

**ANALISIS POTENSI PASAR MARKETPLACE  
TERHADAP PENJUALAN**

**THESIS**

**Oleh :  
SYA'RONI  
NIM. 200605220006**



**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2022**

**ANALISIS POTENSI PASAR MARKETPLACE  
TERHADAP PENJUALAN**

**THESIS**

**Diajukan kepada:  
Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Untuk memenuhi Salah Satu Persyaratan dalam  
Memperoleh Gelar Magister Komputer (M.Kom)**

**Oleh :  
SYA'RONI  
NIM. 200605220006**

**PROGRAM STUDI MAGISTER INFORMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM  
MALANG  
2022**


**ANALISIS POTENSI PASAR MARKETPLACE  
TERHADAP PENJUALAN**

**THESIS**

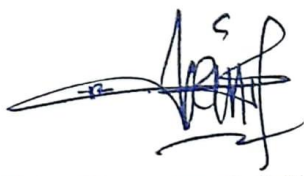
**Oleh :  
SYA'RONI  
NIM. 200605220006**

Telah diperiksa dan disetujui untuk diuji:  
Tanggal: 01 Desember 2022

Pembimbing I,

  
Dr. Bahyo Crysdiان  
NIP. 19740424 200901 1 008

Pembimbing II,

  
Dr. Ririen Kusumawati, S.Si., M.Kom  
NIP. 19830616201101 1 004

Mengetahui,  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Crysdiان  
NIP. 19740424 200901 1 008

# ANALISIS POTENSI PASAR MARKETPLACE TERHADAP PENJUALAN

## THESIS

Oleh :  
**SYA'RONI**  
**NIM. 200605220006**

Telah Dipertahankan di Depan Dewan Penguji Thesis  
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan  
Untuk Memperoleh Gelar Magister Komputer (M.Kom)  
Tanggal: 12 Desember 2022

### Susunan Dewan Penguji


Penguji Utama : Dr. M. Amin Hariyadi, MT  
NIP. 19670018 200501 1 001


Ketua Penguji : Dr. Yunifa Miftachul Arif, M.T  
NIP. 19830616 201101 1 004


Sekretaris Penguji : Dr. Ririen Kusumawati, S.Si., M.Kom  
NIP. 19720309 200501 2 002


Anggota Penguji : Dr. Cahyo Crysdian  
NIP. 19740424 200901 1 008

### Tanda Tangan

(  )

(  )

(  )

(  )

Mengetahui dan Mengesahkan  
Ketua Program Studi Magister Informatika  
Fakultas Sains dan Teknologi  
Universitas Islam Negeri Maulana Malik Ibrahim Malang



Cahyo Crysdian  
19740424 200901 1 008

## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Nama : Sya'Roni  
NIM : 200605220006  
Program Studi : Magister Informatika  
Fakultas : Sains dan Teknologi

Menyatakan dengan sebenarnya bahwa Thesis yang saya tulis ini benar-banar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan Thesis ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 Desember 2022  
Yang membuat pernyataan,



Sya'Roni  
NIM. 200605220006

## MOTTO

فَاصْبِرْ إِنَّ وَعْدَ اللَّهِ حَقٌّ وَلَا يَسْتَخِفَّنَكَ الَّذِينَ لَا يُوقِنُونَ<sup>٤</sup>

Bersabarlah kamu, sesungguhnya janji Allah adalah benar dan sekali-kali janganlah orang-orang yang tidak meyakini (kebenaran ayat-ayat Allah) itu menggelisahkan kamu. (Q.S Ar-Rum: 60)

## **HALAMAN PERSEMBAHAN**

Thesis ini dipersembahkan untuk Bapak dan Ibu yang selalu mencurahkan segala daya dan upayanya selama menjalani pendidikan serta Istri dan putriku untuk doa dan supportnya.

## KATA PENGANTAR

*Assalamu'alaikum Wr. Wb.*

Syukur alhamdulillah penulis haturkan kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan studi di Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang sekaligus menyelesaikan Thesis ini dengan baik.

Selanjutnya penulis haturkan ucapan terima kasih seiring do'a dan harapan jazakumullah ahsanal jaza' kepada semua pihak yang telah membantu terselesaikannya Thesis ini. Ucapan terima kasih ini penulis sampaikan kepada:

1. Bapak Dr. Cahyo Crysdiand dan Ibu Dr. Ririen Kusumawati, S.Si., M.Kom selaku dosen pembimbing Thesis, yang telah banyak memberikan pengarahan dan pengalaman yang berharga.
2. Segenap sivitas akademika Program Studi Magister Informatika, terutama seluruh Bapak/ Ibu dosen, terima kasih atas segenap ilmu dan bimbingannya.
3. Ayahanda dan Ibunda tercinta yang senantiasa memberikan doa dan restunya kepada penulis dalam menuntut ilmu.
4. Istri dan anak penulis yang selalu memberikan semangat kepada penulis untuk menyelesaikan Thesis ini.
5. Semua pihak yang ikut membantu dalam menyelesaikan Thesis ini baik berupa materiil maupun moril.

Penulis menyadari bahwa dalam penyusunan Thesis ini masih terdapat kekurangan dan penulis berharap semoga Thesis ini bias memberikan manfaat kepada para pembaca khususnya bagi penulis secara pribadi. *Amin Ya Rabbal Alamin.*

*Wassalamu'alaikum Wr. Wb.*

Malang, Desember 2022  
Penulis



## DAFTAR ISI

<b>HALAMAN JUDUL</b> .....	i
<b>HALAMAN PENGAJUAN</b> .....	ii
<b>HALAMAN PERSETUJUAN</b> .....	iii
<b>HALAMAN PENGESAHAN</b> .....	iv
<b>HALAMAN PERNYATAAN</b> .....	v
<b>MOTTO</b> .....	vi
<b>HALAMAN PERSEMBAHAN</b> .....	vii
<b>KATA PENGANTAR</b> .....	viii
<b>DAFTAR ISI</b> .....	ix
<b>DAFTAR GAMBAR</b> .....	x
<b>DAFTAR TABEL</b> .....	xi
<b>DAFTAR SIMBOL</b> .....	xii
<b>ABSTRAK</b> .....	xiii
<b>BAB I PENDAHULUAN</b>	
1.1. Latar Belakang .....	1
1.2. Pernyataan Masalah .....	6
1.3. Tujuan Penelitian .....	6
1.4. Batasan Masalah .....	6
1.5. Manfaat Penelitian .....	6
<b>BAB II STUDI LITERATURE</b>	
2.1. Analisis Potensi Pasar Marketplace Terhadap Penjualan .....	7
2.2. Kerangka Teori .....	13
<b>BAB III MODEL PENELITIAN</b>	
3.1. Prosedur Penelitian .....	15
3.2. Pengumpulan dan Persiapan Data .....	15
3.3. Pengembangan Sistem .....	21
3.4. Uji Coba .....	28
<b>BAB IV METODE KNN</b>	
4.1. Desain Metode.....	26
4.2. Uji Coba .....	28
4.3. Kesimpulan .....	36
<b>BAB V METODE C4.5</b>	
5.1. Desain Metode.....	37
5.2. Uji Coba .....	39
5.3. Kesimpulan .....	46
<b>BAB VI PEMBAHASAN</b>	
6.1. Pembahasan .....	47
<b>BAB VII KESIMPULAN</b>	
7.1. Kesimpulan .....	52

7.2. Saran .....	52
<b>DAFTAR PUSTAKA .....</b>	<b>54</b>
<b>LAMPIRAN</b>	

## DAFTAR GAMBAR

Gambar 2.1 Kerangka Teori .....	14
Gambar 3.1 Diagram Blok Metode Penelitian .....	15
Gambar 3.3 Alur Analisis KNN dan C4.5.....	20
Gambar 4.1 Desain Metode KNN .....	25
Gambar 4.2 Dataset Marketplace .....	27
Gambar 4.3 Dataset hasil preprocessing .....	28
Gambar 4.4 Pengolahan data KNN.....	29
Gambar 5.1 Desain Metode C4.5 .....	36
Gambar 5.2 Pohon keputusan C4.5 levels 4 .....	39
Gambar 6.1 Grafik Confusion matrix .....	46

## DAFTAR TABEL

Tabel 3.1 Data input output .....	16
Tabel 3.2 Dataset Awal Sebelum Dinormalisasi.....	21
Tabel 3.3 Dataset Hasil Setelah Dinormalisasi .....	21
Tabel 4.1 Hasil uji coba KNN.....	29
Tabel 5.1 Hasil uji coba C4.5.....	39
Tabel 6.1 Nilai confusion matrix.....	46

## ABSTRAK

Sya'Roni. 2022. **Analisis Potensi Pasar Marketplace Terhadap Penjualan**. Thesis. Program Studi Magister Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Cahyo Crysdiyan, MCS .(II) Dr. Ririen Kusumawati, S.Si., M.Kom

Kata kunci : *Marketplace, KNN, C4.5*.

Menjual suatu barang di zaman yang serba canggih ini sangatlah mudah, berbeda dengan dulu kita harus memiliki tempat atau toko untuk berjualan, sekarang ini kita tidak harus memiliki tempat atau toko untuk berjualan, kita bisa berjualan secara *online*. Marketplace memungkinkan transaksi *Customer to Customer (C2C)* antar konsumen tanpa terikat tempat dan waktu. Perubahan ini juga terjadi pada kebiasaan manusia berbelanja. Sehingga menjadi peluang bagi penjual untuk memasarkan barang dagangannya. Tetapi belum jelas kelompok kategorisasi terhadap pasar orang yang berbelanja. Sehingga dibutuhkan simulasi berbasis klasifikasi menggunakan *K-Nearest Neighbor* dan algoritma *C4.5*, dimana bobot masing-masing kelompok produk penjualan berpengaruh terhadap potensi pasar sehingga memberikan manfaat kepada penjual untuk memilih *marketplace* mana yang cocok dengan barang yang akan dijual. Hasil akurasi algoritma *C4.5* mendapat nilai yang lebih baik dibanding algoritma *K-Nearest Neighbor* yaitu 86% sedangkan hasil akurasi K-NN yaitu 70%, dapat disimpulkan bahwa algoritma *C4.5* lebih akurat untuk analisa potensi pasar *marketplace*.

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Melakukan kegiatan ekonomi merupakan salah satu cara untuk memenuhi setiap kebutuhan hidup manusia. Perdagangan adalah salah satu bentuk usaha yang sering kali dijalankan oleh manusia dimanapun berada. Mulai dari menjual berbagai macam setiap kebutuhan sehari-hari baik kebutuhan primer ataupun kebutuhan sekunder, sampai menjual setiap kebutuhan bahkan barang-barang yang sangat mahal demi kepuasan manusia. Dalam surah Al-Baqarah ayat 275 sudah dijelaskan terkait jual-beli berikut :

الَّذِينَ يَأْكُلُونَ الرِّبَا لَا يَقُومُونَ إِلَّا كَمَا يَقُومُ الَّذِي يَتَخَبَّطُهُ الشَّيْطَانُ مِنَ الْمَسِّ  
ذَلِكَ بِأَنَّهُمْ قَالُوا إِنَّمَا الْبَيْعُ مِثْلُ الرِّبَا وَأَحَلَّ اللَّهُ الْبَيْعَ وَحَرَّمَ الرِّبَا  
فَمَنْ جَاءَهُ مَوْعِظَةٌ مِنْ رَبِّهِ فَانْتَهَى فَلَهُ مَا سَلَفَ وَأَمْرُهُ إِلَى اللَّهِ ۗ وَمَنْ  
عَادَ فَأُولَٰئِكَ أَصْحَابُ النَّارِ ۗ هُمْ فِيهَا خَالِدُونَ

*“Orang-orang yang memakan riba tidak dapat berdiri melainkan seperti berdirinya orang yang kemasukan setan karena gila. Yang demikian itu karena mereka berkata bahwa jual beli sama dengan riba. Padahal Allah telah menghalalkan jual beli dan mengharamkan riba. Barangsiapa mendapat peringatan dari Tuhannya, lalu dia berhenti, maka apa yang telah diperolehnya dahulu menjadi miliknya dan urusannya (terserah) kepada Allah. Barangsiapa mengulangi, maka mereka itu penghuni neraka, mereka kekal di dalamnya” (QS Al-Baqarah : 275).*

Menurut Dalhari (2021) Quraish Shihab dalam Tafsir Al-Mishbah menyatakan bahwa jual beli sangat menguntungkan. Keuntungan bisa didapat dengan dua acara. Pertama diperoleh dengan bekerja dengan orang, sedangkan keuntungan kedua diperoleh dengan jual beli. Keduanya membutuhkan aktivitas manusia. Sedangkan riba adalah salah satu tindakan yang paling tercela dalam kehidupan seorang mukmin. Jika itu yang terjadi di masa lalu, maka harus berhenti menjalani kehidupan yang begitu tercela karena muslim yang baik tidak akan melakukan riba. Jika sudah berhenti melakukan riba dan bertaubat, maka semua pelanggaran sebelumnya akan diampuni oleh Allah.

Berdagang atau menjual suatu barang di era modern ini sudah sangatlah mudah, tidak seperti dulu penjual harus mempunyai suatu tempat atau toko untuk berjualan. Di zaman sekarang penjual tidak harus memiliki tempat atau toko untuk berjualan, penjual bisa berjualan secara *online*. Penjual hanya memerlukan telepon genggam atau laptop dan koneksi jaringan internet serta barang yang akan di jual. Setelah semua yang di perlukan tersedia, penjual hanya perlu mengunjungi beberapa *marketplace* untuk menjual barang dagangan. Penjual juga mendapatkan beberapa keuntungan yang sangat banyak, misalnya penjual tidak perlu menyewa tempat atau toko. Penjual mendapatkan notifikasi saat barang yang di jual akan di beli *customer*. Penjual mendapatkan rekap data transaksi penjualan yang cukup akurat dari *marketplace*. (Artisya dan Saragih, 2022).

Dalam lingkungan bisnis yang global saat ini, setiap perusahaan dituntut untuk mampu bersaing dalam meningkatkan kinerja dan mampu

menyajikan laporan yang sesuai. Sehingga perusahaan dapat mengambil sebuah keputusan dari hasil kinerja dan laporan tersebut. Menurut Pransiska *et al.* (2017) di sisi lain perkembangan sebuah teknologi informasi mempengaruhi pemrosesan data untuk menghasilkan informasi yang lebih cepat, akurat, aktual dan relevan untuk meningkatkan ketepatan atau akurasi dalam pengambilan sebuah keputusan.

*Marketplace* memungkinkan transaksi antara konsumen dengan konsumen atau *Customer to Customer* (C2C) tanpa terikat tempat dan waktu. Menurut *Iprice* (2022) data *iprice* pada kuartal empat tahun 2021 perkembangan data pengguna *marketplace* di indonesia sudah mencapai 382 juta pengguna, dimana penjual masih bingung di *marketplace* mana barang dagangannya akan dijual. Setiap *marketplace* mempunyai potensi pasarnya masing-masing sehingga membutuhkan klasifikasi potensi pasar untuk mengetahui potensi setiap *marketplace*. Salah satu tujuan yang banyak dihasilkan adalah klasifikasi. Tujuan klasifikasi adalah untuk memperkirakan kelas label objek yang tidak diketahui atau untuk membagi sesuatu ke dalam kelasnya menggunakan fungsi pengelompokan atau pengelompokan yang menjelaskan atau membedakan konsep atau kelas data.

Hampir semua toko baik *online* maupun *offline* saat ini menyimpan data penjualan dalam sistem yang terkomputerisasi, maka akan dihasilkan banyak data transaksi. *Volume record* transaksi penjualan yang tersimpan di data transaksi *database* penjualan barang sangat tinggi, sehingga mengakibatkan jumlah data yang terus bertambah. Menginvestigasi akumulasi data untuk



menemukan pola penjualan produk yang dapat digunakan untuk menganalisis pasar dan meramalkan penjualan di masa depan. Menggunakan metode atau teknik untuk memecahkan masalah klasifikasi bertujuan untuk mempermudah proses klasifikasi .

Dalam menghadapi persaingan pasar, para pelaku bisnis juga membutuhkan analisis untuk mengetahui produk apa yang paling laku menurut Pransiska *et al.* (2017). Sedangkan algoritma *K-Nearest Neighbor* pada penelitian Muzani *et al.* (2021) klasifikasi produk pada toko online yang telah diuji pada tiga kategori produk. Dengan menggunakan nilai  $k = 3$ , rata-rata nilai *accuracy*, *precision*, dan *recall* paling rendah saat melakukan klasifikasi produk. Kemudian dengan menggunakan  $k = 5$  atau  $k = 7$  menghasilkan nilai *recall*, *presisi*, dan akurasi tertinggi. Karena memiliki nilai tinggi nilai *precision*, dan *recall* selain nilai *accuracy* yang tinggi, algoritma *K-Nearest Neighbor* dapat digunakan secara efektif. Jarak antara setiap data uji dan seluruh data pelatihan perlu dihitung, proses klasifikasi semakin lama semakin banyak dataset yang digunakan.

*Consumer online rating and review* merupakan faktor yang mempengaruhi terhadap minat beli konsumen (Ichsan *et al.*, 2018). Sedangkan menurut Vilano dan Budi (2020) berdasarkan 54 responden menyatakan bahwa yang mempengaruhi minat pembelian adalah harga, jumlah penjualan, diskon serta deskripsi barang. Islami dan Prasetio (2018) dalam penelitiannya menggunakan algoritma CHAID menyatakan bahwa harga, kota penjual, jenis produk, jumlah penonton, dan peringkat adalah variabel independen.

Artisya dan Saragih, (2022) menarik kesimpulan bahwa pembeli dapat diprediksi dengan cukup akurat menggunakan perhitungan *C4.5*. Efek samping dari perkiraan ketepatan informasi yang diperoleh dari penyusunannya dengan nilai tambah *Confusion Matrix* adalah sebesar 91,33%. Kemudian tingkat ketepatan informasi pengujian mencapai 70,00%.

Berdasarkan beberapa penelitian-penelitian yang telah dilakukan sebelumnya, rumusan permasalahan yang diangkat pada penelitian ini adalah penerapan *K-Nearest Neighbor* dan algoritma *C4.5* untuk melakukan klasifikasi data produk dari beberapa *marketplace* di Indonesia. Penelitian ini mencoba menerapkan teknik data mining metode *K-Nearest Neighbor* dan algoritma *C4.5* pada analisis potensi pasar *marketplace* terhadap penjualan dan diharapkan dapat memberikan informasi berupa klasifikasi penjualan produk (laris dan tidak laris) dimana model *K-Nearest Neighbor* dan algoritma *C4.5* masing-masing mempunyai jarak dan pohon keputusan yang mampu memberikan akurasi klasifikasi yang baik. Sehingga kedepannya penjual dapat melakukan analisa potensi pasar *marketplace* terhadap penjualannya.

## 1.2 Pernyataan Masalah

1. Faktor-faktor apakah yang mempengaruhi kompleksitas *marketplace* yang ada di Indonesia ?
2. Metode apakah yang paling optimal digunakan untuk analisis potensi pasar *marketplace* terhadap penjualan ?

### **1.3 Tujuan Penelitian**

1. Menganalisis faktor-faktor yang mempengaruhi kompleksitas *marketplace* yang ada di Indonesia.
2. Mengevaluasi metode yang paling optimal digunakan untuk analisis potensi pasar *marketplace* terhadap penjualan.

### **1.4 Batasan Masalah**

1. Data diambil dari 5 besar *marketplace* dengan pengunjung terbanyak menurut Iprice pada Q4 2021.
2. Data diambil pada tanggal 20 april 2022.

### **1.5 Manfaat Penelitian**

Output dari penelitian ini dapat dimanfaatkan oleh pihak-pihak sebagai berikut:

1. Direktorat Jenderal Pajak untuk Pemotong/Pemungut Pajak yang sesuai dengan penjualan.
2. Dinas Perindustrian dan Perdagangan untuk mengetahui pertumbuhan ekonomi
3. Asosiasi UMKM Indonesia (Akumindo) untuk peningkatan penjualan UMKM

## BAB II

### STUDI LITERATURE

#### 2.1 Analisis Potensi Pasar *Marketplace* Terhadap Penjualan

Penelitian terkait mengenai pembahasan analisis potensi pasar *marketplace* terhadap penjualan telah banyak dilakukan pada penelitian sebelumnya.

Menurut Pransiska *et al.* (2017) dalam menghadapi persaingan pasar, pelaku bisnis juga harus memiliki analisis yang dapat memprediksi produk yang paling populer. Karena *Naive Bayes* didasarkan pada asumsi penyederhanaan bahwa nilai atribut independen secara kondisional ketika diberi nilai output, penulis menggunakan metode *Naive Bayes*. Pemanfaatan *Naive Bayes* dalam ulasan ini memerlukan jumlah informasi penyiapan yang cukup terbatas untuk menentukan pengukur batas yang diperlukan dalam sistem pengelompokan. Data string bersifat konstan dan data numerik kontinu dibedakan dalam metode *Naive Bayes*. Saat menentukan perbedaan ini akan menentukan.

Penelitian yang dilakukan Muzani *et al.* (2021) bertujuan untuk pembobotan algoritma *K-Nearest Neighbor* menggunakan metode TF-IDF untuk menentukan nilai akurasinya. Dengan menggunakan rumus *cosine similarity*, jarak kemiripan antara dokumen uji dan dokumen training dihitung menggunakan hasil pembobotan TF-IDF. Algoritma *K-Nearest Neighbor*

digunakan untuk mengklasifikasikan dokumen setelah diperoleh hasil *cosine similarity*. Selain itu, hasil menunjukkan bahwa algoritma *K-Nearest Neighbor* efektif sehingga memungkinkan untuk rekomendasinya.

Metode yang digunakan Astri *et al.* (2021) K-Modes Clustering dengan nilai  $k = 2,3,4,\dots,10$ , digunakan untuk mengklasifikasikan tingkat minat belanja online. Jumlah *cluster* terbaik ditentukan dengan menggunakan *Davies-Boulden Index* (DBI). Jumlah *cluster* terbaik diperoleh dari analisis, dengan  $k = 9$  dan nilai DBI 1,3427. Laki-laki muda berusia antara 15 dan 24 tahun membentuk *cluster* ini. Anggota *Cluster 5* adalah *cluster* terbaik, memiliki minat yang besar belanja di media sosial dan *marketplace*. Shopee, Bukalapak, dan Tokopedia adalah *marketplace* yang sangat diminati. Instagram, Facebook, dan Media Chatting adalah platform media sosial yang sangat diminati.

Silalahi (2018) dalam penelitiannya menggunakan algoritma *K-Means* untuk mengelompokkan data menjadi beberapa *cluster* berdasarkan kesamaannya satu sama lain. Silalahi melakukan sembilan iterasi, dengan dua *cluster*, C0 dan C1, dimana C0 merepresentasikan produk yang tidak laku dan C1 merepresentasikan produk yang sangat laris. Pengelompokan ini dapat dilihat pada nilai *centroid* iterasi kedelapan dan kesembilan, dimana nilai *centroid* tersebut berasal dari rata-rata nilai penjualan bulan februari. laris, dan kelompok C1 memiliki produk yang paling. Kelompok produk C0 ada 657 barang yang tidak laris, sedangkan kelompok produk C1 ada 70 barang yang laris.

Yunianto dan Taryadi (2022) melakukan pemilihan *e-marketplace* menggunakan teknik pengambilan keputusan metode *fuzzy* dengan kualitas layanan, kualitas sistem, kualitas sistem informasi dan kualitas penyedia layanan sebagai kriteria. TK, BL, dan SH adalah tiga alternatif pilihan. Urutan nilai integral total untuk setiap derajat optimisme adalah  $SH > TP > BL$  berdasarkan hasil perhitungan.

Dalam penelitian Artisya dan Saragih (2022) pembeli dapat mengantisipasi dengan cukup akurat menggunakan perhitungan *C4.5*. Efek samping memperkirakan presisi informasi yang didapat dari penyiapan informasi dengan nilai tambah *Confusion Matrix* adalah 91,33%. Kemudian tingkat presisi informasi pengujian mencapai 70,00%. Jika dilihat dengan ROC, data testing memiliki presisi 0,838 dan memiliki akurasi 0,750.

Untuk mencari probabilita dari masing-masing kelas Karsito dan Monika (2018) dalam studinya membuat kalkulasi menggunakan algoritma *naive bayes* untuk memutuskan kelas mana yang “Laku”, “Sedang” dan “Tidak Laku” dalam data pelatihan. Untuk menghitung probabilitas akhir setiap kelas, data pelatihan harus dikonversi ke nilai yang telah ditentukan. Nilai probabilitas untuk setiap atribut dikalikan dengan nilai probabilitas kelas untuk setiap atribut agar dapat membandingkan nilai tertinggi berdasarkan hasil yang telah ditentukan untuk setiap kelas. Hasilnya adalah jika kelas “TIDAK LAKU” bernilai lebih besar maka hasilnya “TIDAK LAKU”. Begitu pula dengan “SEDANG” dan “LAKU”.

Islami dan Prasetio (2018) dalam penelitiannya menggunakan algoritma CHAID menyatakan bahwa harga, kota penjual, jenis produk, jumlah penonton, dan peringkat adalah variabel independen. Selain itu, lebih baik menggunakan variabel independen sebagai simpul pemisah. Keputusan dibuat dengan membandingkan nilai  $p$  dari setiap variabel independen dari proses penggabungan dibedakan sebagai variabel utama dalam membentuk pohon keputusan dan menjadikan jumlah *viewer* sebagai elemen yang paling memengaruhi transaksi di Tokopedia.

Algoritma *K-Means* digunakan Yahya dan Mahpuz (2019) untuk menentukan jarak data dari setiap *centroid*. Perhitungan nilai *cluster* dihasilkan dengan menggunakan rumus jarak *Euclidean*. Hasil analisis penelitian ini membawa pada kesimpulan bahwa calon pelanggan SPS Motor Honda Lombok Timur adalah orang-orang berusia antara 33 sampai 46 tahun yang berprofesi sebagai petani dan wiraswasta. Karena nasabah merupakan usia kerja dan berpenghasilan.

Menurut Herdyansyah *et al.* (2020) tentang penerapan *FP-Growth* pada data penjualan. Berdasarkan penelitian ini, dapat ditarik kesimpulan bahwa data penjualan dapat diolah menggunakan teknik data mining untuk mengidentifikasi pola dalam *frequent itemset* dan menggunakannya sebagai informasi penting saat memilih strategi pemasaran dan pengaturan stok. Selain itu, algoritma *FP-Growth* memiliki metode yang sangat efektif untuk mengidentifikasi pola penjualan produk.

Song *et al.* (2019) mengusulkan pendekatan sederhana, terukur, dan non-parametrik untuk klasifikasi teks pendek dengan memilih kata-kata indikatif topik sebagai kata-kata kueri dari segmen teks pendek untuk diklasifikasikan. Pengambilan segmen teks berlabel yang paling cocok dengan kata-kata kueri. Label kelas yang diprediksi adalah suara terbanyak dari hasil pencarian. Hasilnya *Fuzzy* akurasi mencapai 81.27% dan *K-Nearest Neighbor* mencapai 85.17%.

Karthik dan Ganapathy (2021) membagi data uji dibagi menjadi dua bagian. Setengah digunakan untuk merujuk data riwayat pengguna, dan disebut sebagai data referensi masa lalu. Sedangkan sebagian lagi dari data yang digunakan untuk mengevaluasi hasil dari algoritma rekomendasi, disebut sebagai data prediktif. Berdasarkan riwayat pembelian dan skor produk sebelumnya, algoritme ini menghitung produk yang direkomendasikan serupa dan berbeda. Produk-produk diperiksa terhadap kumpulan data prediktif untuk mengetahui apakah produk yang direkomendasikan ada dalam kumpulan data prediktif. Kehadiran produk rekomendasi dalam kumpulan data prediktif memberikan kepercayaan dan cakupan produk serupa dalam daftar rekomendasi. Karena algoritma yang diusulkan juga mencakup produk yang berbeda untuk rekomendasi, keberadaan produk yang tidak diketahui dalam daftar rekomendasi untuk membandingkan kumpulan data prediktif juga dievaluasi.

Menurut Sharma *et al.* (2021) dalam penelitiannya menjelaskan nama produk, harga produk, ulasan dan peringkat produk yang terkait dengan produk



yang disarankan dari *Flipkart* dan *Amazon* menggunakan alat Otomatisasi. Proses Robotik dan data dibersihkan untuk menghilangkan nilai yang mengganggu dan berdasarkan permintaan pengguna, produk dari *Flipkart* dan *Amazon* disarankan di portal *e-commerce* berdasarkan preferensi pengguna pengguna dapat membeli produk dari situs mana pun yang sepenuhnya bergantung pada pengguna. Alat *RPA* Melalui *UiPath* digunakan untuk membuat robot perangkat lunak untuk *Flipkart dan Amazon*, bot ini digunakan untuk mengikis data dari *Flipkart dan Amazon* dan melakukan operasi pembersihan dan membuang data ke *database* yang dapat digunakan untuk penggunaan di masa mendatang. Dan bot dapat dipicu setiap hari untuk pengumpulan data untuk pembaruan basis data. Portal *E-Commerce* ini membantu untuk mengkategorikan dan membedakan perilaku pembelian produk dan memenuhi kebutuhan pelanggan akhir.

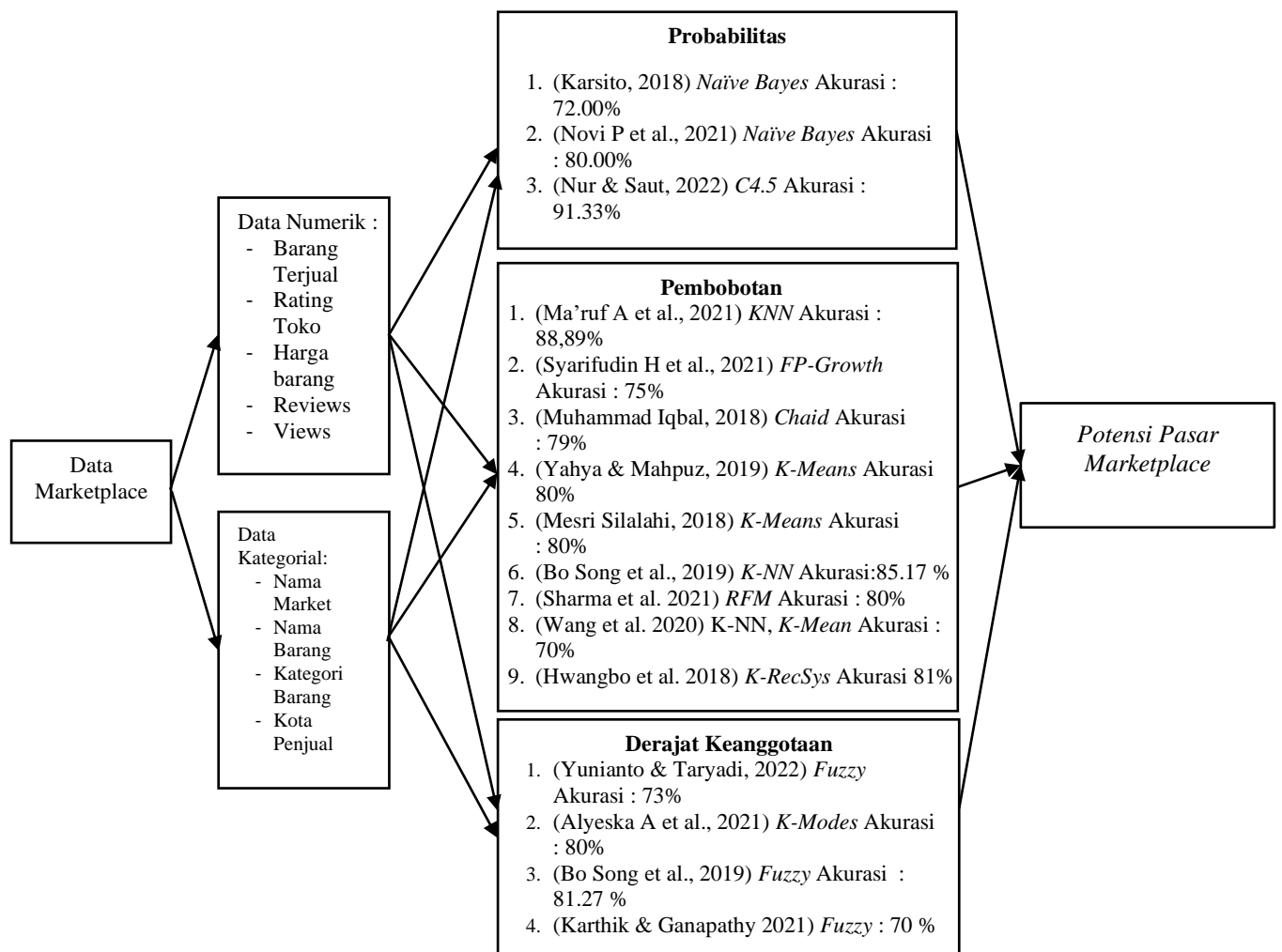
Wang *et al.* (2020) mengusulkan produk *e-commerce* yang dipersonalisasi sistem rekomendasi berbasis pembelajaran representasi *clustering*. Metode *K-Nearest Neighbor* memiliki keterbatasan dalam memilih kumpulan objek yang berdekatan. Wang memperkenalkan faktor tetangga dan fungsi waktu dan memanfaatkan model pemilihan dinamis untuk memilih kumpulan objek yang berdekatan.

Hwangbo *et al.* (2018) menghadirkan sistem penyaringan dan rekomendasi kolaboratif dunia nyata yang digunakan oleh perusahaan mode besar Korea yang menjual produk mode secara *online* dan *offline*. Karakteristik unik dari lingkungan rekomendasi perusahaan adalah sebagai berikut: Pertama,

produk dijual baik di perusahaan *online* maupun toko *offline*. Kedua, item *fashion* biasanya bersifat musiman. Sehingga preferensi umum pelanggan berubah seiring dengan musim. Terakhir, pelanggan biasanya membeli item untuk melengkapi atau mengganti item yang telah mereka beli sebelumnya. Sehingga mengusulkan sistem baru yang disebut *K-RecSys*, yang mengambil karakteristik domain dan memperluas algoritma penyaringan kolaboratif berbasis item standar. *K-RecSys* menggabungkan data klik produk dari penjualan *online* dan *offline* yang diberi bobot untuk mencerminkan preferensi pelanggan dari *online* dan *offline*. Penggunaan fungsi preferensi untuk mencerminkan pergeseran preferensi dari waktu ke waktu, dan menggunakan informasi tentang kategori produk untuk merekomendasikan produk alternatif dan pelengkap.

## **2.2 Kerangka Teori**

Penelitian ini memiliki kerangka teori yang mengacu pada jurnal-jurnal dan tergambar dengan jelas pada Gambar 2.1 berikut ini :



Gambar 2.1 Kerangka teori

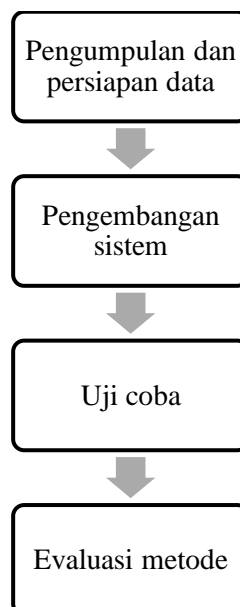
Dalam kerangka teori pada Gambar 2.1 dijelaskan Langkah-langkah yang dilakukan dalam menyusun studi pustaka. Data *marketplace* adalah data yang digunakan berasal dari data *skunder* yaitu hasil dari scraping dari beberapa *marketplace* dan beberapa kategori yang sudah ditentukan. Mengacu pada Gambar 2.1 peneliti memilih metode *K-Nearest Neighbor* dengan nilai akurasi sebesar 88,89% dan *C4.5* dengan nilai akurasi sebesar 91.33% sebagai metode yang digunakan dalam penelitian ini karena nilai akurasi pada metode tersebut memiliki nilai yang tertinggi.

## BAB III

### DESAIN PENELITIAN

#### 3.1. Prosedur Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dalam bentuk pengolahan dataset. Data yang digunakan dalam penelitian ini adalah data yang didapatkan dari salah satu dari 3 besar *marketplace* yang paling banyak dikunjungi berdasarkan *Iprice* Q4 2021 yaitu tokopedia <https://tokopedia.com> yang merupakan data sekunder. Prosedur penelitian dapat digambarkan pada Gambar 3.1.



Gambar 3.1. Diagram Blok Metode Penelitian

#### 3.2. Pengumpulan Dan Persiapan Data

##### A. Pengumpulan Data

Berdasarkan tujuan penelitian dan pendekatan yang digunakan dalam penelitian ini, pengambilan dataset dilakukan secara langsung melalui *website marketplace* tokopedia. Data terdiri dari 145 data dengan masing-masing data berjumlah 150 *record* dengan total 21.750 *record*. dari masing-masing data memiliki 11 atribut dengan nilai yang dimiliki oleh atribut merupakan nilai kategorikal dan nilai angka yaitu Link produk, Kategori 1, Kategori 2, Kategori 3, Nama Barang, Harga, *Reviews*, Terjual, *Views*, *Rating*, Lokasi dan Penjualan. Berikut data *input output* dapat dilihat pada Tabel 3.1.

Tabel 3.1 Data input output

<b>Index Data</b>	<b>Input</b>		<b>Output</b>
Elektronik	Alat Pendingin Ruangan	AC Portable	Sedang
Elektronik	Alat Pendingin Ruangan	AC Reflektor	Sedang
Elektronik	Alat Pendingin Ruangan	AC Standing	Sedang
Elektronik	Alat Pendingin Ruangan	Air Conditioner	Tinggi
Elektronik	Alat Pendingin Ruangan	Air Cooler	Sedang
Elektronik	Alat Pendingin Ruangan	Air Purifier	Tinggi
Elektronik	Alat Pendingin Ruangan	Kipas Angin Listrik	Sedang
Elektronik	Audio	Amplifier	Tinggi
Elektronik	Audio	Earphone	Tinggi
Elektronik	Audio	Headphone	Tinggi
Elektronik	Audio	Kabel & Konektor Audio	Sedang
Elektronik	Audio	Sound System	Tinggi
Elektronik	Audio	Speaker	Sedang
Elektronik	Audio	Voice Recorder	Sedang
Elektronik	Elektronik Dapur	Blender	Sedang
Elektronik	Elektronik Dapur	Juicer	Sedang
Elektronik	Elektronik Dapur	Kompor Listrik	Sedang
Elektronik	Elektronik Dapur	Kulkas	Tinggi

<b>Index Data</b>	<b>Input</b>		<b>Output</b>
Elektronik	Elektronik Dapur	Microwave	Sedang
Elektronik	Elektronik Dapur	Mixer	Tinggi
Elektronik	Elektronik Dapur	Oven	Sedang
Elektronik	Elektronik Dapur	Rice Cooker	Sedang
Elektronik	Elektronik Dapur	Slow Cooker	Tinggi
Elektronik	Elektronik Dapur	Toaster	Rendah
Elektronik	Elektronik Kantor	Kartu RFID	Tinggi
Elektronik	Elektronik Kantor	Laser Pointer	Sedang
Elektronik	Elektronik Kantor	Layar Proyektor	Rendah
Elektronik	Elektronik Kantor	Mesin Absensi	Sedang
Elektronik	Elektronik Kantor	Mesin Fax	Rendah
Elektronik	Elektronik Kantor	Mesin Fotocopy	Sedang
Elektronik	Elektronik Kantor	Mesin Hitung Uang	Sedang
Elektronik	Elektronik Kantor	Mesin Jilid	Tinggi
Elektronik	Elektronik Kantor	Mesin Kasir	Sedang
Elektronik	Elektronik Kantor	Mesin Laminating	Sedang
Elektronik	Elektronik Kantor	Mesin Penghancur Kertas	Sedang
Elektronik	Elektronik Kantor	Reader RFID	Sedang
Elektronik	Elektronik Kantor	LED Proyektor	Tinggi
Elektronik	Elektronik Rumah Tangga	Alat Pengering Baju	Sedang
Elektronik	Elektronik Rumah Tangga	Hand Blower	Tinggi
Elektronik	Elektronik Rumah Tangga	Mesin Cuci	Tinggi
Elektronik	Elektronik Rumah Tangga	Setrika	Sedang
Elektronik	Elektronik Rumah Tangga	Setrika Uap	Sedang
Elektronik	Elektronik Rumah Tangga	Vacuum Cleaners	Tinggi
Elektronik	Kamera Pengintai	Access Control Door	Sedang
Elektronik	Kamera Pengintai	Alarm	Tinggi
Elektronik	Kamera Pengintai	Kamera CCTV	Tinggi
Elektronik	Kamera Pengintai	Motion Detector	Rendah
Elektronik	Kamera Pengintai	Spy Camera	Rendah
Elektronik	Media Player	DVD Player	Sedang
Elektronik	Media Player	MP3 & MP4 Player	Sedang
Elektronik	Media Player	Radio & Tape Player	Sedang
Elektronik	Printer	Printer Ink Jet	Sedang
Elektronik	Printer	Scanner	Tinggi
Elektronik	Printer	Tinta Printer	Sedang
Elektronik	Printer	Toner Printer	Sedang

<b>Index Data</b>	<b>Input</b>		<b>Output</b>
Elektronik	Telepon	Kabel & Konektor Telepon	Tinggi
Elektronik	Telepon	Telepon Kabel	Sedang
Elektronik	Telepon	Telepon Satelit	Sedang
Elektronik	Telepon	Telepon Wireless	Sedang
Elektronik	Telepon	Walkie Talkie	Sedang
Elektronik	Televisi	Antena TV & Parabola	Sedang
Elektronik	Televisi	Kabel & Konektor	Sedang
Elektronik	Televisi	Remote TV	Sedang
Elektronik	Televisi	Televisi	Tinggi
Elektronik	Televisi	TV Box	Sedang
Fashion	Aksesoris Anak	Bando Bandana Anak	Rendah
Fashion	Aksesoris Anak	Dasi Anak	Rendah
Fashion	Aksesoris Anak	Ikat Rambut Anak	Rendah
Fashion	Aksesoris Anak	Kacamata Anak	Sedang
Fashion	Aksesoris Anak	Topi Anak	Rendah
Fashion	Atasan Wanita	Blazer Wanita	Rendah
Fashion	Atasan Wanita	Daster Wanita	Sedang
Fashion	Atasan Wanita	Jaket Parka Wanita	Rendah
Fashion	Atasan Wanita	Kaos Wanita	Sedang
Fashion	Atasan Wanita	Kemeja Wanita	Rendah
Fashion	Jam Tangan Pria	Aksesoris Jam Tangan Pria	Rendah
Fashion	Jam Tangan Pria	Jam Tangan Analog Pria	Tinggi
Fashion	Jam Tangan Pria	Jam Tangan Digital Pria	Sedang
Fashion	Jam Tangan Wanita	Jam Tangan Analog Wanita	Sedang
Fashion	Jam Tangan Wanita	Jam Tangan Digital Wanita	Sedang
Fashion	Jam Tangan Wanita	Strap Jam Wanita	Rendah
Fashion	Muslim	Baju Koko Pria	Sedang
Fashion	Muslim	Bros Jilbab	Sedang
Fashion	Muslim	Gamis Wanita	Sedang
Fashion	Muslim	Hijab Segi Empat	Tinggi
Fashion	Muslim	Mukena	Sedang
Fashion	Muslim	Pashmina	Sedang
Fashion	Muslim	Peci & Kopiah	Tinggi
Fashion	Muslim	Setelan Syari Wanita	Rendah
Fashion	Pakaian Anak Laki Laki	Baju Tidur Anak Laki- Laki	Rendah
Fashion	Pakaian Anak Laki Laki	Celana Panjang Anak Laki-Laki	Rendah
Fashion	Pakaian Anak Laki Laki	Jaket Anak Laki-Laki	Rendah

<b>Index Data</b>	<b>Input</b>		<b>Output</b>
Fashion	Pakaian Anak Laki Laki	Kaos Anak Laki-Laki	Tinggi
Fashion	Pakaian Anak Laki Laki	Kemeja Anak Laki-Laki	Rendah
Fashion	Pakaian Anak Perempuan	Baju Tidur Anak Perempuan	Rendah
Fashion	Pakaian Anak Perempuan	Celana Panjang Anak Perempuan	Rendah
Fashion	Pakaian Anak Perempuan	Dress Anak Perempuan	Rendah
Fashion	Pakaian Anak Perempuan	Kulot Anak Perempuan	Rendah
Fashion	Pakaian Anak Perempuan	Rok Anak Perempuan	Rendah
Fashion	Sepatu Anak	Flat Shoes Anak Perempuan	Rendah
Fashion	Sepatu Anak	Pantofel Anak Laki-Laki	Rendah
Fashion	Sepatu Anak	Sandal Anak Laki Laki	Rendah
Fashion	Sepatu Anak	Sandal Anak Perempuan	Rendah
Fashion	Sepatu Anak	Sepatu Jelly Anak Perempuan	Rendah
Otomotif	Aksesori Body Motor	Aksesori Body Motor	Tinggi
Otomotif	Aksesori Body Motor	Cover Stang Motor	Rendah
Otomotif	Aksesori Body Motor	Jok motor	Rendah
Otomotif	Aksesori Body Motor	Karpet Motor	Rendah
Otomotif	Aksesori Body Motor	Lampu Motor	Tinggi
Otomotif	Aksesori Body Motor	Spion Motor	Rendah
Otomotif	Aksesoris Pengendara Motor	Jaket Motor	Rendah
Otomotif	Aksesoris Pengendara Motor	Jas Hujan	Tinggi
Otomotif	Aksesoris Pengendara Motor	Rompi Motor	Rendah
Otomotif	Aksesoris Pengendara Motor	Sarung Tangan Motor	Tinggi
Otomotif	Aksesoris Pengendara Motor	Sepatu Biker	Sedang
Otomotif	Helm Motor	Helm Anak	Rendah
Otomotif	Helm Motor	Helm Full Face	Rendah
Otomotif	Helm Motor	Helm Retro	Sedang
Otomotif	Helm Motor	Kaca Helm	Rendah
Otomotif	Helm Motor	Kunci Helm	Rendah
Otomotif	Mobil	Mobil Hatchback & City Car	Tinggi



<b>Index Data</b>	<b>Input</b>		<b>Output</b>
Otomotif	Mobil	Mobil Niaga	Tinggi
Otomotif	Mobil	Mobil Sedan	Tinggi
Otomotif	Mobil	Mobil SUV & MPV	Tinggi
Otomotif	Mobil	Velg Mobil	Tinggi
Otomotif	Oli & Penghemat Bbm	Engine Conditioner	Sedang
Otomotif	Oli & Penghemat Bbm	Filter Oli	Sedang
Otomotif	Oli & Penghemat Bbm	Oli mobil	Tinggi
Otomotif	Oli & Penghemat Bbm	Oli motor	Sedang
Otomotif	Oli & Penghemat Bbm	Penghemat BBM	Sedang
Otomotif	Oli & Penghemat Bbm	Tempat Oli	Sedang
Otomotif	Sepeda Motor	Motor Bebek	Sedang
Otomotif	Sepeda Motor	Motor Matic	Tinggi
Otomotif	Sepeda Motor	Motor Sport	Tinggi
Otomotif	Sepeda Motor	Velg Motor	Sedang
Otomotif	Spare Part Mobil	Busi Mobil	Tinggi
Otomotif	Spare Part Mobil	Kampas Rem Mobil	Rendah
Otomotif	Spare Part Mobil	Piston Mobil	Rendah
Otomotif	Spare Part Mobil	Shockbreaker & Kaki Kaki Mobil	Sedang
Otomotif	Spare Part Mobil	Tromol & Piringan Rem Mobil	Sedang
Otomotif	Spare Part Motor	Aki Motor	Sedang
Otomotif	Spare Part Motor	Busi Motor	Sedang
Otomotif	Spare Part Motor	Kampas Rem Motor	Sedang
Otomotif	Spare Part Motor	Rantai & Gir Motor	Tinggi
Otomotif	Spare Part Motor	Handle - Handfat Motor	Rendah

Khusus untuk atribut *output* ditentukan oleh *expert* yaitu Drs. Yusuf, MM. dengan profesi pemasaran di The Taman Dayu Pandaan dengan gelar akademik S1 Pemasaran dan S2 Managemen untuk mengetahui apakah data penjualan tersebut termasuk tinggi, sedang atau rendah.

#### B. Persiapan Data

Pada tahap ini dilakukan pembersihan nilai-nilai dalam atribut yang kosong dan menghapus data yang kosong. Selanjutnya diperlukan proses normalisasi data terlebih dahulu bertujuan untuk menghindari terjadinya anomali data dan tidak konsistensinya sebuah data, yang mana dataset yang didapat pada penelitian ini terdapat 2 tipe data yaitu *integer* dan data *real* sehingga perlu dilakukan normalisasi data dengan menggunakan metode *min-max*. Rumus *min-max normalization* sebagai berikut.

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \dots \dots \dots (3.1)$$

dengan

$X^*$  : data baru

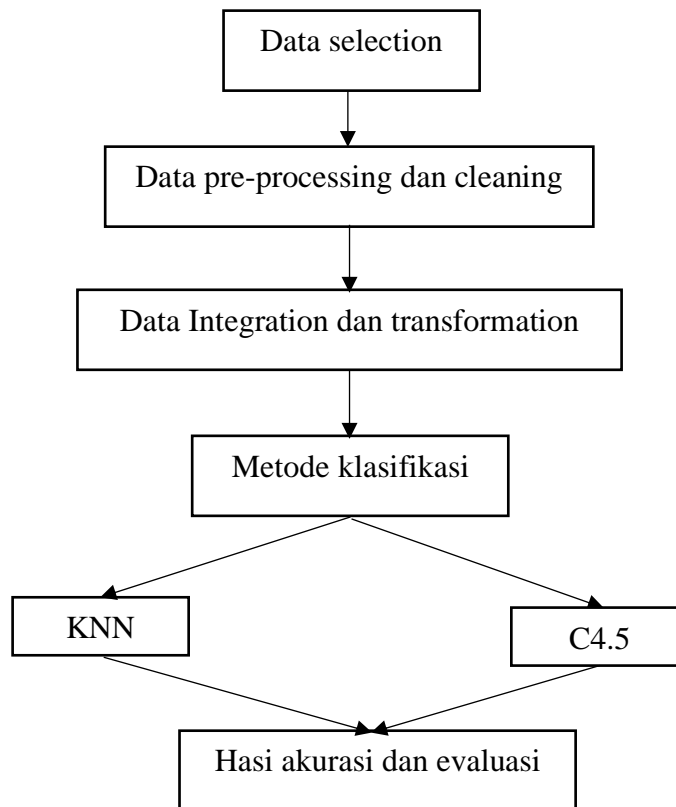
$X$  : data lama

$\min(X)$  : nilai minimum dari data per kolom

$\max(X)$  : nilai maksimum dari data per kolom

### 3.3. Pengembangan Sistem

Untuk melakukan klasifikasi menggunakan algoritma KNN dan C.45 pada penelitian ini, tahapan-tahapan yang digunakan dari metode KDD (*Knowledges Discovery in Databases*). Metode KDD yang digunakan adalah *data selection*, *data pre-processing*, *data integration*, *data transformation*, data mining, dan *evaluation* (Putri, 2021). Berikut alur analisis data klasifikasi KNN dan C4.5 pada data penjualan *marketplace*.



Gambar 3. 3 Alur Analisis KNN dan C4.5

#### 1. Data Selection

Dataset awal yang digunakan dalam penelitian ini menggunakan data penjualan. Dari dataset data awal selanjutnya dilakukan proses seleksi untuk mencari atribut yang dibutuhkan klasifikasi data penjualan.

#### 2. Data Pre-Processing dan cleaning

Pada tahap ini dilakukan pembersihan nilai-nilai dalam atribut yang kosong dan menghapus data yang kosong (*missing values* dan *noisy*).

#### 3. Data Integration dan Transformation

Setelah tahap data *pre-processing* dilakukan maka selanjutnya melakukan data *intergration* yaitu dengan menggabungkan semua variabel dataset ke dalam satu tabel. Selanjutnya proses normalisasi data terlebih

dahulu bertujuan untuk menghindari terjadinya anomali data dan tidak konsistensinya sebuah data menggunakan Rumus 3.1.

Tabel 3.2 : Dataset Awal Sebelum Dinormalisasi

NO	X1	X2	X3	X4	X5
1	160.000	3	5,0	10	0,00
2	18.000	50	5,0	376	0,00
3	19.150	19000	4,2	63000	0,65
4	19.500	196	4,8	574	0,00
5	7.000	183	4,9	10000	0,00
6	55.200	84000	4,7	10000	0,08
7	100.000	88	4,5	192	0,00
8	34.250	21000	4,7	59000	0,00
9	100.000	12000	4,7	25000	0,00
10	110.890	0	0,0	2	0,28
.....	.....	.....	.....	.....	.....
21747	300.000	0	0,0	0	0,00
21748	45.000	1	5,0	1	0,10
21749	67.500	0	0,0	1	0,10
21750	58.500	10	4,9	15	0,10

Sedangkan untuk hasil normalisasi dapat dilihat pada tabel 3.2 sebagai berikut.

Tabel 3.3 : Dataset Hasil Setelah Dinormalisasi

NO	X1	X2	X3	X4	X5
1	0,001848569	6,31579E-06	1	0,000101	0
2	0,000206948	0,000105263	1	0,003798	0
3	0,000220243	0,04	0,84	0,636364	0,698925
4	0,000224289	0,000412632	0,96	0,005798	0
5	7,97804E-05	0,000385263	0,98	0,10101	0
6	0,000637007	0,176842105	0,94	0,10101	0,086022
7	0,001154926	0,000185263	0,9	0,001939	0
8	0,00039481	0,044210526	0,94	0,59596	0
9	0,001154926	0,025263158	0,94	0,252525	0
10	0,001280822	0	0	2,02E-05	0,301075
.....	.....	.....	.....	.....	.....

21747	0,003467068	0	0	0	0
21748	0,000519087	2,10526E-06	1	1,01E-05	0,107527
21749	0,000779203	0	0	1,01E-05	0,107527
21750	0,000675157	2,10526E-05	0,98	0,000152	0,107527

Dari tabel diatas dapat disimpulkan bahwa dataset awal sebelumnya memiliki 2 tipe data setelah dilakukan proses normalisasi maka didapat 1 tipe data, sehingga kemudian dapat dilakukan proses.

#### 4. Klasifikasi

Setelah dilakukan transformasi, maka selanjutnya dilakukan proses klasifikasi dengan algoritma KNN dimana cara kerjanya melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Sedangkan C4.5 didasarkan pada nilai *gain ratio* tertinggi dari variabel-variabel yang ada.

### 3.4. Uji Coba

Pada tahap uji coba, data penjualan akan di proses menggunakan desain metode KNN dan C4.5 yang sudah disiapkan untuk mengetahui apakah data tersebut termasuk dalam kategori penjualan tinggi, sedang dan rendah. Sehingga dapat dicocokkan dengan *output* yang sudah ditentukan oleh *expert* atau pakar sebelumnya.

Untuk pengujian metode pada penelitian ini menggunakan *confusion matrix* yang dapat melakukan pengukuran kinerja dari klasifikasi. Menurut Febrian *et al.* (2022) metode evaluasi yang menggambarkan hasil dari proses

klasifikasi adalah *confusion matrix*: *true positives*, *true negatives*, *false positives*, dan *false negatives*.

1. *Accuracy*, persentase jumlah data yang terklasifikasi dengan benar oleh sistem menggunakan rumus berikut.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots\dots (3.2)$$

2. *Precision*, presentase akurasi antara data yang diminta dengan hasil klasifikasi menggunakan rumus berikut.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (3.3)$$

3. *Recall*, keberhasilan model klasifikasi dalam menentukan kembali informasi atau hasil nilai dengan menggunakan rumus berikut.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (3.4)$$

4. *F<sub>1</sub>-Score*, suatu perbandingan nilai rata-rata dari *precision* dan *recall* dengan menggunakan rumus berikut.

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \dots\dots\dots (3.5)$$

## BAB IV

### K-NEAREST NEIGHBOR (KNN)

#### 4.1. Desain Metode

Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data training yang jaraknya paling dekat dengan objek tersebut. Menurut Kurniawan dan Romadhon (2021) algoritma KNN sangat nonlinier karena model pembelajaran mesin nonparametrik sehingga garis keputusan kelas yang dihasilkan sangat fleksibel dan mudah diimplementasikan. KNN adalah sebuah metode untuk mencari kasus dengan menghitung kedekatan antara kasus baru dan kasus lama. KNN adalah salah satu metode yang digunakan untuk analisis klasifikasi, namun metode KNN juga digunakan untuk prediksi. (Noviana *et al.*, 2019).

Jarak antara dua titik pada data training dan titik pada data testing dapat didefinisikan dengan rumus *Euclidean*, seperti berikut.

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \dots\dots\dots (4.1)$$

Dengan

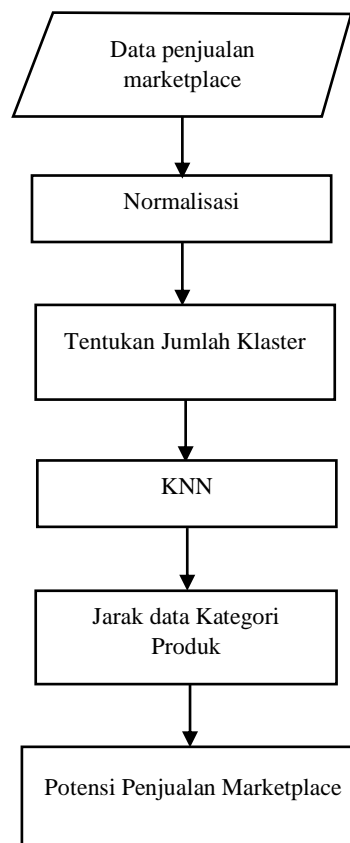
$d$  : jarak *Euclidean*

$x_{2i}$  : nilai pada data testing ke -  $i$

$x_{1i}$  : nilai pada data training ke -  $i$

$p$  : banyaknya atribut

Selanjutnya tahapan merancang dan mengimplementasikan sistem dimana sistem dirancang dan diimplemetasikan sesuai dengan tujuan penelitian yang telah difokuskan.



Gambar 4. 1 Desain Metode KNN

Pada Gambar 4.1 menjelaskan tentang rancangan desain metode KNN dengan tahapan-tahapan sebagai berikut:

- a. Pada tahap pertama, data pada tahap ini data yang digunakan secara langsung melalui *website marketplace* tokopedia tersedia melalui laman <https://tokopedia.com>.



- b. Setelah itu melakukan normalisasi data untuk meningkatkan performanya. Salah satu normalisasi yang umum digunakan adalah dengan *min-max* menggunakan Rumus 3.1.
- c. Langkah selanjutnya yaitu menentukan jumlah kluster berdasarkan kategori potensi pasar yaitu {Rendah, Sedang, Tinggi}
- d. Setelah itu akan di lakukan proses klasifikasi KNN . Proses inti dari sistem ini adalah proses perhitungan jarak terdekat antar data yang kemudian dijadikan dasar untuk klasifikasi menggunakan Rumus 4.1.
- e. Dari hasil KNN yang berupa perhitungan jarak terdekat antar data, dari jarak tersebut digunakan untuk mengetahui jarak potensi pasar menggunakan Rumus 4.3

$$\text{Jarak potensi} = \frac{d_i}{d_i+p} \dots\dots\dots (4.3)$$

- f. Setelah mengetahui jarak potensi pasar, data tersebut digunakan untuk melihat potensi pasar menggunakan Rumus 4.4.

$$\text{Potensi} = \text{Max}(\text{Jarak potensi}_i) \dots\dots\dots (4.4)$$

## 4.2. Uji Coba

Adapun untuk uji coba KNN sesuai dengan desain yang telah dikembangkan untuk menjalankan tahapan-tahapan sesuai perancangan yang telah dibuat sebelumnya berdasarkan Gambar 4.1.

### A. Data Penjualan Marketplace

Dataset yang digunakan pada penelitian ini seperti dijelaskan pada bab sebelumnya terdiri dari 145 data dengan masing-masing data berjumlah 150 *record* dengan total 21.750 *record*. dari masing-masing data memiliki 11 atribut dengan nilai yang dimiliki oleh atribut merupakan nilai kategorikal dan nilai angka yaitu Link produk, Kategori 1, Kategori 2, Kategori 3, Nama Barang, Harga, Reviews, Terjual, Views, Rating, Lokasi dan Penjualan yang dapat dilihat pada Gambar 4.2.

Product Name	Location	Category 3	Price	Reviews	Sold	Views	Rating	
1	C'PENDI...	Jakarta Selatan	AC Portable	1220000	4	7	172	5.0
2	ngin Po...	Jakarta Pusat	AC Portable	21500	10	76	2807	4.5
3	TUFE Po...	Jakarta Utara	AC Portable	3850000	21	54	32671	5.0
4	g Ruang...	Jakarta Barat	AC Portable	160000	38	51	2617	4.8
5	wered L...	Jakarta Barat	AC Portable	83900	59	78	6494	4.8
6	ir Purifi...	Jakarta Pusat	AC Portable	1849000	75	169	7435	5.0
7	i Purifier...	Jakarta Pusat	AC Portable	1085000	125	196	7853	5.0
8	in Robot...	Kota Cimahi	AC Portable	110000	22	41	1433	5.0
9	ii [Tanpa ...	Jakarta Pusat	AC Portable	82900	246	489	70066	4.6
10	0.5 - 1...	Kota Surabaya	AC Portable	80000	5	28	435	5.0
11	ii portab...	Kota Medan	AC Portable	86000	126	229	3862	4.7
12	MPF2-12...	Jakarta Selatan	AC Portable	4699000	10	47	3969	4.8
13	ngin leh...	Kota Surabaya	AC Portable	65000	6	11	665	5.0
14	n Pump ...	Jakarta Pusat	AC Portable	650000	139	284	18663	4.9
15	ipas ang...	Kota Surabaya	AC Portable	27210	1	1	107	5.0
16	tpa Khu...	Kota Bekasi	AC Portable	88900	2	39	750	5.0
17	isa Cuci ...	Kota Bekasi	AC Portable	70000	87	188	5843	4.9
18	ii PORTA...	Kota Malang	AC Portable	92000	142	243	5927	4.8
19	oo Air C...	Jakarta Selatan	AC Portable	299000	109	206	27931	4.8
20	Y-PRO ...	Kab. Bekasi	AC Portable	3930000	15	35	6165	4.7
21	P7807 L...	Jakarta Timur	AC Portable	249900	79	179	7216	5.0
22	i portabl...	Kota Tangerang	AC Portable	76500	42	79	1268	4.9
23	ng Hepa ...	Kota Surabaya	AC Portable	949630	56	136	3823	4.9
24	LOGO iu...	Jakarta Barat	AC Portable	5000	34	92	1248	4.8
25	lter Air ...	Kota Surabaya	AC Portable	699600	55	135	3326	4.9
26	eon R13...	Jakarta Barat	AC Portable	31500	47	120	6692	4.8
27	lEX HEP...	Jakarta Barat	AC Portable	350000	62	279	17844	4.9
28	d Coil Si...	Kota Malang	AC Portable	30000	2	9	131	5.0
29	ngin Gr...	Jakarta Barat	AC Portable	24900	5	26	299	5.0
30	ngin Ge...	Jakarta Utara	AC Portable	25000	21	65	5588	4.9
31	e Fan pl...	Kota Tangerang	AC Portable	21900	110	755	6272	4.5
32	lPk1 520...	Kab. Semarang	AC Portable	160000	4	25	855	4.8

Gambar 4.2 Dataset Marketplace

## B. Normalisasi

Pada tahapan ini dilakukan normalisasi data untuk menghindari terjadinya anomali data dan tidak konsistensinya sebuah data dengan tujuan untuk menghindari kesalahan. Perhitungan normalisasi menggunakan Rumus 3.1.

Berikut ini dataset hasil normalisasi data menggunakan Rumus 3.1 seperti diperlihatkan pada Gambar 4.3.

URL	Location	Category 3	Product Name	Price	Reviews	Sold	Views	Rating
13271	vv.to...	Bekasi Timur	Setelan Syari W...	-2284903.12	-269.88	-1733.81	-16796.76	-4.707
17195	vv.to...	Jakarta Barat	Sepatu Biker	-2084903.12	-85.88	-1271.81	8617.24	0.293
6681	vv.to...	Jakarta Barat	Alarm	-2378003.12	-14.88	1406.19	-10480.76	0.093
11074	vv.to...	Jakarta Barat	Kaos Wanita	-2359903.12	-174.88	-1138.81	-13120.76	0.193
1063	vv.to...	Jakarta Barat	Amplifier	-1923098.12	-236.88	-1684.81	-13019.76	0.193
7562	vv.to...	Jakarta Barat	Radio & Tape Pl...	-2277255.12	-246.88	-1699.81	-15812.76	0.293
2054	vv.to...	Jakarta Barat	Voice Recorder	-2191662.12	-254.88	-1699.81	-16391.76	-0.007
7451	vv.to...	Jakarta Barat	MP3 & MP4 Pla...	-2195473.12	-210.88	-1643.81	-14159.76	0.093
3991	vv.to...	Jakarta Barat	Layar Proyektor	-2304953.12	60.12	-1118.81	1441.24	0.193
13652	vv.to...	Jakarta Barat	Jaket Anak Laki...	-2339903.12	-230.88	-1662.81	-13905.76	0.193
13673	vv.to...	Jakarta Barat	Jaket Anak Laki...	-2339903.12	-230.88	-1662.81	-13905.76	0.193
7147	vv.to...	Jakarta Barat	Spy Camera	-2239903.12	-98.88	-1332.81	365.124	-0.107
8522	vv.to...	Jakarta Barat	Telepon Kabel	-2378403.12	-40.12	-300.81	-7056.76	0.193
7007	vv.to...	Jakarta Barat	Motion Detector	-2324903.12	-232.88	-1649.81	-14379.76	-0.007
6617	vv.to...	Jakarta Barat	Alarm	-2249903.12	-58.88	-1036.81	6508.24	0.093
7510	vv.to...	Jakarta Barat	Radio & Tape Pl...	-2344903.12	-259.88	-1673.81	-16057.76	0.293
8572	vv.to...	Jakarta Barat	Telepon Satelit	-2264903.12	-265.88	-1725.81	-16077.76	0.293
4734	vv.to...	Jakarta Barat	Mesin Jilid	-2320883.12	-95.88	-1410.81	-12709.76	0.193
18874	vv.to...	Jakarta Barat	Engine Conditio...	-2369903.12	-116.88	-1313.81	-12673.76	-0.007
18840	vv.to...	Jakarta Barat	Engine Conditio...	-2369903.12	-96.88	-1271.81	-9362.76	0.093
16209	vv.to...	Jakarta Barat	Lampu Motor	-2382418.12	1144.12	103240.19	50280.24	-0.107
17041	vv.to...	Jakarta Barat	Sarung Tangan ...	-2364903.12	374.12	-29.81	5413.24	0.093
16894	vv.to...	Jakarta Barat	Rompi Motor	-2367903.12	-119.88	-1212.81	-11915.76	0.093
16893	vv.to...	Jakarta Barat	Rompi Motor	-2367903.12	-73.88	-1167.81	-9571.76	0.093
16241	vv.to...	Jakarta Barat	Lampu Motor	-2382903.12	-167.88	-506.81	-13608.76	0.193
16362	vv.to...	Jakarta Barat	Spion Motor	-2379103.12	-4.88	6832.19	-3100.24	0.093
17089	vv.to...	Jakarta Barat	Sarung Tangan ...	-2379103.12	331.12	17194.19	25017.24	0.093
17068	vv.to...	Jakarta Barat	Sarung Tangan ...	-2379103.12	869.12	17194.19	25017.24	0.093
1194	vv.to...	Jakarta Barat	Amplifier	-2181203.12	114.12	-1060.81	23563.24	0.193
18445	vv.to...	Jakarta Barat	Mobil Sedan	-2334903.12	-261.88	-1689.81	-16552.76	0.293
12668	vv.to...	Jakarta Barat	Hijab Seqi Empat	-2362783.12	-244.88	-1607.81	-14264.76	0.293

Gambar 4.3 Dataset hasil preprocessing

### C. Tentukan Kluster

Jumlah kluster yang digunakan adalah berjumlah 3 kluster berdasarkan kategori potensi pasar yaitu {Rendah, Sedang, Tinggi} dengan tujuan mengetahui setiap potensi *marketplace* pada kategori produk yang sudah ditentukan.

### D. KNN

Setelah menentukan kluster yaitu proses KNN. Untuk pengolahan data menggunakan KNN dengan menghitung kuadrat jarak *euclid* (*Euclidean distance*) masing-masing obyek terhadap data yang diberikan dengan Rumus 4.1.

	Harga	Jumlah Ulasan	Diskon	Terjual
233	367607.18	-2384.09	-0.0909	-2257.48
234	-221992.82	-2369.09	0.0291	-2224.48
235	-207442.82	-2384.09	-0.0909	-2257.48
236	-116665.82	-2384.09	-0.0909	-2257.48
237	-226392.82	-2383.09	-0.0909	-2252.48
238	-99592.82	-2378.09	-0.0909	-2220.48
239	-88392.82	-2381.09	0.0791	-2252.48
240	-26392.82	-2384.09	-0.0909	-2257.48
241	-193392.82	-2363.09	-0.0909	-2225.48
242	-128892.82	-2384.09	0.0691	-2256.48
243	-165392.82	-2384.09	-0.0909	-2257.48
244	-99392.82	-2384.09	-0.0909	-2257.48
245	-225792.82	-2382.09	-0.0909	-2255.48
246	-194392.82	-2383.09	-0.0909	-2256.48
247	-108392.82	-2383.09	-0.0909	-2256.48
248	-188392.82	-2347.09	-0.0909	-2181.48
249	21607.18	-2383.09	-0.0909	-2256.48
250	-221893.82	-2376.09	0.2691	-2236.48
251	-43392.82	-993.09	0.3291	2742.52
252	-227493.82	-2377.09	-0.0909	-1257.48
253	-178492.82	-2380.09	-0.0909	-2248.48
254	-219502.82	-2001.09	-0.0909	-1257.48
255	-177342.82	-2314.09	-0.0409	-2157.48
256	-195392.82	-1561.09	-0.0909	-257.48
257	-202392.82	-1556.09	-0.0909	-257.48
258	-216492.82	-2276.09	-0.0909	-1257.48
259	-54392.82	-2284.09	0.4591	-2157.48
260	-137892.82	-2384.09	-0.0909	-2257.48
261	-84392.82	-2315.09	0.4091	-2157.48
262	-144392.82	-2138.09	0.4091	-1757.48

Gambar 4.4 Pengolahan data KNN

Pada Gambar 4.4 dapat dilihat bahwa kluster sudah terbentuk berdasarkan jarak *euclidean* yang ditandai warna pada data.

#### E. Jarak Potensi Pasar

Selanjutnya sistem akan melakukan perhitungan jarak untuk menentukan *centroid* sehingga menghasilkan kordinat *centroid* baik dari kategori produk maupun dari kategori potensi pasar menggunakan Rumus 4.2. Kordinat *centroid* tersebut digunakan untuk mengetahui jarak potensi pasar menggunakan Rumus 4.3 dapat dilihat pada Tabel 4.1.

Tabel 4.1 Hasil uji coba KNN

NO	Kategori	Rendah (KNN)	Sedang (KNN)	Tinggi (KNN)	Prediksi (KNN)	Aktual (Output)	TP	TN	FP	FN
1	AC Standing	0	0	1	Tinggi	Sedang	0	1	1	1
2	Kipas Angin Listrik	0	1	0	Sedang	Sedang	1	2	0	0
3	Amplifier	0	1	0	Sedang	Tinggi	0	1	1	1
4	Headphone	0	0.650294	0.349706	Sedang	Tinggi	0	1	1	1
5	Speaker	0.328908	0.671092	0	Sedang	Sedang	1	2	0	0
6	Voice Recorder	0	0.917142	0.0828582	Sedang	Sedang	1	2	0	0
7	Kulkas	0	0.0684411	0.931559	Tinggi	Tinggi	1	2	0	0

NO	Kategori	Rendah (KNN)	Sedang (KNN)	Tinggi (KNN)	Prediksi (KNN)	Aktual (Output)	TP	TN	FP	FN
8	Microwave	0.719679	0.280321	0	Rendah	Sedang	0	1	1	1
9	Rice Cooker	0	1	0	Sedang	Sedang	1	2	0	0
10	Slow Cooker	0	0.206661	0.793339	Tinggi	Tinggi	1	2	0	0
11	Kartu RFID	0	1	0	Sedang	Tinggi	0	1	1	1
12	Layar Proyektor	0.563239	0.207391	0.229371	Rendah	Rendah	1	2	0	0
13	Mesin Fotocopy	0	1	0	Sedang	Sedang	1	2	0	0
14	Mesin Hitung Uang	0	0.129288	0.870712	Tinggi	Sedang	0	1	1	1
15	Mesin Laminating	0	0.208175	0.791825	Tinggi	Sedang	0	1	1	1
16	Mesin Penghancur Kertas	0	1	0	Sedang	Sedang	1	2	0	0
17	Alat Pengering Baju	0	0.174443	0.825557	Tinggi	Sedang	0	1	1	1
18	Vacuum Cleaners	0.251171	0	0.748829	Tinggi	Tinggi	1	2	0	0
19	Access Control Door	0.487593	0.258987	0.25342	Rendah	Sedang	0	1	1	1
20	Alarm	0	0.646507	0.353493	Sedang	Tinggi	0	1	1	1
21	Kabel & Konektor Telepon	0.442379	0.557621	0	Sedang	Tinggi	0	1	1	1
22	Telepon Satelit	0	0	1	Tinggi	Sedang	0	1	1	1
23	Remote TV	0	1	0	Sedang	Sedang	1	2	0	0
24	Topi Anak	0.699251	0.300749	0	Rendah	Rendah	1	2	0	0
25	Kemeja Wanita	1	0	0	Rendah	Rendah	1	2	0	0
26	Baju Koko Pria	0.473077	0.526923	0	Sedang	Sedang	1	2	0	0
27	Bros Jilbab	0.719032	0.280968	0	Rendah	Sedang	0	1	1	1
28	Gamis Wanita	0.569413	0.430587	0	Rendah	Sedang	0	1	1	1
29	Baju Tidur Anak Laki-Laki	0.773038	0.226962	0	Rendah	Rendah	1	2	0	0
30	Kemeja Anak Laki-Laki	0.717071	0.282929	0	Rendah	Rendah	1	2	0	0
31	Kulot Anak Perempuan	0.7362	0.2638	0	Rendah	Rendah	1	2	0	0
32	Rok Anak Perempuan	0.715603	0.284397	0	Rendah	Rendah	1	2	0	0
33	Sepatu Jelly Anak Perempuan	0.614347	0.385653	0	Rendah	Rendah	1	2	0	0
34	Jok motor	0.39596	0	0.60404	Tinggi	Rendah	0	1	1	1
35	Karpet Motor	0.603218	0	0.396782	Rendah	Rendah	1	2	0	0
36	Lampu Motor	0	0.291174	0.708826	Tinggi	Tinggi	1	2	0	0
37	Spion Motor	0.25383	0.318649	0.42752	Tinggi	Rendah	0	1	1	1
38	Kaca Helm	0.694933	0.305067	0	Rendah	Rendah	1	2	0	0
39	Mobil Sedan	0	0	1	Tinggi	Tinggi	1	2	0	0
40	Velg Mobil	0.672442	0.327558	0	Rendah	Tinggi	0	1	1	1
41	Filter Oli	0.807631	0.192369	0	Rendah	Sedang	0	1	1	1
42	Oli mobil	0.333054	0.666946	0	Sedang	Tinggi	0	1	1	1
43	Penghemat BBM	0.278607	0.721393	0	Sedang	Sedang	1	2	0	0
44	Motor Matic	0	0	1	Tinggi	Tinggi	1	2	0	0
45	Velg Motor	0	1	0	Sedang	Sedang	1	2	0	0

NO	Kategori	Rendah (KNN)	Sedang (KNN)	Tinggi (KNN)	Prediksi (KNN)	Aktual (Output)	TP	TN	FP	FN
46	Kampas Rem Mobil	0.320874	0.27494	0.404185	Tinggi	Rendah	0	1	1	1
47	Piston Mobil	0.720072	0.279928	0	Rendah	Rendah	1	2	0	0
48	Shockbreaker & Kaki Kaki Mobil	0.767747	0.232253	0	Rendah	Sedang	0	1	1	1
49	Aki Motor	0	1	0	Sedang	Sedang	1	2	0	0
50	Air Purifier	0.161225	0.190274	0.648502	Tinggi	Tinggi	1	2	0	0
51	Juicer	0	0.956747	0.0432528	Sedang	Sedang	1	2	0	0
52	Kompore Listrik	0.256624	0.743376	0	Sedang	Sedang	1	2	0	0
53	Mixer	0	0.5235	0.4765	Sedang	Tinggi	0	1	1	1
54	Oven	0	0.842683	0.157317	Sedang	Sedang	1	2	0	0
55	Toaster	0.71388	0.28612	0	Rendah	Rendah	1	2	0	0
56	Laser Pointer	0.559798	0.440202	0	Rendah	Sedang	0	1	1	1
57	Mesin Absensi	0.122485	0.289897	0.587618	Tinggi	Sedang	0	1	1	1
58	Mesin Jilid	0.614785	0.385215	0	Rendah	Tinggi	0	1	1	1
59	Mesin Kasir	0.418589	0.381153	0.200258	Rendah	Sedang	0	1	1	1
60	Hand Blower	0	0.645389	0.354611	Sedang	Tinggi	0	1	1	1
61	Mesin Cuci	0	0.626746	0.373254	Sedang	Tinggi	0	1	1	1
62	Setrika Uap	0.785701	0.214299	0	Rendah	Sedang	0	1	1	1
63	Kamera CCTV	0	0.828294	0.171706	Sedang	Tinggi	0	1	1	1
64	Tinta Printer	1	0	0	Rendah	Sedang	0	1	1	1
65	Telepon Wireless	0	0.533814	0.466186	Sedang	Sedang	1	2	0	0
66	Walkie Talkie	0	0.726028	0.273972	Sedang	Sedang	1	2	0	0
67	Antena TV & Parabola	0	0.588247	0.411753	Sedang	Sedang	1	2	0	0
68	Televisi	0	0	1	Tinggi	Tinggi	1	2	0	0
69	TV Box	0.334886	0.354835	0.310279	Sedang	Sedang	1	2	0	0
70	Bando Bandana Anak	0.41344	0.58656	0	Sedang	Rendah	0	1	1	1
71	Dasi Anak	0	1	0	Sedang	Rendah	0	1	1	1
72	Ikut Rambut Anak	0.403683	0.356765	0.239553	Rendah	Rendah	1	2	0	0
73	Blazer Wanita	0.206027	0.793973	0	Sedang	Rendah	0	1	1	1
74	Jaket Parka Wanita	0.741143	0.137905	0.120952	Rendah	Rendah	1	2	0	0
75	Kaos Wanita	0	0.623871	0.376129	Sedang	Sedang	1	2	0	0
76	Aksesoris Jam Tangan Pria	1	0	0	Rendah	Rendah	1	2	0	0
77	Jam Tangan Analog Pria	0	0.280529	0.719471	Tinggi	Tinggi	1	2	0	0
78	Jam Tangan Digital Pria	0.300052	0.699948	0	Sedang	Sedang	1	2	0	0
79	Jam Tangan Analog Wanita	0.276786	0.723214	0	Sedang	Sedang	1	2	0	0
80	Strap Jam Wanita	1	0	0	Rendah	Rendah	1	2	0	0
81	Pashmina	0	0.285542	0.714458	Tinggi	Sedang	0	1	1	1
82	Celana Panjang Anak Laki-Laki	0.319952	0.680048	0	Sedang	Rendah	0	1	1	1
83	Celana Panjang Anak Perempuan	0	1	0	Sedang	Rendah	0	1	1	1

NO	Kategori	Rendah (KNN)	Sedang (KNN)	Tinggi (KNN)	Prediksi (KNN)	Aktual (Output)	TP	TN	FP	FN
84	Sandal Anak Perempuan	0.445966	0.554034	0	Sedang	Rendah	0	1	1	1
85	Cover Stang Motor	0.690008	0	0.309992	Rendah	Rendah	1	2	0	0
86	Sarung Tangan Motor	0	0.436256	0.563744	Tinggi	Tinggi	1	2	0	0
87	Sepatu Biker	0.30451	0.300804	0.394686	Tinggi	Sedang	0	1	1	1
88	Helm Full Face	0.476027	0.286855	0.237119	Rendah	Rendah	1	2	0	0
89	Helm Retro	0	0.841841	0.158159	Sedang	Sedang	1	2	0	0
90	Mobil SUV & MPV	0	0	1	Tinggi	Tinggi	1	2	0	0
91	Oli motor	0	0.731844	0.268156	Sedang	Sedang	1	2	0	0
92	Motor bebek	0.15799	0.328936	0.513074	Tinggi	Sedang	0	1	1	1
93	Motor Sport	0	0.761183	0.238817	Sedang	Tinggi	0	1	1	1
94	Busi Mobil	1	0	0	Rendah	Tinggi	0	1	1	1
95	Tromol & Piringan Rem Mobil	0.297085	0.702915	0	Sedang	Sedang	1	2	0	0
96	Busi Motor	0.666103	0.333897	0	Rendah	Sedang	0	1	1	1
97	Handle - Handfat Motor	0.269213	0.730787	0	Sedang	Rendah	0	1	1	1
98	AC Portable	0	0.569168	0.430832	Sedang	Sedang	1	2	0	0
99	AC Reflektor	0.262954	0.737046	0	Sedang	Sedang	1	2	0	0
100	Air Conditioner	0	0.769658	0.230342	Sedang	Tinggi	0	1	1	1
101	Air Cooler	0.355741	0.644259	0	Sedang	Sedang	1	2	0	0
102	Earphone	0	1	0	Sedang	Tinggi	0	1	1	1
103	Kabel & Konektor Audio	0	1	0	Sedang	Sedang	1	2	0	0
104	Sound System	0	0.755972	0.244028	Sedang	Tinggi	0	1	1	1
105	Blender	0.25462	0.74538	0	Sedang	Sedang	1	2	0	0
106	Mesin Fax	0	0.84669	0.15331	Sedang	Rendah	0	1	1	1
107	Reader RFID	0	0.678066	0.321934	Sedang	Sedang	1	2	0	0
108	LED Proyektor	0	0.800177	0.199823	Sedang	Tinggi	0	1	1	1
109	Setrika	0	0.299423	0.700577	Tinggi	Sedang	0	1	1	1
110	Motion Detector	0.370936	0.629064	0	Sedang	Rendah	0	1	1	1
111	Spy Camera	0.320523	0.679477	0	Sedang	Rendah	0	1	1	1
112	DVD Player	0	0.763624	0.236376	Sedang	Sedang	1	2	0	0
113	MP3 & MP4 Player	0.325039	0.674961	0	Sedang	Sedang	1	2	0	0
114	Radio & Tape Player	0.83196	0.16804	0	Rendah	Sedang	0	1	1	1
115	Printer Ink Jet	0	1	0	Sedang	Sedang	1	2	0	0
116	Scanner	0	0.290247	0.709753	Tinggi	Tinggi	1	2	0	0
117	Toner Printer	0.361154	0.638846	0	Sedang	Sedang	1	2	0	0
118	Telepon Kabel	0	0.382307	0.617693	Tinggi	Sedang	0	1	1	1
119	Kabel & Konektor	0.656797	0	0.343203	Rendah	Sedang	0	1	1	1
120	Kacamata Anak	0.488512	0.511488	0	Sedang	Sedang	1	2	0	0
121	Daster Wanita	0	1	0	Sedang	Sedang	1	2	0	0

NO	Kategori	Rendah (KNN)	Sedang (KNN)	Tinggi (KNN)	Prediksi (KNN)	Aktual (Output)	TP	TN	FP	FN
122	Jam Tangan Digital Wanita	0	0.273063	0.726937	Tinggi	Sedang	0	1	1	1
123	Hijab Segi Empat	0	1	0	Sedang	Tinggi	0	1	1	1
124	Mukena	0	0.661032	0.338968	Sedang	Sedang	1	2	0	0
125	Peci & Kopiah	0.241522	0.758478	0	Sedang	Tinggi	0	1	1	1
126	Setelan Syari Wanita	0.512919	0.487081	0	Rendah	Rendah	1	2	0	0
127	Jaket Anak Laki-Laki	1	0	0	Rendah	Rendah	1	2	0	0
128	Kaos Anak Laki-Laki	0.255478	0.744522	0	Sedang	Tinggi	0	1	1	1
129	Baju Tidur Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
130	Dress Anak Perempuan	0.598606	0	0.401394	Rendah	Rendah	1	2	0	0
131	Flat Shoes Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
132	Pantofel Anak Laki-Laki	0.821152	0.178848	0	Rendah	Rendah	1	2	0	0
133	Sandal Anak Laki Laki	1	0	0	Rendah	Rendah	1	2	0	0
134	Aksesori Body Motor	0	1	0	Sedang	Tinggi	0	1	1	1
135	Jaket Motor	0.224912	0.775088	0	Sedang	Rendah	0	1	1	1
136	Jas Hujan	0	0.663086	0.336914	Sedang	Tinggi	0	1	1	1
137	Rompi Motor	0.790209	0.209791	0	Rendah	Rendah	1	2	0	0
138	Helm Anak	1	0	0	Rendah	Rendah	1	2	0	0
139	Kunci Helm	0.570024	0	0.429976	Rendah	Rendah	1	2	0	0
140	Mobil Hatchback & City Car	0	0.254006	0.745994	Tinggi	Tinggi	1	2	0	0
141	Mobil Niaga	0	0	1	Tinggi	Tinggi	1	2	0	0
142	Engine Conditioner	1	0	0	Rendah	Sedang	0	1	1	1
143	Tempat Oli	0.529095	0.470905	0	Rendah	Sedang	0	1	1	1
144	Kampas Rem Motor	1	0	0	Rendah	Sedang	0	1	1	1
145	Rantai & Gir Motor	0.766578	0	0.233422	Rendah	Tinggi	0	1	1	1
Jumlah							79	224	66	66

## F. Evaluasi

Evaluasi pengujian metode pada penelitian ini menggunakan *confusion matrix* yang dapat melakukan pengukuran kinerja dari klasifikasi sehingga diperoleh nilai *accuracy* menggunakan Rumus 3.2 seperti dijabarkan berikut ini:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$



$$Accuracy = \frac{79 + 224}{79 + 66 + 66 + 224} \times 100\%$$

$$Accuracy = \frac{303}{435} \times 100\%$$

$$Accuracy = 70\%$$

Sementara untuk *precision* menggunakan Rumus 3.3

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{79}{79 + 66} \times 100\%$$

$$Precision = 54\%$$

Untuk *recall* menggunakan Rumus 3.4

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{79}{79 + 66} \times 100\%$$

$$Recall = 54\%$$

Selanjutnya untuk *F<sub>1</sub>-Score* menggunakan Rumus 3.5

$$F_1 - Score = 2 \times \frac{0.54 \times 0.54}{0.54 + 0.54} \times 100\%$$

$$F_1 - Score = 54\%$$

### 4.3. Kesimpulan

Berdasarkan hasil potensi pasar pengolahan metode KNN dan menggunakan *confusion matrix* menunjukkan bahwa *accuracy* 70%, *precision* 54%, *recall* 54% dan *F<sub>1</sub>-Score* 54%.

**BAB V**  
**ALGORITMA C4.5**

**5.1. Desain Metode**

Algoritma C4.5 merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal (Noviana et al., 2019). Secara umum Algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut (Zainuddin dan Noercholis, 2019) :

- 1) Memilih variabel sebagai akar.
- 2) Membuat cabang untuk masing-masing nilai.
- 3) Membagi kasus dalam cabang.
- 4) Mengulangi proses untuk masing-masing cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pemilihan variabel sebagai akar, didasarkan pada nilai *gain ratio* tertinggi dari variabel-variabel yang ada. *Gain ratio* diperoleh dari perhitungan nilai *entropy*, *gain*, dan *split info*. Perhitungan *entropy* digunakan rumus sebagai berikut :

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i \dots \dots \dots (5.1)$$

Keterangan:

- S : Himpunan Kasus
- n : Jumlah partisi S
- $p_i$  : Proporsi dari  $S_i$  terhadap S

Nilai *entropy* yang telah dihitung digunakan dalam perhitungan nilai *gain* sebagai berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots \dots \dots (5.2)$$

dengan

S : himpunan kasus.

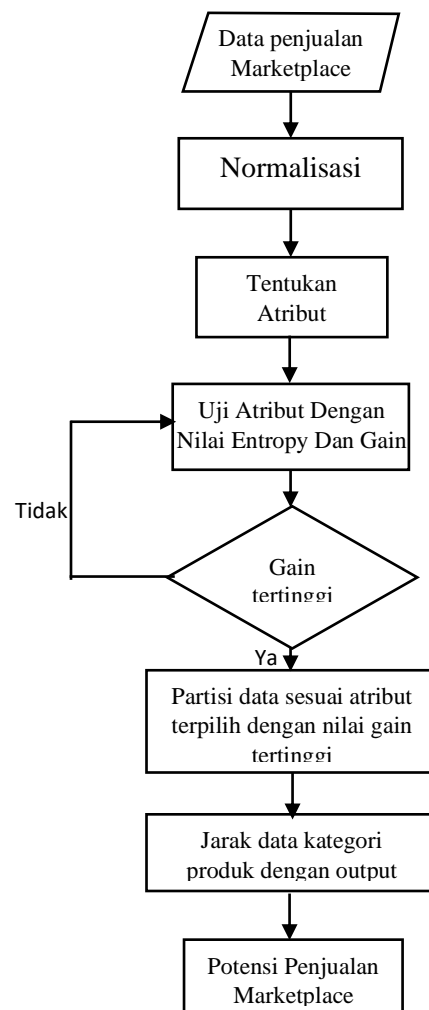
A : Atribut.

n : jumlah partisi atribut A.

|S<sub>i</sub>| : jumlah kasus pada partisi ke i.

|S| : jumlah kasus dalam S.

Selanjutnya tahapan merancang dan mengimplementasikan sistem dimana sistem dirancang dan diimplemetasikan sesuai dengan tujuan penelitian yang telah difokuskan.



Gambar 5.1 Desain Metode C4.5

Pada Gambar 5.1 menjelaskan tentang rancangan desain metode C4.5 dengan tahapan-tahapan sebagai berikut diantaranya:

- a. Pada tahap pertama, data pada tahap ini data yang digunakan secara langsung melalui *website marketplace* tokopedia tersedia melalui laman <https://tokopedia.com>.
- b. Setelah itu melakukan normalisasi data untuk meningkatkan performanya. Salah satu normalisasi yang umum digunakan adalah dengan *min-max* menggunakan Rumus 3.1.
- c. Setelah proses normalisasi, selanjutnya pemilihan atribut output kategori potensi pasar yaitu {Rendah, Sedang, Tinggi}
- d. Setelah itu akan di lakukan proses pemilihan variabel sebagai akar, didasarkan pada nilai *gain ratio* tertinggi dari variabel-variabel yang ada. *Gain ratio* diperoleh dari perhitungan nilai *entropy* menggunakan Rumus 5.1 dan *gain* menggunakan Rumus 5.2.
- e. Selanjutnya dari nilai *gain* dan *entropy*, digunakan untuk mengetahui jarak potensi pasar menggunakan Rumus 4.3.
- f. Setelah mengetahui jarak potensi pasar, data tersebut digunakan untuk melihat potensi pasar menggunakan Rumus 4.4.

## 5.2. Uji Coba

Adapun untuk uji coba algoritma C4.5 sesuai dengan desain yang telah dikembangkan untuk menjalankan tahapan-tahapan sesuai perancangan yang telah dibuat sebelumnya berdasarkan Gambar 5.1.

#### A. Memasukan Dataset

Seperti yang dijelaskan pada bab sebelumnya, terdiri dari 145 data dengan masing-masing data berjumlah 150 *record* dengan total 21.750 *record*. dari masing-masing data memiliki 11 atribut dengan nilai yang dimiliki oleh atribut merupakan nilai kategorikal dan nilai angka yaitu Link produk, Kategori 1, Kategori 2, Kategori 3, Nama Barang, Harga, Reviews, Terjual, Views, Rating, Lokasi dan Penjualan yang dapat dilihat pada Gambar 4.2.

#### B. Preprocessing

Pada tahapan ini yang harus dilakukan untuk normalisasi data dengan tujuan untuk menghindari terjadinya anomali data dan tidak konsistensinya sebuah data untuk menghindari kesalahan pada perhitungan menggunakan Rumus 3.1. dapat dilihat pada Gambar 4.3

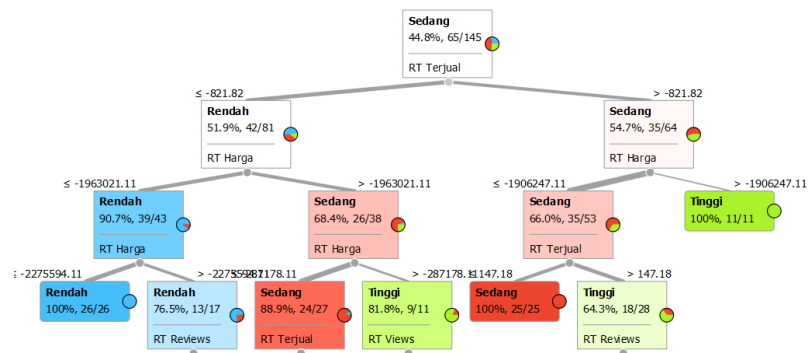
#### C. Tentukan Atribut

Atribut yang digunakan sebagai target adalah *output* penjualan dengan tujuan mengetahui setiap potensi kategori produk yang sudah ditentukan.

#### D. Algoritma C4.5

Setelah menentukan atribut yaitu proses algoritma C4.5. Untuk pengolahan data menggunakan C4.5 terdapat dua proses yaitu menghitung nilai *entropy* terlebih dahulu menggunakan rumus 5.1. Selanjutnya nilai *entropy* digunakan untuk menghitung nilai *gain*

menggunakan Rumus 5.2 menghasilkan pohon keputusan seperti gambar berikut.



Gambar 5.2 Pohon keputusan C4.5 levels 4

Pada Gambar 5.2 dapat dilihat bahwa nilai *gain* tertinggi adalah atribut terjual sebanyak 44,8% atau 65 dari 145 kategori produk terjual dari total keseluruhan kategori produk yang dijual sebanyak 145 kategori dengan 150 produk setiap kategori.

#### E. Jarak Data

Setelah proses perhitungan nilai *gain* dan membuat pohon keputusan, berikut tabel yang sudah terbentuk.

Tabel 5.1 Hasil uji coba C4.5

NO	Kategori	Rendah (C4.5)	Sedang (C4.5)	Tinggi (C4.5)	Prediksi (C4.5)	Aktual (Output)	TP	TN	FP	FN
1	AC Standing	0	0	1	Tinggi	Sedang	0	1	1	1
2	Kipas Angin Listrik	0	0.6	0.4	Sedang	Sedang	1	2	0	0
3	Amplifier	0	0.25	0.75	Tinggi	Tinggi	1	2	0	0
4	Headphone	0	0.25	0.75	Tinggi	Tinggi	1	2	0	0
5	Speaker	0	0.25	0.75	Tinggi	Sedang	0	1	1	1
6	Voice Recorder	0	1	0	Sedang	Sedang	1	2	0	0
7	Kulkas	0	0	1	Tinggi	Tinggi	1	2	0	0
8	Microwave	0.4	0.6	0	Sedang	Sedang	1	2	0	0
9	Rice Cooker	0	1	0	Sedang	Sedang	1	2	0	0
10	Slow Cooker	0	0	1	Tinggi	Tinggi	1	2	0	0

NO	Kategori	Rendah (C4.5)	Sedang (C4.5)	Tinggi (C4.5)	Prediksi (C4.5)	Aktual (Output)	TP	TN	FP	FN
11	Kartu RFID	0	0.25	0.75	Tinggi	Tinggi	1	2	0	0
12	Layar Proyektor	0	1	0	Sedang	Rendah	0	1	1	1
13	Mesin Fotocopy	0.4	0.6	0	Sedang	Sedang	1	2	0	0
14	Mesin Hitung Uang	0	0	1	Tinggi	Sedang	0	1	1	1
15	Mesin Laminating	0	1	0	Sedang	Sedang	1	2	0	0
16	Mesin Penghancur Kertas	0	0	1	Tinggi	Sedang	0	1	1	1
17	Alat Pengereng Baju	0	0	1	Tinggi	Sedang	0	1	1	1
18	Vacuum Cleaners	0	0	1	Tinggi	Tinggi	1	2	0	0
19	Access Control Door	0	1	0	Sedang	Sedang	1	2	0	0
20	Alarm	0	0.25	0.75	Tinggi	Tinggi	1	2	0	0
21	Kabel & Konektor Telepon	0	0	1	Tinggi	Tinggi	1	2	0	0
22	Telepon Satelit	0	0	1	Tinggi	Sedang	0	1	1	1
23	Remote TV	0	0	1	Tinggi	Sedang	0	1	1	1
24	Topi Anak	1	0	0	Rendah	Rendah	1	2	0	0
25	Kemeja Wanita	1	0	0	Rendah	Rendah	1	2	0	0
26	Baju Koko Pria	0	1	0	Sedang	Sedang	1	2	0	0
27	Bros Jilbab	0	1	0	Sedang	Sedang	1	2	0	0
28	Gamis Wanita	0	1	0	Sedang	Sedang	1	2	0	0
29	Baju Tidur Anak Laki-Laki	1	0	0	Rendah	Rendah	1	2	0	0
30	Kemeja Anak Laki-Laki	1	0	0	Rendah	Rendah	1	2	0	0
31	Kulot Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
32	Rok Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
33	Sepatu Jelly Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
34	Jok motor	0	1	0	Sedang	Rendah	0	1	1	1
35	Karpet Motor	1	0	0	Rendah	Rendah	1	2	0	0
36	Lampu Motor	0	0	1	Tinggi	Tinggi	1	2	0	0
37	Spion Motor	1	0	0	Rendah	Rendah	1	2	0	0
38	Kaca Helm	1	0	0	Rendah	Rendah	1	2	0	0
39	Mobil Sedan	0	0	1	Tinggi	Tinggi	1	2	0	0
40	Velg Mobil	0	0.6	0.4	Sedang	Tinggi	0	1	1	1
41	Filter Oli	0	1	0	Sedang	Sedang	1	2	0	0
42	Oli mobil	0	0.6	0.4	Sedang	Tinggi	0	1	1	1
43	Penghemat BBM	0	1	0	Sedang	Sedang	1	2	0	0
44	Motor Matic	0	0	1	Tinggi	Tinggi	1	2	0	0
45	Velg Motor	0	0	1	Tinggi	Sedang	0	1	1	1
46	Kampas Rem Mobil	0.2	0.8	0	Sedang	Rendah	0	1	1	1
47	Piston Mobil	0.2	0.8	0	Sedang	Rendah	0	1	1	1
48	Shockbreaker & Kaki Kaki Mobil	0	1	0	Sedang	Sedang	1	2	0	0

NO	Kategori	Rendah (C4.5)	Sedang (C4.5)	Tinggi (C4.5)	Prediksi (C4.5)	Aktual (Output)	TP	TN	FP	FN
49	Aki Motor	0	1	0	Sedang	Sedang	1	2	0	0
50	Air Purifier	0	0	1	Tinggi	Tinggi	1	2	0	0
51	Juicer	0	1	0	Sedang	Sedang	1	2	0	0
52	Kompore Listrik	0	1	0	Sedang	Sedang	1	2	0	0
53	Mixer	0	0	1	Tinggi	Tinggi	1	2	0	0
54	Oven	0	1	0	Sedang	Sedang	1	2	0	0
55	Toaster	0.333333	0.666667	0	Sedang	Rendah	0	1	1	1
56	Laser Pointer	0	1	0	Sedang	Sedang	1	2	0	0
57	Mesin Absensi	0	1	0	Sedang	Sedang	1	2	0	0
58	Mesin Jilid	0	0.25	0.75	Tinggi	Tinggi	1	2	0	0
59	Mesin Kasir	0	1	0	Sedang	Sedang	1	2	0	0
60	Hand Blower	0	0	1	Tinggi	Tinggi	1	2	0	0
61	Mesin Cuci	0	0	1	Tinggi	Tinggi	1	2	0	0
62	Setrika Uap	0	1	0	Sedang	Sedang	1	2	0	0
63	Kamera CCTV	0	0	1	Tinggi	Tinggi	1	2	0	0
64	Tinta Printer	0	0.666667	0.333333	Sedang	Sedang	1	2	0	0
65	Telepon Wireless	0	1	0	Sedang	Sedang	1	2	0	0
66	Walkie Talkie	0	1	0	Sedang	Sedang	1	2	0	0
67	Antena TV & Parabola	0	0.666667	0.333333	Sedang	Sedang	1	2	0	0
68	Televisi	0	0	1	Tinggi	Tinggi	1	2	0	0
69	TV Box	0	1	0	Sedang	Sedang	1	2	0	0
70	Bando Bandana Anak	1	0	0	Rendah	Rendah	1	2	0	0
71	Dasi Anak	1	0	0	Rendah	Rendah	1	2	0	0
72	Ikat Rambut Anak	1	0	0	Rendah	Rendah	1	2	0	0
73	Blazer Wanita	1	0	0	Rendah	Rendah	1	2	0	0
74	Jaket Parka Wanita	0.666667	0.333333	0	Rendah	Rendah	1	2	0	0
75	Kaos Wanita	0	1	0	Sedang	Sedang	1	2	0	0
76	Aksesoris Jam Tangan Pria	1	0	0	Rendah	Rendah	1	2	0	0
77	Jam Tangan Analog Pria	0	0	1	Tinggi	Tinggi	1	2	0	0
78	Jam Tangan Digital Pria	0	0.25	0.75	Tinggi	Sedang	0	1	1	1
79	Jam Tangan Analog Wanita	0.333333	0.666667	0	Sedang	Sedang	1	2	0	0
80	Strap Jam Wanita	1	0	0	Rendah	Rendah	1	2	0	0
81	Pashmina	0	1	0	Sedang	Sedang	1	2	0	0
82	Celana Panjang Anak Laki-Laki	1	0	0	Rendah	Rendah	1	2	0	0
83	Celana Panjang Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
84	Sandal Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
85	Cover Stang Motor	1	0	0	Rendah	Rendah	1	2	0	0
86	Sarung Tangan Motor	0	0.666667	0.333333	Sedang	Tinggi	0	1	1	1



NO	Kategori	Rendah (C4.5)	Sedang (C4.5)	Tinggi (C4.5)	Prediksi (C4.5)	Aktual (Output)	TP	TN	FP	FN
87	Sepatu Biker	0	1	0	Sedang	Sedang	1	2	0	0
88	Helm Full Face	0.666667	0.333333	0	Rendah	Rendah	1	2	0	0
89	Helm Retro	1	0	0	Rendah	Sedang	0	1	1	1
90	Mobil SUV & MPV	0	0	1	Tinggi	Tinggi	1	2	0	0
91	Oli motor	0	1	0	Sedang	Sedang	1	2	0	0
92	Motor Bebek	0	1	0	Sedang	Sedang	1	2	0	0
93	Motor Sport	0	0	1	Tinggi	Tinggi	1	2	0	0
94	Busi Mobil	0	0.25	0.75	Tinggi	Tinggi	1	2	0	0
95	Tromol & Piringan Rem Mobil	0	1	0	Sedang	Sedang	1	2	0	0
96	Busi Motor	0	1	0	Sedang	Sedang	1	2	0	0
97	Handle - Handfat Motor	1	0	0	Rendah	Rendah	1	2	0	0
98	AC Portable	0	1	0	Sedang	Sedang	1	2	0	0
99	AC Reflektor	0	0.428571	0.571429	Tinggi	Sedang	0	1	1	1
100	Air Conditioner	0	0	1	Tinggi	Tinggi	1	2	0	0
101	Air Cooler	0	1	0	Sedang	Sedang	1	2	0	0
102	Earphone	0	0.75	0.25	Sedang	Tinggi	0	1	1	1
103	Kabel & Konektor Audio	0	0.428571	0.571429	Tinggi	Sedang	0	1	1	1
104	Sound System	0	1	0	Sedang	Tinggi	0	1	1	1
105	Blender	0	0	1	Tinggi	Sedang	0	1	1	1
106	Mesin Fax	0.666667	0.333333	0	Rendah	Rendah	1	2	0	0
107	Reader RFID	0	1	0	Sedang	Sedang	1	2	0	0
108	LED Proyektor	0	0.5	0.5	Sedang	Tinggi	0	1	1	1
109	Setrika	0	0.428571	0.571429	Tinggi	Sedang	0	1	1	1
110	Motion Detector	1	0	0	Rendah	Rendah	1	2	0	0
111	Spy Camera	1	0	0	Rendah	Rendah	1	2	0	0
112	DVD Player	0	1	0	Sedang	Sedang	1	2	0	0
113	MP3 & MP4 Player	0	1	0	Sedang	Sedang	1	2	0	0
114	Radio & Tape Player	1	0	0	Rendah	Sedang	0	1	1	1
115	Printer Ink Jet	0	1	0	Sedang	Sedang	1	2	0	0
116	Scanner	0	0.5	0.5	Sedang	Tinggi	0	1	1	1
117	Toner Printer	0	1	0	Sedang	Sedang	1	2	0	0
118	Telepon Kabel	0	1	0	Sedang	Sedang	1	2	0	0
119	Kabel & Konektor	0	0.428571	0.571429	Tinggi	Sedang	0	1	1	1
120	Kacamata Anak	0	1	0	Sedang	Sedang	1	2	0	0
121	Daster Wanita	0	1	0	Sedang	Sedang	1	2	0	0
122	Jam Tangan Digital Wanita	0	1	0	Sedang	Sedang	1	2	0	0
123	Hijab Segi Empat	0	0.428571	0.571429	Tinggi	Tinggi	1	2	0	0
124	Mukena	0	1	0	Sedang	Sedang	1	2	0	0
125	Peci & Kopiah	0	0.428571	0.571429	Tinggi	Tinggi	1	2	0	0

NO	Kategori	Rendah (C4.5)	Sedang (C4.5)	Tinggi (C4.5)	Prediksi (C4.5)	Aktual (Output)	TP	TN	FP	FN
126	Setelan Syari Wanita	1	0	0	Rendah	Rendah	1	2	0	0
127	Jaket Anak Laki-Laki	1	0	0	Rendah	Rendah	1	2	0	0
128	Kaos Anak Laki-Laki	0	0.428571	0.571429	Tinggi	Tinggi	1	2	0	0
129	Baju Tidur Anak Perempuan	0.75	0.25	0	Rendah	Rendah	1	2	0	0
130	Dress Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
131	Flat Shoes Anak Perempuan	1	0	0	Rendah	Rendah	1	2	0	0
132	Pantofel Anak Laki-Laki	1	0	0	Rendah	Rendah	1	2	0	0
133	Sandal Anak Laki Laki	0.75	0.25	0	Rendah	Rendah	1	2	0	0
134	Aksesori Body Motor	0	0.428571	0.571429	Tinggi	Tinggi	1	2	0	0
135	Jaket Motor	1	0	0	Rendah	Rendah	1	2	0	0
136	Jas Hujan	0	0.75	0.25	Sedang	Tinggi	0	1	1	1
137	Rompi Motor	0.75	0.25	0	Rendah	Rendah	1	2	0	0
138	Helm Anak	1	0	0	Rendah	Rendah	1	2	0	0
139	Kunci Helm	1	0	0	Rendah	Rendah	1	2	0	0
140	Mobil Hatchback & City Car	0	0	1	Tinggi	Tinggi	1	2	0	0
141	Mobil Niaga	0	0	1	Tinggi	Tinggi	1	2	0	0
142	Engine Conditioner	0	1	0	Sedang	Sedang	1	2	0	0
143	Tempat Oli	1	0	0	Rendah	Sedang	0	1	1	1
144	Kampas Rem Motor	0	1	0	Sedang	Sedang	1	2	0	0
145	Rantai & Gir Motor	0	0.428571	0.571429	Tinggi	Tinggi	1	2	0	0
Jumlah							115	260	30	30

## G. Evaluasi

Evaluasi pengujian metode pada penelitian ini menggunakan *confusion matrix* yang dapat melakukan pengukuran kinerja dari klasifikasi sehingga diperoleh nilai *accuracy* menggunakan Rumus 3.2 seperti dijabarkan berikut ini:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100\%$$

$$Accuracy = \frac{115 + 260}{115 + 30 + 30 + 260} \times 100\%$$

$$Accuracy = \frac{375}{435} \times 100\%$$

$$Accuracy = 86\%$$

Sementara untuk *precision* menggunakan Rumus 3.3

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

$$Precision = \frac{115}{115 + 30} \times 100\%$$

$$Precision = 79\%$$

Untuk *recall* menggunakan Rumus 3.4

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

$$Recall = \frac{115}{115 + 30} \times 100\%$$

$$Recall = 79\%$$

Selanjutnya untuk *F<sub>1</sub>-Score* menggunakan Rumus 3.5

$$F_1 - Score = 2 \times \frac{0.79 \times 0.79}{0.79 + 0.79} \times 100\%$$

$$F_1 - Score = 79\%$$

### 5.3. Kesimpulan

Berdasarkan hasil potensi pasar pengolahan metode algoritma C4.5 dan menggunakan *confusion matrix* menunjukkan bahwa *accuracy* 86%, *precision* 79%, *recall* 79% dan *F<sub>1</sub>-Score* 79%.

## BAB VI

### PEMBAHASAN

Pada pembahasan dilakukan evaluasi dan analisa hasil klasifikasi dari dua metode yaitu KNN dan C4.5 terhadap data *marketplace* tokopedia yang diperoleh dari website resmi <https://tokopedia.com> pada bulan maret 2022. Untuk dapat diimplementasikan pada metode KNN dan C4.5 maka data yang diperoleh tidak bisa langsung diproses klasifikasi. Oleh karena itu data harus diolah atau melalui *preprocessing* yang dilakukan antara lain menghilangkan data yang kosong, menghilangkan data NA, dan mengkonversi tipe agar dapat diproses oleh kedua metode. Dataset yang digunakan untuk analisis ini adalah data *marketplace* tokopedia setelah melalui preprocessing sebanyak 145 data dengan masing-masing data berjumlah 150 *record* dengan total 21.750 *record*. dari masing-masing data memiliki 11 atribut dengan nilai yang dimiliki oleh atribut merupakan nilai kategorikal dan nilai angka yaitu Link produk, Kategori, Nama Barang, Harga, Reviews, Terjual, Views, Rating, Lokasi dan Penjualan.

Dalam menentukan potensi penjualan, model yang digunakan dalam klasifikasi yaitu KNN dan C4.5, serta mencari berapa tingkat akurasi dari kedua metode ini yaitu dari data ini akan diprediksi apakah potensi penjualan yang baru terdeteksi dengan potensi penjualan tinggi, sedang, ataupun rendah.

Pada pengujian klasifikasi menggunakan metode KNN dan C4.5 menggunakan 145 data dengan masing-masing data berjumlah 150 *record* dengan total 21.750 *record*. Hasil *accuracy precision, recall* dan *F<sub>1</sub>-Score* dipengaruhi

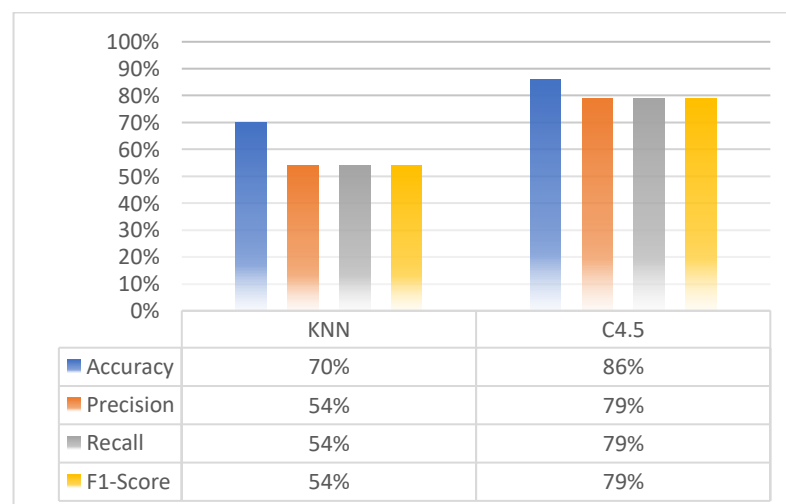
oleh nilai *true positives*, *true negatives*, *false positives*, dan *false negatives* bisa dilihat pada Tabel 6.1 pada setiap model.

Tabel 6.1 Nilai confusion matrix

	TP	TN	FP	FN
KNN	79	224	66	66
C4.5	115	260	30	30

Dari Tabel 6.1 menunjukkan nilai *true positives*, *true negatives*, *false positives*, dan *false negatives* sangat berpengaruh terhadap hasil akurasi. Dari tabel tersebut ada beberapa perbedaan antara nilai aktual dengan nilai prediksi dari kedua metode. Perbedaan tersebut diakibatkan karena nilai aktual didapat berdasarkan nilai harga barang dan nilai terjual suatu produk. Sehingga ada perbedaan antara barang yang terjual harga tinggi dengan barang terjual harga murah. Karena barang dengan harga tinggi terjual maka keuntungan juga akan semakin tinggi.

Berikut ini hasil *confusion matrix* pada hasil uji coba dengan akurasi dapat dilihat pada Gambar 6.1.



Gambar 6.1 Grafik Confusion matrix

Berdasarkan Gambar 6.1 menunjukkan hasil ujicoba yang dilakukan untuk nilai akurasi klasifikasi potensi pasar antara algoritma KNN dan C4.5 bahwa nilai *accuracy* C4.5 mengungguli dari KNN yaitu 86%. Sedangkan untuk nilai *precision*, *recall* dan *F1-score* lebih tinggi C4.5 sebesar 79%.

Dalam penelitiannya Idris dan Enghariano (2020) menjelaskan bahwa secara secara tersirat dan tersurat Allah SWT membagi manusia menjadi beberapa klaster yaitu *al-insan*, *al-basyar* dan *bani Adam* atau *zurriyat Adam*. Sehingga perlu memberikan klaster-klaster terhadap potensi pasar dengan tujuan memudahkan dalam proses analisa potensi pasar tersebut.

Pada masa Nabi Muhammad, salah satu sistem ekonomi diamati. Pasar (*al-suq*) harus menjadi contoh bagaimana kehidupan modern harus dijalani. Tempat dimana penjual dan pembeli berkumpul untuk membeli dan menjual barang dan jasa disebut pasar. Perekonomian sangat dipengaruhi oleh pasar. Konsep Islam menekankan bahwa harus ada persaingan yang sehat di pasar. Namun, ini tidak berarti bahwa kebebasan ini tidak terbatas, sebaliknya kerangka syari'ah mencakup kebebasan. Transaksi bersifat sukarela dalam Islam (*antaradim minkum*) (Abd Ghafur, 2019). Sebagaimana disebutkan dalam al-Qur'an surat An-Nisa': 29.

يَا أَيُّهَا الَّذِينَ آمَنُوا لَا تَأْكُلُوا أَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ إِلَّا أَنْ تَكُونَ  
تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْ ۖ وَلَا تَقْتُلُوا أَنْفُسَكُمْ ۗ إِنَّ اللَّهَ كَانَ بِكُمْ رَحِيمًا

“Hai orang-orang yang beriman, janganlah kamu saling memakan harta sesamamu dengan jalan yang batil, kecuali dengan jalan perniagaan yang Berlaku dengan suka sama-suka di antara kamu. dan janganlah kamu membunuh

*dirimu.Sesungguhnya Allah adalah Maha Penyayang kepadamu". (Q.S. An-Nisa':29)*

Dalam penelitian Cahyono (2020) menjelaskan tempat berkumpulnya orang-orang untuk menukar kepemilikan barang atau jasa dengan uang adalah pasar itu sendiri. Kekuatan penawaran dan permintaan, dimana penjual ingin menukarkan barang atau jasa dengan uang dan pembeli ingin menukarkan uang dengan barang atau jasa, dapat juga diartikan sebagai pasar. Orang bisa mendapatkan apa yang mereka butuhkan dari pasar, dan tidak ada orang yang tidak membutuhkannya. Al-Qur'an sudah menjelaskan terkait pasar terdapat dalam surat Al-Furqan ayat 20 yang berbunyi:

وَمَا أَرْسَلْنَا قَبْلَكَ مِنَ الْمُرْسَلِينَ إِلَّا إِنَّهُمْ لَيَأْكُلُونَ الطَّعَامَ  
وَيَمْشُونَ فِي الْأَسْوَاقِ وَجَعَلْنَا بَعْضَكُمْ لِبَعْضٍ فِتْنَةً ۗ أَتَصْبِرُونَ  
وَكَانَ رَبُّكَ بَصِيرًا ۝

*“Dan Kami tidak mengutus rasul-rasul sebelummu (Muhammad), melainkan mereka pasti memakan makanan dan berjalan di pasar-pasar. Dan Kami jadikan sebagian kamu sebagai cobaan bagi sebagian yang lain. Maukah kamu bersabar? Dan Tuhanmu Maha Melihat.” (Q.S. An-Furqan:20)*

Menurut Abd Ghafur (2019) pengertian istilah pasar dalam Islam dapat disinggung dari hadits Rasulullah SAW. menurut Anas r.a. sebagai akibat dari naiknya harga barang-barang di Kota Madinah. Dari hadits tersebut terlihat bahwa Islam banyak mengajarkan tentang mekanika pasar lebih awal dari Adam Smith (lebih dari 1160 tahun).

*“Hanya barang dagangan pernah melambung tinggi di Madinah pada zaman Nabi SAW. lalu orang-orang berkata, ‘Wahai Rasulullah, harga barang melambung, maka tetapkanlag standar harga untuk kami.’Maka Rasulullah SAW. bersabda, ‘ Sesungguhnya Allah-lah al-Musa’ir (Yang Maha Menetapkan Harga), al-Qabidh, al-Basith, dan ar-Raziq. Dan sungguh aku benar-benar berharap berjumpa dengan Allah dalam keadaan tidak ada seorangpun dari kalangan yang menuntutku dengan kezaliman dalam masalh darah (nyawa) dan harga’.” (H.R. Al-Khamsah, kecuali an-Nasa’i dan disahihkan oleh Ibnu Hibban).*

Berdasarkan hadis sebelumnya, mekanisme pasar adalah harga dapat ditentukan secara bebas. Pasar menentukan harga namun Islam tidak menganut harga yang ditentukan oleh pasar bebas. Ketika pasar memiliki monopoli harga, islam akan turun tangan. Hal ini menunjukkan bahwa dari perspektif Islam, mekanisme pasar memiliki komponen sosial dan teologis karena pasar diatur oleh syariah. Hal tersebut sangat sesuai dengan salah satu faktor yang mempengaruhi kompleksitas *marketplace* yaitu harga pasar.



## **BAB VII**

### **KESIMPULAN**

#### **7.1. Kesimpulan**

Pada perbandingan KNN dan C4.5 dalam menganalisa potensi pasar marketplace, adalah sebagai berikut:

- 1) Faktor yang mempengaruhi kompleksitas *marketplace* yang ada di Indonesia adalah rentan tinggi rendahnya penjualan di setiap kategori produk, reteng toko dan diskon.
- 2) Pada penelitian ini untuk menentukan potensi pasar *marketplace*, metode C4.5 memiliki performa ketepatan prediksi lebih baik C4.5 dapat digunakan sebagai salah satu metode prediksi potensi pasar *marketplace* dengan memperoleh *accuracy* 86%, *precision* 79%, *recall* 79% dan *F1-score* 79% dibanding dengan metode KNN yang hanya dapat memperoleh *accuracy* 70%, *precision* 54%, *recall* 54% dan *F1-score* 54%.

#### **7.2. Saran**

Saran penelitian ini dapat dikembangkan lagi sehingga hasilnya lebih baik diantaranya:

- 1) Pentingnya mengetahui struktur dataset yang digunakan dalam penelitian karena data tersebut dapat mengubah keakuratan hasil.
- 2) Pada penelitian selanjutnya dapat dilakukan dengan metode klasifikasi yang lainnya.

- 3) Pada penelitian kedepannya dapat dikembangkan dengan menggunakan lebih dari satu *marketplace*.

## DAFTAR PUSTAKA

- Abd Ghafur. 2019. "Mekanisme Pasar Perspektif Islam." *Iqtishodiyah : Jurnal Ekonomi dan Bisnis Islam* 5(1): 1–19.
- Artisya, Nur, and Saut Pintubipar Saragih. 2022. "Prediction Tingkat Potensi Pembeli Pada Hari Special Di Toko Mitra Shopee." *JURNAL COMASIE PREDICTION E-ISSN 2715-6265* 06(04): 99–108. <http://ejournal.upbatam.ac.id/index.php/comasiejournal>.
- Astri, Alyeska et al. 2021. "Penerapan Algoritma K-Modes Clustering Dengan Validasi Davies Bouldin Index Pada Pengelompokan Tingkat Minat Belanja Online Di Provinsi Daerah Istimewa Yogyakarta." *Jurnal MSA ( Matematika dan Statistika serta Aplikasinya )* 9(1): 24.
- Cahyono, Heru. 2020. "Konsep Pasar Syariah Dalam Perspektif Etika Bisnis Islam." *Ecobankers : Journal of Economy and Banking* 1(2): 14.
- Churniawan, Agus Dwi, and Yoppy Mirza Maulana. 2022. "Menganalisa Cara Mendapatkan Nilai Klaster Pada K Nearest Neighbor Dan K Means." *Jurnal Ilmiah Scroll: Jendela Teknologi Informasi (P-ISSN: 2338-8625 ; E-ISSN: 2829-3304)* 10(2019): 42–45. <https://univ45sby.ac.id/ejournal/index.php/informatika>.
- Dalhari. 2021. "Jual Beli Dalam Kajian Fiqih." *At-Tujjar* 09(02): 2302–6235. <http://ejournal.kopertais4.or.id/mataraman/index.php/atTujjar/article/view/4683>.
- Febrian, Andhika Karulyana, Yulison Herry Chrisnanto, and Puspita Nurul Sabrina. 2022. "Studi Komparasi Metode Klasifikasi K-Nearest Neighbor Dan Naïve Bayes Dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk." *Snestik DOI : 10.31284/p.snestik.2022.2717*: 333–38.
- Herdyansyah, Syarifudin, Eni Heni Hermaliani, Laela Kurniawati, and Sri Rahayu. 2020. "Analisa Metode Association Rule Menggunakan Algoritma Fp-Growth Terhadap Data Penjualan (Study Kasus Toko Berkah)." *JURNAL KHATULISTIWA INFORMATIKA, p-ISSN: 2339-1928 I e-ISSN: 2579-633X* VIII(2): 2579–2633. [www.nusamandiri.ac.id](http://www.nusamandiri.ac.id).
- Hwangbo, Hyunwoo, Yang Sok Kim, Kyung Jin Cha, and -. 2018. "Recommendation System Development for Fashion Retail E-Commerce." *Electronic Commerce Research and Applications* 28: 94–101. <https://doi.org/10.1016/j.elerap.2018.01.012>.

- Ichsan M, Jumhur ,HH, um MDharmoputra S. 2018. “Pengaruh Consumer Online Rating and Review Terhadap Minat Beli Konsumen Pada Marketplace Tokopedia Di Wilayah Dki Jakarta Effect of Consumer Online Rating and Review To Buying.” *e-Proceeding of Management* 5(2): 1828–35.
- Idris, Muhammad, and Desri Ari Enghariano. 2020. “Karakteristik Manusia Dalam Perspektif Al-Qur’an.” *AL-FAWATIHI Jurnal Kajian Al-Qur’an dan Hadis fawatih : Jurnal kajian Al-quran dan hadist* 1(1): 1–15.
- Iprice. 2022. “Peta E-Commerce Indonesia - Q4 2021.” [Accessed Maret 2022]. [Online]. Available: <https://iprice.co.id/insights/mapofecommerce/>.
- Islami, Muhammad Iqbal, and Adhi Prasetyo. 2018. “Analisis Dan Prediksi Penjualan Pada Marketplace Berdasarkan Pendekatan Klasifikasi Dengan Metode Pohon Keputusan ( Studi Kasus Pada Data Katalog Dan Penjualan Tas Pria & Wanita Di Tokopedia Secara Nasional ).” *e-Proceeding of Management ISSN : 2355-9357* 5(1): 30–37.
- Karsito, and Winda Monika Sari. 2018. “Prediksi Potensi Penjualan Produk Delifrance Dengan Metode Naive Bayes Di Pt. Pangan Lestari.” *Jurnal Teknologi Pelita Bangsa ISSN : 2407-3903* 9(1): 67–78.
- Karthik, R. V., and Sannasi Ganapathy. 2021. “A Fuzzy Recommendation System for Predicting the Customers Interests Using Sentiment Analysis and Ontology in E-Commerce.” *Applied Soft Computing* 108: 107396. <https://doi.org/10.1016/j.asoc.2021.107396>.
- Kurniawan, Fachrul, and Manzilur Rahman Romadhon. 2021. “A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia.” *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIconCIT 2021*: 41–44.
- Mesri Silalahi. 2018. “Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Pada PT Batamas Niaga Jaya.” *Computer Based Information System Journal ISSN (Print): 2337-8794 | E- ISSN : 2621-5292* 02: 20–35. <http://ejournal.upbatam.ac.id/index.php/cbis>.
- Muzani, Ma’ruf Aziz et al. 2021. “Data Mining Untuk Klasifikasi Produk Menggunakan Algoritma K-Nearest Neighbor Pada Toko Online.” *Prosiding SISFOTEK ISSN 2597-3584* 5(1): 141–45. <http://seminar.iaii.or.id/index.php/SISFOTEK/article/view/273>.
- Noviana, Dita et al. 2019. “Analisis Rekomendasi Penerima Beasiswa Menggunakan Algoritma K-Nearest Neighbor (K-NN) Dan Algoritma C4.5.” *Seminar Nasional Penelitian Pendidikan Matematika (SNP2M) 2019 UMT*: 79–87.

- Pransiska, Novi, A.Haidar Mirza, Andri., and -. 2017. "Penerapan Data Mining Prediksi Penjualan Barang Elektronik Terlaris Menggunakan Algoritma Naïve Bayes ( Study Kasus : Planet Cash And Credit Cabang Muara Enim )." *Bina Darma Conference on Computer Science e-ISSN: 2685-2683p-ISSN: 2685-2675*: 2157–69.
- Putri, Ayu Azlina. 2021. "Penerapan Data Mining Untuk Memprediksi Penjualan Buah Dan Sayur Menggunakan Metode K-Nearest Neighbor ( Studi Kasus : PT . Central Brastagi Utama )." *RESOLUSI : Rekayasa Teknik Informatika dan Informasi ISSN 2745-7966* 1(6): 354–61. <https://djournals.com/resolusi>.
- Sharma, Dilip Kumar et al. 2021. "E-Commerce Product Comparison Portal for Classification of Customer Data Based on Data Mining." *Materials Today: Proceedings* 51(xxxx): 166–71. <https://doi.org/10.1016/j.matpr.2021.05.068>.
- Song, Bo, Wei Yan, Tianjiao Zhang, and -. 2019. "Cross-Border e-Commerce Commodity Risk Assessment Using Text Mining and Fuzzy Rule-Based Reasoning." *Advanced Engineering Informatics* 40(January): 69–80. <https://doi.org/10.1016/j.aei.2019.03.002>.
- Vilano, Nucky, and Setia Budi. 2020. "Penerapan Kansei Engineering Dalam Perbandingan Desain Aplikasi Mobile Marketplace Di Indonesia." *Jurnal Teknik Informatika dan Sistem Informasi* 6(2): 354–64.
- Wang, Kai et al. 2020. "E-Commerce Personalized Recommendation Analysis by Deeply-Learned Clustering." *Journal of Visual Communication and Image Representation* 71.
- Yahya, and Mahpuz. 2019. "Penggunaan Algoritma K-Means Untuk Menganalisis Pelanggan Potensial Pada Dealer SPS Motor Honda Lombok Timur Nusa Tenggara Barat." *Infotek : Jurnal Informatika dan Teknologi e-ISSN 2614-8773* 2(2): 109–18.
- Yunianto, Era, and Taryadi Taryadi. 2022. "Implementasi Fuzzy Decision Making Untuk Pemilihan Marketplace." *Smart Comp: Jurnalnya Orang Pintar Komputer P-ISSN: 2089-676X E-ISSN: 2549-0796* 11(1): 100–104.
- Zainuddin, Moh., and Achmad Noercholis. 2019. "Studi Komparasi Algoritma Decesion Tree (3. C4.5) Dengan Algoritma K-Nn Dalam Memprediksi Kelulusan Tepat Waktu Mahasiswa." *Jurnal Teknologi Informasi ISSN 2086-2989* 10. <http://ejurnal.stimata.ac.id/index.php?journal=ti&page=article&op=view&path%5B%5D=364>.