

Analisis Dan Implementasi Algoritma K-Nearest Neighbors Untuk Pengelompokan Member Potensial Produk Market

Dadan Nugraha

Fakultas Ilmu Komputer Universitas Kuningan

Jalan Tjut Nyak Dhien Cijoho Kuningan Telepon (0232) 873696 Fax. (0232) 874824

dadannugraha2013@gmail.com

Abstraksi

Waralaba merupakan salah satu usaha yang banyak digeluti oleh para pengusaha, dan salah satunya adalah minimarket. Kunci sukses menjamurnya mini market tersebut adalah terletak pada promosi. Minimarket dengan nama-nama besar tidak ragu lagi mengeluarkan biaya yang besar hanya untuk melakukan promosi. Salah satu langkah promosi yang banyak digunakan saat ini adalah dengan menyebarkan brosur –brosur bergambar dan berwarna yang berisi penawaran produk yang sedang dipromosikan. Akan tetapi promosi yang dilakukan selama ini kurang mengenai sasaran yang diinginkan, hal itu dikarenakan tidak semua customer yang mendapatkan brosur tersebut termasuk customer yang berpotensi untuk membeli produk-produk yang sedang dipromosikan. Selain itu biaya yang dikeluarkan untuk membuat brosur berwarna juga sangat mahal. Akan tetapi hal tersebut dapat diatasi dengan menggunakan algoritma K-Nearest Neighbors dimana hanya customer yang berpotensi saja yang akan menjadi obyek promosi. Dalam hal ini customer potensial dapat diketahui dengan cara melihat pola pembelian yang dilakukan oleh customer tersebut pada periode tertentu yang selanjutnya akan dianalisa dengan menggunakan metode K-Nearest Neighbors sehingga dapat diketahui siapa saja customer yang akan dikirim promo penawaran barang melalui SMS. Jadi dengan aplikasi ini tidak semua customer yang terdaftar dikirim promo produk yang sedang dipromosikan. Sehingga promosi dapat dilakukan secara efektif dan efisien.

Kata kunci : Pengelompokan, Member potensial, K-Nearest Neighbors, Promosi.

Bab I

Pendahuluan

1.1 Latar Belakang Masalah

Dewasa ini perkembangan dunia usaha terutama bidang waralaba sangat diminati banyak kalangan. Selain dapat memberikan keuntungan dalam waktu yang relatif singkat juga memungkinkan untuk pengembangan usaha. Salah satu bidang usaha yang banyak diminati adalah minimarket. Kunci sukses menjamurnya minimarket tersebut adalah terletak pada promosi. Minimarket dengan nama-nama besar tidak ragu lagi mengeluarkan biaya yang besar hanya untuk melakukan promosi. Masalah tersebut dapat diatasi dengan menggunakan Algoritma K-Nearest Neighbors dimana hanya member

yang berpotensi saja yang akan menjadi obyek promosi. Dalam hal ini member potensial dapat diketahui dengan cara melihat pola pembelian yang dilakukan oleh member tersebut pada periode tertentu yang selanjutnya akan dianalisa dengan menggunakan metode K-Nearest Neighbors sehingga dapat diketahui siapa saja member yang akan diberi penawaran promo. Sehingga dengan menggunakan aplikasi ini pemberian penawaran promo akan tepat sasaran kepada member yang potensial.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas maka permasalahan yang akan dibahas pada skripsi ini adalah :

1. Bagaimana mengelompokkan member potensial menggunakan Algoritma K-Nearest Neighbors
2. Bagaimana membuat aplikasi pengelompokan member potensial untuk keperluan promosi.

1.3 Batasan Masalah

Batasan masalah yang digunakan dalam skripsi ini meliputi :

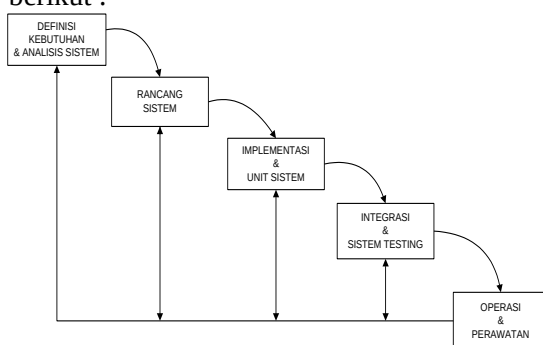
1. Member yang diperhitungkan dalam pengelompokan hanya member yang tercatat sebagai member dari minimarket dan ditentukan berdasarkan total belanja
2. Aplikasi ini dibangun dengan menggunakan PHP
3. Menggunakan Algoritma K-Nearest Neighbors untuk menentukan member potensial
4. Aplikasi ini tidak membahas pengiriman promosi kepada member potensial.

1.4 Metodologi Penelitian

Dalam penulisan Skripsi ini penulis mengumpulkan data dengan menggunakan 2 (dua) metode penelitian, diantaranya : Metode Pengumpulan data dan Metode Pendekatan Sistem.

1.6.1 Metode Pendekatan Sistem

Dalam Pengembangan Perangkat Lunak, Penulis menggunakan metode waterfall. Adapun langkah-langkahnya adalah sebagai berikut :



Gambar 1.1 Diagram Waterfall

Bab II

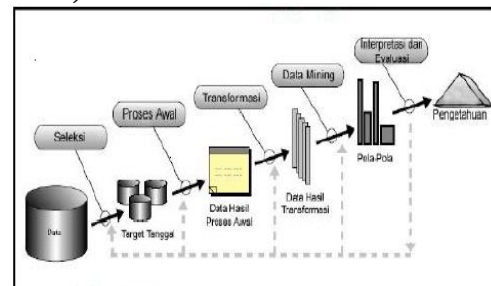
Landasan Teori

2.1 Data Mining

Data mining, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan

keterangan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan masa depan. Sehingga istilah *pattern recognition* sekarang jarang digunakan karena ia termasuk bagian dari data mining.

Faktor penentu bagi bentuk usaha atau bisnis apapun pada masa sekarang ini adalah kemampuan untuk menggunakan informasi secara seefektif mungkin. Penggunaan data yang strategis ini bisa diakibatkan oleh kesempatan-kesempatan yang dihasilkan atau ditimbulkan karena menemukan fakta-fakta sangat berharga yang cukup sering, tersembunyi dan tidak terdeteksi sebelumnya mengenai konsumen, *retailer* dan *supplier*, tren-tren bisnis, dan factor-faktor penunjukan yang lain. (Budi Santosa, 2007)



Gambar 2.1 Proses Data Mining
(sumber : Brachman dan Anand, 1996)

Dari gambar di atas dapat dijelaskan tiap-tiap tahap sebagai berikut :

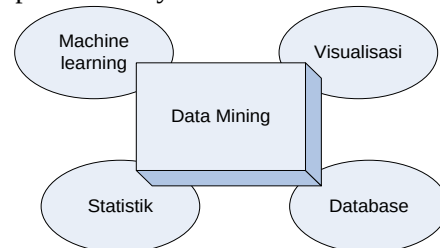
1. Tahap seleksi (*Selection*). Tahap memilih atau memilih-milih data berdasarkan kepada sejumlah kriteria.
2. Tahap proses awal (*Preprocessing*). Tahap pembersihan data dimana informasi-informasi tertentu dipindahkan karena dianggap tidak penting dan bisa memperlambat *queries* sebagai contoh : tidak perlu untuk mencatat jenis kelamin seorang pasien ketika mempelajari kehamilan. Data tersebut juga dikonfigurasi kembali untuk memastikan format yang konsisten karena adanya kemungkinan format-format inkonsisten dikarenakan data-data yang ada diambil dari beberapa sumber, seperti contoh jenis kelamin bisa dicatat dengan simbol f atau m dan juga dengan simbol 1 dan 0.

3. Tahap transformasi (*Transformation*). Dalam tahap ini data tidak hanya dipindahkan saja tetapi diubah sehingga dapat ditambahkan lapisan-lapisan seperti contoh lapisan demografis yang biasanya digunakan dalam penelitian. Data dibuat agar berguna dan dinamis.
4. Tahap datamining (*Datamining*). Tahap ini mengenai ekstraksi pola dari data.
5. Tahap implementasi dan evaluasi (*Interpretation and evaluation*). Pola-pola yang diidentifikasi oleh system yang ada diinterpretasikan ke dalam pengetahuan yang kemudian bisa digunakan untuk mendukung pengambilan keputusan. Seperti contoh : tugas-tugas klasifikasi dan prediksi, merangkum isi dari database atau menjelaskan fenomena yang diteliti.

Machine learning adalah suatu area dalam *artificial intelligence* atau kecerdasan buatan yang berhubungan dengan pengembangan teknik-teknik yang bisa diprogramkan dan belajar dari data masa lalu. Pengenalan pola, data mining dan machine learning sering dipakai untuk menyebut sesuatu yang sama. Bidang ini bersinggungan dengan ilmu probabilitas dan statistic kadang juga optimasi. Machine learning menjadi alat analisis dalam data mining. Bagaimana bidang-bidang ini berhubungan bisa dilihat dalam Gambar 2.2. Berikut ini adalah perbedaan-perbedaan yang bisa diberikan. Ini pun tidak secara jelas membedakan karena masih ada area abu-abu :

- a. Statistik : lebih berdasarkan teori, lebih focus pada pengujian hipotesis
- b. Machine learning : lebih bersifat heuristic, focus pada perbaikan performansi dari suatu teknik learning, juga meliputi real-time learning dan robotic area yang tidak termasuk dalam data mining.
- c. Data mining : gabungan teori dan heuristic, focus pada seluruh proses penemuan knowledge / pola termasuk data cleaning, learning, dan visualisasi dari hasilnya.
- d. Visualisasi : konversi data ke dalam format visual atau tabel sehingga karakteristik dari data dan relasi diantara item data atau atribut dapat di analisis atau dilaporkan.

- e. Database : koleksi data-data yang saling berhubungan mengenai suatu organisasi / enterprise dengan macam-macam pemakaiannya.



Gambar 2.2 Data mining merupakan irisan dari berbagai disiplin

2.4.1 Klastering

Teknik klaster termasuk teknik yang sudah cukup dikenal dan banyak dipakai dalam *data mining*. Sampai sekarang para ilmuwan dalam bidang data mining masih melakukan berbagai usaha untuk melakukan perbaikan model klaster karena metode yang dikembangkan sekarang masih bersifat heuristik. Usaha-usaha untuk menghitung jumlah klaster yang optimal dan pengklasteran yang paling baik masih terus dilakukan. Dengan demikian menggunakan metode yang sekarang, kita tidak bisa menjamin hasil pengklasteran kita sudah merupakan hasil yang optimal. Namun, hasil yang dicapai biasanya sudah cukup bagus dari segi praktis.

Suatu kumpulan dari entity yang hampir sama. Metode ini pada dasarnya melakukan segmentasi atau pengelompokan suatu populasi data yang heterogen menjadi beberapa sub grup atau *cluster*, metode ini dikategorikan kedalam teknik *Undirect Knowledge* atau *Unsupervised Learning* karena tidak membutuhkan proses pelatihan untuk klasifikasi awal data dalam masing-masing grup atau *cluster*.

Tujuan utama dari metode klaster adalah pengelompokan sejumlah data / objek kedalam klaster (group) sehingga dalam setiap klaster akan berisi data yang semirip mungkin. Dalam klastering kita berusaha untuk menempatkan obyek yang mirip (jaraknya dekat) dalam satu klaster dan membuat jarak antara klaster sejauh mungkin. Ini berarti obyek dalam satu klaster sangat mirip satu sama lain dan berbeda dengan obyek dalam klaster-klaster

yang lain. Klastering adalah salah satu teknik *unsupervised learning* dimana kita tidak perlu melatih metode tersebut atau dengan kata lain, tidak fase learning. Masuk dalam pendekatan unsupervised learning adalah metode-metode yang tidak membutuhkan label atau pun keluaran dari setiap data yang kita investigasi. Sebaliknya, supervised learning adalah metode yang memerlukan training (melatih) dan testing (menguji). Masuk dalam kategori ini adalah regresi, neural network (ANN), analisis diskriminan (LDA) dan support vector machine (SVM). Ada dua pendekatan dalam klastering : partisioning dan hirarki. Dalam partisioning kita mengelompokkan objek $X_1, X_2, \dots, X_n$ kedalam k klaster. ini bisa dilakukan dengan menentukan pusat klaster awal, lalu dilakukan realokasi obyek berdasarkan kriteria tertentu sampai dicapai pengelompokan yang optimum. Dalam klaster hirarki, kita mulai dengan membuat m klaster dimana setiap klaster beranggotakan satu obyek dan berakhir dengan satu klaster dimana anggotanya adalah m obyek.

2.4.2 Klastering Hirarki

Dalam klastering hirarki kita hitung jarak masing-masing obyek dengan setiap obyek yang lain. Selanjutnya kita temukan pasangan obyek yang jaraknya terdekat. Sehingga tiap obyek akan berpasangan dengan satu obyek atau kelompok obyek yang lain yang paling dekat jaraknya. Langkah-langkah yang perlu dilakukan untuk melakukan klastering dengan cara klastering hirarki adalah:

1. Kelompokkan setiap obyek ke dalam kelompok atau klastering sendiri
2. Temukan pasangan paling mirip untuk dimasukan kedalam klaster yang sama dengan melihat data dalam matriks kemiripan
3. Gabungkan kedua obyek dalam satu klaster
4. Ulangi sampai tersisa hanya satu klaster

Dalam klaster, untuk menggabungkan dua atau lebih obyek menjadi satu klaster, biasanya digunakan ukuran kemiripan atau ketidak miripan. Semakin mirip dua obyek semakin tinggi peluang untuk dikelompokkan dalam satu klaster. Sebaliknya semakin tidak

mirip semakin rendah peluang untuk dikelompokkan dalam satu klaster. untuk mengukur kemiripan (*similarity*) dan ketidak miripan (*dissimilarity*) diantara data atau obyek bisa dipakai beberapa ukuran. untuk ukuran kemiripan bisa dipakai cosines, kovarian dan korelasi. Sedangkan untuk ukuran ketidak miripan bisa dipakai jarak. Dalam ukuran kemiripan, semakin besar nilainya berarti semakin mirip. Sebaliknya dalam ketidak miripan semakin besar nilainya semakin tidak mirip.

1. Cosinus antara dua titik x dan y didefinisikan sebagai :

$$(3.1)$$

Dimana didefinisikan sebagai

Dari data pertama di Tabel 2.1, [10 5], dan data kedua, [20 20], kita bisa menghitung besarnya cosinus diantara keduanya.

$$(3.2)$$

2. Kovarian

Kovarian antara dua data didefinisikan sebagai :

$$(3.3)$$

Dimana x adalah data pertama dan y data kedua. Dari tabel 3.1 kita bisa menghitung kovarian antara data ke satu dan ke dua. Rata-rata dari data ke satu dan ke dua adalah :

3. Koefisien korelasi.

Misalkan kita punya data sebagai berikut dan kita ingin melakukan klastering dengan klastering hirarki.

Tabel 2.1 Contoh set data dengan dua variabel atau atribut

No	X1	X2
1	10	5
2	20	20
3	30	10
4	30	15
5	5	10

Dari tabel 2.1 di atas bisa dibuat matriks ketidak miripan (angka di belakang koma dibulatkan), yaitu jarak. Selanjutnya kita susun dalam bentuk matriks. Dalam matriks ini jarak dari setiap pasang obyek dihitung. Nilai dalam satu entri menunjukkan jarak

antara obyek dari indeks dari kolom dan baris. Misalnya entri dari kolom 1 baris 2, berarti jarak antara obyek 1 dan obyek 2. Karena matriks jarak ini simetris, maka kita hanya menampilkan separuh dari seluruh matriks.

	1	2	3	4
5				
1	0	.	.	.
.				
2	18.0	0	.	.
.				
3	20.6	14.1	0	.
.				
4	22.4	11.2	5	0
.				
5	7.1	18.0	25	
25.5	0			

Biasanya untuk menunjukan jarak antara dua klaster A dan B digunakan salah satu dari beberapa ukuran :

- Jarak maksimum antara elemen dalam klaster (complete linkage clustering). Dengan cara ini jarak antara dua klaster didefinisikan sebagai :

Dimana d_{AB} adalah jarak dua data x dan y masing – masing dari klaster A dan B.

- Jarak minimum antara elemen dari setiap klaster (*single linkage clustering*). Dengan cara ini jarak antara dua klaster didefinisikan sebagai :
- Rata-rata jarak antara elemen dari setiap klaster (*average linkage clustering*).

Dimana n_A dan n_B masing-masing adalah banyaknya data dalam set A dan B.

- Centroid Linkage, dengan cara ini jarak antara dua klaster didefinisikan sebagai :

Dimana :

,

- Ward linkage

2.4.2 Algoritma K-Nearest Neighbors

Algoritma *K-Nearest Neighbors* merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana masing-masing dimensi merepresentasikan fitur dari data. Ruang ini dibagi menjadi bagian-bagian berdasarkan klasifikasi data pembelajaran. Sebuah titik pada ruang ini ditandai kelas c jika kelas c merupakan klasifikasi yang paling banyak ditemui pada k buah tetangga terdekat titik tersebut. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *Euclidean*.

Teknik ini termasuk dalam kelompok klasifikasi nonparametric. Disini kita tidak memperhatikan distribusi dari data yang ingin kita kelompokkan. Teknik ini sangat sederhana dan mudah diimplementasikan. Mirip dengan teknik klastering, kita mengelompokkan suatu data baru berdasarkan jarak data baru itu ke beberapa data / tetangga (neighbor) terdekat. Dalam hal ini jumlah data / tetangga terdekat ditentukan oleh user yang dinyatakan dengan k . Misalkan kita tentukan nilai $k = 5$, maka setiap data testing dihitung jaraknya terhadap data training dan kita pilih 5 data training yang jaraknya paling dekat ke data testing. Setelah kita ketahui 5 data ini kita periksa output atau labelnya masing-masing. Lalu kita tentukan output mana yang frekuensinya paling banyak (3.6). Kita masukan suatu data testing ke kelompok dengan output paling banyak. Misalkan dalam kasus klasifikasi dengan 3-kelas, dari lima data itu, tiga data dari kelas 1, satu data dari kelas 2 dan satu data dari kelas 3, kita simpulkan output dengan label 1 adalah yang paling banyak. Maka kita kelompokkan data baru kita ke kelas 1. Kita lakukan prosedur ini untuk semua data testing. (Budi Santosa, 2007) (3.8)

Langkah-langkah yang dilakukan pada metode K-Nearest Neighbors adalah sebagai berikut :

- Menentukan data training, label data training, k , dan data testing (3.9)

Data input yang digunakan adalah data training yaitu seluruh data yang akan diproses dengan algoritma K-Nearest

Neighbors, label data training yaitu kategori data yang akan dijadikan sebagai penentu dalam mengelompokkan data, parameter k yaitu jumlah data training yang jaraknya paling dekat dengan data testing, dan data testing yaitu data yang dijadikan sebagai data penentu yang diinginkan.

2. Menghitung jarak antara data testing ke setiap data training
Untuk menghitung jarak antara data testing dan data training dilakukan dengan menggunakan rumus jarak Euclidean.
3. Menentukan k data training yang jaraknya paling dekat dengan data testing
Parameter k merupakan parameter yang harus dimasukkan untuk menentukan berapa jumlah maksimum data training yang akan dipilih berdasarkan jaraknya.
4. Memeriksa label dari k data
Label data yang dimasukkan akan dijadikan sebagai penentu dalam proses klasifikasi.
5. Menentukan label yang frekuensinya paling banyak
Dari sejumlah k data yang telah terpilih kemudian dipilih satu data yang memiliki label yang frekuensinya paling banyak
6. Masukkan data testing ke label dengan frekuensi paling banyak
Dengan melihat frekuensi label yang paling banyak kita dapat memprediksi bahwa data testing dimasukkan termasuk kedalam kelompok label tersebut.
7. Stop

2.4.3 Metode Euclidean Distance

Jarak Euclidean paling sering digunakan untuk menghitung jarak. Jarak Euclidean berfungsi menguji ukuran yang biasa digunakan sebagai interpretasi kedekatan jarak antara dua objek. Adapun rumus untuk mencari jarak Euclidean adalah sebagai berikut :

dengan d adalah jarak euclidean antara vektor i dan vektor k ;
 d_i adalah komponen ke j dari vektor i ;

adalah komponen ke j dari vektor k ;
 D adalah jumlah komponen pada vektor i dan vektor k .
Sehingga jarak Euclid antara I dan K adalah

2.4 Teori Perancangan Perangkat Lunak

2.5.1 Definisi Basis Data

Basis data adalah kumpulan data (elementer) yang secara logik berkaitan dalam mempresentasikan fenomena atau fakta secara terstruktur dalam domain tertentu untuk mendukung aplikasi pada sistem tertentu. (Bambang Hariyanto, Ir., MT., 2004). Basis data adalah suatu susunan/kumpulan data operasional terlengkap dari suatu organisasi/perusahaan yang diorganisir/dikelola dan disimpan secara terintegrasi dengan menggunakan metode tertentu menggunakan komputer sehingga mampu menyediakan informasi optimal yang diperlukan pemakainya.

2.6 Perangkat Lunak Yang Digunakan

2.6.1 Macromedia Dreamweaver 8

Macromedia Dreamweaver 8 merupakan sebuah HTML professional untuk mendesain secara visual dan mengelola situs web maupun halaman web fasilitas editing secara visual dari dreamweaver memungkinkan untuk dapat menambahkan desain program secara manual.

Hanya saja dikarenakan perkembangan perangkat lunak yang semakin cepat Macromedia Dreamweaver semakin membutuhkan spesifikasi komputer yang semakin baik yang cepat sehingga jika dalam penggunaannya menggunakan perangkat keras yang tidak mencukupi akan menjadi hambatan melakukan penelitian.

2.6.2 PHP

PHP singkatan dari PHP Hypertext Preprocessor. PHP adalah bahasa scripting yang menyatu dengan HTML dan dijalankan pada server side, Artinya semua sintak PHP yang kita berikan akan sepenuhnya dijalankan pada server sedangkan yang

dikirimkan ke browser hanya hasilnya saja, PHP digunakan untuk pengembangan aplikasi diatas teknologi web karena memiliki banyak kelebihan, yang salah satunya adalah didukung kebanyakan konektifitas sistem database server yang ada dipasaran, sehingga sangat cocok digunakan untuk aplikasi web yang terkoneksi ke sistem database seperti on-line shopping. Sistem database yang telah didukung oleh PHP antara lain : Oracle, dBase, MySQL Interface ODBC, PHP juga mendukung komunikasi dengan layanan lain melalui protocol IMAP, SNMP, NNTP, POP3, atau bahkan HTTP.

Bab III

Analisis Dan Perancangan

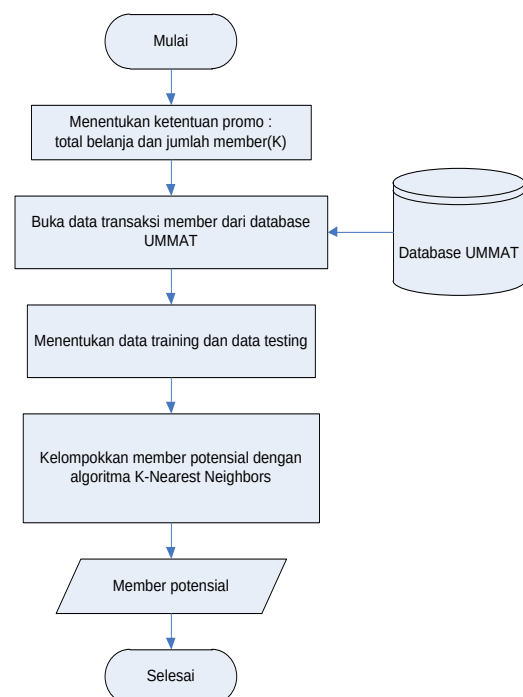
3.1 Analisis Dokumen

Sistem informasi yang terdapat di minimarket saat ini pada umumnya masih sebatas pada aplikasi Point of Sales (POS) saja atau aplikasi yang mencatat tentang penjualan, pembelian dan stock barang saja. Dalam system yang ada saat ini tidak terdapat aplikasi yang bisa membantu dalam hal promosi minimarket tersebut, promosi saat ini hanya menggunakan cara-cara manual yang tentu saja menurut penulis kurang efektif karena kebanyakan sasaran atau objek promosi yang dituju kurang jelas. Kebanyakan cara yang sering dilakukan oleh pihak minimarket adalah dengan mencetak dan menyebarkan brosur-brosur yang berisi tentang produk-produk yang sedang dipromosikan, gambar-gambar yang ada pada brosur-brosur tersebut memang sangat menarik, akan tetapi kebanyakan orang sering tidak memperhatikan keberadaan brosur-brosur tersebut, sering kita melihat brosur-brosur yang berisi tentang produk-produk yang sedang dipromosikan oleh pihak minimarket tersebut bergeletak begitu saja di jalan-jalan atau bahkan ada yang menggunakan kertas brosur-brosur tersebut untuk membungkus sesuatu tanpa memperhatikan isi dari brosur tersebut. Aplikasi yang dibuat ini nantinya berupa prototipe, dimana aplikasi ini hanya merupakan bagian dari sebuah sistem informasi minimarket yang besar dan lengkap dimana terdapat berbagai aplikasi yang dapat menunjukan dari kinerja

minimarket tersebut seperti promosi, penilaian kinerja karyawan, sistem pakar untuk menentukan barang yang akan dijual dan lain sebagainya. Kalau dilihat dari segi investasi, membangun sebuah sistem informasi minimarket ini sangatlah menguntungkan karena dampak yang dihasilkan sangatlah menguntungkan jika dibandingkan dengan cara-cara manual, sebagai contoh kita ambil saja prototype aplikasi yang dibuat oleh penulis. Sasaran atau objek promosi jauh lebih mengenai karena diperoleh dari hasil perhitungan sebuah metode clustering yang sudah teruji.

3.2 Analisis Sistem

3.2.1 Analisis Pengelompokan Member Potensial

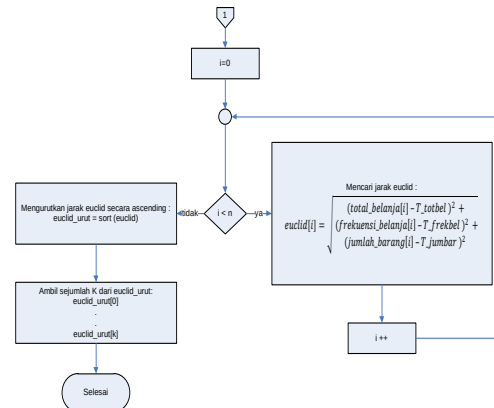
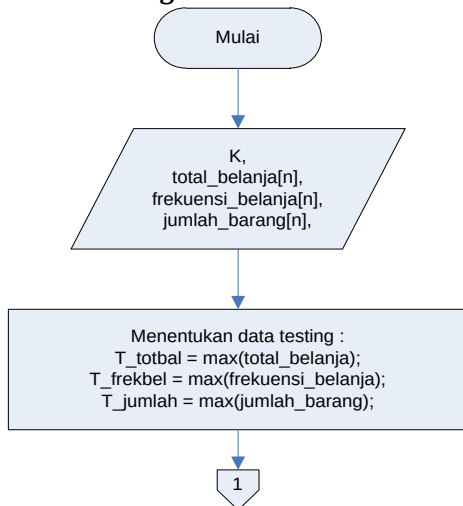


Gambar 3.1 Sistem Pengelompokan Member Potensial

Tahapan proses yang dijalankan berdasarkan gambar 3.1 di atas adalah :

1. Menentukan ketentuan promo yaitu dengan memasukkan nilai total belanja minimal yang akan digunakan untuk menyaring member, dan nilai jumlah member yang akan digunakan untuk mengambil jumlah member potensial
2. Membuka data histori transaksi member yang terdapat pada database UMMAT
3. Menentukan data training yaitu data total belanja, frekuensi belanja dan jumlah barang pada setiap member, dan data testing yaitu data maksimum dari masing-masing data training
4. Mengelompokkan member potensial dengan menggunakan algoritma K-Nearest neighbors
5. Hasil yang didapat adalah sejumlah K data member potensial.

3.2.2 Analisis Proses K-Nearest Neighbors



Gambar 3.2 Proses Algoritma K-Nearest Neighbors

Tahapan proses yang dijalankan berdasarkan gambar 3.2 di atas adalah :

1. Menentukan K (jumlah member yang akan diambil sebagai member potensial)
2. Menentukan total belanja, frekuensi belanja, dan jumlah barang dari seluruh member
3. Menentukan data testing sebagai data yang dijadikan penentu dalam pengelompokan. Data testing yang digunakan adalah data maksimum dari total belanja, maksimum dari frekuensi belanja, dan maksimum dari jumlah barang. Data maksimum ini digunakan karena data maksimum merupakan data yang paling berpotensi untuk dikelompokkan sebagai member potensial
4. Menghitung jarak Euclid antara data training dan data testing yang dimasukkan sebanyak jumlah member (n)
5. Data jarak Euclid yang didapat kemudian diurutkan secara ascending, yang berarti bahwa jarak Euclid yang berada pada urutan teratas (jarak Euclid terkecil) adalah data training yang memiliki tingkat keserupaan paling tinggi terhadap data testing

Dari hasil pengurutan kemudian diambil data sebanyak K dari mulai data teratas sebagai data member potensial.

3.2.3 Analisis Tahapan Proses Penghitungan K-Nearest Neighbors

Tahapan proses penghitungan K-Nearest Neighbors yang dilakukan untuk

mengelompokkan member potensial adalah sebagai berikut :

Misalkan akan dipilih 3 orang member (K=3) dari 6 member dengan rincian transaksi sebagai berikut :

X = Total belanja

Y = Seringnya belanja

Z = Total barang yang dibeli

H = Hasil penjumlahan jarak euclid

Proses pencarian jarak euclid antara data training dengan data testing

Dibawah ini merupakan hasil pengurutan secara ascending jarak euclid data training dan data testing :

Setelah melakukan pengurutan kemudian dipilih 3 data sebagai data (k) dari mulai data teratas, yaitu :

3.2.4 Analisis Kebutuhan Sistem Fungsional

Kebutuhan fungsional dari sistem ini adalah sebagai berikut :

1. Dapat mengelompokkan member potensial dari seluruh member yang terdapat pada database UMMAT
2. Dapat menampilkan hasil pengelompokan berupa member potensial yang sudah dipilih
3. Dapat menampilkan hasil analisis pengelompokan member potensial termasuk penghitungan matematis algoritma K-Nearest Neighbors

3.2.5 Analisis Kebutuhan Sistem Nonfungsional

3.2.5.1 Perangkat Keras

Spesifikasi perangkat keras yang digunakan selama pengerjaan skripsi ini adalah sebagai berikut :

1. Processor Pentium(R) Dual-Core CPU E5400 @ 2.70GHz
2. RAM 1 GB
3. Hard Disk 150 GB
4. Modem Huawei E173 digunakan untuk mengirim SMS.

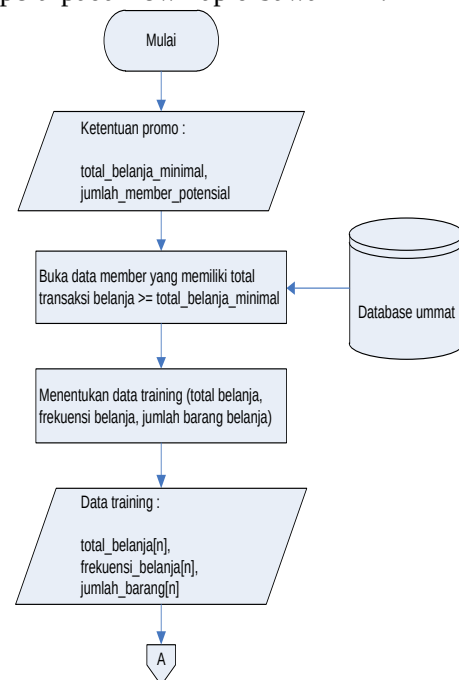
3.2.5.2 Perangkat Lunak

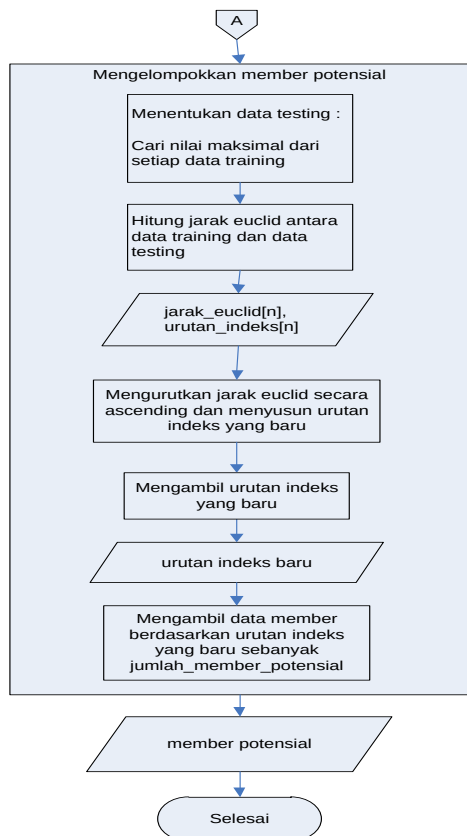
Adapun perangkat lunak yang di gunakan :

1. Sistem operasi Microsoft Windows XP Professional Service Pack 2
2. Xampp_Win 1.6.4
3. Macromedia Dreamweaver 8.0
4. Mozilla Firefox 8.0

3.3 Perancangan Sistem

Dari analisa permasalahan yang telah dilakukan maka akan dirancang Sistem Pengelompokan Member Potensial untuk promosi produk minimarket dengan menggunakan algoritma K-Nearest Neighbors. Adapun sistem yang dibangun seperti pada flowmap dibawah ini :





Gambar 3.3 Perancangan Sistem Pengelompokan Member Potensial

Tahapan proses yang dijalankan berdasarkan gambar 3.3 di atas adalah :

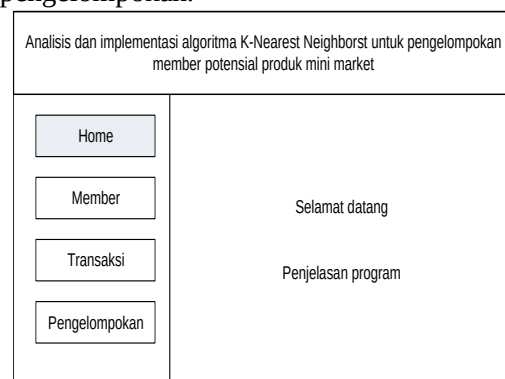
1. Memasukkan ketentuan promo. Ada dua ketentuan promo yang terlebih dahulu harus dimasukan oleh pengguna yaitu total belanja minimal, dan jumlah member potensial. Total belanja minimal merupakan batas terendah dari total transaksi belanja member, dimana member yang memiliki total belanja mencapai total belanja minimal akan diikuti sertakan dalam proses pengelompokan member potensial. Jumlah member potensial merupakan jumlah member yang akan diberikan penawaran promo.
2. Membuka data member dari database ummat. Member yang akan diikuti sertakan dalam proses pengelompokan adalah member yang memiliki total belanja minimal sebesar total belanja minimal yang sudah ditetapkan.

3. Menentukan data training. Data training yang digunakan dari data member adalah data total belanja, frekuensi belanja, dan jumlah barang yang dibeli.
4. Ketiga variabel data training dari setiap member yang sudah disaring tersebut akan digunakan sebagai nilai acuan pada proses pengelompokan member potensial.
5. Menentukan data testing. Data testing yang digunakan pada sistem ini adalah nilai maksimum dari setiap data training yang dimasukkan.
6. Menghitung jarak euclid antara data training dan data testing
7. Mengurutkan jarak euclid. Dari jarak euclid yang dihasilkan kemudian diurutkan secara ascending untuk mendapatkan nilai jarak euclid yang kecil. Ketika urutan jarak euclid berubah, hal itu berarti urutan indeks member juga berubah.
8. Mengambil urutan indeks yang baru sebagai urutan dari member yang paling berpotensi.
9. Mengambil data member berdasarkan urutan indeks yang baru sebanyak jumlah member yang sudah ditentukan.
10. Sejumlah K member yang sudah dipilih merupakan member potensial.

3.4 Perancangan Antar Muka

3.4.1 Halaman Beranda

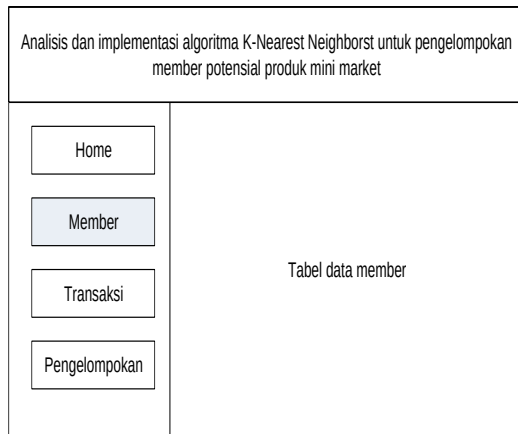
Halaman utama yang berisi beberapa menu yaitu Home, member, transaksi, pengelompokan.



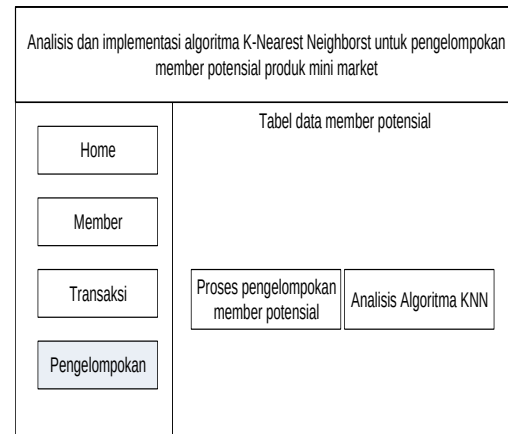
Gambar 3.4 Halaman Beranda

3.4.2 Halaman Member

Halaman ini digunakan untuk menampilkan data member yang terdapat pada database UMMAT.



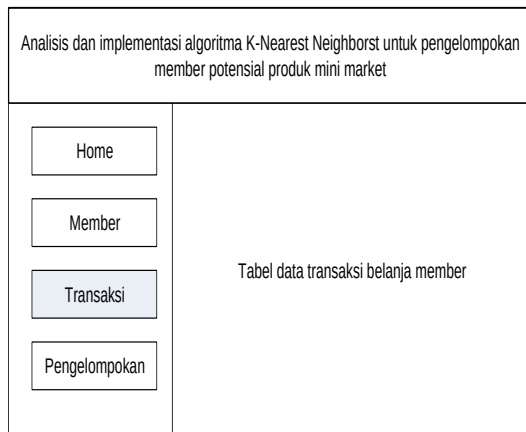
Gambar 3.5 Halaman Member



Gambar 3.7 Halaman Pengelompokan

3.4.3 Halaman Transaksi

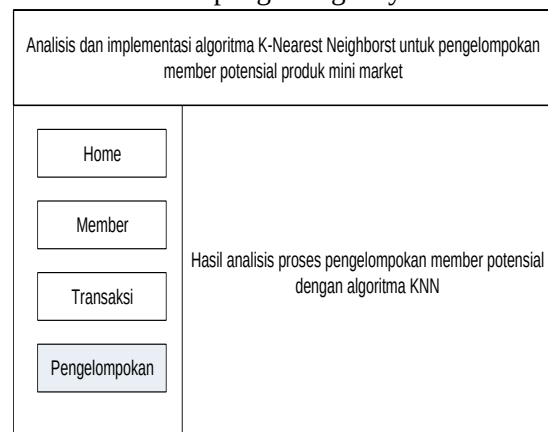
Halaman ini di gunakan untuk menampilkan data transaksi belanja yang dilakukan oleh member.



Gambar 3.6 Halaman Transaksi

3.4.5 Halaman Analisis Algoritma K-Nearest Neighbors

Halaman ini digunakan untuk menampilkan analisis pengelompokan member potensial dengan algoritma K-Nearest Neighbors termasuk rincian penghitungannya.



Gambar 3.8 Halaman Analisis K-Nearest Neighbors

3.4.4 Halaman Pengelompokan Member Potensial

Halaman ini berisi tombol proses pengelompokan member potensial dan tombol analisis K-Nearest Neighbors. Pada halaman ini juga akan ditampilkan data member potensial jika sudah dilakukan proses pengelompokan member potensial.

Bab IV Implementasi Dan Pengujian

4.1 Implementasi Sistem

4.1.1 Prosedural Sistem

Pada sistem pengelompokan member potensial dengan Algoritma K-Nearest Neighbors ini akan dilakukan beberapa tahapan proses sebagai berikut :

1. Memasukan nilai minimum total belanja member dan jumlah maksimum member potensial yang akan dipilih
2. Melakukan pengelompokan member potensial menggunakan Algoritma K-Nearest Neighbors berdasarkan total belanja, frekuensi belanja dan jumlah barang yang dibeli member

3. Penghitungan secara lengkap dapat dilihat pada hasil analisis Algoritma K-Nearest Neighbors.

4.1.2 Algoritma K-Nearest Neighbors

Tahapan proses yang dilakukan pada Algoritma K-Nearest Neighbors adalah sebagai berikut :

- a. Proses inisiasi

Pada proses ini dilakukan pemberian nilai total belanja, frekuensi belanja, jumlah barang, sebagai nilai yang akan diproses pada Algoritma K-Nearest Neighbors.

//input total belanja, frekuensi belanja dan jumlah barang

```
public function
__construct($totBelanja,$frekBelanja,$
jumBarang) {
    $this->totBelanja=$totBelanja;
    $this->frekBelanja=$frekBelanja;
    $this->jumBarang=$jumBarang;
    $this-
    >jumData=count($totBelanja);
    $this->maxTotBelanja= $this-
    >getMaxTotBelanja();
    $this->maxFrekBelanja= $this-
    >getMaxFrekBelanja();
    $this->maxJumBarang=
    $this->getMaxJumBarang();
}
```

- b. Pencarian nilai maksimum dari total belanja

Pada proses ini dilakukan pencarian nilai maksimum dari setiap total belanja member.

//cari nilai maksimum dari besar total belanja

```
public function getMaxTotBelanja(){
return max($this->totBelanja);
}
```

- c. Pencarian nilai maksimum dari frekuensi belanja

Pada proses ini dilakukan pencarian nilai maksimum dari setiap frekuensi belanja member.

//cari nilai maksimum dari seringnya belanja

```
public function getMaxFrekBelanja(){
return max($this->frekBelanja);
}
```

- d. Pencarian nilai maksimum dari banyak jenis barang

Pada proses ini dilakukan proses pencarian nilai maksimum dari banyak jenis barang yang diberi oleh member.

//cari nilai maksimum dari banyak jenis barang

```
public function getMaxJumBarang(){
return max($this->jumBarang);
}
```

- e. Pencarian nilai jarak euclid

Pada proses ini dilakukan proses pencarian jarak euclid antara data training dan data testing (nilai maksimum setiap variabel).

//cari jarak euclid antara data training dan data testing

```
public function getJarakEuclid(){
    $euclid=array();
    for($i=0;$i<$this->jumData;$i++){
        $euclid[$i][0]=sqrt(pow($this-
        >totBelanja[$i]-$this-
        >maxTotBelanja, 2)+ pow($this-
        >frekBelanja[$i]-$this-
        >maxFrekBelanja, 2)+ pow($this-
        >jumBarang[$i]-$this-
        >maxJumBarang, 2));
        $euclid[$i][1]=$i;
    }
    return $euclid;
}
```

- f. Proses pengurutan jarak euclid secara ascending

Pada proses ini dilakukan pengurutan jarak euclid secara ascending yaitu dari mulai nilai terkecil sampai nilai yang terbesar.

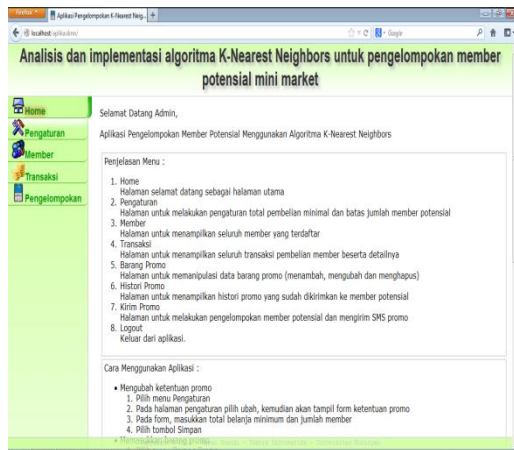
//urutkan nilai euclid yang didapat secara ascending

```
public function
urutAscending($data){
    sort($data);
    return $data;
}
```

4.2 Implementasi Tampilan

4.2.1 Tampilan Awal Aplikasi

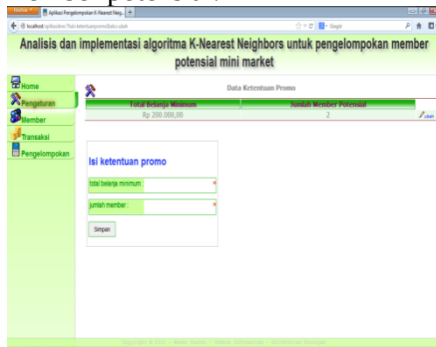
Halaman selamat datang dan sebagai halaman utama.



Gambar 4.1 Tampilan Awal Aplikasi

4.2.2 Tampilan Pengaturan

Halaman untuk melakukan pengaturan total pembelian minimal dan batas jumlah member potensial.



Gambar 4.2 Tampilan Pengaturan

4.2.3 Tampilan Member

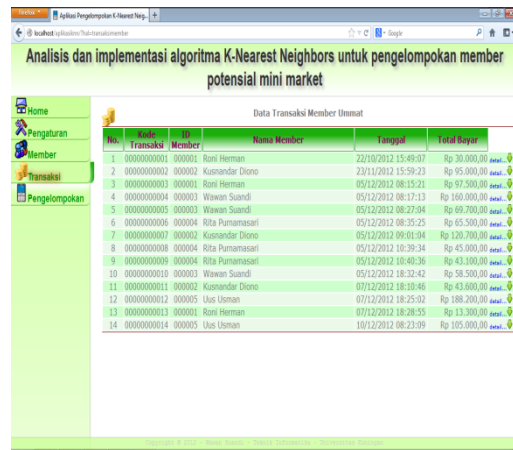
Halaman untuk menampilkan seluruh member yang terdaftar



Gambar 4.3 Tampilan Member

4.2.4 Tampilan Transaksi

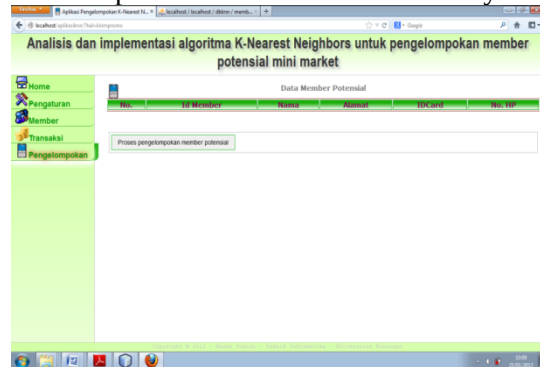
Halaman untuk menampilkan seluruh transaksi pembelian member beserta detailnya.



Gambar 4.4 Tampilan Transaksi

4.2.5 Tampilan Pengelompokan

Halaman untuk melakukan pengelompokan member potensial beserta hasil analisisnya.



Gambar 4.5 Tampilan Pengelompokan

4.3 Pengujian Sistem

4.3.1 Pengelompokan Member Potensial

Data transaksi yang digunakan untuk proses pengelompokan member potensial menggunakan algoritma K-Nearest Neighbors adalah sebagai berikut :

No.	Kode Transaksi	ID Member	Nama Member	Tanggal	Total Bayar
1	0000000001	000001	Roni Herman	22/10/2012 15:49:07	Rp 30.000,00
2	0000000002	000002	Kusnandar Diono	23/11/2012 15:59:23	Rp 95.000,00
3	0000000003	000001	Roni Herman	05/12/2012 08:15:21	Rp 97.500,00
4	0000000004	000003	Wawan Suandi	05/12/2012 08:17:13	Rp 160.000,00
5	0000000005	000003	Wawan Suandi	05/12/2012 08:27:04	Rp 69.700,00
6	0000000006	000004	Rita Purnamasari	05/12/2012 08:35:25	Rp 65.500,00
7	0000000007	000002	Kusnandar Diono	05/12/2012 09:01:04	Rp 120.700,00
8	0000000008	000004	Rita Purnamasari	05/12/2012 10:39:34	Rp 45.000,00
9	0000000009	000004	Rita Purnamasari	05/12/2012 10:40:36	Rp 43.100,00
10	0000000010	000003	Wawan Suandi	05/12/2012 18:32:42	Rp 58.500,00
11	0000000011	000002	Kusnandar Diono	07/12/2012 18:10:46	Rp 43.600,00
12	0000000012	000005	Uus Usman	07/12/2012 18:25:02	Rp 188.200,00
13	0000000013	000001	Roni Herman	07/12/2012 18:28:55	Rp 13.300,00
14	0000000014	000005	Uus Usman	10/12/2012 08:23:09	Rp 105.000,00

Gambar 4.6 Data Transaksi Member

Tahapan proses pengelompokkan member potensial dari data transaksi di atas adalah sebagai berikut :

1. Mengatur ketentuan promo dengan memasukkan total belanja minimum dan jumlah member potensial yang akan dipilih. Pada pengujian ini total belanja minimum yang digunakan adalah Rp 200.000 dan jumlah member potensial yang akan dipilih adalah
2. Menyaring data member berdasarkan total belanja yang mencapai Rp 200.000 sehingga dari data transaksi di atas setelah dikalkulasi keseluruhan, member yang memiliki total belanja mencapai Rp 200.000 adalah seperti pada tabel berikut :

No	ID Member	Nama	Total Belanja (X)	Frekuensi Belanja (Y)	Banyak Barang (Z)
1	000002	Kusnandar Diono	Rp 259.300,00	3	28
2	000003	Wawan Suandi	Rp 288.200,00	3	20
3	000005	Uus Usman	Rp 293.200,00	2	10

Gambar 4.7 Data member hasil penyaringan

3. Melakukan penghitungan K-Nearest Neighbors dengan menghitung jarak Euclid antara data training dan data testing. Dari data training di atas kemudian ditentukan data testing yaitu nilai maksimum dari setiap variabel sehingga diperoleh data testing sebagai berikut :
 Total Belanja (X) : 293200
 Frekuensi Belanja (Y) : 3
 Banyak Barang (Z) : 28
 Secara lengkap penghitungan K-Nearest Neighbors dapat diperlihatkan pada hasil penghitungan di bawah :

Analisis Proses Pengelompokan KNN

→ Data member yang memiliki jumlah total belanja mencapai Rp 200.000,00 :

No.	ID Member	Nama	Total Belanja (X)	Frekuensi Belanja (Y)	Banyak Barang (Z)
1	000002	Kusnandar Diono	Rp 259.300,00	3	28
2	000003	Wawan Suandi	Rp 288.200,00	3	20
3	000005	Uus Usman	Rp 293.200,00	2	10
Data Testing			293200	3	28

→ Jarak euclid antara data training dan data testing :

No.	ID Member	Total Belanja (X)	Frekuensi Belanja (Y)	Banyak Barang (Z)	Jarak Euclid
1	000002	$\text{sqr}(259300 - 293200) = 114921,0000$	$\text{sqr}(3 - 3) = 0$	$\text{sqr}(28 - 28) = 0$	$\text{sqr}(X + Y + Z) = 33900$
2	000003	$\text{sqr}(288200 - 293200) = 25000000$	$\text{sqr}(3 - 3) = 0$	$\text{sqr}(20 - 28) = 64$	$\text{sqr}(X + Y + Z) = 5000,0064$
3	000005	$\text{sqr}(293200 - 293200) = 0$	$\text{sqr}(2 - 3) = 1$	$\text{sqr}(10 - 28) = 324$	$\text{sqr}(X + Y + Z) = 18,0277563773$

Keterangan:

sqr = kuadrat

sqrt = akar pangkat dua

→ Hasil pengurutan secara ascending Jarak euclid :

No.	ID Member	Jarak Euclid
1	000005	18.0277563773
2	000003	5000.0064
3	000002	33900

→ Hasil pengurutan dipilih sejumlah K= 2 sebagai Member Potensial :

No.	ID Member	Nama	Jarak Euclid
1	000005	Uus Usman	18.0277563773
2	000003	Wawan Suandi	5000.0064

Gambar 4.8 Perhitungan Algoritma K-Nearest Neighbors

Dari hasil penghitungan di atas didapat 2 member potensial yang memiliki jarak Euclid paling kecil yaitu Uus Usman dengan jarak Euclid 18,0277563773 dan Wawan Suandi dengan jarak Euclid 5000,0064.

4.3.2 Pengujian Black Box

Tabel 4.1 Tabel pengujian black box

No.	Skenario Uji	Hasil Yang Di Harapkan	Hasil Kenyataan	Kesimpulan
1.	Mengubah ketentuan promo	ketentuan promo dapat diubah	Sesuai	Valid
2.	Menampilkan data member	Data member tampil di layar	Sesuai	Valid
3.	Menampilkan	Data	Sesuai	Valid

	kan data transaksi member	transaksi tampil di layar	i	
4.	Mengelompokkan member potensial	Data member potensial tampil di layar	Sesuai	Valid
5.	Menampilkan analisis Algoritma K-Nearest Neighbors	Analisis penghitungan Algoritma K-Nearest Neighbors tampil di layar	Sesuai	Valid

4.3.3 Pengujian White Box

Pengujian *white-box* berfokus pada struktur *control* program. *Test case* dilakukan untuk memastikan bahwa semua *statement* pada program telah dieksekusi paling tidak satu kali selama pengujian dan bahwa semua kondisi logis telah diuji.

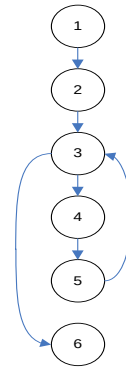
Pada pengujian *white-box* ini diambil contoh pada proses pencarian jarak euclid pada algoritma K-Nearest Neighbors. Kode programnya adalah sebagai berikut :

```

public function getJarakEuclid(){
    $seuclid=array();
    for($i=0;$i<$this->jumData;$i++){

        $seuclid[$i][0]=sqrt(pow($this->totBelanja[$i]-$this->maxTotBelanja, 2)+
        pow($this->frekBelanja[$i]-$this->maxFrekBelanja, 2)+ pow($this->jumBarang[$i]-$this->maxJumBarang, 2));
        $seuclid[$i][1]=$i;
    }
    return $seuclid;
}

```



Gambar 4.9 Flowgraph Pengujian *white-box* jarak euclid

Rumus :

$$V(G) = E - N + 2$$

$$= 6 - 6 + 2$$

$$V(G) = 2$$

$$V(G) = 1 + 1$$

$$= 1 + 1$$

$$V(G) = 2$$

Keterangan : - $V(G)$: Jumlah jalur yang bisa dilalui

- E : Edge (Banyak busur)

- N : Node (Banyak titik)

- P : Path (Node yang memiliki cabang lebih dari satu)

Jalur dari proses pencarian jarak euclid, yaitu :

1-2-3-4-5. *Node* 1 merupakan awal method, *node* 2 merupakan proses inisiasi (pemberian nilai awal), *node* 3 proses pengulangan yang berisi proses penghitungan jarak Euclid pada *node* 4 dan *node* 5.

1-2-3-6. Pada jalur ini, ketika kondisi pengulangan pada *node* 3 bernilai false akan langsung ke *node* 6 yaitu pengembalian nilai Euclid.

3

Bab V

Penutup

5.1 Kesimpulan

Setelah melakukan penelitian atau observasi pada mini market ummat, maka saya dapat menarik kesimpulan :

1. Algoritma K-Nearest Neighbors dapat digunakan untuk mengelompokkan member potensial
2. Algoritma K-Nearest Neighbors dapat digunakan untuk menghitung jarak Euclid antara data testing dan data training yang dimasukkan

3. Hasil pengelompokan member potensial dapat digunakan untuk keperluan promosi produk minimarket.

Saran atau masukan untuk pengembangan lebih lanjut dari sistem ini adalah sebagai berikut :

1. Untuk pengembangan berikutnya, akan lebih baik lagi jika dapat dikembangkan menjadi SMS Gateway, agar pelanggan dapat langsung menanyakan barang apa saja yang sedang dipromosikan melalui SMS.

5.2 Saran

Daftar Pustaka

- Santosa, B., 2007, Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis, Graha Ilmu, Yogyakarta.
- Sidik, Ir. B., 2006, Pemrograman Web dengan PHP, Informatika Bandung, Bandung.
- K-Nearest Neighbors, 20 Oktober 2012, ([http://www.statsoft.com/textbook/k-nearest neighbors](http://www.statsoft.com/textbook/k-nearest-neighbors))
- MySQL AB. 2008, MySQL 5.1 Reference Manual, Sun Microsystems Inc., (<http://dev.mysql.com/doc/refman/5.1/en/index.html>)
- The PHP Documentation Group. 2012, PHP Manual, (php.net/download-docs.php)
- Jogiyanto, Hartono, M.Akt, MBA, Ph.d. 2002. Analisis Desain dan Sistem Informasi, Andi Offset,
- Andri, Krisnanto. 2003. Perancangan Sistem Informasi. Gava Media, Yogyakarta