

Klasifikasi Teks Mining Terhadap Analisa Isu Kegiatan Tenaga Lapangan Menggunakan Algoritma K-Nearest Neighbor (KNN)

Nur Ajijah¹, Adi Kurniawan², Susilawati³

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Singaperbangsa Karawang, Indonesia

e-mail: nur.ajijah18036@student.unsika.ac.id¹, adi.kurniawan17035@student.unsika.ac.id², susilawati.sobur@staff.unsika.ac.id³

Abstract

Information and communication technology at this time is developing very rapidly, including in the easier dissemination of information. With the sophisticated technology in conveying information, it is possible that there is information that is not certain of its truth. The issue that occurs is due to the discrepancy of expectations expected by distributors, kiosks, and farmers. The amount of issue data obtained greatly affects the efficiency of the results that will be obtained. Therefore, it is necessary to have a text analysis to find out the issues spread in the field regarding the services of products and services provided by PT XYZ. In this study, it applied the CRISP-DM research stages and the application of the K-Nearest Neighbor (KNN) algorithm which showed that the resulting accuracy rate was 93.88% with data of 2,500 data. And the highest precision value is obtained by the payment qualification of 98.67%.

Keywords : Text mining, classification, algorithms K-Nearest Neighbor (KNN).

Abstrak

Teknologi informasi dan komunikasi pada saat ini berkembang dengan sangat pesat, termasuk dalam penyebaran informasi yang semakin mudah. Dengan canggihnya teknologi dalam menyampaikan sebuah informasi tidak menutup kemungkinan akan adanya informasi yang belum pasti kebenarannya. Isu yang terjadi dikarenakan adanya ketidaksesuaian harapan yang diharapkan oleh distributor, kios, maupun petani. Jumlah banyaknya data isu yang diperoleh sangat mempengaruhi efisiensi terhadap hasil yang akan di dapat. Oleh karena itu perlu adanya sebuah Analisa teks untuk mengetahui isu-isu yang tersebar di lapangan mengenai pelayanan produk maupun jasa yang disediakan oleh PT XYZ. Pada penelitian ini mengaplikasikan tahapan penelitian CRISP-DM serta penerapan algoritma K-Nearest Neighbor (KNN) yang menunjukkan bahwa tingkat akurasi yang dihasilkan sebesar 93.88% dengan data sebanyak 2.500 data. dengan nilai recall tertinggi diperoleh oleh klasifikasi harga sebesar 96, 91% yang termasuk klasifikasi harga. Dan nilai precision tertinggi diperoleh oleh kualifikasi pembayaran sebesar 98.67%.

Kata Kunci: Text mining, klasifikasi, algoritma K-Nearest Neighbor (KNN).

1. PENDAHULUAN

Teknologi informasi dan komunikasi pada saat ini berkembang dengan sangat pesat, termasuk dalam penyebaran informasi yang semakin mudah. Seiring dengan perkembangan ilmu pengetahuan dan teknologi, maka informasi yang disampaikan semakin berkembang dan semakin luas. Dengan canggihnya teknologi dalam menyampaikan sebuah informasi tidak menutup kemungkinan akan adanya informasi yang belum pasti kebenarannya.

Sales Representative (SR) merupakan salah satu bagian dari struktur organisasi yang ada pada PT XYZ yang bertugas untuk bertanggung jawab dalam



melakukan penjualan dan mempromosikan produk secara langsung [1]. *Asisten Sales Representative* (ASR) adalah jabatan seseorang yang memiliki tugas untuk melakukan kegiatan marketing seperti kunjungan dinas, kunjungan kepada para petani, kunjungan terhadap distributor ataupun kios di wilayah kerja [2]. Banyaknya kegiatan yang dilakukan oleh ASR baik di wilayah kantor maupun dilapangan tidak terlepas dari adanya isu yang tersebar terkait pelayanan produk maupun jasa.

Isu menurut Kamus Besar Bahasa Indonesia (KBBI) merupakan sebuah masalah yang bersifat opini pernyataan atau berupa kata-kata isyarat yang tidak terjamin kebenarannya. Pada Suatu kunjungannya terhadap Distributor, Kios, maupun petani sering kali SR dan ASR menemukan isu yang menjadi suatu kendala dalam melakukan proses penjualan hal ini di sebabkan oleh beberapa faktor yang belum di ketahui secara pasti. Isu yang terjadi dikarenakan adanya ketidaksesuaian harapan yang diharapkan oleh distributor, kios, maupun petani. Jumlah banyaknya data isu yang diperoleh sangat mempengaruhi efisiensi terhadap hasil yang akan di dapat. Oleh karena itu perlu adanya sebuah Analisa *teks* untuk mengetahui isu-isu yang tersebar di lapangan mengenai pelayanan produk maupun jasa yang disediakan oleh PT XYZ. Analisa teks atau klasifikasi dokumen teks adalah suatu aktivitas menganalisis data teks, tulisan yang terkandung dalam sebuah dokumen. Pada penelitian yang telah dilakukan dengan judul Klasifikasi Dokumen Temu Kembali Informasi dengan *K-Nearest Neighbor* (KNN), evaluasi sistem dilakukan dengan melihat hasil perhitungan *recall*, *precision*, F-Measure dengan menggunakan uji coba 20 dokumen testing dan menggunakan nilai $K = \{37, 41, 43\}$ nilai akurasi F-Measure paling tinggi yang dihasilkan dengan menggunakan nilai $k = 43$ yaitu sebesar 0.539 [3]. Penelitian sebelumnya yang dilakukan oleh didapatkan hasil akurasi sebesar 67.2% dengan nilai $k = 5$, nilai presisi tertinggi sebesar 56,94%, dan nilai recall sebesar 78.24% dengan menggunakan nilai $k = 15$ [6].

Pada penelitian ini mengaplikasikan tahapan penelitian CRISP-DM serta penerapan algoritma *K-Nearest Neighbor* (KNN) dalam membahas mengenai klasifikasi *teks* mining terhadap Analisa isu kegiatan tenaga lapangan di PT XYZ. Adapun Manfaat dari penelitian ini dapat digunakan untuk menganalisis dan mengetahui jenis *issue* atau ulasan pelanggan, distributor, kios, dan petani terhadap pelayanan baik barang maupun jasa di PT XYZ yang nantinya informasi yang didapatkan bisa jadi suatu keputusan dalam menanggapi isu-isu yang ada.

2. METODOLOGI PENELITIAN

2.1. Klasifikasi

Klasifikasi adalah salah satu proses pada data mining yang bertujuan untuk menemukan hubungan dan menentukan atribut atau kelas label dari sampel yang akan diklasifikasi [7]. Klasifikasi merupakan suatu proses yang bersifat supervised learning dan digunakan untuk membedakan kelas label data dengan melalui pencarian model atau fitur yang dapat memprediksi kelas dari suatu objek dengan tepat yang [8].

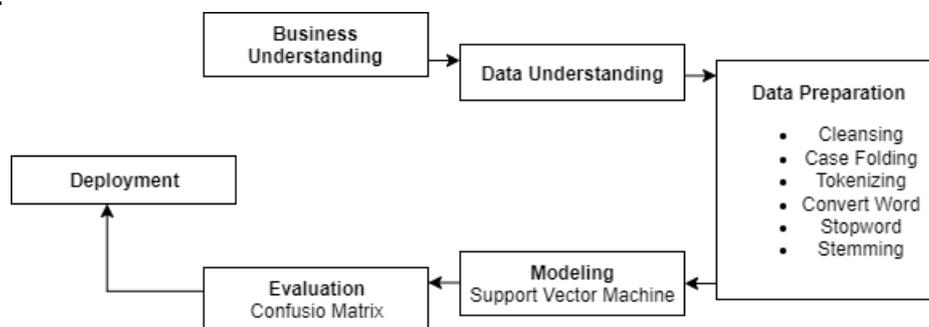


2.2. Teks Mining

Teks mining merupakan proses penggalian informasi berdasarkan sumber data berbentuk teks yang bersifat *unstructured data* [9]. *Teks mining* adalah penelitian yang mencoba untuk memecahkan sebuah permasalahan informasi yang berlebih dengan menggunakan sebuah teknik data mining, dan melibatkan tahapan preprocessing dokumen seperti seleksi teks, kategorisasi [10].

2.3. Cross Industry Standard for Data mining CRIPS-DM

Untuk pendekatan metode yang digunakan pada penelitian ini menggunakan metodologi *Cross Industry Standard for Data mining* (CRISP-DM) dengan menerapkan metode klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN).



Gambar 1. Metodologi CRISP-DM

Menurut Laroske CRIPS-DM adalah standarisasi dari proses *data mining* sebagai pemecahan masalah secara umum dalam suatu penelitian ataupun bisnis dengan enam tahapan proses yaitu [11]:

a) **Business Understanding (Pemahaman Bisnis)**

Tahap pertama dilakukan sebuah pemahaman atau analisis terhadap masalah apa yang akan diselesaikan dalam bisnis. Dalam hal ini perlu dilakukan dengan pencarian solusi dan kajian secara berulang kali.

b) **Data Understanding (Pemahaman Data)**

Tahap kedua yaitu pemahaman struktur data dilakukan pengumpulan data, penggambaran data, atau deskripsi data, selanjutnya mengidentifikasi masalah yang berkaitan dengan data yang tersedia. Pada penelitian ini data didapatkan dari Departemen Project Retail Manajemen PT XYZ berdasarkan laporan keluhan dari pelanggan periode Tahun 2021.

c) **Data Preparation (Persiapan Data)**

Pada tahap ketiga, pemilihan data, pemberian label pada data atau atribut yang akan digunakan, selanjutnya akan dilakukan preprocessing sebagai berikut:

- 1) *Cleansing* (Pembersihan Data), yaitu pembersihan karakter pada teks dokumen yang berupa seperti hastag, symbol, dan tanda tanya.
- 2) *Case folding*, yaitu proses mengubah huruf kapital yang ada pada dokumen menjadi huruf kecil.
- 3) *Tokenizing*, yaitu tahapan potongan kata yang bertujuan untuk mengubah huruf kapital menjadi kumpulan kata atau kalimat tunggal.



- 4) *Convert to word*, yaitu memperbaiki penulisan kata.
- 5) *Stopwords*, pembersihan atau penghilangan kata yang tidak digunakan terhadap hasil klasifikasi.
- 6) *Stemming*, proses transformasi kata imbuhan menjadi kata dasarnya.

d) Modeling (Pemodelan Data)

Pada tahap modeling dilakukan proses penerapan model, algoritma, serta tools yang akan digunakan pada proses data mining. Pada penelitian ini model data mining yang akan digunakan yaitu metode klasifikasi dengan algoritma *K-Nearest Neighbor* (KNN) terhadap data keluhan pelanggan terhadap kinerja produk dan jasa di PT XYZ. Hasil pemodelan tersebut selanjutnya dapat digunakan sebagai pertimbangan dalam mengambil sebuah keputusan.

e) Evaluation (Evaluasi)

Pada tahap evaluation ini dilakukan sebuah analisis terhadap hasil dari proses pengolahan data yang telah dilakukan. Proses interpretasi termasuk ke dalam tahap evaluasi. Evaluasi dilakukan untuk mengetahui kesesuaian model yang diterapkan dalam proses pembelajaran data.

f) Deployment (Penyebaran)

Tahap terakhir yaitu proses penyebaran pengetahuan dan informasi yang didapatkan dari hasil pemodelan dan evaluasi pada proses data mining. Hasil yang didapatkan akan diberikan kepada pihak PT XYZ yang digunakan untuk menentukan pertimbangan atau keputusan yang tepat dalam mengatasi keluhan pelanggan terhadap pelayanan yang diberikan oleh PT XYZ.

2.4. K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) adalah algoritma yang digunakan dalam klasifikasi teks atau data yang bersifat *supervised*. Algoritma *K-Nearest Neighbor* (KNN) bekerja berdasarkan jarak terpendek dari query instance ke data uji atau data train untuk menentukan klasifikasi ketetanggaannya [12]. Klasifikasi menggunakan algoritma *K-Nearest Neighbor* (KNN)

2.5. Confusion Matrix

Confusion matrix merupakan hasil dari klasifikasi yang berupa tabel berdasarkan jumlah data uji yang benar dan salah [13]. *Confusion matrix* juga merupakan suatu proses untuk mengetahui hasil kinerja dan bahan evaluasi dari metode klasifikasi [14].

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Isu merupakan sebuah informasi, rumor yang belum pasti kebenarannya. Isu terjadi karena adanya ketidaksesuaian harapan yang diharapkan oleh para distributor atau pelanggan produk yang dipasarkan. Pada hal ini isu yang terjadi di lapangan dalam proses pemasaran sebuah produk dan jasa yang dilakukan di PT XYZ salah satunya isu mengenai kenaikan harga produk, kualitas kemasan, pengiriman produk yang tidak sesuai dengan *Purchase Order* (PO) yang dilakukan



oleh pelanggan. Maka berdasarkan hal tersebut, perlu adanya analisis isu yang dapat memvalidasi isu atau informasi yang tersebar di lapangan.

3.2. Data Understanding

Jumlah data isu yang terjadi di lapangan yang berhasil dikumpulkan adalah sebanyak 2.500 data. Data tersebut merupakan sampel dari data isu tahun 2021. Untuk tahap selanjutnya akan dilakukan pelabelan secara manual pada data dengan mengkategorikan isu-isu menjadi kategori harga, kualitas produk, packaging, program promosi, pembayaran, dan isu lainnya.

3.3. Data Preparation

Tahap data preparation dilakukan pemilihan data, seleksi data dengan menghapus data duplikat yang memiliki isu yang sama.

a) *Cleansing* (Pembersihan data)

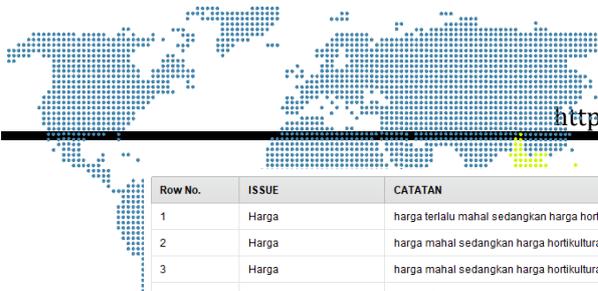
Pembersihan karakter pada teks dokumen yang berupa seperti hastag, simbol, dan tanda tanya. Pada tahap ini dilakukan penghapusan atribut yang tidak dibutuhkan dalam proses klasifikasi *text mining* analisis isu, selanjutnya pembersihan pada data simbol dan jenis-jenis tanda tanya yang terdapat pada dokumen. Jumlah atribut sebelum dilakukan pembersihan data sebanyak 13 atribut data, selanjutnya atribut yang tidak akan digunakan dalam proses klasifikasi akan dihapus. Atribut yang akan digunakan hanya 2 atribut yaitu atribut issue dan catatan.

Row No.	ID	VISIT PLAN ID	Tanggal	Tahun	Bulan	Tanggal(1)	Nama SE	Nama ASE	Kode Kios	Nama Kios	CATATAN
1	6d7544d1-63...	3319357-c2...	2022-06-03	2022	6	3	saepuluh	m.askarudin	RT00000026...	SERBAGUNA...	pupuk non su...
2	7500f5e-53...	d5ede2bf-1b...	2022-06-06	2022	6	6	saepuluh	m.askarudin	RT00000030...	SINAR TANI J...	harga non su...
3	cf41b016-9cc...	6b19b8d5-67...	2022-06-06	2022	6	6	saepuluh	m.askarudin	RT00000026...	SUBUR TANI	harga pupuk ...
4	efaa579c-f9e...	11cdf158-1f6...	2022-06-06	2022	6	6	saepuluh	m.askarudin	RTN0000001...	Rahayu tani	harga pupuk...
5	cf195459-70...	4e6e4a46-84...	2022-06-08	2022	6	8	saepuluh	m.askarudin	RT00000028...	MEKAR TANI	pupuk nonsu...
6	55a08d05-f6...	ed0c02ef-16...	2022-06-08	2022	6	8	saepuluh	m.askarudin	RT00000028...	TANI SUGIH	harga pupuk...
7	33e984fb-22...	a872f97f-5d0...	2022-06-08	2022	6	8	saepuluh	m.askarudin	RT00000028...	BAGJA PS	harga nonsu...
8	22e5288b-dd...	955b47e9-a2...	2022-06-08	2022	6	8	saepuluh	m.askarudin	RT00000537...	TANI MULYA ...	harga pupuk ...
9	52bfb275-07...	c47315b1-d0...	2022-06-08	2022	6	8	saepuluh	m.askarudin	RT00000028...	FAHIRA	harga pupuk...
10	ab1f3587-3c...	f72519c5-1b...	2022-06-09	2022	6	9	saepuluh	m.askarudin	RT00000026...	SETIA MANG...	harga pupuk ...
11	e46a4956-0c...	526da87b-8d...	2022-06-09	2022	6	9	saepuluh	m.askarudin	RT00000026...	ANUGRAH T...	harga pupuk...
12	cd86415b-d5...	37992cc7-18...	2022-06-09	2022	6	9	saepuluh	m.askarudin	RT00000026...	TANI BERKA...	pupuk non su...
13	d8af5f45-c52...	b7ea4d0c-b3...	2022-06-09	2022	6	9	saepuluh	m.askarudin	RT00000026...	WG TANI	Harga pupuk ...
14	f5064e0b-bc...	8709c064-b3...	2022-06-10	2022	6	10	saepuluh	m.askarudin	RTN0000001...	Dulur tani	harga pupuk...
15	f678c347-edc...	b894ba69-e8...	2022-06-10	2022	6	10	saepuluh	m.askarudin	RT00000026...	SIMPANG TANI	haroa eupuk ...

ExampleSet (53 examples, 0 special attributes, 11 regular attributes)

Gambar 1. Dataset awal Isu Tenaga di Lapangan

Pada Gambar 1 menunjukkan dataset awal terkait isu yang terjadi pada tenaga kerja di Lapangan. Namun untuk memudahkan dalam proses klasifikasi pada data mining, akan dilakukan tahapan pengurangan jumlah atribut yang terlibat dalam menentukan suatu nilai kelas target. Proses pemilihan data ini adalah data yang telah melalui proses *text mining*.



Row No.	ISSUE	CATATAN
1	Harga	harga terlalu mahal sedangkan harga hortikultura murah
2	Harga	harga mahal sedangkan harga hortikultura murah
3	Harga	harga mahal sedangkan harga hortikultura murah
4	Harga	harga nitrea dulu terjangkau, sekarang harganya mahal.
5	Harga	harga mahal, harga hortikultura murah
6	Harga	harga mahal, harga hortikultura murah
7	Harga	harga nitrea mahal
8	Harga	harga mahal sedangkan harga hortikultura murah
9	Isu lainnya	kios lagi blm stabil untuk menjual pupuk non subsidi lagi, dikarenakan harga yg tinggi permodalan yg lgi trun
10	Isu lainnya	pihak kios keberatan dgn harga nitrea yg mahal, sehingga tdk bs order, terlebih persediaan/kuota pupuk subsidi...
11	Isu lainnya	pihak kios keberatan dgn harga nitrea yg mahal, sehingga tdk bs order, terlebih persediaan/kuota pupuk subsidi...
12	Harga	sekarang harga nitrea mahal
13	Harga	sekarang harga nitrea mahal
14	Harga	terkait harga Nitrea saat ini dikeluhkan Petani km Kondisi harga Yg melonjak tinggi
15	Haroa	terkait haroa Nitrea saat ini dikeluhkan Petani km Kondisi haroa Ya meloniak tinoai

ExampleSet (2,510 examples, 0 special attributes, 2 regular attributes)

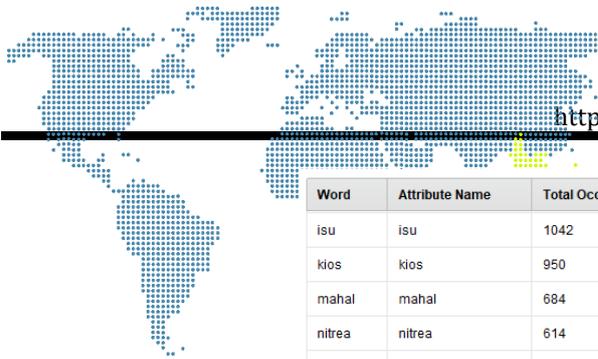
Gambar 2. Dataset setelah pembersihan atribut dan ekstraksi data

- b) *Case folding*, proses mengubah huruf kapital yang ada pada dokumen menjadi huruf kecil. Selanjutnya merubah semua karakter huruf pada catatan isu menjadi karakter huruf kecil semua.
- c) *Tokenizing*, yaitu tahapan potongan kata yang bertujuan untuk mengubah huruf kapital menjadi kumpulan kata atau kalimat tunggal. Tokenisasi berfungsi untuk membagi catatan isu dari kalimat menjadi per kata. Pada tahap ini proses tokenisasi dilakukan dengan melihat kalimat per spasi dan dipisahkan menjadi per kata.

Name	Type	Missing	Statistics	Filter (1,166 / 1,166 attributes)
Abal	Real	0	Min: 0, Max: 0.671, Average: 0.000	
Ada	Real	0	Min: 0, Max: 0.565, Average: 0.006	
Akan	Real	0	Min: 0, Max: 0.478, Average: 0.000	
Applikasi	Real	0	Min: 0, Max: 0.481, Average: 0.000	
Ava	Real	0	Min: 0, Max: 0.457, Average: 0.001	
Avena	Real	0	Min: 0, Max: 0.399, Average: 0.001	
Bagi	Real	0	Min: 0, Max: 0.444, Average: 0.000	
Balk	Real	0	Min: 0, Max: 0.524, Average: 0.003	

Gambar 3. Hasil Proses *Tokenizing*

- d) *Convert to word*, yaitu memperbaiki penulisan kata.
- e) *Stopwords*, pembersihan atau penghilangan kata yang tidak digunakan terhadap hasil klasifikasi. Pada tahap ini dilakukan penghapusan kata yang tidak penting seperti yang, dan, di, adalah dan seterusnya.

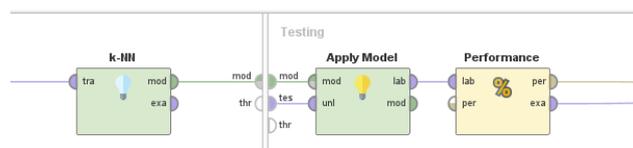


Word	Attribute Name	Total Occure... ↓	Document Occurences
isu	isu	1042	1042
kios	kios	950	866
mahal	mahal	684	682
nitrea	nitrea	614	605
petani	petani	597	525
subsidi	subsidi	549	472
pupuk	pupuk	466	350
non	non	378	371
rms	rms	366	341
beli	beli	335	328
produk	produk	328	288
daya	daya	317	313
menurun	menurun	214	213

Gambar 4. Proses pembersihan atau penghilangan kata (*Stopword*)

- f) *Stemming*, proses transformasi kata imbuhan menjadi kata dasarnya.
- g) *Modeling*

Tahap penerapan teknik data mining yang akan digunakan sesuai dengan permasalahan dan tujuan penelitian. Pada penelitian model yang akan digunakan dalam proses data mining klasifikasi adalah algoritma *K-Nearest Neighbor (KNN)*. Data yang akan di klasifikasi merupakan data terkait isu yang terjadi pada tenaga kerja di lapangan. Yang memiliki dua atribut yaitu catatan dan issu, atribut yang akan menjadi acuan dan dijadikan sebagai label yaitu atribut isu yang berisi jenis-jenis isu meliputi harga, pembayaran, program promosi, kualitas produk, packaging, dan isu lainnya.



Gambar 5. Pemodelan menggunakan algoritma *K-Nearest Neighbor (KNN)*

```

SimpleDistribution
Distribution model for label attribute ISSUE

Class HARGA (0.500)
927 distributions

Class ISU LAINNYA (0.415)
927 distributions

Class PEMBAYARAN (0.038)
927 distributions

Class PROGRAM PROMOSI (0.032)
927 distributions

Class PACKAGING (0.002)
927 distributions

Class KUALITAS PRODUK (0.012)
927 distributions
  
```

Gambar 6. Pemodelan atribut isu dengan algoritma *K-Nearest Neighbor (KNN)*



Klasifikasi *text mining* yang akan dibuat adalah 6 klasifikasi, penentuan klasifikasi *text mining* akan dibagi 6 *class* kriteria yaitu kelas harga, pembayaran, program promosi, packaging, kualitas produk, dan isu lainnya.

h) Evaluation

Setelah data di *performance* maka hasil perhitungan pemodelan perlu dilakukan evaluasi untuk mengetahui tingkat kualitas dari pada pemodelan yang telah dilakukan. Hasil pemodelan selanjutnya dilakukan evaluasi menggunakan *confusion matrix*.

accuracy: 93.88% +/- 1.80% (micro average: 93.88%)

	true HARGA	true ISU LAINNYA	true PEMBAYARAN	true PROGRAM PR...	true PACKAGING	true KUALITAS PR...	class precision
pred. HARGA	973	77	2	3	1	1	92.05%
pred. ISU LAINNYA	28	753	1	0	0	1	96.17%
pred. PEMBAYARAN	1	0	74	0	0	0	98.67%
pred. PROGRAM P...	2	2	0	60	0	0	93.75%
pred. PACKAGING	0	1	0	0	4	0	80.00%
pred. KUALITAS PR...	0	1	0	2	0	22	88.00%
class recall	96.91%	90.29%	96.10%	92.31%	80.00%	91.67%	

Gambar 7. Hasil perhitungan akurasi algoritma *K-Nearest Neighbor* (KNN)

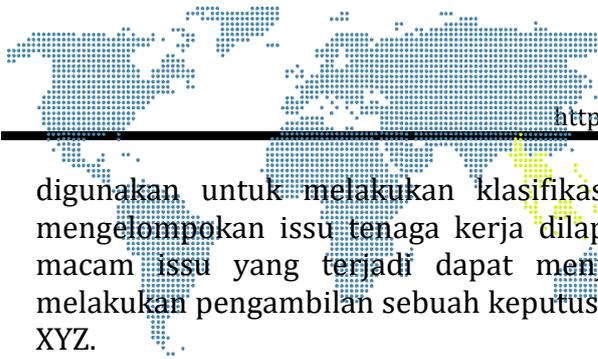
Pada Gambar 7 merupakan hasil pengujian menggunakan algoritma *K-Nearest Neighbor* (KNN) yang menghasilkan nilai akurasi sebesar 93.88% +/- 1.80. Evaluasi sistem dilakukan dengan melihat hasil perhitungan nilai *recall*, *precision* terhadap hasil uji coba yang telah dilakukan menggunakan 2.500 data dengan menggunakan nilai K = 5. Dari hasil uji coba yang telah diterapkan diperoleh nilai tertinggi *recall* sebesar 96, 91% yang termasuk dalam klasifikasi harga, Untuk nilai akurasi *precision* tertinggi diperoleh oleh kelas pembayaran sebesar 98.67%, sedangkan nilai akurasi *recall* dan *precision* terendah yaitu kelas klasifikasi packaging sebesar 80.00%. Dibawah ini merupakan nilai akurasi dari kriteria isu-isu berdasarkan catatan.

Tabel 1. Tabel penjelasan kriteria issue

Catatan Issue	Jumlah Kata	Kata-kata	Nilai class Recall	Nilai class Precision
Harga	973	Harga, mahal, daya beli, menurun, rendah	96.91%	92.05%
Pembayaran	74	Bayar, langsung, jangka,	96.10%	98.67%
Program Promosi	60	Hari, promo, penjualan, turun, naik	92.31%	93.75%
Packaging	4	Kemasan, baik, rusak,	80.00%	80.00%
Kualitas Produk	22	Bagus, terjamin, kadaluarsa	91.67%	80.00%
Isu Lainnya	753	Kios, sudah non aktif, tidak, menjual pupuk	90.29%	96.17%

i) Deployment

Setelah proses pemodelan data mining klasifikasi dengan menerapkan algoritma *K-Nearest Neighbor* (KNN), selanjutnya adalah tahapan implementasi dari proses data mining yang dilakukan yaitu dengan penyusunan laporan. Berdasarkan hasil yang didapat dengan algoritma *K-Nearest Neighbor* (KNN) dengan pemodelan CRISP-DM menunjukkan bahwa tingkat akurasi yang dihasilkan sebesar 93.88% dengan data sebanyak 2.500 data. Nilai tersebut membuktikan bahwa model yang diterapkan dalam penelitian ini dapat



digunakan untuk melakukan klasifikasi teks mining dan ketepatan dalam mengelompokkan isu tenaga kerja dilapangan. Pemahaman terhadap macam-macam isu yang terjadi dapat menjadi salah satu pertimbangan dalam melakukan pengambilan sebuah keputusan dalam memasarkan produk pada PT. XYZ.

4. SIMPULAN

Berdasarkan hasil dan pembahasan yang telah dilaksanakan maka dapat diambil kesimpulan bahwa berdasarkan perhitungan dengan menerapkan model data mining CRISP-DM algoritma K-Nearest Neighbor (KNN) dapat ditarik kesimpulan bahwa tingkat akurasi yang diperoleh yaitu 93.88%, dengan nilai recall tertinggi diperoleh oleh klasifikasi harga sebesar 96, 91% yang termasuk klasifikasi harga. Dan nilai precision tertinggi diperoleh oleh kualifikasi pembayaran sebesar 98.67%. Dari hasil observasi terhadap data isu tenaga kerja di lapangan dengan melalui tahapan klasifikasi menggunakan algoritma K-Nearest Neighbor (KNN) didapatkan jumlah nilai karakter yang dibuat oleh para distributor, kios, maupun petani yaitu harga.

DAFTAR PUSTAKA

- [1] Alfiani, A. M., Saptono, R., & Anggraningsih, R. (2015, Juni 01). Sistem Klasifikasi Feedback Pelanggan Dan Rekomendasi Solusi Atas Keluhan Di Upt Puskom Uns Dengan Algoritma Naïve Bayes Classifier Dan Cosine Similarity. *Jurnal Itsmart*, 4(1), 36 - 42.
- [2] Darma, D. S., & Saragih, L. (2019, 05). Peranan Sales Promotion Representative Terhadap Volume Penjualan Shinzui Pada Pt. Focus Distribusi Nusantara. *Manajemen : Jurnal Ekonomi*, 1(1), 1 - 8. Doi:<https://doi.org/10.36985/Manajemen.V1i1.76>
- [3] Deviyanto, A., & Wahyudi, M. R. (2018, Mei). Penerapan Analisis Sentimen Pada Pengguna Twitter Menggunakan Metode K-Nearest Neighbor (Knn). *Jisika (Jurnal Informatika Sunan Kalijaga)*, 3(1), 1-13.
- [4] Gata, W., & Purnomo. (2017). Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbor Pada Data Content Berita Sms. *Jurnal Format*, 6(1), 1-13.
- [5] Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017, Desember). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat Media Sosial Twitter Menggunakan Metode K-Nearest Neighbor Dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(12), 1750-1757.
- [6] Putra, M. R., & Nova Wardani, K. R. (2020, Juni). Penerapan Text Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial. *Jutim (Jurnal Teknik Informatika Musirawas)*, 05(01), 63-71.
- [7] Ruhyana, N., & Rosiyadi, D. (2019). Klasifikasi Komentar Instagram Untuk Identifikasi Keluhan Pelanggan Jasa Pengiriman Barang Dengan Teknik Smote. *Faktor Exacta*, 12(4), 280 - 290. Doi: [10.30998/Faktorexacta.V12i4.4981](https://doi.org/10.30998/Faktorexacta.V12i4.4981)
- [8] Sukma, A., Zaman, B., & Purwati, E. (2015, Desember). Klasifikasi Temu Kembali Informasi Dengan K-Nearest Neighbor. *Record And Library Journal*, 1(2), 129-138.
- [9] T. Rismawan, "Aplikasi Klasifikasi Jenis Tumbuhan Mangrove Berdasarkan Karakteristik Morfologi Menggunakan Metode K-Nearest Neighbor (Knn) Berbasis Web," 2016.