

Implementasi Data Mining dengan Algoritma Fuzzy C-Means Studi Kasus Penjualan di UD Subur Baru

Alvian Kusuma Wijaya¹

^{1,3} Jurusan Teknik Informatika, FASILKOM UDINUS
Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA

¹111201005192@mhs.dinus.ac.id

Abstrak - Perusahaan retail tiap hari mencatat data transaksi yang cukup banyak. Hal ini menimbulkan tumpukan data yang dapat diolah lebih lanjut agar menghasilkan informasi yang bermanfaat. Salah satu cara pengolahan tersebut adalah dengan datamining. Pada penelitian ini membahas bagaimana teknik datamining diimplementasikan pada perusahaan retail, yaitu UD Subur Baru yang bergerak dibidang perdagangan furniture untuk memperoleh informasi mengenai tingkat penjualan produk. Untuk mengimplementasikan datamining pada UD Subur Baru, maka dirancang sistem aplikasi berbasis java yang dapat mengolah data transaksi penjualan seperti nota dan buku laku menjadi dataset baru dengan variabel jumlah invoice, jumlah terjual dan kode barang dengan menggunakan metode clustering fuzzy c-means. Data yang digunakan adalah data transaksi selama bulan Januari sampai Maret 2013 dan 2014. Pengujian validitas cluster dilakukan dengan menggunakan algoritma modified partition coefficient (MPC). Hasil proses mining membuktikan bahwa pada periode 1 tahun tertentu mempunyai kemiripan produk potensial yang sama, sementara pada cluster lain nilainya bervariasi. Untuk hasil pengujian validitas, jumlah cluster 2 mempunyai nilai lebih tinggi dari jumlah cluster yang lain.

Kata kunci: *retail, datamining, fuzzy c-means, modified partition coefficient*

Abstract - Each day, Retail company recorded data transaction quite a lot. This has led to a pile of data that can be further processed to produce useful information. One way is by datamining processing. In this study discusses how data mining techniques are implemented in a retail company, namely UD Subur Baru engaged in the furniture trade to obtain information on the level of sales products. For implement datamining on UD Subur Baru, then the system is designed based java application that can process data such as sales transactions memorandum and sell books into a new dataset with a variable amount of the invoice, the amount of goods sold and code by using fuzzy c-means clustering. The data used is the transaction data for January to March 2013 and 2014 cluster validity test performed using the modified algorithm partition coefficient (MPC). The results prove that the mining process in a certain period of 1 year have the same potential product similarity, while the other clusters vary in value. To test the validity of the results, 2 cluster has a higher value than the other clusters.

Keywords— *retail, datamining, fuzzy c-means, modified partition coefficient*

I. PENDAHULUAN

(Perusahaan ritel setiap harinya mencatat transaksi penjualan yang sangat banyak. Hal ini berdampak pada pertumbuhan jumlah data yang sangat pesat dan menimbulkan tumpukan data yang berjumlah besar dalam basis data. Jika data sedikit, masih memungkinkan digunakannya kemampuan manusia untuk melakukan analisa dan interpretasi data menggunakan metode tradisional [1].

Dari data-data yang ada pada sebuah perusahaan ritel dapat digali informasi-informasi baru yang berguna untuk menunjang proses bisnis perusahaan. Data-data tersebut digali dengan metode yang disebut dengan *datamining* [1]. *Datamining* merupakan suatu proses pengekstrakan informasi baru yang diambil dari bongkahan data besar

yang membantu dalam pengambilan keputusan [2]. UD Subur Baru adalah perusahaan ritel yang bergerak dibidang *furniture* yang berlokasi di Jalan Bakti no 84 Kudus. Sampai saat ini, semua pencatatan penjualan dilakukan dengan cara tradisional (tanpa komputer). Penggunaan metode tradisional untuk memelakukan identifikasi produk dari data membutuhkan kemampuan manusia untuk menganalisa dan menginterpretasikan data. Dengan volume data yang berkembang sangat pesat, analisa terhadap data tidak mungkin dilakukan secara manual oleh manusia [1]. Penerapan *datamining* dapat membantu untuk menganalisa data yang diperoleh dari transaksi penjualan sehingga dapat menggali pola-pola yang dapat dijadikan pengetahuan baru untuk proses identifikasi produk di UD Subur Baru. *Datamining* pada kasus penjualan di UD Subur Baru diharapkan dapat memudahkan perusahaan dalam mengidentifikasi produk-produk terbaik mereka.

Oleh karena itu, penulis membuat sebuah sistem aplikasi *datamining* untuk membantu proses analisa data yang diperoleh dari *record-record* transaksi pada UD Subur Baru. Pada penelitian ini digunakan metode *fuzzy cluster*, yaitu dengan algoritma *fuzzy c-means* (FCM). Algoritma ini dipilih karena dengan metode ini, data-data beserta parameter-parameternya dapat dikelompokkan dalam *cluster-cluster* sesuai dengan kecenderungannya. Selain itu, dengan metode ini bisa ditentukan jumlah *cluster* yang akan dibentuk. Dengan penentuan jumlah *cluster* diawal, bisa diatur keragaman nilai akhir sesuai dengan *cluster*-nya. Kelebihan algoritma ini adalah penempatan pusat *cluster* yang lebih tepat dibandingkan dengan metode *cluster* lain. Caranya adalah dengan memperbaiki pusat *cluster* secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat. FCM juga memiliki tingkat akurasi yang tinggi dan waktu komputasi yang cepat [3]. Dalam penelitian sebelumnya, algoritma *fuzzy c-means* dibandingkan dengan algoritma *fuzzy k-means*. Hasilnya adalah algoritma *k-means* membagi data menjadi beberapa kelompok berdasarkan data yang ada, dimana data dalam satu kelompok memiliki karakteristik yang sama dengan lainnya dan memiliki karakteristik yang berbeda dengan data yang ada pada kelompok yang lain. Secara umum, perbedaan mendasar antara *k-means* dengan *c-means* terletak pada data yang di *cluster*. Pada *k-means*, data ter-*cluster* secara tegas, sedangkan pada *c-means* data ter-*cluster* pada semua *cluster* dengan derajat keanggotaan yang berbeda. Dibandingkan dengan *k-means*, algoritma *c-means* lebih unggul dalam mengatasi data *outlier*. Pada *c-means*, data *outlier* bergabung menjadi satu *cluster* dengan pusat *cluster*, sedangkan pada *k-means*, data *outlier* cenderung membentuk *cluster* tersendiri. Hal ini dikarenakan *c-means* memiliki derajat keanggotaan yang berguna untuk mengelompokkan data kedalam *cluster* yang semestinya [1]. Dengan algoritma ini akan dilakukan penggalian informasi mengenai tingkat penjualan produk pada UD Subur Baru. Proses penggalian data ini menggunakan *tools* yang dibuat sendiri dengan menggunakan bahasa pemrograman Java.

II. TUJUAN

Mengimplementasikan teknik *data mining clustering* pada data transaksi UD Subur Baru untuk mengidentifikasi tingkat penjualan produk yang dibagi menjadi 3 katagori, yaitu sangat laku, laku dan kurang laku dengan menggunakan algoritma *fuzzy c-means*.

III. METODE YANG DIUSULKAN

A. Sumber Data

Sumber data dari penelitian ini diambil dari data transaksi UD Subur Baru Kudus, yang beralamat di Jalan Bakti 84 Kudus. Waktu pengumpulan data pada bulan Mei-Juni 2014. Metode pengumpulan data menggunakan metode observasi dan jenis data adalah data sekunder, yaitu data yang diambil langsung dari nota transaksi penjualan

dan buku laku pada UD Subur Baru mulai dari bulan Januari sampai Maret tahun 2013 dan 2014. Data-data ini masih berupa tumpukan buku dan nota sehingga memerlukan proses lebih lanjut untuk mendapatkan jumlah data dan memprosesnya.

B. Variabel Penelitian

Variabel yang digunakan untuk melakukan proses *mining* pada penelitian ini adalah :

1. Kode barang.
2. Jumlah barang terjual dalam periode bulan tertentu.
3. Jumlah invoice atau jumlah nota atau faktur yang mencatat penjualan pada barang tersebut untuk periode bulan tertentu.

C. Metode Perancangan Data

Rancangan tahapan penelitian mengacu pada enam tahap CRISP-DM, atau siklus hidup pengembangan *data mining* sebagai *framework* dari proyek *data mining*. [4]

1. Business Understanding

Dalam aplikasinya pada UD Subur Baru, perlu didahului dengan pemahaman terhadap tujuan bisnis yang akan menentukan pola yang akan dicari dalam proses *data mining*. Beberapa tujuan bisnis yang berkaitan dengan penelitian ini adalah :

- Mengetahui tingkat penjualan produk pada UD Subur Baru.
- Meminimalkan pembelian produk yang kurang laku.
- Memaksimalkan pembelian produk yang sering laku.

Untuk mencapai tujuan tersebut maka perlu dilakukan proses *clustering* data, dimana data penjualan akan dikelompokkan menjadi tiga kelompok, yaitu : sangat laku, laku, kurang laku, yang akan berguna untuk pemilihan produk-produk potensial.

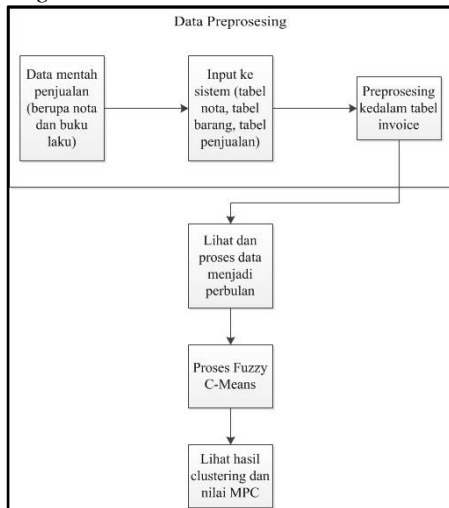
2. Data Understanding

Pada tahap ini akan dilakukan pemahaman terhadap kebutuhan data terkait dengan pencapaian tujuan dalam penelitian ini. Proses yang dilakukan adalah memahami tentang data dan mengambil beberapa data yang diperlukan dalam penelitian ini. Data yang akan digunakan sudah dijelaskan pada bab sebelumnya yaitu pada pengumpulan data.

3. Data Preparation

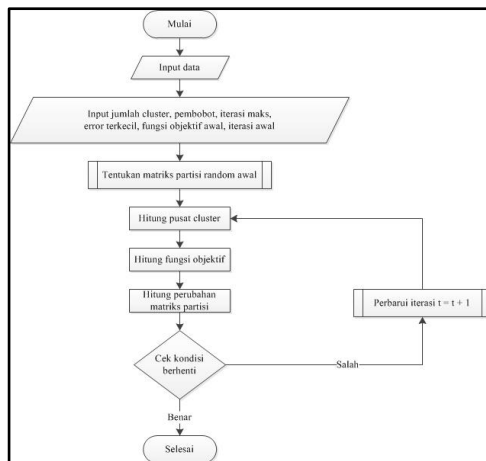
Setelah dilakukan pengumpulan dan input data, kemudian data diolah dan disatukan kedalam satu tabel, yaitu tabel *invoice*. Atribut dari tabel *invoice* lah yang nantinya akan digunakan untuk proses *clustering* dengan menggunakan algoritma *fuzzy c-means*. Atribut yang digunakan adalah atribut *jmlh_invoice* dan *jmlh_barang*.

4. Modeling



Gambar 3.1 Diagram proses data

Dataset yang telah dinormalisasi dan ditransformasi dari tahap sebelumnya, selanjutnya dimodelkan menggunakan metode *fuzzy c-means* untuk melakukan proses *clustering* pada data. *Fuzzy c-means* merupakan metode *clustering* yang terawasi (*supervised*). Pada FCM jumlah *cluster* ditentukan.



Gambar 3.2 Flowchart Algoritma *Fuzzy C-means*

5. Evaluation

Pada tahap ini, hasil *clustering* diuji tingkat validitasnya dengan menggunakan metode pengukuran *Modified Partition Coefficient* untuk menentukan jumlah *cluster* yang paling optimal. Hasil segmentasi *cluster* yang diperoleh dari hasil *clustering* algoritma *Fuzzy C-means* dibandingkan dengan *dataset* periode yang lain untuk mengukur kinerja proses *mining*.

6. Deployment

Pada tahap ini akan dilakukan pembuatan aplikasi *data mining* yang dapat melakukan proses mulai dari input data sampai menampilkan hasil *clustering*.

D. Metode Analisa Data

Pada penelitian ini, metode analisis data yang digunakan adalah *fuzzy c-means* (FCM). FCM adalah suatu teknik pengclusteringan data yang keberadaan tiap-tiap titik data suatu *cluster* ditentukan oleh nilai keanggotaan. Nilai keanggotaan tersebut akan mencakup bilangan real pada interval 0-1 [4].

FCM adalah salah satu metode *optimizing partitioned cluster*. Kelebihan metode FCM adalah penempatan pusat *cluster* yang lebih tepat dibandingkan dengan metode *cluster* lain. Caranya adalah dengan memperbaiki pusat *cluster* secara berulang, maka akan dapat dilihat bahwa pusat *cluster* akan bergerak menuju lokasi yang tepat [4].

Algoritma dari *fuzzy c-means* adalah sebagai berikut (Yan, 1994) [5]:

1. Input data yang akan dikelompokkan, yaitu X berupa matriks berukuran n x m (n = jumlah sampel data, m = atribut setiap data). X_{ij} data sampel ke-i ($i=1,2,\dots,n$), atribut ke-j ($j=1,2,\dots,m$).
2. Tentukan jumlah *cluster* (c), pangkat untuk matriks partisi (w), maksimum iterasi (MaxIter), error terkecil yang diharapkan (ξ), fungsi objektif awal ($P_0=0$), dan iterasi awal ($t=1$).
3. Bangkitkan bilangan random μ_{ik} , $i=1,2,\dots,n$; $k=1,2,\dots,c$ sebagai elemen matriks partisi awal U.
4. Hitung pusat *cluster* ke-k: V_{kj} , dengan $k=1,2,\dots,c$; dan $j=1,2,\dots,m$, menggunakan persamaan berikut (Yan, 1994) :

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \cdot X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (1)$$

dengan :

V_{kj} = pusat *cluster* ke-k untuk atribut ke-j

μ_{ik} = derajat keanggotaan untuk data sampel ke-i pada *cluster* ke-k

X_{ij} = data ke-i, atribut ke-j

5. Hitung fungsi objektif pada iterasi ke-t menggunakan persamaan berikut (Yan, 1994) :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (2)$$

dengan :

V_{kj} = pusat *cluster* ke-k untuk atribut ke-j

μ_{ik} = derajat keanggotaan untuk data sampel ke-i pada *cluster* ke-k

X_{ij} = data ke-i, atribut ke-j

P_t = fungsi objektif pada iterasi ke-t

6. Hitung perubahan matriks partisi menggunakan persamaan berikut (Yan, 1994) :

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{1}{w-1}}} \quad (3)$$

dengan $i=1,2,\dots,n$; dan $k=1,2,\dots,c$

Dimana :

V_{kj} = pusat *cluster* ke-k untuk atribut ke-j

X_{ij} = data ke-i, atribut ke-j

μ_{ik} = derajat keanggotaan untuk data sampel ke-i pada *cluster* ke-k

7. Cek kondisi berhenti :

Jika : $(|P_t - P_{t-1}| < \xi)$ atau $(t > MaxIter)$ maka berhenti. Jika tidak : $t = t + 1$, ulangi langkah ke-4.

E. Metode Pengujian

Untuk pengujian hasil *cluster*, digunakan metode *modified partition coefficient* (MPC). MPC merupakan metode yang digunakan untuk menguji validitas jumlah *cluster*. MPC sendiri merupakan pengembangan dari metode *partition coefficient* (PC). *Partition coefficient* (PC) merupakan metode yang mengukur jumlah *cluster* yang mengalami overlap. Nilai PC berada dalam batas $\frac{1}{c} \leq PC(c) \leq 1$. Pada umumnya jumlah *cluster* yang paling optimal ditentukan dari nilai PC yang paling besar ($\max_{2 \leq c \leq n-1} PC(c)$). Berikut adalah algoritma metode PC : [5]

$$PC(c) = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N (\mu_{ij}^2) \quad (3)$$

Dimana:

c = jumlah *cluster*

N = jumlah data

μ_{ij} = derajat keanggotaan data k - j pada *cluster* ke- i

$PC(c)$ = nilai indeks PC pada *cluster* ke- c

Partition coefficient cenderung mengalami perubahan yang monoton terhadap beragam nilai c (jumlah *cluster*). Modifikasi dari indeks PC (*Modified Partition Coefficient* / MPC) dapat mengurangi perubahan yang monoton tersebut. Nilai MPC berada dalam batas $0 \leq PC(c) \leq 1$. Pada umumnya jumlah *cluster* yang paling optimal ditentukan dari nilai MPC yang paling besar ($\max_{2 \leq c \leq n-1} PC(c)$). Berikut adalah algoritma metode MPC : [5]

$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC(c)) \quad (5)$$

Dimana:

c = jumlah *cluster*

$MPC(c)$ = nilai indeks MPC pada *cluster* ke- c

F. Metode Perancangan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan dalam penelitian ini adalah metode pengembangan perangkat lunak secara prototipe. Prototipe merupakan suatu pendekatan membuat model yang memperlihatkan fitur-fitur produk, layanan, atau sistem usulan. Modelnya dikenal dengan prototipe. Langkah-langkahnya adalah sebagai berikut:

1. Mengidentifikasi kebutuhan-kebutuhan dasar perangkat lunak. Dalam hal ini perancang sistem bekerja dengan pemakai (*user*) untuk menangkap dasar yang diperlukan pemakai (*user*).
2. Membuat prototipe dari identifikasi kebutuhan-kebutuhan dasar perangkat lunak sesuai dengan apa yang *user* inginkan.
3. Menggunakan prototipe, pada tahap ini *user* diminta untuk bekerja dengan sistem untuk menentukan cocok tidaknya prototipe terhadap kebutuhan *user* dan

diharapkan *user* memberi masukan untuk memperbaiki prototipe.

4. Memperbaiki dan meningkatkan prototipe. Prototipe diperbaiki sesuai dengan semua perubahan yang diminta oleh *user*.

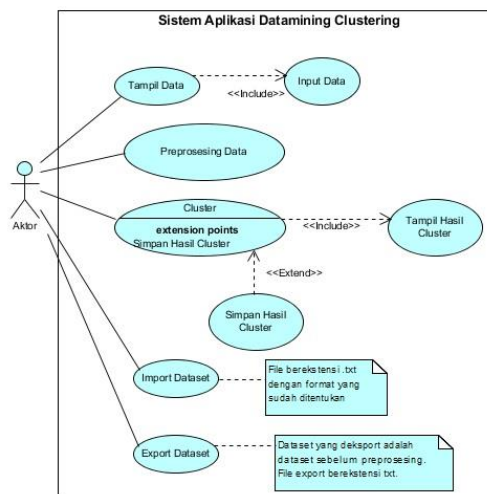
Setelah itu langkah tiga dan empat dilakukan secara terus menerus sampai *user* merasa cocok dan puas.



Gambar 3.3 Penerapan Model Prototipe

IV. IMPLEMENTASI

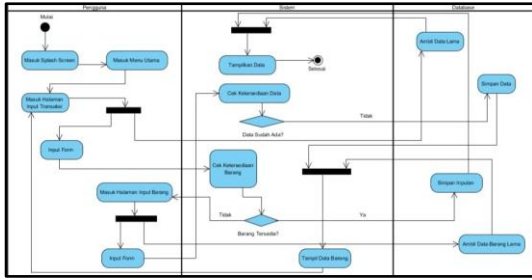
A. Use Case Diagram



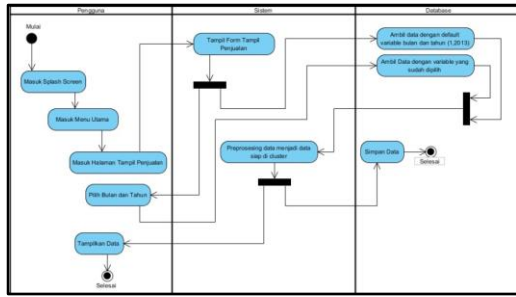
Gambar 4.1 Use Case Diagram

Diagram *use case* diatas menggambarkan aktivitas yang dapat dilakukan pengguna pada sistem. Pengguna dapat melakukan input data, menampilkan data yang sudah tersimpan, preprosesing data, proses *clustering*, melakukan *import dataset* sebagai alternatif inputan data berupa file berekstensi txt dan *export data* mentah kedalam file berekstensi txt sebagai media *backup data*. Untuk dapat menampilkan hasil *cluster* dan menyimpannya, pengguna diwajibkan melakukan proses *cluster* pada data yang sudah dipilih berdasarkan bulan dan tahun sebelumnya. Setelah pengguna melakukan proses *clustering*, hasil dari *clustering* akan tampil secara otomatis, sedangkan untuk fungsi simpan hasil adalah optional, artinya pengguna dapat melakukan simpan ataupun tidak sesuai dengan keinginan.

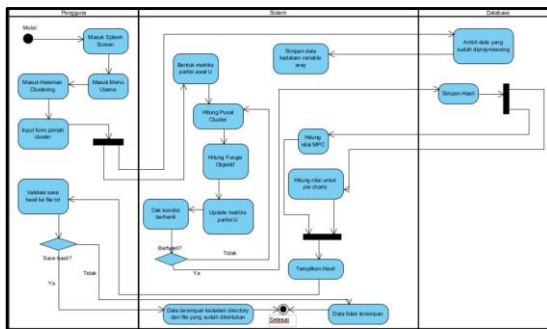
B. Activity Diagram



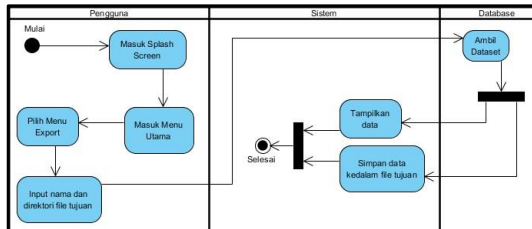
Gambar 4.2 Activity Diagram Tampil Data



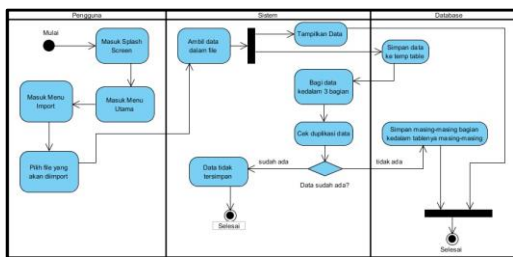
Gambar 4.3 Activity Diagram Preprocessing Data



Gambar 4.4 Activity Diagram Clustering



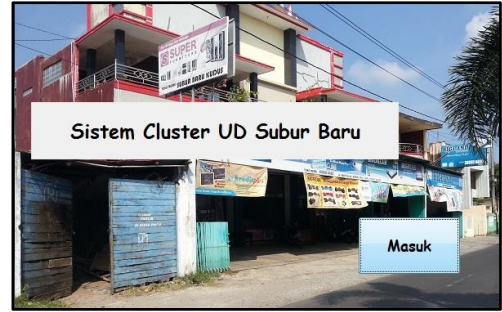
Gambar 4.5 Activity Export Data



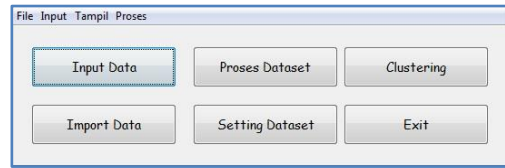
Gambar 4.6 Activity Import Data

C. Desain Antarmuka

Berikut ini adalah beberapa tampilan dari desain antarmuka pada sistem *datamining* ini :



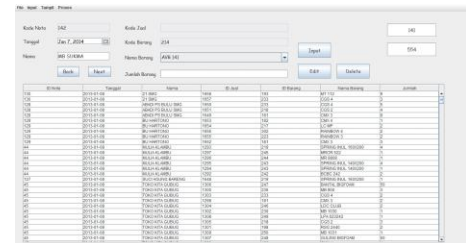
Gambar 4.7 Splash Screen



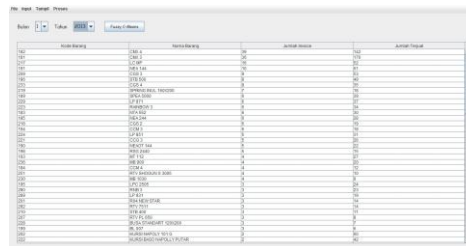
Gambar 4.8 Menu Utama



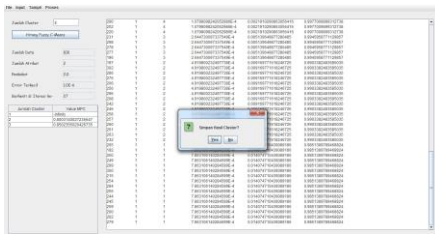
Gambar 4.9 Halaman Input Barang



Gambar 4.10 Halaman Input Transaksi



Gambar 4.11 Halaman Preprocessing Data



Gambar 4.12 Halaman Proses Cluster



Gambar 4.13 Halaman Backup dan Import Data

V. HASIL & PEMBAHASAN

A. Pembahasan

Untuk mengetahui tingkat keberhasilan pada penelitian ini dapat dilihat dari hasil kerja sistem yang dibuat. Pada penelitian ini, proses *mining* dilakukan dengan menggunakan *dataset* awal, yaitu dataset untuk periode Januari sampai Maret 2013 dan 2014. Hasil *mining* akan dibandingkan dengan hasil *mining* dari *dataset* yang lain untuk memperoleh kesimpulan tentang produk potensial apa yang dimiliki UD Subur Baru pada satu periode bulan tertentu. Untuk menguji validitas *cluster*, digunakan metode *Modified Partition Coefficient* (MPC).

B. Evaluasi Hasil Data Mining Menggunakan Dataset UD Subur Baru

Proses *clustering* menggunakan algoritma FCM diuji coba dengan enam periode *dataset*. Proses *clustering* dilakukan dengan menetapkan nilai awal sebagai berikut:

1. Jumlah *cluster* yang akan dibentuk = 3.
2. Pangkat pembobot = $w = 2$
3. Max iterasi = 100
4. Kriteria penghentian = $\xi = 10^{-6}$
5. Fungsi objektif awal = 0
6. Nilai iterasi awal = 1

Untuk hasil *clustering* dan nilai dari indeks MPC dari masing-masing periode akan dijelaskan pada tabel berikut (sebagai catatan, nilai fungsi objektif dibulatkan tiga angka dibelakang koma):

Tabel 5.1 Hasil Clustering

Bulan	Berhenti diiterasi ke-	Fungsi Objektif
Januari 13	26	5409,68
Febuari 13	12	1120,98
Maret 13	16	769,79
Januari 14	20	578,48
Febuari 14	99	800,83
Maret 14	99	3235,68

Nilai-nilai pada tabel diatas tidak akan sama bila dilakukan proses *cluster* ulang. Hal ini disebabkan oleh nilai matriks partisi U awal yang dibangkitkan secara *random*. Namun perubahan dari nilai-nilai tersebut tidak terlalu signifikan dan tidak mempengaruhi keanggotaan *cluster*.

Untuk detail dari produk potensial perbulan akan dijabarkan pada gambar berikut ini:



Gambar 5.1 Pie Charts Januari 2013

Pada periode bulan Januari 2013, produk yang masuk kedalam katagori cluster sangat laku adalah : 181, 182. Untuk katagori cluster laku : 217, 191, 209, 186, 233, 189, 229, 223, 183, 185, 224, 221, 193, 195, 280, 345, 306, 305, 248, 247. Untuk katagori cluster kurang laku : 219, 218, 194, 190, 198, 236, 184, 251, 230, 289, 282, 201, 210, 207, 228, 199, 281, 205, 283, 234, 297, 295, 284, 288, 238, 227, 225, 212, 253, 204, 214, 213, 287, 203, 188, 292, 301, 226, 285, 239, 211, 291, 293, 243, 216, 264, 220, 290, 252, 278, 277, 377, 231, 196, 246, 286, 206, 296, 261, 232, 242, 208, 302, 263, 260, 276, 197, 258, 257, 256, 299, 300, 192, 200, 294, 215, 262, 265, 259, 255, 254, 250, 249, 245, 244, 279.



Gambar 5.2 Pie Charts Febuari 2013

Pada periode bulan Febuari 2013, produk yang masuk kedalam katagori cluster sangat laku adalah : 181, 182. Untuk katagori cluster laku : 223, 217, 189, 186, 191, 183, 226. Untuk katagori cluster kurang laku : 242, 225, 233, 230, 185, 218, 219, 238, 190, 237, 198, 220, 209, 221, 236, 272, 224, 228, 229, 199, 232, 253, 275, 234, 210, 246, 270, 193, 251, 274, 201, 266, 263, 267, 269, 227, 239, 231, 240, 268, 271, 273, 235, 196, 241, 250.



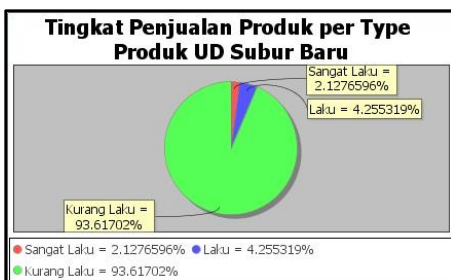
Gambar 5.3 Pie Charts Maret 2013

Pada periode bulan Maret 2013, produk yang masuk kedalam katagori cluster sangat laku adalah : 181, 323. Untuk katagori cluster laku : 219, 315, 182, 317, 306, 307. Untuk katagori cluster kurang laku : 231, 305, 345, 364, 319, 311, 309, 324, 326, 328, 191, 233, 201, 220, 213, 341, 325, 343, 313, 356, 353, 327, 186, 329, 183, 335, 336, 337, 344, 339, 190, 322, 259, 237, 308, 310, 312, 209, 314, 208, 316, 204, 320, 321, 253.



Gambar 5.4 Pie Charts Januari 2014

Pada periode bulan Januari 2014, produk yang masuk kedalam katagori cluster sangat laku adalah : 345. Untuk katagori cluster laku : 219, 181, 305. Untuk katagori cluster kurang laku : 182, 190, 315, 349, 311, 306, 343, 355, 231, 339, 317, 321, 189, 304, 364, 255, 253, 356, 220, 319, 351, 201, 213, 362, 367, 352, 353, 354, 368, 377, 366, 365, 361, 363, 350, 214, 242, 243, 252, 303, 309, 316, 210, 326, 346, 347, 348, 191.



Gambar 5.5 Pie Charts Februari 2014

Pada periode bulan Februari 2014, produk yang masuk kedalam katagori cluster sangat laku adalah : 345. Untuk katagori cluster laku : 306, 307, 378. Untuk katagori cluster kurang laku : 219, 343, 311, 305, 367, 326, 181, 316, 368, 376, 317, 315, 355, 381, 371, 353, 309, 218, 233, 206, 190, 182, 351, 380, 375, 379, 374, 382, 377, 372, 373, 369, 254, 329, 336, 253, 243, 234, 349, 337, 362, 209, 384.



Gambar 5.6 Pie Charts Maret 2014

Pada periode bulan Maret 2014, produk yang masuk kedalam katagori cluster sangat laku adalah : 345. Untuk katagori cluster laku : 219, 306, 182, 371, 305. Untuk katagori cluster kurang laku : 181, 378, 186, 190, 319, 253, 233, 311, 217, 321, 343, 393, 382, 372, 353, 351, 367, 361, 364, 213, 209, 260, 270, 204, 390, 307, 394, 354, 349, 304, 326, 189, 392, 391, 396, 388, 387, 398, 386, 397, 201, 377, 316, 309, 320, 325, 303, 246, 236, 220, 368, 315, 400.

Dari hasil diatas dapat ditarik hipotesa awal bahwa produk potensial pada tahun 2013 mengalami penurunan tingkat potensial dari katagori sangat laku menjadi katagori laku pada tahun 2014. Sedangkan untuk produk yang kurang potensial pada tahun 2013 menjadi sangat potensial pada tahun 2014. Untuk produk yang dikatakan sebagai *mascot* oleh pemilik UD Subur Baru, yaitu produk dengan kode barang 181 dan 182, memang memiliki tingkat penjualan yang tinggi ditahun 2013, dan itu bertahan selama tiga bulan pada *dataset* yang digunakan. Namun ditahun 2014 produk tersebut mengalami penurunan. Ditahun 2014, produk dengan kode barang 345 masuk kedalam katagori sangat laku di tiga bulan pada *dataset*. Hal ini disebabkan produk tersebut memiliki data yang jauh (perbandingan nilai terjual dengan nilai invoice berbeda jauh) dari data yang lain atau dapat disebut sebagai data *outlier*. Untuk mengatasi data seperti ini, perlu ditambahkan jumlah cluster atau bisa juga dilakukan penghapusan data tersebut.

C. Evaluasi Validitas Cluster

Pengujian validitas *cluster* menggunakan metode *Modified Partition Coefficient* (MPC). Tiap periode *dataset* diuji tingkat validitas dengan jumlah *cluster* dua sampai lima. Penjelasan tentang tingkat validitas jumlah *cluster* akan dijelaskan pada tabel berikut :

Tabel 5.2 Hasil Validitas Cluster

Periode	2 Cluster	3 Cluster	4 Cluster	5 Cluster
Januari 13	0.825	0.850	0.809	0.804
Februari 13	0.967	0.879	0.818	0.836
Maret 13	0.962	0.910	0.930	0.863
Januari 14	0.966	0.946	0.869	0.865
Februari 14	0.992	0.960	0.347	0.810
Maret 14	0.998	0.933	0.902	0.775

Nilai dari indeks MPC berkisar antar 0 sampai 1. Tingkat validitas tertinggi dimiliki oleh jumlah *cluster* dengan

indeks MPC tertinggi. Pada periode *dataset* Januari 2013, jumlah *cluster* dengan indeks MPC tertinggi dimiliki oleh jumlah *cluster* 3. Berbeda dengan periode *dataset* yang lain, rata-rata indeks MPC tertinggi dimiliki oleh jumlah *cluster* 2. Jadi kesimpulan dari tabel diatas adalah jumlah *cluster* 2 memiliki tingkat validitas tertinggi, namun jumlah 2 *cluster* tentu tidak dapat mengatasi masalah data *outlier*, sehingga perlu ditambahkan satu *cluster* lagi untuk menampung data *outlier*. Atau juga dapat dilakukan penghapusan data *outlier*.

VI. PENUTUP

A. Kesimpulan

Dari penelitian yang sudah dilakukan, maka dapat ditarik kesimpulan sebagai berikut:

1. Anggapan bahwa metode *fuzzy c-means* cocok digunakan pada aplikasi segmentasi barang pada perusahaan retail pada penelitian sebelumnya adalah benar, hal ini dikarenakan dengan menggunakan metode ini bisa menghasilkan output berupa tingkat penjualan produk-produk pada penelitian ini.
2. Metode *fuzzy c-means* dapat mengatasi data *outlier*. Hal ini terbukti pada pengclustoran *dataset* pada periode bulan Januari, Februari dan Maret tahun 2014 pada penelitian ini. Data *outlier* masuk pada *cluster* yang ada dan tidak membentuk *cluster* tersendiri.
3. Jumlah *cluster* 2 memiliki tingkat validitas yang lebih tinggi dibandingkan dengan jumlah *cluster* 3 pada rata-rata *dataset* yang digunakan.
4. Diharapkan dengan melihat hasil dari proses *clustering* pada sistem ini perusahaan dapat mengambil langkah untuk meningkatkan efektivitas penyimpanan barang dengan cara meningkatkan jumlah produk potensial dan mengurangi jumlah produk yang kurang potensial.

B. Saran

Sebagai saran, dapat ditambahkan *dataset* untuk periode yang lain sehingga perbedaan hasil *clustering* dari beberapa periode dapat lebih terlihat. Kemudian dapat ditambahkan grafik mengenai perkembangan tingkat penjualan pada produk tertentu berdasarkan kurun waktu tertentu.

REFERENCES

- [1] Cakra Ramadhana, Yohana Dewi Lulu W, and Kartina Diah K. W., "Data Mining dengan Algoritma Fuzzy C-Means Clustering Dalam Kasus Penjualan di PT Sepatu Bata," in *Seminar Nasional Teknologi Informasi & Komunikasi Terapan 2013 (SEMANTIK 2013)*, Semarang, November 2013.
- [2] Eko Prasetyo, *Data Mining - Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta, Indonesia: C.V Andi Offset, 2012.
- [3] Pang Ning Tan, Steinbach Michael, and Kumar Vipin, *Introduction to Data Mining*, 1st ed. Boston, USA: Addison-Wesley Longman Publishing Co., 2005.
- [4] Nurhikmah Megawati, Moch. Abdul Mukid, and Rita Rahmawati, "Segmentasi Pasar Pada Pusat Perbelanjaan Menggunakan Fuzzy C-Means (Studi Kasus : Rita Pasaraya Cilacap)," *Jurnal Gaussian*, vol. 2, no. 4, pp. 343-350, 2013.
- [5] Yohana Nugraheni, "Data Mining dengan Metode Fuzzy untuk Customer Relationship Management (CRM) pada Perusahaan Ritel," Universitas Udayana, Denpasar, Thesis 2011.
- [6] Eko Priyo Utomo, *Panduan Mudah Mengenal Bahasa Java*. Bandung, Indonesia: CV. Yrama Widya, 2009.
- [7] Ian Sommerville, *Software Engineering*, 9th ed. Boston, United States: Pearson Education, 2011.
- [8] Roger S. Pressman, *Rekayasa Perangkat Lunak : Pendekatan Praktisi*, 2nd ed., L. N. Harnaningrum, Ed. Yogyakarta, Indonesia: Andi, 2003.
- [9] Hendri Ma'ruf, *Pemasaran Ritel*. Jakarta, Indonesia: PT Gramedia Pustaka Utama, 2005.
- [10] Sri Kusumadewi and Sri Hartati, *Neuro-Fuzzy : Integrasi Sistem Fuzzy dan Jaringan Syaraf*, 2nd ed. Yogyakarta, Indonesia: Graha Ilmu, 2010.
- [11] Raymond McLeod Jr. and George P. Schell, *Management Information Systems*, 10th ed. Upper Saddle River, New Jersey: Pearson Education, 2008.
- [12] Joseph P. Cannon, William D. Perreaut, and Jr., E. Jerome McCarthy, *Pemasaran Dasar*, 16th ed. Jakarta, Indonesia: Salemba Empat, 2009.