighbor*, karena algoritma ini dianggap memiliki karakteristik yang presisi untuk menyelesaikan persoalan yang terjadi. Hasil klafsifikasi yang dilakukan, akan disimpan dan dijadikan acuan dalam pembuatan rekomendasi produk sebagai *strategy* meningkatkan penjualan menggunakan konsep *up-selling*. Pemanfaatan konsep *up-selling* disini lebih mengarahkan pelanggan untuk membeli produk yang direkomendasikan dengan menutupi produk jus kurang laku hingga produk jus tersebut dapat menghasilkan pendapatan bagi perusahaan yang semulanya tidak bisa terjual banyak.

II. METODE

Metode penelitian adalah suatu cara yang dipergunakan untuk melakukan penelitian sehingga mampu menjawab identifikasi masalah dan tujuan penelitian. Pada metode penelitian terdapat beberapa tahapan pekerjaan untuk menyelesaikan penelitian yang sedang dilakukan. Ada beberapa pekerjaan yang utama dalam penelitian ini. Pertama, *input* data penjualan untuk menentukan data yang akan

digunakan. Kemudian dilakukan proses data mining menggunakan algoritma *k-nearest neighbor* dan hasil dari klasifikasi tersebut untuk menerapkan konsep *up-selling*. Lebih jelasnya diberikan oleh Gambar 1.



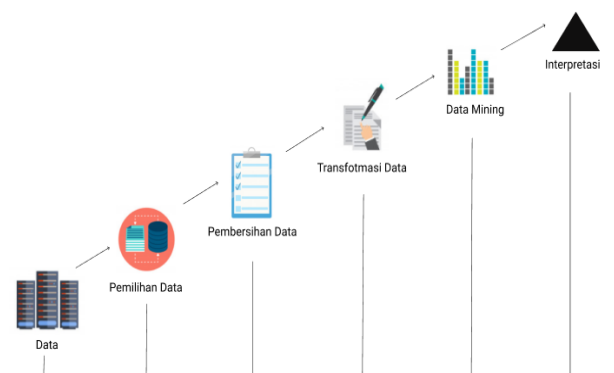
Gambar 1. Metode penelitian

A. Data Penelitian

Dalam data penelitian merupakan data yang diambil dari sebuah databases perusahaan PT. Rafid Intan Gemilang. Data penelitian yang didapatkan tersebut, digunakan untuk merumuskan sebuah permasalahan dan menentukan data atau objek yang akan digunakan dalam penelitian yang sedang dilakukan.

B. Proses KDD

Penambangan data juga dikenal dengan proses Knowledge Discovery in Databases (KDD) [16] merupakan proses pencarian pengetahuan tersembunyi dalam sejumlah data besar yang mampu menghasilkan dan menyimpan suatu informasi [17]. Pada tahap ini memutuskan teknik yang akan digunakan dalam melakukan pengelolaan data. Tahapan pada proses KDD seperti pada Gambar 2.



Gambar 2. Proses Knowledge Discovery in Databases (KDD) [18]

1) Pemilihan Data

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam klasifikasi dimulai. Data hasil seleksi yang digunakan untuk

proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

2) *Pembersihan Data*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi focus untuk klasifikasinya. Langkah pada proses *cleaning* atau pembersihan data antara lain membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

3) *Transformasi Data*

Proses transformasi data adalah proses melakukan pengkodean pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Pada transformasi data terdapat beberapa langkah, yaitu :

- Pemetaan Data: Menetapkan elemen dari sumber data untuk melakukan klasifikasi.
- Pembuatan Kode: Pembuatan program transformasi aktual [16].

4) *Proses Mining*

Pada mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam data mining sangat bervariasi.

5) *Interpretasi*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Hasil dari aturan – aturan ini kemudian dianalisa untuk dijadikan keputusan pemilihan strategi bagi perusahaan dalam kebutuhan bisnis [19].

C. *Klasifikasi*

Klasifikasi adalah proses dari mencari suatu himpunan model (fungsi) yang dapat mendeskripsikan dan membedakan kelas – kelas data atau konsep – konsep [20], dengan tujuan dapat menggunakan model tersebut untuk memprediksi kelas dari suatu objek yang mana kelasnya belum diketahui [21].

D. *Algoritma K-Nearest Neighbor*

K-Nearest Neighbor (K-NN) adalah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat atau memiliki persamaan ciri paling banyak dengan objek tersebut [24]. Langkah – langkah untuk menghitung algoritma K-NN:

- 1) Menentukan nilai k.
- 2) Menghitung kuadrat jarak euclid (query instance) masing – masing objek terhadap training data.
- 3) Mengurutkan objek – objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil.
- 4) Mengumpulkan label class Y (Klasifikasi Nearest Neighborhood).

Untuk menggunakan jarak *Euclidean* digunakan standarisasi penggunaan formula dalam pengerjaannya. Formula yang digunakan ada pada persamaan (1).

$$d(X, Y) = \sqrt{\sum_{i=1}^k (X_i - Y_i)^2} \tag{1}$$

Keterangan:

- d(X,Y) : jarak antara data uji dengan data latih,
- X : data uji,
- Y : data latih.

Metode K-Nearest Neighbor (K-NN) dibagi menjadi dua tahap yaitu pembelajaran (pelatihan) dan klasifikasi.

E. *Up-Selling*

Up-Selling merupakan program penjualan yang digunakan untuk menawarkan produk yang memiliki produk komplementer dengan nilai (*value*) yang lebih tinggi dari produk yang bersangkutan. Salah satu konsep *up-selling* adalah untuk menghasilkan tingkat penjualan yang tinggi dengan menawarkan rangkaian produk baru ataupun menawarkan produk terbaik lainnya. Penerapan metode *up-selling* bisa dikatakan menjadi 5 bagian yaitu:

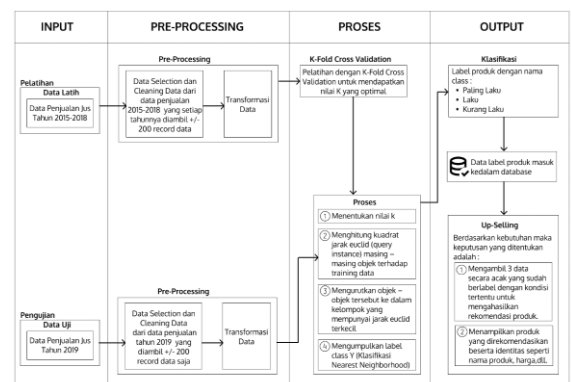
- 1) Menawarkan produk secara mendalam.
- 2) Membaca karakteristik atau kebutuhan pelanggan.
- 3) Memberikan kesempatan pelanggan untuk memilih.
- 4) Berikan penawaran yang wajar dari segi harga atau kualitas yang sesuai kepada pelanggan.
- 5) Memberikan nilai tambah pada produk yang akan ditawarkan.

III. HASIL DAN DISKUSI

Hasil dan diskusi ini membahas bagian isi dari tahapan yang dilakukan pada metode penelitian.

A. *Perancangan Data Mining*

Pada perancangan Data Mining yang didalamnya menjelaskan alur proses pengerjaan algoritma *k-nearest neighbor* untuk mendapatkan keputusan *up-selling* seperti pada Gambar 3.



Gambar 3. Alur Perancangan Aplikasi Data Mining

B. Pre-Processing

Pada tahap *pre-processing* adalah tahap di mana data menjadi selaras dengan kebutuhan untuk proses klasifikasi yang akan terjadi.

1) Pemilihan Data

Pada tahapan ini dilakukan pemilihan sekumpulan data untuk menentukan variabel dari data yang akan diklasifikasikan. Dalam tahap ini data yang digunakan adalah data laporan penjualan jus dari PT. Rafid Kintan Gemilang dari tahun 2015 sampai tahun 2019. Pada data laporan jus tersebut menyebutkan banyaknya jus yang terjual berdasarkan produk dan customers seperti pada Gambar 3. Data laporan penjualan jus yang diberikan oleh perusahaan memiliki atribut berupa kode *customer*, nama *customer*, kode barang, tanggal pemesanan, nomor bukti, *qty*, harga satuan, diskon, total. Dari ± 39.127 record data yang diberikan oleh perusahaan, penelitian ini menggunakan ± 1.062 record data yang terbilang lengkap untuk dijadikan sampel penelitian, seperti pada Gambar 4.

LAPORAN HISTORY PENJUALAN

DARI TGL : 01/01/2015 S/D 31/12/2019
 DARI KODE BARANG : S/D ZZZ
 DARI KODE CUSTOMER : S/D ZZZ

KODE CUSTOMER		NAMA CUSTOMER		PENJUALAN				
TANGGAL	NO BUKTI	QTY	HS	DISC	TOTAL			
KODE BARANG : 1010101		RTD Premium 1 Ltr Apple						
CUS-0067	D CONSULATE	19/01/2015	GF1500055	12,00	16.950,00	0,00	203.400,00	
CUS-0067	D CONSULATE	20/02/2015	GF1500173	12,00	16.950,00	0,00	203.400,00	
CUS-0067	D CONSULATE	05/03/2015	GF1500229	12,00	16.950,00	0,00	203.400,00	
CUS-0067	D CONSULATE	09/04/2015	GF1500360	12,00	16.950,00	0,00	203.400,00	
CUS-0067	D CONSULATE	07/05/2015	GF1500497	12,00	16.950,00	0,00	203.400,00	
CUS-0067	D CONSULATE	10/06/2015	GF1500673	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	01/07/2015	GF1500768	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	13/07/2015	GF1500835	24,00	16.645,00	0,00	447.480,00	
CUS-0067	D CONSULATE	09/09/2015	GF1501124	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	31/10/2015	GF1501444	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	19/12/2015	GF1501784	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	05/01/2016	GF1600015	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	29/01/2016	GF1600153	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	04/04/2016	GF1600556	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	07/04/2016	GF1600576	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	20/04/2016	GF1600674	12,00	16.645,00	0,00	223.740,00	
CUS-0067	D CONSULATE	27/04/2016	GF1600687	17,00	16.645,00	0,00	271.760,00	

Gambar 4. Bentuk data yang terpilih dan belum diperbaiki

2) Pembersihan Data

Pada tahapan ini dilakukan pembersihan data dengan cara memusnahkan, menghilangkan atau menghapus data yang tidak konsisten.

No	Customer	Jenis Jus	Rasa	Ukuran	Qty	Total Harga	No Bukti
1	020-0007	020-0007	020-0007	020-0007	12	203.400,00	GF1500055
2	020-0009	020-0009	020-0009	020-0009	12	203.400,00	GF1500173
3	020-0014	020-0014	020-0014	020-0014	12	203.400,00	GF1500229
4	020-0142	020-0142	020-0142	020-0142	12	203.400,00	GF1500360
5	020-0142	020-0142	020-0142	020-0142	12	203.400,00	GF1500497
6	020-0142	020-0142	020-0142	020-0142	12	223.740,00	GF1500673
7	020-0142	020-0142	020-0142	020-0142	12	223.740,00	GF1500768
8	020-0142	020-0142	020-0142	020-0142	24	447.480,00	GF1500835
9	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1501124
10	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1501444
11	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1501784
12	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1600015
13	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1600153
14	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1600556
15	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1600576
16	020-0080	020-0080	020-0080	020-0080	12	223.740,00	GF1600674
17	020-0080	020-0080	020-0080	020-0080	17	271.760,00	GF1600687

Gambar 5. Data penjualan jus yang sudah diperbaiki

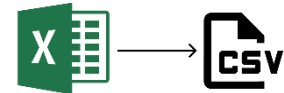
Data *noise* dan data *double* (duplikasi data) dihilangkan untuk digantikan dengan data lain yang sesuai agar data lebih lengkap dan siap untuk di *mining*. Implementasi pembersihan data ini seperti menghapus beberapa atribut yang tidak terpakai dan kurang cocok untuk dijadikan data *set*. Pada data *cleaning* juga terdapat penyempitan variabel yang akan diproses pada data *mining*-nya. Variabel tersebut adalah jenis jus, nama rasa, ukuran dan *qty* yang ada pada tabel. Keputusan tersebut dilakukan karena adanya normalisasi data untuk merincikan produk lebih detail seperti pada Gambar 5. dan Gambar 6.

No	Nama Rasa	Tersedia				Terdapat			
		Ukuran	Ukuran	Ukuran	Ukuran	Jenis Jus	Jenis Jus	Jenis Jus	Jenis Jus
1	Apple	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
2	Lemon	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
3	Lime	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
4	Mango	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
5	Orange	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
6	Kedondong	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
7	Martebe	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
8	Dutch Eggplant	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
9	Honeydew	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
10	Pineapple	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
11	Kiwi	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
12	Kurma	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
13	Lychee	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
14	Pink Guava	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
15	Soursop	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
16	Star Fruit	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
17	Strawberry	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium
18	Grape	250 ml	270 ml	1 Ltr	5 Ltr	Freshly	Junior	Buffet	Premium

Gambar 6. Produk yang dipakai untuk klasifikasi

3) Transformasi Data

Pada tahapan ini terjadi dua transformasi data, yaitu yang pertama adalah pada saat proses *import database* dengan mengubah terlebih dahulu format data dari MS Excel (.xlsx) ke dalam bentuk data *text* berupa (.csv) agar lebih mudah, seperti pada Gambar 7.



Gambar 7. Import file

Transformasi kedua yang dilakukan adalah dilakukannya perubahan skala skor pada data lain menjadi skala skor lain yang dapat memenuhi ketentuan data *mining*. Pada penelitian ini mengubah nama jenis jus yang semula bernama *freshly*, *junior*, *buffet* dan *premium* menjadi angka 1, 2, 3 dan 4. Kemudian merubah nilai ukuran yang semula 250 ml, 270 ml, 1 Ltr dan 5 Ltr menjadi 1, 2, 3 dan 4 juga. Hal tersebut dilakukan dengan asumsi untuk menghitung kedekatan jarak semua data pada setiap tahunnya. Untuk menunjang data yang sesuai, di tahap ini dilakukan penyempitan *record* data karena pada atribut *qty* nilai jumlah *qty*-nya menjadi pertahun. *Record* data latih yang siap di proses menjadi 387 *record* data dari tahun 2015 – 2018, sedangkan data uji berjumlah 74 *record* data dari tahun 2019. Contoh dari bentuk data yang digunakan dapat dilihat pada Tabel 1. dan Tabel 2.

TABEL 1. DATA LATIH SEBELUM DI UBAH MENJADI NUMERIK

No	Data Sebelum Diubah				
	Nama Rasa	Jenis Jus	Ukuran	Qty/Thn	Kelas

1	Apel	Premium	1 Ltr	42	Paling laku
2	Lemon	Premium	1 Ltr	22	Laku
3	Lime	Premium	1 Ltr	28	Laku
4	Mango	Premium	1 Ltr	41	Paling laku
5	Pink Guava	Premium	1 Ltr	72	Paling laku
6	Kedondong	Junior	1 Ltr	20	Laku
7	Martebe	Junior	1 Ltr	20	Laku
8	Dustch Eggplant	Junior	1 Ltr	20	Laku
9	Honeydew	Junior	1 Ltr	20	Laku
10	Apel	Buffet	1 Ltr	18	Laku
11	Orange	Premium	1 Ltr	33	Paling Laku
12	Kedondong	Premium	1 Ltr	20	Laku
13	Martebe	Premium	1 Ltr	15	Laku
14	Dutch Eggplant	Premium	1 Ltr	15	Laku
15	Kurma	Premium	1 Ltr	17	Laku

TABEL 2. DATA LATIH SESUDAH DIUBAH MENJADI BENTUK NUMERIK

No	Data Setelah Diubah				
	Nama Rasa	Jenis Jus	Ukuran	Qty/Tahun	Kelas
1	Apel	4	3	42/Qty terbesar	Paling laku
2	Lemon	4	3	22/Qty terbesar	Laku
3	Lime	4	3	28/Qty terbesar	Laku
4	Mango	4	3	41/Qty terbesar	Paling laku
5	Pink Guava	4	3	72/Qty terbesar	Paling laku
6	Kedondong	2	3	20/Qty terbesar	Laku
7	Martebe	2	3	20/Qty terbesar	Laku
8	Dustch Eggplant	2	3	20/Qty terbesar	Laku
9	Honeydew	2	3	20/Qty terbesar	Laku
10	Apel	3	3	18/Qty terbesar	Laku
11	Orange	4	3	33/Qty terbesar	Paling Laku
12	Kedondong	4	3	20/Qty terbesar	Laku
13	Martebe	4	3	15/Qty terbesar	Laku
14	Dutch Eggplant	4	3	15/Qty terbesar	Laku
15	Kurma	4	3	17/Qty terbesar	Laku

C. K-Fold Cross Validation

Pada bagian ini merupakan penentuan nilai k untuk melakukan perhitungan jarak terdekat klasifikasi K-NN. Pada penelitian ini menggunakan percobaan 5-fold cross validation dengan nilai 3, 5, 7, 9 dan 11. Pengambilan nilai ini bernilai ganjil karena nilai ganjil dapat menghilangkan ambiguitas data saat mencari nilai jarak terdekat antar tetangga yang paling dekatnya. Dari percobaan sederet nilai k yang dilakukan, hasilnya akan baik untuk melakukan K-NN tetapi tergantung pada data yang ada. Nilai 5-fold ini di uji menggunakan metode

cross validation dengan rumus pada persamaan (2), sebagai berikut :

$$\text{Uji Akurasi} = \frac{\text{Jumlah Hasil Data Yang Sama}}{\text{Jumlah Keseluruhan Data}} = 100 \% \quad (2)$$

Hasil uji akurasi yang telah dilakukan terhadap nilai 5-fold untuk digunakan pada sejumlah data, dengan jumlah data 387 record data di representasikan dalam Tabel 3.

TABEL 3. HASIL UJI AKURASI

	PERCOBAAN				
	Ke-1	Ke-2	Ke-3	Ke-4	Ke-5
Nilai K-Fold	k = 3	k = 5	k = 7	k = 9	k = 11
Uji Akurasi (%)	74,16	76,74	86,04	85,52	69,50

Dilihat dari hasil rangkuman perhitungan uji akurasi terhadap 5-fold dengan nilai 3, 5, 7, 9 dan 11 seperti pada tabel di atas, terlihat bahwa k = 11 menghasilkan nilai akurasi terendah sebesar 69,50 %. Dan nilai akurasi paling tinggi dilakukan oleh k = 7 dengan hasil kurasi 86,04 %. Maka kesimpulan dari percobaan nilai k ke-n, diambil k = 7 untuk penelitian ini karena k = 7 menghasilkan nilai akurasi paling tinggi.

D. Perhitungan Menggunakan Euclidian Distance

Pada bagian ini melakukan percobaan perhitungan terhadap data latih yang ada menggunakan rumus Euclidian Distance dengan data uji yang dimisalkan seperti pada Tabel 4. dan Tabel 5.

TABEL 4. DATA UJI SEBELUM DI NORMALISASI

No	Nama Rasa	Jenis Jus	Ukuran	Qty/Tahun	Kelas
1.	Soursop	Premium	1 Ltr	72/300	?

TABEL 5. DATA UJI SESUDAH DI NORMALISASI

No	Nama Rasa	Jenis Jus	Ukuran	Qty/Tahun	Kelas
1.	Soursop	4	3	0,24	?

Data uji merupakan sekumpulan data baru yang akan diklasifikasikan terhadap data latih untuk mencari jarak terdekatnya hingga mendapatkan kelas paling laku, laku dan kurang laku. Cara menghitung jarak terdekat adalah dengan Euclidean Distance. Perhitungan tersebut seperti :

$$U1L1 = \sqrt{(4 - 4)^2 + (3 - 3)^2 + (0,24 - 0,58)^2} = 0,34$$

$$U1L2 = \sqrt{(4 - 4)^2 + (3 - 3)^2 + (0,24 - 0,30)^2} = 0,06$$

$$U1L3 = \sqrt{(4 - 4)^2 + (3 - 3)^2 + (0,24 - 0,25)^2} = 0,01$$

..... (dsb)

$$U1L15 = \sqrt{(4 - 4)^2 + (3 - 3)^2 + (0,24 - 0,23)^2} = 0,01$$

dengan keterangan :

L_n : Data latih ke-n,

U_1 : Data uji ke-1 (karena hanya 1 data saja yang diujikan)

E. Implementasi Antarmuka Data Mining

Pada bagian ini menggambarkan dan menjelaskan implementasi dari bagian – bagian perangkat lunak klasifikasi yang telah dibuat.

1) Antarmuka Data Latih Siap Proses

Pada bagian implementasi antarmuka ini menampilkan data yang variabelnya sudah dilakukan sortir dan siap dipakai saat proses klasifikasi, seperti pada Gambar 8.

NO	Nama	Jenis	Ukuran	QTY/Tahun	Kelas
1	Apple	4	3	0,140	Paling Laku
2	Lemon	4	3	0,073	Laku
3	Lime	4	3	0,060	Laku
4	Mango	4	3	0,137	Paling Laku
5	Orange	4	3	0,110	Paling Laku
6	Kedondong	4	3	0,067	Laku
7	Martabe	4	3	0,050	Laku
8	Dutch Eggplant	4	3	0,030	Laku
9	Honeydew	4	3	0,050	Laku
10	Pineapple	4	3	0,050	Laku

Gambar 8. Data latih siap di proses

2) Antarmuka Input Data Uji

Halaman perhitungan K-NN terdapat suatu form untuk menginputkan nilai variabel data uji sebelum fase *training* terhadap data latih dilakukan, seperti pada Gambar 9.

Gambar 9. Input Data Uji

3) Hasil Pelatihan

Pada bagian implementasi antarmuka ini menampilkan hasil *training* dengan menampilkan 5 tetangga terdekat dan hasil kelas pengelompokannya, seperti pada Gambar 10.

NO	Jarak	Kelas
114	0,000	Paling Laku
115	0,007	Paling Laku
113	0,017	Paling Laku
218	0,020	Kurang Laku
277	0,020	Kurang Laku

Total	
Paling Laku	3
Laku	0
Kurang Laku	2

Gambar 10. Hasil perhitungan

4) Uji Akurasi Nilai K

Pada halaman implementasi antarmuka pengujian ini adalah menampilkan nilai hasil uji akurasi dengan menekan *button* uji, seperti pada Gambar 11.

Gambar 11. Hasil uji akurasi nilai K

5) Rekomendasi Produk

Halaman implementasi antarmuka *up-selling* ini menampilkan rekomendasi produk bila *button* rekomendasi produk ditekan, seperti pada Gambar 12.

Produk Terpilih 1	NO	Nama	Jenis	Ukuran	Kelas
1	Appel	Junior	270ml	Laku	
2	Honeydew	Buffet	5lt	Kurang Laku	
3	Mango	Buffet	1lt	Paling Laku	

Produk Terpilih 2	NO	Nama	Jenis	Ukuran	Kelas
1	Lime	Junior	250ml	Laku	
2	Lychee	Buffet	5lt	Kurang Laku	
3	Lemon	Junior	250ml	Paling Laku	

Produk Terpilih 3	NO	Nama	Jenis	Ukuran	Kelas
1	Sourcapp	Junior	1lt	Laku	
2	Kedondong	Premium	1lt	Kurang Laku	
3	Mango	Buffet	1lt	Paling Laku	

Gambar 12. Pilihan rekomendasi produk

IV. KESIMPULAN

Setelah melakukan penelitian ini, dapat disimpulkan bahwa dengan menggunakan algoritma k-nearest neighbor untuk klasifikasi produk jus dapat dilakukan dan menghasilkan kelas produk paling laku, laku dan kurang laku yang sesuai. Penentuan kelas produk jus dilakukan dengan mencari jarak tetangga terdekat menggunakan rumus Euclidian Distance. Penggunaan nilai k dapat memberikan hasil klasifikasi yang berbeda – beda. Nilai k yang dapat memberikan nilai hasil klasifikasi yang baik adalah nilai k-optimal (nilai k paling

optimal). Maka k-fold pada penelitian ini adalah k=7 dengan hasil tingkat akurasi sebesar 86,04 % yang mampu mengenali data uji untuk masuk ke dalam kelas produk jus paling laku, laku atau kurang laku. Hasil dari klasifikasi tersebut disimpan dan digunakan untuk mengambil keputusan up-selling dan dapat berhasil diterapkan dengan merekomendasikan 3 produk jus yang memiliki ketentuan, terdapat 1 produk jus paling laku, 1 produk jus laku dan 1 produk jus kurang laku serta memiliki rasa yang berbeda. Hal tersebut dilakukan sesuai dengan latar belakang penelitian yang telah disampaikan pada bab sebelumnya.

DAFTAR PUSTAKA

- [1] Muladi, U. Pujiyanto, and U. Qomaria, "Predicting high school graduates using Naive Bayes in State University Entrance Selections," *4th Int. Conf. Vocat. Educ. Training, ICOVET 2020*, pp. 155–159, 2020, doi: 10.1109/ICOVET50258.2020.9230336.
- [2] S. Sumarlin, "Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM," *J. Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 52–62, 2015, doi: 10.21456/vol5iss1pp52-62.
- [3] B. M. Metisen and H. L. Sari, "Analisis clustering menggunakan metode K-Means dalam pengelompokan penjualan produk pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015.
- [4] D. L. Fithri, "Model Data Mining Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Menggunakan Metode Naive Bayes," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 7, no. 2, p. 725, 2016, doi: 10.24176/simet.v7i2.787.
- [5] A. G. Muttaqin, K. Auliasari, and F. S. Wahyuni, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada Pt. Wika Industry Energy," vol. 4, no. 2, 2020.
- [6] A. Qur'ania, "Analisis Tektur Dan Ekstraksi Fitur Warna Untuk Klasifikasi Apel Berbasis Citra," pp. 296–304, 2012.
- [7] Y. F. Safri, R. Arifudin, and M. A. Muslim, "K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor," *Sci. J. Informatics*, vol. 5, no. 1, p. 18, 2018, doi: 10.15294/sji.v5i1.12057.
- [8] D. A. Nasution, H. H. Khotimah, and N. Chamidah, "Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *Comput. Eng. Sci. Syst. J.*, vol. 4, no. 1, p. 78, 2019, doi: 10.24114/cess.v4i1.11458.
- [9] F. Ardiansyah, D. Syaquy, and G. E. Setyawan, "Klasifikasi Tingkat Kematangan Susu Kefir dengan Metode K-Nearest Neighbor (KNN) menggunakan Sensor Cahaya dan Sensor Warna," vol. 3, no. 9, 2019.
- [10] S. Asmiatun, V. Vydia, and E. Widodo, "Penerapan Metode K-Nearest Neighbor pada Aplikasi E-Kuliner untuk Strategi Marketing Wisata Kuliner Indonesia," vol. XI, no. 2, 2020.
- [11] N. A. Putra, A. T. Putri, D. A. Prabowo, L. Surtiningsih, R. Arniantya, and I. Cholissodin, "Klasifikasi Sepeda Motor Berdasarkan Karakteristik Konsumen Dengan Metode K-Nearest Neighbour Pada Big Data Menggunakan Hadoop Single Node Cluster," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 2, p. 81, 2017, doi: 10.25126/jtiik.201742242.
- [12] D. Andriani, R. Veronika, and N. S. Achmadi, "The Implementation of Waiters And Waitresses' Upselling Techniques At Ticket's Cafe of Horison Ultima Makassar Hotel," *PUSAKA (Journal Tour. Hosp. Travel Bus. Event)*, vol. 2, no. 1, pp. 59–65, 2020, doi: 10.33649/pusaka.v2i1.41.
- [13] R. Imbar and D. Gunawan, "Aplikasi Penjualan Komputer dengan Metode Crossselling dan Upselling Dilengkapi Algoritma Greedy Dalam Pengambilan Keputusan," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, pp. 95–111, 2013.
- [14] F. Purwaningtiyas, "Strategi Up Selling Pada Website Penjualan," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 109–120, 2018, doi: 10.24176/simet.v9i1.1910.
- [15] U. B. Darma, J. Jenderal, A. Yani, and N. Palembang, "Implementasi Metode Up Selling Pada Sistem," no. 3, pp. 192–198, 2015.
- [16] J. Han, M. Kamber, and J. Pei, *Introduction*. 2012.
- [17] U. Fayyad, "Knowledge discovery in databases: An overview," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 15, no. 12, pp. 1–16, 2017, doi: 10.1007/3540635149_30.
- [18] U. Shafique and H. Qaiser, "A Comparative Study of Data Mining Process Models (KDD , CRISP-DM and SEMMA)," *Int. J. Innov. Sci. Res.*, vol. 12, no. 1, pp. 217–222, 2014, [Online]. Available: <http://www.ijisr.issr-journals.org/>.
- [19] Y. Yahya and W. Puspita Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 104–114, 2020, doi: 10.29408/jit.v3i2.2279.
- [20] G. F. Mandias, "Konferensi Nasional Sistem dan Informatika 2015 Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Di Universitas Klabat Dengan Metode Klasifikasi," *Konf. Nas. Sist. Inform.*, p. 20, 2015.
- [21] L. R. Angga Ginanjar Mabur, "Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit," *J. Komput. dan Inform.*, vol. 1, no. 1, pp. 53–57, 2012.