**A Hybrid Fuzzy C-Means And Ant Colony Optimization For Customer Segmentation Using Data Mining Techniques**



**OLEH :**

**PANDITO DEWA PUTRA**

**09042681721002**

**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

**A Hybrid Fuzzy C-Means And Ant Colony Optimization For Customer Segmentation Using Data Mining Techniques**

# **HALAMAN JUDUL**

**TESIS**

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Magister



**OLEH :**

**PANDITO DEWA PUTRA**

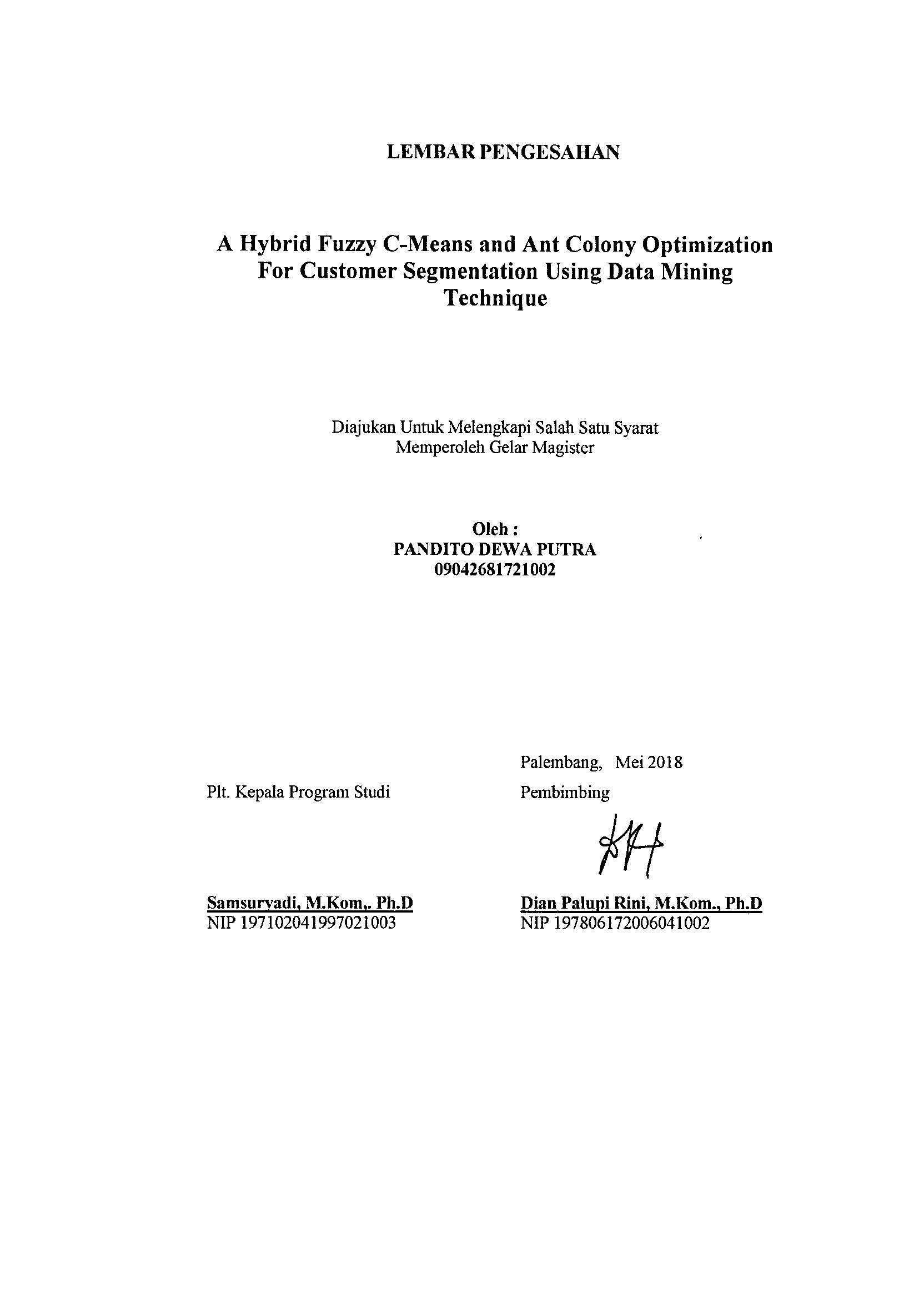
**09042681620002**

**PROGRAM STUDI MAGISTER TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

****

# **DAFTAR GAMBAR**

[**HALAMAN JUDUL** i](#_Toc507588184)

[**LEMBAR PENGESAHAN** ii](#_Toc507588185)

[**HALAMAN PERSETUJUAN** iii](#_Toc507588186)

[**DAFTAR ISI** iv](#_Toc507588187)

[**DAFTAR GAMBAR** v](#_Toc507588188)

[**DAFTAR TABEL** vi](#_Toc507588189)

[**BAB I PENDAHULUAN** 1](#_Toc507588190)

[**1.1 Latar Belakang** 1](#_Toc507588191)

[**1.2 Perumusan Masalah** 2](#_Toc507588192)

[**1.3 Tujuan Penelitian** 2](#_Toc507588193)

[**1.4 Manfaat Penelitian** 2](#_Toc507588194)

[**1.5 Ruang Lingkup Penelitian** 3](#_Toc507588195)

[**1.6 Batasan Masalah** 3](#_Toc507588196)

[**BAB II TINJAUAN PUSTAKA** 4](#_Toc507588197)

[**2.1 Customer Segmentation Using Data Mining Techniues** 4](#_Toc507588198)

[**2.2 Hybrid Metode Fuzzy C-means and Ant Colony Optimization**](#_Toc507588199) 5

[**2.3 Fuzzy C-means Clustering**](#_Toc507588200) 6

[**2.4 Ant Colony Optimization**](#_Toc507588200) 7

[**BAB III METODE PENELITIAN**](#_Toc507588201) 9

[**3.1 Analisis Metodologi**](#_Toc507588202) 9

[**3.2 Jadwal Penelitian**](#_Toc507588203) 11

[**DAFTAR PUSTAKA**](#_Toc507588206) 12

# **DAFTAR GAMBAR**

[**Gambar 3.1** Kerangka Penelitian 1](#_Toc503851549)0

# **DAFTAR TABEL**

[**Tabel 2.1** Jadwal Penelitian. 11](#_Toc503852288)

**BAB I PENDAHULUAN**

**1.1 Latar Belakang**

Segmentasi pelanggan pernah diterapkan ketika menangani kasus pada bank yang membagi pasar kedalam kelompok pelanggan berdasarkan kebutuhan dan karakteristik yang berbeda dengan memerlukan kebutuhan produk pemasaran secara terpisah (Maghfirah dan Te 2015). Segmentasi yang efektif adalah segmentasi yang lebih mengarah pada persaingan keuntungan, pengakuan dan pengeksploitasi dari peluang pasar yang terbaru, pemilihan target pasar yang tepat sasaran, serta melakukan peningkatan profitabilitas (Ni Putu Putri Yuliari, I Ketut Gede Darma Putra dan Ni Kadek Dwi Rusjayanti, 2015). Segmentasi pelanggan juga pernah diterapkan menggunakan teknik data mining pada suatu objek toko onlinemaka pengidentifikasian pelanggan dapat dilakukan dengan menggunakan manajemen hubungan pelanggan, dengan konsep tersebut dapat mengidentifikasi pelanggan yang berpotensial melalui segmentasi pelanggan (Lisna Zahrotun, 2017).

Proses segmentasi pelanggan dilakukan dengan membagi pelanggan ke beberapa *cluster* dengan kategori pelanggan yang memiliki loyalitas untuk membangun strategi pasar, kemudian karakter segmentasi dibuat dengan aturan bisnis, serta algoritma *clustering* yang bisa menganalisa karakteristik data yang menghasilkan model data untuk *monitoring* (Ni Putu Putri Yuliari, I Ketut Gede Darma Putra dan Ni Kadek Dwi Rusjayanti, 2015).

Saat ini dalam melakukan pendekatan pemasaran yaitu dengan berorientasi kepada produk yang akan dipasarkan sehingga menyebabkan mayoritas perusahaan terlalu fokus ke produk yang mereka hasilkan saja, maka dengan hanya mengandalkan produk terjadi kesulitan saat akan dilakukan pengidentifikasian pelanggan secara benar (Sudriyanto, 2017). Data – data para pelanggan lebih didominan ada di data transaksi pelanggan yang memiliki banyak atribut sehingga kesulitan dan kurang optimal saat dilakukan pengolahan data (Hu dan Yeh, 2014).

Analisis clustering merupakan salah satu metode atau teknik yang lebih berguna untuk mengelola dan mengumpulkan pengetahuan berdasarkan dari sebuah *dataset*, Kemudian dapat mengelompokan berdasarkan kemiripan data yang dapat mendeteksi secara heterogentias tinggi diantara kluster dan homogenitas tinggi tanpa kluster. (Chiu, Kuo, dan Shih, tahun 2014).

Akibat pertumbuhan data transaksi yang cepat dan terbatasnya kemampuan segmentasi pelanggan secara manual maka digunakanlah teknik data mining menggunakan beberapa metode dengan melakukan *hybrid* algoritma. Dalam penelitian ini akan dilakukan *clustering* dengan menggunakan teknik data mining dengan objek *datasets* pelanggan. Kemudian akan dilakukan *hybrid* metode untuk proses clustering yaitu fuzzy c-means dan ant colony optimizition.

**1.2 Perumusan Masalah**

Dari latar belakang diatas terdapat beberapa isu yang akan dibahas dalam penelitian ini :

1. Bagaimana melakukan proses hybrid algoritma fuzzy c-means dengan algoritma ant colony optimization.
2. Bagaimana melakukan proses *clutering dataset* pelanggan.
3. Bagaimana melakukan proses validasi terhadap hasil *cluter* data yang baik atau buruk.

**1.3 Tujuan Penelitian**

Tujuan dari penelitian ini adalah :

1. Melakukan proses clustering dengan objek datasets pelanggan
2. Mengimplementasi hybrid algoritma fuzzy c-means dengan algoritma ant colony optimization.
3. Melakukan proses validasi terhadap hasil *cluster* data yang baik ataupun buruk

**1.4 Manfaat**

Sedangkan manfaat yang ingin dicapai dari penelitian ini adalah :

1. Dapat menghasilkan kelompok pelanggan yang memiliki loyalitas yang tinggi ataupun rendah terhadap perusahaan.
2. Dapat memberikan informasi mengenai tingkat akurasi dari algoritma yang diterapkan.
3. Dapat memberikan rekomendasi terkait segmentasi pelanggan yang terbaik pada bidang data mining.

**1.5 Ruang Lingkup Penelitian**

Ruang lingkup penelitian ini mengaju pada penelitian sebelumnya dimana masalah utama pada penelitian proses mensegmentasi pelanggan dengan teknik data mining *clustering*.

**1.6 Batasan Masalah :**

1. Data yang digunakan adalah datasets pelanggan.
2. Normalisasi data dengan analisis RFM (recency, frequncy, monetary).
3. Proses segmentasi pelanggan dengan teknik clustering.
4. Proses clustering dengan *hybrid* algoritma.
5. Validasi kelompok data yang baik atau buruk.

**BAB II TINJAUAN PUSTAKA**

Banyak penelitian data mining yang dilakukan pada segmentasi pelanggan, namun sebagian besar saat proses clustering dengan metode fuzzy c-means saja atau dikombinasikan dengan analisis RFM, namun belum ada yang melakukan *hybrid* algoritma untuk menemukan akurasi yang lebih baik dalam bidang data mining clustering segmentasi pelanggan. Pengujian proses *clustering* menggunakan purity test untuk mengetahui cluster baik atau buruk.

**2.1 Customer Segmentation Using Data Mining Techniues**

Penelitian yang dilakukan (Lisna, 2017) fokus segmentasi pelanggan dengan konsep manajemen hubungan pelanggan (CRM), dengan objek penelitian *online shop* dengan menggunakan 132 data pelanggan. Proses data *cleanup* dari 132 data menjadi 108 data yang digunakan dan proses data *selection* dari 7 atribut menjadi 4 atribut. Kemudian dianalisis menggunakan analisis RFM (recency, frequency, dan monetary). Dimana recency yaitu transaksi terakhir yang dilakukan pelanggan, frequency yaitu jumlah total transaksi perpriode, dan monetary yaitu rata – rata jumlah pembelian pelangga. Dengan tool yang digunakan yaitu php my admin

Penelitian yang dilakukan (Ni Putu Putri Yuliari, I Ketut Gede Darma Putra dan Ni Kadek Dwi Rusjayanti, 2015) segmentasi pelanggan dengan menerapkan analisis RFM (recency, frequency, dan monetary) dan validasi data menggunakan metode MPC (modified partition coefficient). Kemudian tool yang digunakan untuk melakukan segmentasi yaitu Matlab dan SQL server.

Dalam Pemaparan (Ali, Reza, dan Samira, 2011) fokus penelitian melakukan segmentasi pelanggan berdasarkan loyalitas pelanggan pada objek iranian food industry. Proses preparation menggunakan analisis model RFM (recency, frequency, dan monetary) yang ditambahkan 2 parameter D dan C (diversity dan Continuousness), dimana diversity yaitu berbagai produk yang dibeli pelanggan dalm suatu periode sedangkan Continuousness mengacu pada urutan belanja yang besar secara berkelanjutan yang dilakukan konsumen pada suatu priode. Objek data menggunakan data pelanggan seperti *Customer Code, Customer Description, Transaction Date, Product Type and Value Buy.*

Penelitian yang dilakukan (naime dan sasan, 2013) melakukan segmentasi pelanggan untuk membantu pelanggan membeli kebutuhannya, kemudian setelah segmentasi pelanggan digunakan fuzzy forecasting untuk memperkirakan kebutuhan pelanggan tersebut. Kontribusi pada penelitian ini yaitu pencapaian terbaru dalam *multi-variable fuzzy forecasting* untuk mengantisipasi kebutuhan pelanggan. Kemudian metodologi baru ini memiliki 2 bagian : pertama fuzzy clustering berbasis fuzzy c-means untuk segmentasi pelanggan dan kedua fuzzy forecasting dengan teknik interpolasi *multi-variable rule.* Tujuannya dapat menyarankan pelanggan terhadap produk yang dapat meningkatkan kepuasan pelanggan.

**2.2 *Hybrid* Metode Fuzzy C-means and Ant Colony Optimization**

Dalam pemaparan, (Saptarshi, Sanchita, dan Richard, 2018) Penelitian dengan teknik data mining menggunakan metode Fuzzy c-means telah dilakukan *hybrid* dengan metode Quantum-behaved Particle Swarm Optimization (QPSO), dimana QPSO merupakan variasi dari particle swarm optimization (PSO) dengan kemampuan pencarian yang lebih baik. Objek penelitian menggunakan 5 datasets (Fisher’s Iris, Breast Cancer Wisconsin, Seeds, Mammographic Mass, Sonar) yang berasal dari UCI Machine Learning Repository.

(Hesam, Ajith, dan Václav, 2009) Pada penelitian ini teknik *clustering* dengan melakukan *hybrid* metode fuzzy c-means dengan metode particle swarm optimization, kemudian objek penelitian yang digunakan yaitu 6 datasets (Fisher’s iris, Glass, Wisconsin breast cancer, Wine, Vowel, dan Contraceptive Method Choice.

Pada penelitian yang dilakukan (Sankar dan Krishnamoorthy, 2010) teknik *clustering* pernah dilakukan perbandingan metode Hard c-means dengan Fuzzy c-means clustering, dimana untuk metode fuzzy c-means di paralel kan dengan ant colony optimization dengan objek penelitian menggunakan 3 datasets yaitu Glass, Iris, dan Wine.

Pada penelitian ini dilakukan (Hao, Yan, dan Danyun, 2010) pendeteksian jaringan intrusion untuk mendeteksi aktivitas yang mencurigakan atau mencari bukti dari percobaan penyusupan dalam sebuah sistem jaringan dengan menggunakan hybrid metode fuzzy c-means clustering dengan metode quantum particle swarm optimization algorithm (QPSO). Objek penelitian menggunakan 38500 *records* dari "kddcup.data\_10.percent" sebagai datasets, dimana 38125 data yang normal sedangkan 375 data abnormal.

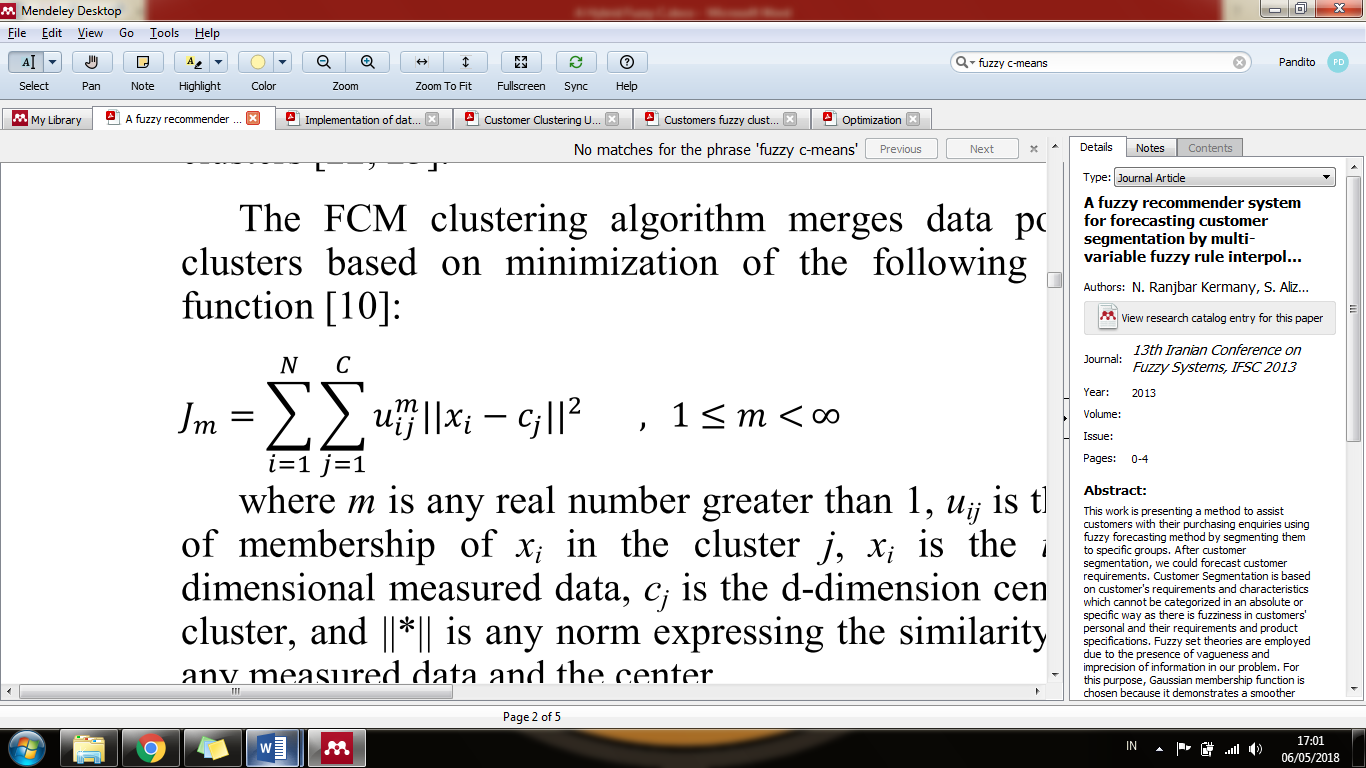
(gao, 2008) fokus melakukan hybrid yang menggunakan metode Ant Colony Optimization (ACO) dengan K-means untuk kasus *clustering* dengan objek datanya menggunakan Iris plant dataset. Hasil menunjukkan bahwa algoritma ant colony algoritma sangat optimal saat digunakan untuk *optimization problem* dan memiliki *performace* yang sama dengan metode sebelumnya..

Pada penelitian lain (YiJun, Man-Leung, dan Haibing, 2014) Penerapan Ant colony optimiztion (ACO) pada permasalahan data mining belum menyelesaikan pada kasus *stacking configuration*, maka dalam penelitian ini dilakukanlah pengembangan untuk masalah tersebut yang disebut dengan metode ACO-Stacking. ACO-Stacking di implementasikan dengan 3 versi berbeda ACO-S1, ACO-S2, ACO-S3. Hasilnya ACO-S3 lebih baik dalam 11 dataset benchmark dan 12 dataset benchmark dibanding metode variasi ACO-Stacking lainnya.

**2.3 Fuzzy C-means Clustering**

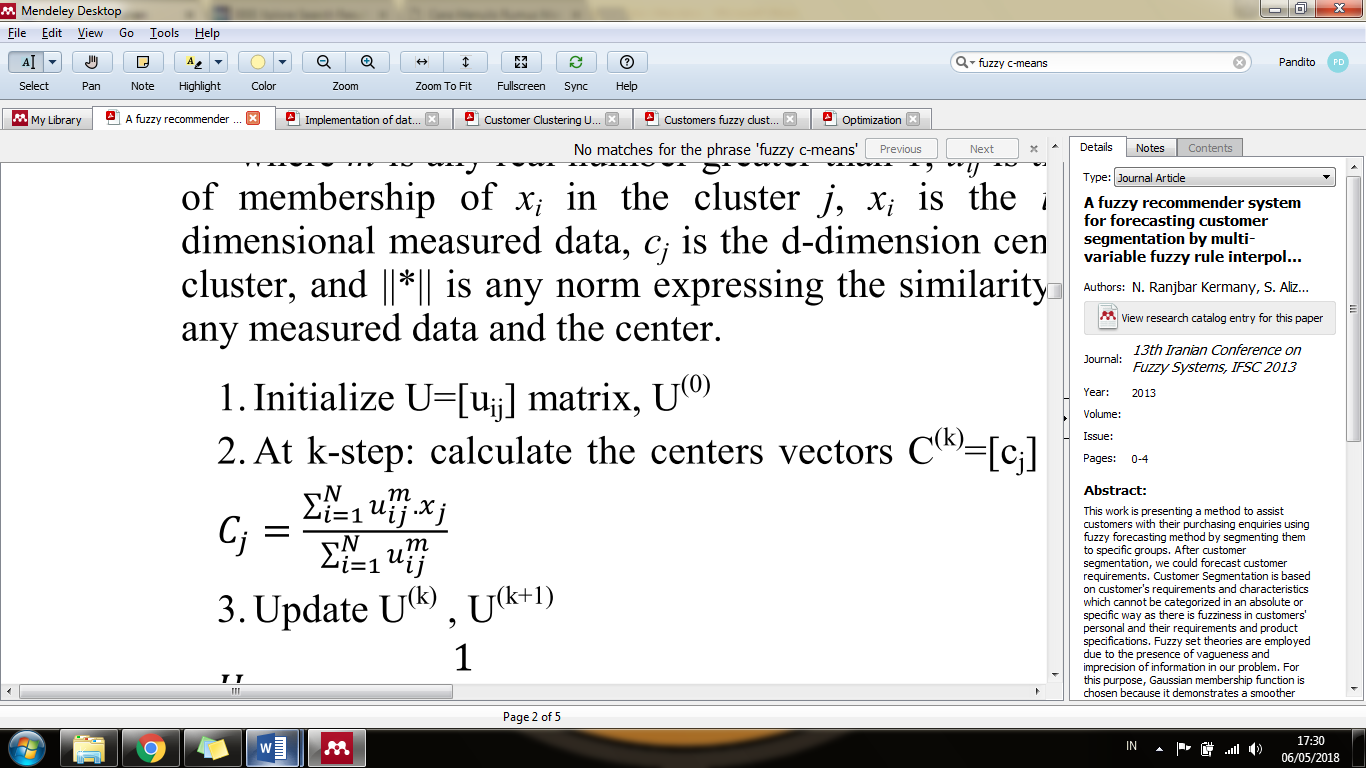
Fuzzy c-means (FCM) adalah sebuah metode pengelompokkan yang memungkinkan dari satu bagian data menjadi dua atau lebih kelompok. Metode dikembangkan Dunn pada tahun 1973 dan diperbaiki oleh Bezdek pada tahun 1981. Metode ini sering digunakan pengenalan pola (Mehdi, Kamran, Amir, dan Amir, 2009).

Dalam pemaparan, (Naime dan Sasan, 2013) Algoritma FCM *clutering* menggabungkan data *points* kedalam *cluster* berdasarkan pada minimisasi fungsi objek berikut :

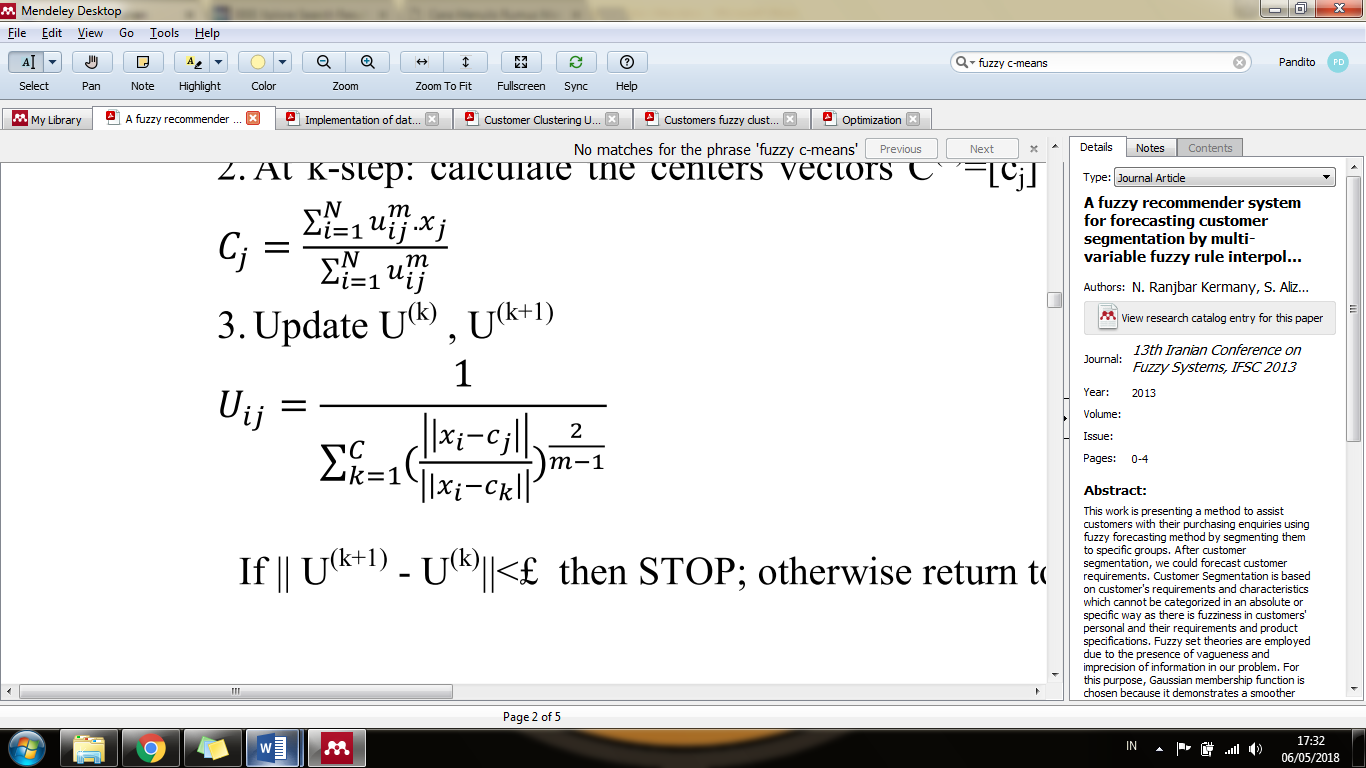


Dimana *m* adalah bilangan *real* yang lebih dari 1, adalah tingkat anggota dari dalam *cluster j*, adalah dari data yang diukur d-dimensional, adaah pusat d-dimension dari *cluster*, dan ||\*|| adalah banyak nya norma yang menyatakan kesamaan antara data terukur dan data pusat.

1. Inisialisasi U=[] matrix,
2. Pada k-step : menghitung vektor pusat =[] dengan



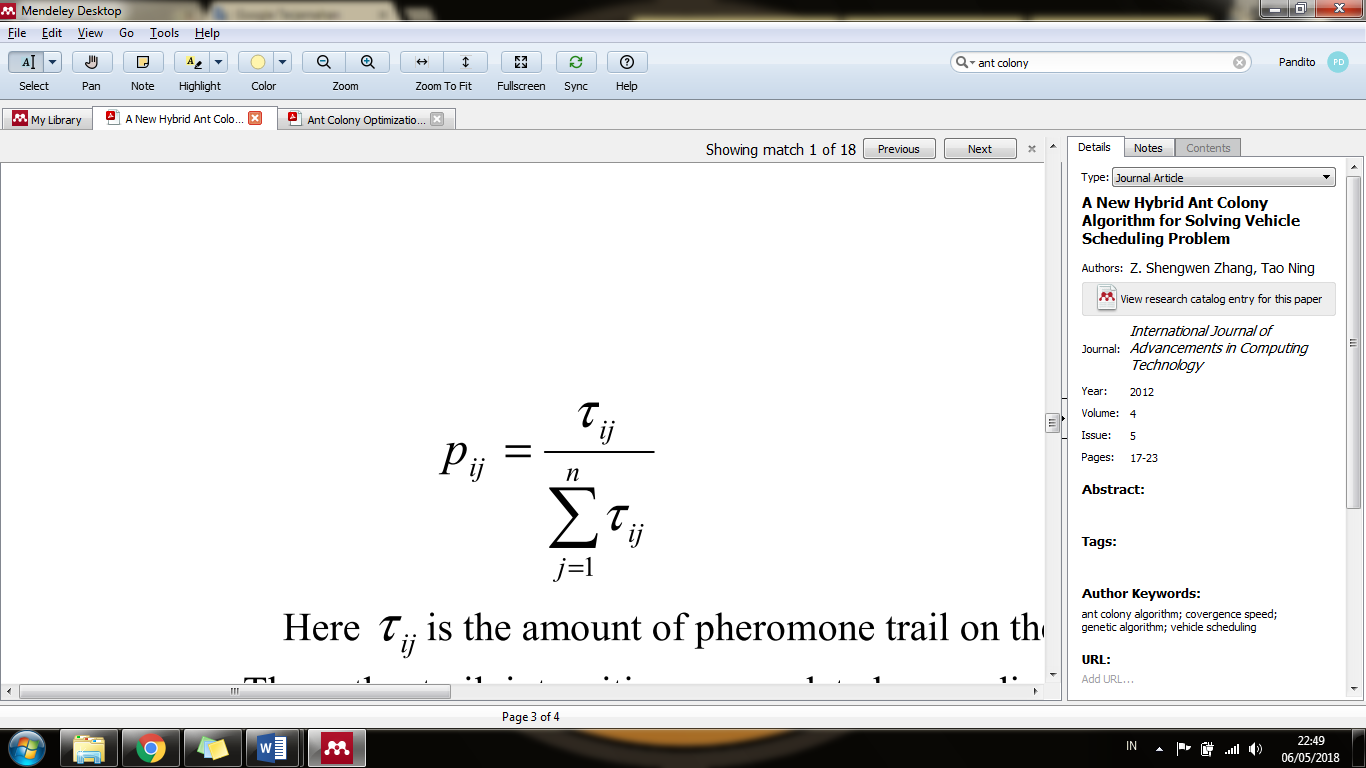
1. *Update* ,



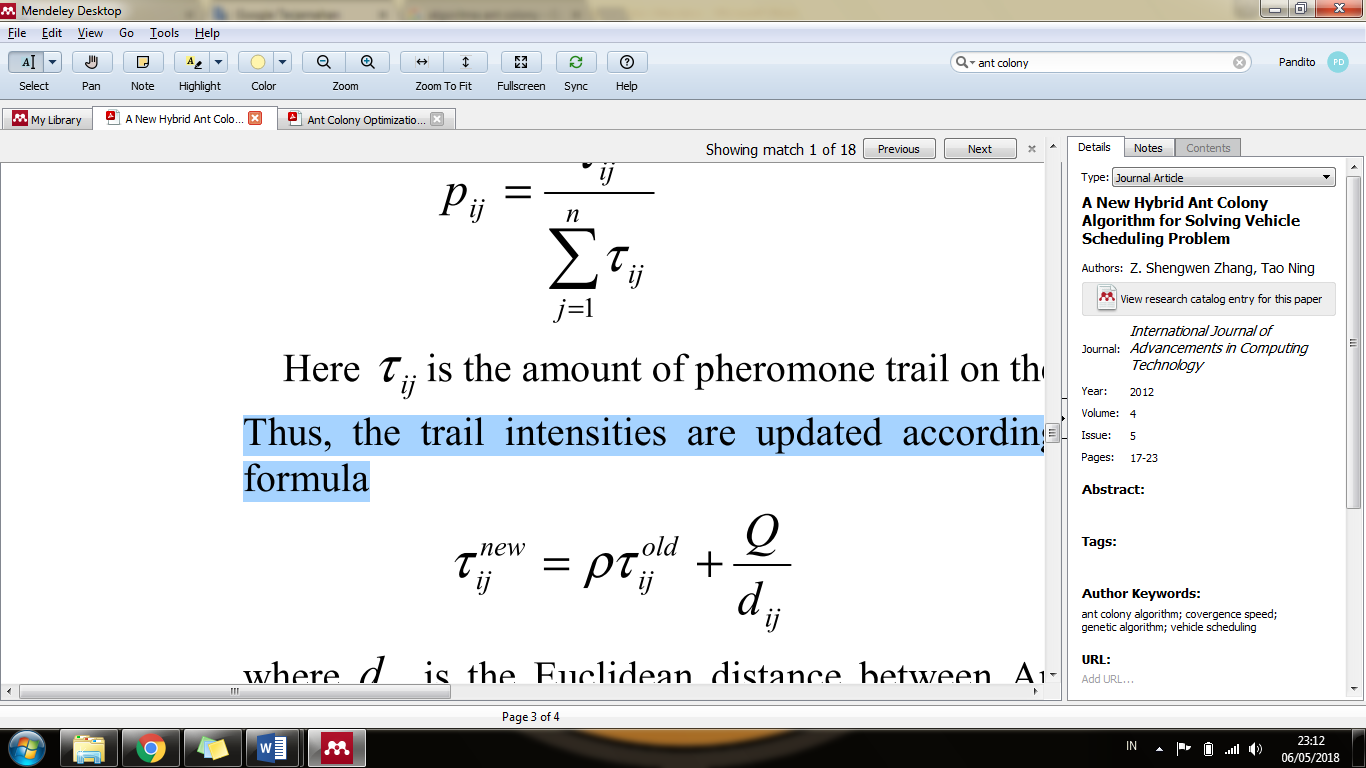
Jika || - || < £ maka STOP; jika tidak kembali ke langkah 2

**2.4 Ant Colony Optimization**

Penelitian yang dilakukan (Gao, Zhang, Zhang, dan Cao, 2008) terinspirasi oleh prilaku koloni semut secara nyata, Marco Dorigo pertama kali memperenalkan Ant System (AS) tahun 1992. Algoritma ant colony optimization (ACO) merupakan salah satu metode optimisasi meta-heuristik adaptif yang mencakup *simulated annealing* (SA)*, the genetic algorithms* (GA), *the evolutionary strategies* (ES), *dan Tabu Searcher* (TS). Setiap sampel memeliki seekor semut, jika mendistribusikan sampel *i* ke pusat *cluster* (*j* = 1,2,...K), *ant* akan meletakkan beberapa *pheromone* ditepinya. Kemungkinan transisi untuk *ant i* termasuk ke pusat *cluster*  adalah :



Disini adalah jumlah jejak *pheromone* ditepinya. Dengan demikian intensitas jejak di perbarui sesuai dengan rumus



dimana adalah jarak euclidean diantara *ant i*  dan pusat *cluster* , dengan 0 < < 1 adalah kegigihan dari jejak, demikian 1 - menunjukan *evaporation. Q* adalah sebuah konstan terkait dengan tingkat jejak yang ditaruh oleh semut. Rumus algoritma ant colony adalah :

1. Step pertama, Set *nc* = 0, dan menetapkan nilai yang sama ke .
2. Step kedua, memilih pusat *cluster* untuk berpindah dengan kemungkinan transisi .
3. Step ketiga, memperbarui nilai jejak dan kemungkinan transisi, serta set *nc* = *nc* + 1.
4. Step keempat, jika (*nc <* ) kemudian menuju ke step 3, jika tidak cetak pemilihan komponen dan stop.

# **BAB III METODE PENELITIAN**

## **3.1 Analisis Metodologi**

Eksperimen yang di lakukan mengacu pada pengelompokan data – data pelanggan (data akun pelanggan dan data transaksi pelanggan) dimana teknik pengelompokan tersebut sering disebut sebagai *clustering*. Data yang akan digunakan dalam penelitian ini yaitu menggunakan Telco Customer Churn dataset.

Dataset Telco Customer Churn berisi informasi tentang :

1. Pelanggan yang tersisa sebulan terakhir
2. Pelayanan yang telah disepakati pelanggan
3. Informasi tentang akun pelanggan
4. Informasi demografis tentang pelanggan

Jumlah total atribut dari dataset Telco Customer Churn ada 21 atribut (customerID, gender, Senior\_Citizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, TotalCharges, Churn). Kemudian melakukan normalisasi data untuk membersihkan data – data yang mengandung noise, memiliki nilai kosong, tidak konsisten, dan tidak valid.

Proses transformasi data adalah untuk menyajikan data – data yang siap digunakan sebelum dilakukan proses *clustering*. Dalam transformasi data menggunakan analisis model RFM (Recency, Frequency, Monetary), Maka akan dilakukan proses preprocessing dengan menyeleksi atribut data dari 21 atribut menjadi 7 atribut (customerID, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges, TotalCharges, Churn).

Kemudian dilakukan proses hybrid algoritma fuzzy c-means dengan algoritma ant colony optimization, sehingga *formula* yang didapatkan dari hasil hybrid algoritma tersebut dapat digunakan pada proses *clustering*. *Tools* yang digunakan untuk proses *clustering* yaitu Python.

Selanjutnya melakukan validasi dari hasil proses *clustering* yang telah dilakukan dengan menggunakan purity test, tujuan nya untuk mengetahui *cluster* yang baik ataupun buruk dari hasil proses *clustering*. Terakhir membuat kesimpulan hasil penelitian. Gambar 3.3 diagram alir dari kerangka kerja penelitian

Mulai

Telco Customer Churn Dataset

Normalisasi Data

Transformasi Data

Hybrid

Fuzzy c-means

Ant Colony Optimization

Purity Test

Kesimpulan

Selesai

Data

Preprocessing

Clustering

Validasi

Evaluasi

**Gambar 3.1** Kerangka Kerja Penelitian

## **3.2 Jadwal Penelitian**

Penelitian dilakukan mulai semester 3 tahun ajaran 2018/2019 dengan penjadwalan sebagai berikut :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| No | Daftar Kegiatan | Tahun 2018 | | | | |
| Agu | Sep | Okt | Nov | Des |
| 1 | Kajian Pustaka |  |  |  |  |  |
| 2 | Pengumpulan dan Pengolahan Data |  |  |  |  |  |
| 3 | Perancangan Perangkat Lunak |  |  |  |  |  |
| 4 | Implementasi |  |  |  |  |  |
| 5 | Testing dan Hasil |  |  |  |  |  |
| 6 | Dokumentasi dan Laporan Akhir |  |  |  |  |  |

**Tabel 2.1** Jadwal Penelitian

**DAFTAR PUSTAKA**

Maghfirah and Te. B. Adji, “Appropriate Data mining Technique and Algorithm for Using in Analysis of Customer Relationship Management ( CRM ) in Bank Industry,” in Seminar Nasional Aplikasi Teknologi Informasi (SNATi), 2015, pp. 7–10.

N. P. P. Yuliari, I. K. G. D. Putra, and N. K. D. Rusjayanti, “Customer Segmentation Through Fuzzy C-Means and Fuzzy RFM Method,” J. Theor. Appl. Inf. Technol., vol. 78, no. 3, pp. 380–385, 2015.

Lisna Zahrotun, “Implementation of Data Mining Technique for Customer Relationship Management (CRM) on Online Shop Tokodiapers.com With Fuzzy C-Means Clustering,” in Conferences Information Systems and Electrical Engineering (ICITISEE), 2017.

Y.-H. Hu and T.-W. Yeh, "Discovering valuable frequent patterns based on RFM analysis without," Knowledge-Based Systems, p. 76–88, 2014.

Sudriyanto, “Clustering Loyalitas Pelanggan Dengan Metode Rfm (Recency, Frequency, Monetary ) Dan Fuzzy C-Means, ”in Universitas Muria Kudus, 2017.

C. Chiu, H. Ku, I. Kuo, and P. Shih, “Customer information system using fuzzy query and cluster analysis,” J. Ind. Prod. Eng., vol. 31, no. 3, pp. 134– 145, 2014.

Hesam Izakian, Ajith Abraham, and Václav Snášel, “Fuzzy Clustering Using Hybrid Fuzzy c-means and Fuzzy Particle Swarm Optimization,” Nature & Biologically Inspired Computing, 2009.

K.Sankar and Dr. K.Krishnamoorthy, “Cluster Mining with Ant Colony Optimizer using Fuzzy inferences,” Second International conference on Computing, 2010.

Saptarshi Sengupta, Sanchita Basak, and Richard Alan Peters II, “Data Clustering using a Hybrid of Fuzzy C-Means and Quantum-behaved Particle Swarm Optimization,” Computing and Communication Workshop and Conference (CCWC), 2018.

Hao Wang, Yan Zhang, Danyun Li, “Network Intrusion Detection Based on Hybrid Fuzzy C-Mean Clustering,” Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010.

Gao Shang, “A New Hybrid Ant Colony Algorithm for Clustering Problem,” International Symposium on Intelligent Information Technology Application Workshops, 2008.

YiJun Chen, Man-Leung Wong, Haibing Li, “Applying Ant Colony Optimization to configuring stacking ensembles for data mining,” in Lingnan University, 2014.

Ali Hajiha, Reza Radfar, Samira Sarafi Malayeri, ” Data Mining Application for Customer Segmentation Based on Loyalty: An Iranian Food Industry Case Study,” Industrial Engineering and Engineering Management (IEEM), 2011

Mehdi Fathi, Kamran Kianfar, Amir Hasanzadeh, Amir Sadeghi, “Customers Fuzzy Clustering and Catalog Segmentation In Customer Relationship Management,” International Conference on Industrial Engineering and Engineering Management, 2009.

Naime Ranjbar Kermany dan Sasan H. Alizadeh, “A fuzzy recommender system for forecasting customer segmentation by multi-variable fuzzy rule interpolation,” Iranian Conference on Fuzzy Systems (IFSC), 2013.

Gao Shang, Zhang Zaiyue, Zhang Xiaoru, Cao Cungen, “A New Hybrid Ant Colony Algorithm for Clustering Problem,” International Workshop on Education Technology and Training, 2008.

Telco Customer Churn dataset (2015). Customer dataset. Retrieved from https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/