

**YUDI ISTIANTO**  
**NIM. 18206050003**



**KLASIFIKASI KEBUTUHAN JUMLAH PRODUK MAKANAN *CUSTOMER*  
MENGUNAKAN K-MEANS *CLUSTERING*  
DENGAN OPTIMASI PUSAT AWAL *CLUSTER* ALGORITMA GENETIKA**

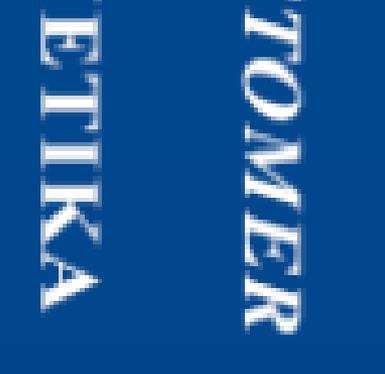
STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

**DATA  
SCIENCE**

**2020**

YUDI ISTIANTO  
NIM.18206050003

KLASIFIKASI KEBUTUHAN  
JMLAH PRODUK MAKANAN  
CUSTOMER  
MENGGUNAKAN K-MEANS  
CLUSTERING  
DENGAN OPTIMASI  
PUSAT GRAVITASI



2020



**PROGRAM STUDI INFORMATIKA  
PROGRAM MAGISTER FAKULTAS SAINS DAN TEKNOLOGI  
UIN SUNAN KALIJAGA YOGYAKARTA**

**KLASIFIKASI KEBUTUHAN JUMLAH PRODUK  
MAKANAN *CUSTOMER*  
MENGUNAKAN K-MEANS *CLUSTERING*  
DENGAN OPTIMASI PUSAT AWAL *CLUSTER*  
ALGORITMA GENETIKA**

**TESIS**

**Disusun oleh:**

**Yudi Istianto  
NIM. 18206050003**

**PROGRAM STUDI INFORMATIKA  
PROGRAM MAGISTER  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA  
YOGYAKARTA  
2020**

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama : Yudi Istianto

NIM : 18206050003

Jenjang : Magister

Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan adalah hasil penelitian/karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sumbernya.

Yogyakarta, 10 Januari 2020

Saya yang menyatakan,



**Yudi Istianto**

NIM: 18206050003

## PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama : Yudi Istianto

NIM : 18206050003

Jenjang : Magister

Program Studi : Informatika

menyatakan bahwa naskah tesis ini secara keseluruhan benar-benar bebas dari plagiasi. Jika di kemudian hari terbukti melakukan plagiasi, maka saya siap ditindak sesuai ketentuan hukum yang berlaku.

Yogyakarta, 10 Januari 2020

Saya yang menyatakan,



**Yudi Istianto**

NIM: 18206050003



## PENGESAHAN TUGAS AKHIR

Nomor : B-1120/Un.02/DST/PP.00.9/05/2020

Tugas Akhir dengan judul : **KLASIFIKASI KEBUTUHAN JUMLAH PRODUK MAKANAN CUSTOMER MENGGUNAKAN K-MEANS CLUSTERING DENGAN OPTIMASI PUSAT AWAL CLUSTER ALGORITMA GENETIKA**

yang dipersiapkan dan disusun oleh:

Nama : YUDI ISTIANTO  
Nomor Induk Mahasiswa : 18206050003  
Telah diujikan pada : Jumat, 24 April 2020  
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

### TIM UJIAN TUGAS AKHIR



Ketua Sidang

Dr. Shofwatul 'Uyun, S.T., M.Kom.  
SIGNED

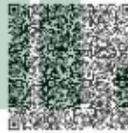
Valid ID: Sed1086257860



Penguji I

Dr. Bambang Sugiantoro, S.Si., M.T.  
SIGNED

Valid ID: Sed1048500aa



Penguji II

Maria Ulfah Siregar, S.Kom. MIT., Ph.D.  
SIGNED

Valid ID: Sed1305ed77c



Yogyakarta, 24 April 2020  
UIN Sunan Kalijaga  
Dekan Fakultas Sains dan Teknologi

Dr. Murtono, M.Si.  
SIGNED

Valid ID: Sed8968201ee4

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## PERSETUJUAN TIM PENGUJI UJIAN TESIS

Tesis berjudul : KLASIFIKASI KEBUTUHAN JUMLAH PRODUK  
MAKANAN *CUSTOMER* MENGGUNAKAN K-  
MEANS *CLUSTERING* DENGAN OPTIMASI  
PUSAT AWAL *CLUSTER* ALGORITMA

GENETIKA

Nama : Yudi Istianto

NIM : 18206050003

Prodi : Informatika

telah disetujui tim penguji ujian munaqosah

Ketua Penguji/Pembimbing : Dr. Shofwatul 'Uyun, S.T., M.Kom. (\_\_\_\_\_)

Penguji 1 : Dr. Bambang Sugiantoro, M.T. (\_\_\_\_\_)

Penguji 2 : Maria Ulfah Siregar, S.Kom., MIT., Ph.D. (\_\_\_\_\_)

Diuji di Yogyakarta pada tanggal : 24 April 2020

Waktu : 14:00 – 15.00

Hasil/Nilai : ~~A-~~

Predikat : Memuaskan/Sangat Memuaskan/Cumlaude\*

\* Coret yang tidak perlu

## NOTA DINAS PEMBIMBING

Kepada Yth.,  
Dekan Fakultas Sains dan Teknologi  
UIN Sunan Kalijaga Yogyakarta

*Assalamu 'alaikum wr. wb.*

Setelah melakukan bimbingan, arahan, dan koreksi terhadap penulisan tesis yang berjudul:

**KLASIFIKASI KEBUTUHAN JUMLAH PRODUK  
MAKANAN *CUSTOMER*  
MENGUNAKAN *K-MEANS CLUSTERING*  
DENGAN OPTIMASI PUSAT AWAL *CLUSTER*  
ALGORITMA**

Yang ditulis oleh:

Nama : Yudi Istianto  
NIM : 18206050003  
Jenjang : Magister  
Program Studi : Informatika

Saya berpendapat bahwa tesis tersebut sudah dapat diajukan kepada Magister Informatika UIN Sunan Kalijaga untuk diujikan dalam rangka memperoleh gelar Magister Informatika.

*Wassalamu'alaikum wr. wb.*

Yogyakarta, 10 Januari 2020

Pembimbing,

Dr. Shofwatul 'Uyun, S.T., M.Kom  
NIP. 19820511 200604 2 002

## ABSTRAK

PT.XYZ adalah salah satu perusahaan di Yogyakarta yang bergerak pada bidang produksi dan distribusi produk makanan roti. Setiap konsumen memiliki jumlah kebutuhan roti yang tidak teratur sedangkan roti hanya dapat bertahan dalam waktu dua hari. Roti yang sudah berusia lebih dari dua hari akan diganti dengan yang baru oleh distributor menimbulkan kerugian bagi perusahaan. Penelitian ini mencoba untuk menerapkan data mining untuk klasifikasi kebutuhan jumlah produk makanan *customer* menggunakan k-means *clustering* dengan optimasi pusat awal *cluster* algoritma genetika. Pada penelitian ini digunakan 210 data dari penjualan produk selama tiga Minggu. Data tersebut akan diproses dengan menerapkan metode data mining dengan *preprocessing* sebelum melalui tahap klasifikasi. *Preprocessing* yang dilakukan antara lain data transformation dan k-means *clustering*. Hasil dari *clustering* yang membutuhkan aturan tertentu lebih efektif dengan optimasi karena dari 210 data terdapat 200 data yang layak masuk tahap klasifikasi. Hasil dari pengujian mendapatkan akurasi terbaik sebesar 58.50 % dan *crossvalidation* untuk lima *fold* berhasil mendapatkan rata-rata akurasi sebesar 50.58% lebih besar 2.51 % dari KNN tanpa *preprocessing*.

**Kata Kunci:** *K-means, Clustering, Algoritma Genetika.*

## **ABSTRACT**

*PT.XYZ is one of the companies in Yogyakarta engaged in the production and distribution of bakery food products. Every consumer has an irregular amount of bread needs while bread can only last for two days. Bread that is more than two days old will be replaced by a new one by the distributor which causes losses for the company. This study tries to apply data mining to classify the number of customer needs for food products using k-means clustering with optimization initial cluster center genetic algorithm. In this study 210 data were used from product sales for three weeks. The data will be processed by applying the data mining method with preprocessing before going through the classification stage. Preprocessing includes data transformation and k-means clustering. The results of clustering that require certain rules are more effective with optimization because of 210 data there are 200 data that are worth entering the classification stage. The results of the test get the best accuracy of 58.50% and crossvalidation for five fold managed to get an average accuracy of 50.58% greater than 2.51% of KNN without preprocessing.*

**Keywords:** *K-means, Clustering, Genetic Algorithm.*

## KATA PENGANTAR

Segala puji bagi Allah Subhanahu wata'ala, yang telah memberi kesempatan dan kekuatan sehingga thesis yang berjudul “Klasifikasi Kebutuhan Jumlah Produk Makanan *Customer* menggunakan K-Means *Clustering* dengan Optimasi Pusat Awal *Cluster* Algoritma Genetika” ini dapat diselesaikan sebagai salah satu persyaratan mencapai derajat Magister Informatika. Penyusun mengucapkan terima kasih kepada semua pihak yang telah memberikan dorongan, semangat, dan ide-ide kreatif sehingga tahap demi tahap penyusunan tesis ini telah selesai. Ucapan terima kasih tersebut secara khusus disampaikan kepada:

1. Bapak Dr. Phil. Sahiron, M.A., sebagai Plt. Rektor UIN Sunan Kalijaga Yogyakarta.
2. Dr. Bambang Sugiantoro, M.T., selaku Ketua Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Dr. Shofwatul 'Uyun, S.T., M.Kom., selaku dosen Pembimbing Thesis yang telah memberikan motivasi dan pengarahan selama studi sehingga dapat menyelesaikan penyusunan tesis ini.
4. Dosen-dosen Program Studi Magister Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang sudah membagi ilmu yang sangat bermanfaat.
5. Seluruh Staf Karyawan Fakultas Sains dan Teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta yang telah membantu sehingga penyusunan tesis ini dapat berjalan dengan lancar.

6. Bapak dan Ibu selaku orangtua yang telah memberikan doa, motivasi dan segalanya selama studi di Magister Informatika Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
7. Teman-teman magister informatika angkatan 2018 Universitas Islam Negeri Sunan Kalijaga atas kerjasama, saran, dan bantuannya.
8. Teman-teman kerja dari Kominfo Kota Jogja, Gamatechno Indonesia, dan Riztra atas suportnya.
9. Pihak PT.XYZ yang telah telah membantu dalam observasi dan pengambilan data.
10. Semua pihak yang tidak bisa penulis sebutkan satu persatu atas bantuannya dalam penyelesain thesis ini.

Akhirnya penulis hanya dapat bersyukur kepada Allah SWT semoga yang telah dilakukan selamama ini dapat menjadi amal dan bekal di akhirat nanti. Penulis menyadari sepenuhnya masih banyak kesalahan dan kekurangan dalam thesis ini, maka dari itu berbagai saran dan kritik sangat diharapkan demi perbaikan. Semoga tesis ini dapat bermanfaat bagi penulis sendiri khususnya dan bagi pembaca pada umumnya, terimakasih.

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

Yogyakarta, 28 Januari 2020

Yudi Istianto  
NIM. 18206050003

## HALAMAN PERSEMBAHAN

Karya tesis ini saya persembahkan kepada:

- ❖ Bapak Ngadiyo dan Ibu Sumilah tercinta, orang paling berpengaruh dalam hidupku yang tak pernah jenuh mendo'akan dan menyangiku, atas semua pengorbanan dan kesabaran mengantarku sampai kini. Tak pernah cukup ku membalas namun berharap semoga karya ini dapat membanggakan Bapak dan Ibu.
- ❖ Keluarga besar Bapak Marto Rejo dan Soeparto di kampung halaman (Ngasem, Botodayaan Rongkop Gunungkidul) yang selalu mendo'akan, memberi motivasi dan memberi arahan agar bisa menjadi seorang magister dan membagikan keluarga.
- ❖ Sahabat-sahabatku Anindya Purba, Vivi Noviyanti, Dwi Ari Wahyuni, Tulus Kurniawan, dkk yang selalu ada dimana saya bersedih dan tertawa.
- ❖ Teman-teman seperjuangan sejak masih kecil dan masih menjadi teman meski bekerja di daerah yang berbeda-beda Faris Leo, Serdy Aditya, Gani Ahmad, Restu Soleh, dkk.
- ❖ Teman-teman satu tim kerja dan satu project aplikasi dalam pekerjaan di Kominfo Kota Jogja, Gamatechno, dan Riztra yang selalu solid.

- ❖ Teman-teman selama menempuh program sarjana yang masih tetap kompak kumpul ngopi bareng Joni, Rizki, Setyo, Aji, Yuha, Anggoro, dkk.
- ❖ Teman-teman seperjuangan thesis Magister Informatika 2018 khususnya yang telah membantu sehingga thesis ini dapat diselesaikan Daru, Gatra, Sutriman, Rizki, dkk.
- ❖ Beberapa orang yang mungkin membuat saya berjuang keras untuk membuktikan saya bisa dan tak seperti apa yang mereka pikirkan terimakasih atas semuanya.



## HALAMAN MOTTO

“Jika kita lunak di dalam, maka dunia luar akan keras kepada kita. Tapi jika kita keras di dalam, maka dunia luar akan lunak kepada kita”

(Andrie Wongso)

“Tugas seorang manusia adalah menjadi manusia”

(Multatuli)

“Tidak ada kebetulan yang terjadi di dunia ini, sepertinya semua pertemuan dan kejadian sudah direncanakan sejak awal”

(Silvers Rayleigh: One Piece)

“Masa muda adalah anugerah dalam hidup. Buat masa mudamu penuh kenangan dan penuh arti”

(Crows Zero)

“Jika kebaikan hari ini hanya akan membuat keburukan dimasa mendatang, maka lebih baik menahan sakit atas keburukan hari ini dan menunggu kebaikan dimasa yang akan datang”

(Yudi Istianto)

## DAFTAR ISI

HALAMAN JUDUL.....	iv
PERNYATAAN KEASLIAN .....	v
PERNYATAAN BEBAS PLAGIASI.....	vi
HALAMAN PENGESAHAN .....	vii
PERSETUJUAN TIM PENGUJI UJIAN TESIS .....	viii
NOTA DINAS PEMBIMBING .....	ix
ABSTRAK.....	x
<i>ABSTRACT</i> .....	xi
KATA PENGANTAR .....	xii
HALAMAN PERSEMBAHAN.....	xiv
HALAMAN MOTTO.....	xvi
DAFTAR ISI.....	xvii
DAFTAR TABEL.....	xx
DAFTAR GAMBAR .....	xxiii
DAFTAR LAMPIRAN .....	xxiv
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
A. Latar Belakang Masalah.....	1
B. Rumusan Masalah .....	4
C. Batasan Masalah.....	5
D. Tujuan Penelitian.....	5
E. Manfaat Penelitian.....	6
F. Sistematika Penulisan.....	6
<b>BAB II KAJIAN PUSTAKA DAN LANDASAN TEORI.....</b>	<b>7</b>
A. Kajian Pustaka.....	7
B. Landasan Teori .....	11
1. Data Mining.....	12

2. K-Means .....	12
3. Algoritma Genetika .....	13
4. K-Nearest Neighbor (K-NN).....	17
5. HTML.....	18
6. PHP .....	19
7. Javascript .....	19
8. MySQL.....	20
<b>BAB III METODE PENELITIAN.....</b>	<b>22</b>
A. Jenis dan Desain Penelitian.....	22
B. Sumber Data dan Metode Pengumpulan Data .....	23
C. Data .....	24
D. <i>Preprocessing</i> .....	25
1. <i>Data Transformation</i> .....	25
2. <i>K-Means Clustering</i> .....	26
E. K-Nearest Neighbor (K-NN) <i>Classification</i> .....	32
F. Pengujian dan Evaluasi .....	32
<b>BAB IV HASIL DAN PEMBAHASAN.....</b>	<b>35</b>
A. Data .....	35
B. <i>Preprocessing</i> .....	36
1. <i>Data Transformation</i> .....	36
2. <i>K-Means Clustering</i> .....	37
C. K-Nearest Neighbor (K-NN) <i>Classification</i> .....	54
D. Pengujian dan Evaluasi .....	58
<b>BAB V PENUTUP .....</b>	<b>67</b>
A. Kesimpulan.....	67
B. Saran.....	68
<b>DAFTAR PUSTAKA .....</b>	<b>69</b>



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	10
Tabel 3.1 Atribut Data.....	24
Tabel 3.2 Kode dan Transformasi Atribut.....	25
Tabel 3.3 <i>Rule Cluster</i> .....	27
Tabel 3.4 Rasio Normalisasi Atribut.....	27
Tabel 3.5 Pembentukan Kromosom.....	29
Tabel 4.1 Data Latih dan Uji.....	35
Tabel 4.2 Kode Data Latih dan Uji.....	36
Tabel 4.3 Rasio Normalisasi Atribut <i>Customer</i> .....	38
Tabel 4.4 Rasio Normalisasi Atribut Jenis Hari.....	39
Tabel 4.5 Rasio Normalisasi Atribut Cuaca.....	39
Tabel 4.6 Rasio Normalisasi Atribut Hari.....	40
Tabel 4.7 Normalisasi Data Latih dan Uji.....	40
Tabel 4.8 Bobot Data Latih dan Uji.....	42
Tabel 4.9 Populasi Awal.....	42
Tabel 4.10 <i>Fitness</i> Populasi Awal.....	43
Tabel 4.11 <i>Fitness</i> dan <i>Probabilitas</i> .....	44
Tabel 4.12 <i>Roulette Wheel Selection</i> .....	44

Tabel 4.13 <i>Crossover</i> .....	46
Tabel 4.14 Mutasi .....	46
Tabel 4.15 Elitisme.....	47
Tabel 4.16 Hasil Algoritma Genetika .....	48
Tabel 4.17 Hasil Percobaan Algoritma Genetika .....	48
Tabel 4.18 Pusat Awal <i>Cluster</i> .....	50
Tabel 4.19 Normalisasi Pusat Awal <i>Cluster</i> .....	50
Tabel 4.20 Jarak <i>Cluster</i> Iterasi I.....	51
Tabel 4.21 Jumlah Data <i>Cluster</i> Iterasi I.....	51
Tabel 4.22 Pusat <i>Cluster</i> Iterasi II.....	52
Tabel 4.23 Jarak <i>Cluster</i> Iterasi II.....	53
Tabel 4.24 Hasil <i>Preprocessing</i> .....	53
Tabel 4.25 Transformasi <i>Cluster</i> dan <i>Class</i> .....	54
Tabel 4.26 Data Latih.....	55
Tabel 4.27 Transformasi Data Latih.....	56
Tabel 4.28 Normalisasi Data Latih.....	56
Tabel 4.29 Jarak Klasifikasi .....	57
Tabel 4.30 Percobaan Klasifikasi .....	58
Tabel 4.31 Pengujian <i>Confusion Matrix</i> .....	59
Tabel 4.32 Hasil Perbandingan Pengujian <i>Confusion Matrix</i> .....	61
Tabel 4.33 Perbandingan Waktu Eksekusi K-Means .....	62

Tabel 4.34 Perbandingan Jumlah Data K-Means .....	63
Tabel 4.35 Perbandingan Waktu Eksekusi Klasifikasi KNN .....	63
Tabel 4.36 Pengujian <i>Crossvalidation</i> KNN .....	64
Tabel 4.37 Pengujian <i>Crossvalidation</i> KNN ( <i>Preprocessing</i> K-Means) ..	65



## DAFTAR GAMBAR

Gambar 2.1	<i>Flowchart</i> Algoritma Genetika.....	17
Gambar 3.1	Tahapan Penelitian.....	23
Gambar 4.1	Grafik Perbandingan Pengujian <i>Confusion Matrix</i> .....	61
Gambar 4.2	Grafik Pengujian <i>Crossvalidation</i> KNN.....	65
Gambar 4.3	Grafik Pengujian <i>Crossvalidation</i> KNN ( <i>Preprocessing</i> K-Means).....	66
Gambar 4.	Grafik Perbandingan Pengujian <i>Crossvalidation</i> .....	66



STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

## DAFTAR LAMPIRAN

LAMPIRAN A : DATA PENJUALAN.....	70
LAMPIRAN B : DATA LATIH DAN UJI.....	74
LAMPIRAN C : KODE DATA LATIH DAN UJI.....	92
LAMPIRAN D : NORMALISASI DATA LATIH DAN UJI .....	99
LAMPIRAN E : BOBOT DATA LATIH DAN UJI.....	106
LAMPIRAN F : JARAK <i>CLUSTER</i> ITERASI I .....	113
LAMPIRAN G : JARAK <i>CLUSTER</i> ITERASI II .....	121
LAMPIRAN H : HASIL <i>PREPROCESSING</i> .....	129
LAMPIRAN I : DATA LATIH KLASIFIKASI .....	149
CURRICULUM VITAE .....	157

STATE ISLAMIC UNIVERSITY  
SUNAN KALIJAGA  
YOGYAKARTA

# BAB I

## PENDAHULUAN

### A. Latar Belakang Masalah

Perkembangan teknologi telah membawa masyarakat dalam perubahan hidup yang lebih baik. Teknologi tidak hanya sekedar pelengkap namun telah menjadi kebutuhan hidup manusia yang tidak dapat terpisahkan. Komputer adalah salah satu teknologi yang digunakan untuk mempermudah kehidupan manusia dalam semua bidang tidak terkecuali bidang industri. Pemanfaatan ilmu komputer sangat berperan dalam penggunaan komputer yang lebih baik. Salah satu ilmu komputer yang sangat berperan ini adalah data mining.

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data. Salah satu metode yang digunakan dalam pengolahan data adalah pengelompokan / *clustering* (Rustiyan & Mustakim, 2018).

Algoritma k-means adalah metode *clustering* yang paling populer digunakan karena beberapa kelebihan yang dimiliki antara lain algoritma ini sederhana dan mudah diterapkan (Widiarina, 2013). Selain itu algoritma k-means tidak terpengaruh dengan adanya urutan objek (Aranda & Natasya, 2016). Namun, k-means mempunyai kelemahan yang diakibatkan oleh penentuan pusat awal *cluster*. Hal ini menyebabkan hasil *cluster*-nya berupa solusi yang sifatnya *local optimal*. Menggunakan metode hierarki tertentu dapat dikolaborasikan untuk optimasi pusat awal *cluster* k-means. Salah satu algoritma yang dapat digunakan untuk optimasi adalah algoritma genetika. Algoritma Genetika merupakan metode pencarian solusi yang sesuai dengan kriteria dari banyak kombinasi solusi yang ada tanpa harus menguji satu-persatu untuk mendapatkannya. Tahap proses siklus genetika diawali dengan membangkitkan populasi awal yaitu pembangkitan sejumlah individu secara acak.

Klasifikasi merupakan salah satu teknik dalam data mining. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu (Java, Wisdayani, Nur, Wasono, & Semarang, n.d.). Salah satu algoritma klasifikasi yang cukup populer adalah k-nearest neighbor (KNN). Dalam pengenalan pola, algoritma k-nearest neighbor

adalah salah satu metode non-parametrik yang paling terkenal dan berguna untuk mengelompokkan objek berdasarkan fitur-fitur yang dekat. k-nearest neighbor dirancang dengan konsep bahwa label atau kelas ditentukan oleh suara mayoritas tetangganya (Won Yoon & Friel, 2015).

Distribusi merupakan sebuah aktivitas yang dianggap sangat penting di dalam sebuah perusahaan. Sistem distribusi merupakan salah satu pilar yang penting dalam kelangsungan dan keberhasilan perusahaan, oleh karena itu diperlukan manajemen yang baik untuk mengatur sistem distribusi agar distribusi bekerja dengan baik. Distributor dituntut menyalurkan produk sesuai kebutuhan *customer* namun juga harus melakukan manajemen stok yang baik untuk mencegah kekosongan stok. *Customer* yang potensial adalah pelanggan yang memiliki tingkat kepuasan yang tinggi sehingga memiliki loyalitas yang tinggi dengan frekwensi dan nilai permintaan yang tinggi pula (Musanto, 2004). Transaksi retur penjualan terjadi jika perusahaan menerima pengembalian barang dari *customer* (Chandrashekar, Rotte, Tax, & Grewal, 2007). Semakin tinggi nilai retur terhadap penjualan maka kepuasan *customer* semakin menurun dan hal tersebut akan menurunkan loyalitas *customer* (Ngai, Xiu, & Chau, 2009).

PT.XYZ adalah salah satu perusahaan di Yogyakarta yang bergerak pada bidang produksi dan distribusi makanan roti. PT.XYZ pada

saat ini mampu memproduksi kurang lebih 1000 roti dalam sehari, sebelum hari selanjutnya mereka distribusikan ke toko-toko dan pedagang jajanan pasar. Pada saat ini jumlah konsumen telah mencapai lebih dari 50 konsumen. Setiap konsumen memiliki jumlah kebutuhan roti yang tidak teratur setiap hari sedangkan roti hanya dapat bertahan dalam waktu dua hari. Roti yang sudah berusia lebih dari dua hari akan diganti dengan yang baru oleh distributor yang bisa menimbulkan kerugian bagi perusahaan. Berdasarkan latar belakang di atas maka peneliti tertarik untuk melakukan penelitian mengenai data mining untuk klasifikasi kebutuhan jumlah produk makanan *customer* menggunakan *k-means clustering* dengan optimasi pusat awal *cluster* algoritma genetika yang bermanfaat untuk mempermudah prediksi jumlah kebutuhan *customer* dimasa mendatang berdasarkan data penjualan masa lampau pada PT.XYZ.

## **B. Rumusan Masalah**

Berdasarkan uraian diatas dapat dirumuskan beberapa pertanyaan penelitian yang dinyatakan dalam bentuk pertanyaan yang nantinya akan dipecahkan pada penelitian ini :

1. Bagaimana kinerja algoritma *k-means clustering* dengan optimasi pusat awal *cluster* algoritma genetika pada pengelompokkan data penjualan produk PT.XYZ ?

2. Bagaimana kinerja algoritma klasifikasi k-nearest neighbors (KNN) yang menggunakan *preprocessing* k-means *clustering* pada data penjualan produk PT.XYZ ?
3. Bagaimana kinerja data mining dengan menerapkan usulan-usulan dalam memprediksi kebutuhan jumlah produk dimasa depan untuk *customer* PT.XYZ ?

### C. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah :

1. Jumlah konsumen yang akan digunakan dalam penelitian dibatasi yakni sepuluh *customer*.
2. Aturan pengelompokan, kelas, dan atribut yang digunakan pada penelitian ditentukan oleh pihak PT.XYZ.
3. Produk makanan *customer* dalam penelitian adalah produk makanan roti yang hanya bertahan dalam dua hari.

### D. Tujuan Penelitian

Adapun tujuan penelitian dalam penelitian ini adalah :

1. Mengetahui kinerja algoritma k-means *clustering* dengan optimasi pusat awal *cluster* algoritma genetika pada pengelompokan data penjualan produk PT.XYZ.

2. Mengetahui kinerja algoritma klasifikasi k-nearest neighbors (KNN) yang menggunakan *preprocessing* k-means *clustering* pada data penjualan produk PT.XYZ.
3. Mengetahui kinerja data mining dengan menerapkan usulan-usulan yang dalam memprediksi kebutuhan jumlah produk dimasa depan untuk *customer* PT.XYZ.

#### **E. Manfaat Penelitian**

Penelitian tentang data mining untuk klasifikasi kebutuhan jumlah produk makanan menggunakan k-means *clustering* dengan optimasi pusat awal *cluster* algoritma genetika dan klasifikasi k-nearest neighbors (KNN) mempunyai manfaat untuk mempermudah prediksi jumlah kebutuhan *customer* dimasa mendatang berdasarkan data penjualan masa lampau.

#### **F. Sistematika Penulisan**

Penyusunan laporan penelitian ini mengkaji beberapa bab pembahasan dalam pengerjaannya. Sistematika penulisan dimulai dari bab 1 sampai bab 5 yaitu :

**BAB I** Pendahuluan

**BAB II** Kajian Pustaka dan Landasan Teori

**BAB III** Metode Penelitian

**BAB IV** Hasil dan Pembahasan

**BAB V** Penutup

## BAB V

### PENUTUP

#### A. Kesimpulan

Kinerja algoritma k-means *clustering* dengan optimasi pusat awal *cluster* algoritma genetika pada pengelompokan data penjualan produk PT.XYZ cukup baik karena dapat meningkatkan iterasi yang dilakukan pada tahap *clustering*. Optimasi pusat awal *cluster* dapat mengatasi kelemahan algoritma k-means yang diakibatkan oleh penentuan pusat awal *cluster* yang tidak tepat menyebabkan hasil *cluster*-nya berupa solusi yang sifatnya *local* optimal. Hasil dari *clustering* yang membutuhkan aturan tertentu pada *cluster*-nya juga lebih efektif dengan optimasi ini karena dari 210 data terdapat 200 data yang layak untuk masuk tahap klasifikasi.

Kinerja algoritma klasifikasi k-nearest neighbors (KNN) yang menggunakan *preprocessing* k-means *clustering* pada data penjualan produk PT.XYZ cukup baik karena hasil dari pengujian menggunakan *confusion matrix* berhasil mendapatkan akurasi terbaik sebesar 58.50 % dan menggunakan *crossvalidation* untuk lima *fold* berhasil mendapatkan rata-rata akurasi sebesar 50.58% lebih besar 2.51 % dari KNN tanpa *preprocessing*.

Kinerja data mining dengan menerapkan usulan-usulan yang dalam memprediksi kebutuhan jumlah produk dimasa depan untuk konsumen PT.XYZ cukup baik karena berdasarkan uji kasus yang dilakukan dapat memprediksi kebutuhan dengan tepat. Berdasarkan uji jumlah  $k = 1 - 6$  hasil prediksi tetap menunjukkan kelas yang sama.

## **B. Saran**

Pada penelitian mengenai data mining untuk klasifikasi kebutuhan jumlah produk makanan *customer* menggunakan *k-means clustering* dengan optimasi pusat awal *cluster* algoritma genetika berhasil dilakukan namun akurasi yang didapat menggunakan pengujian *crossvalidation* dengan lima *fold* masih belum maksimal. Hal ini sangat dipengaruhi oleh jumlah data yang masih kurang dan memiliki jumlah kelas yang tidak merata, sehingga perlu dilakukan penambahan jumlah data dan pemerataan kelas untuk mendapatkan akurasi yang lebih baik.