

IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGUNAKAN ALGORITMA K-MEANS

IMPLEMENTATION OF DATA MINING FOR POTENTIAL CUSTOMER SELECTION USING K-MEANS ALGORITHM

Randi Rian Putra¹, Cendra Wadisman²
Universitas Pembangunan Panca Budi Medan¹
Universitas Putra Indonesia “YPTK”²
Randirian@dosen.pancabudi.ac.id¹

ABSTRACT

This research was conducted to study data mining using K-Means algorithm which is implemented in the selection of potential customers in MC Laundry. data collected through observation and interviews conducted to the owner of the company concerned. Further data is analyzed to determine the data needed in the process of determining potential customers. From experiments conducted using K-Means algorithm and using Tanagra Software can provide accurate results in classifying potential customers.

Keywords: *Data Mining, K-Means Algorithm, Tanagra Software*

ABSTRAK

Penelitian ini dilakukan untuk mempelajari data mining menggunakan algoritma *K-Means* yang diimplementasikan dalam pemilihan pelanggan potensial di MC Laundry. data dikumpulkan melalui observasi dan interview yang dilakukan kepada pemilik perusahaan yang bersangkutan. Selanjutnya data dianalisa untuk menentukan data yang dibutuhkan dalam proses penentuan pelanggan potensial. Dari eksperimen yang dilakukan dengan menggunakan algoritma *K-Means* serta menggunakan *Software Tanagra* dapat memberikan hasil yang akurat dalam mengelompokkan pelanggan potensial.

Kata kunci: *Data Mining, Algoritma K-Means, Software Tanagra*

PENDAHULUAN

Hubungan antara perusahaan dengan pelanggan sangat penting untuk menunjang perkembangan dan kelangsungan perusahaan. Pihak manajemen perusahaan harus mampu untuk mengenali pelanggan potensialnya dan mempercayainya dengan meningkatkan pemahaman perusahaan akan kebutuhan mereka sebagai individu sehingga dapat mempertahankan loyalitasnya terhadap perusahaan.

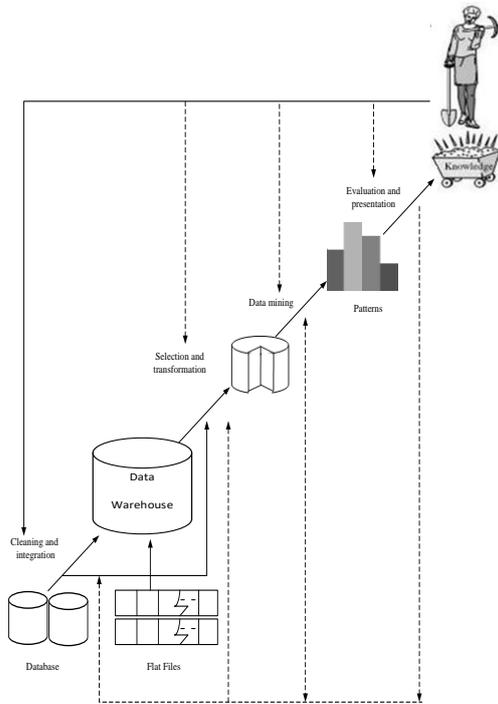
Salah satu hal yang penting dalam pengelolaan pelanggan adalah bagaimana suatu perusahaan dapat mempertahankan pelanggan yang dimilikinya. Usaha mempertahankan

pelanggan ini menjadi hal yang penting bagi MC Laundry mengingat semakin banyaknya usaha yang sama. Untuk mencegah terjadinya perpindahan pelanggan ini, maka perlu diketahui kelompok pelanggan yang potensial, sehingga perusahaan bisa melindungi pelanggan potensial tersebut dengan cara memberikan pelayanan prima dan memberikan hadiah. Hadiah yang diberikan perusahaan kepada pelanggan berupa cinderamata dan dilakukan secara kontiniu di setiap tahunnya.

Penelitian yang akan dibuat penulis penerapan Data Mining menggunakan Algoritma *K-Means*. Penelitian mengenai penerapan Data

Minning menggunakan Algoritma *K-Means* sudah banyak dilakukan oleh orang-orang yang berada dibidang *Information Technology* atau pada bidang lainnya.

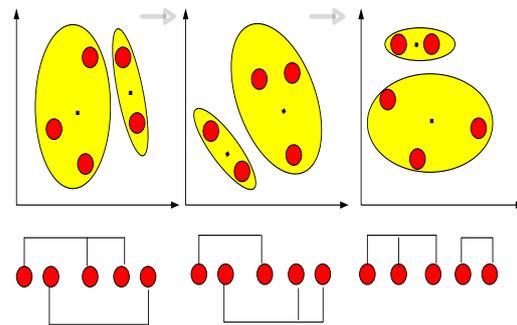
METODE PENELITIAN



Gambar 1. Skema Dari Proses KDD

***K-means* Clustering**

K-Means merupakan algoritma clustering yang berulang-ulang. Algoritma *K-Means* dimulai dengan pemilihan secara acak *K*, *K* disini merupakan banyaknya cluster yang ingin dibentuk. Kemudian tetapkan nilai-nilai *K* secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus *Euclidian* hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil) (Pralita, 2010).



Gambar 2. Langkah-Langkah Metode *k-means*

Pelanggan Potensial

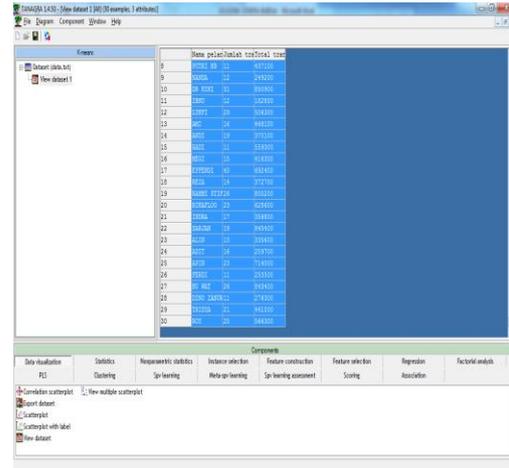
Pelanggan menduduki posisi penting dalam pengembangan strategi bisnis, pelanggan juga merupakan salah satu sumber keuntungan dalam perusahaan. Untuk itu diperlukan suatu pemahaman yang baik tentang pelanggan. Pemahaman yang baik terhadap pelanggan dapat digunakan perusahaan untuk berinvestasi pelanggan yang potensial. Masalah yang sering dihadapi adalah kesulitan dalam menganalisa nilai pelanggan. Banyak pemasar mengalami kesulitan untuk mengidentifikasi pelanggan atau nasabah yang tepat hal tersebut dapat mengakibatkan perusahaan kehilangan nasabah potensial dan tentunya akan sangat merugikan perusahaan.

Segmentasi pelanggan adalah metode yang populer yang digunakan untuk memilih pelanggan atau nasabah yang tepat untuk memulai promosi. Dengan segmentasi nasabah berdasarkan prilakunya, kita dapat menargetkan tindakan mereka dengan lebih baik. Seperti peluncuran produk yang disesuaikan, target pemasaran dan untuk memenuhi harapan pelanggan. Namun untuk menganalisa data pelanggan atau nasabah dalam jumlah besar memerlukan tenaga dan waktu yang banyak (Widiarina dan Romi S.W., 2015).

HASIL DAN PEMBAHASAN

A. Data Pengujian

Proses implementasi dan analisa data dilakukan dengan menggunakan program aplikasi *Tanagra*, dengan melakukan pengujian pada seluruh data yang telah didapatkan pada bulan oktober 2015 dan maret 2016 di MC Laundry sebanyak 20 *sample* data dimana sumber data yang pada awalnya disimpan dalam database yang menggunakan *Microsoft access* dirubah kebentuk format dokumen program aplikasi *excel*.



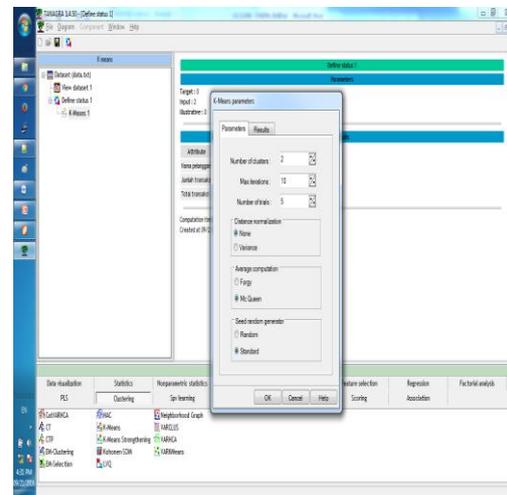
Gambar 4. Tampilan View dataset

B. Langkah Pengujian Tanagra

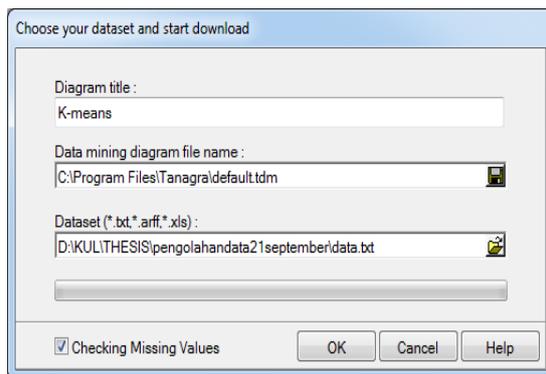
Untuk membuktikan kebenaran pada tahap analisa dan pengujian secara manual, maka diperlukan pengujian dengan menggunakan *Software Tanagra*, Adapun langkah-langkah yang dilakukan dalam pengujian ini yaitu :

1. Import Data

Setelah selesai melakukan pengumpulan data serta pengolahan data yang telah didapat dari hasil penelitian pada MC Laundry dalam format *.txt* pada *Notepad++*.

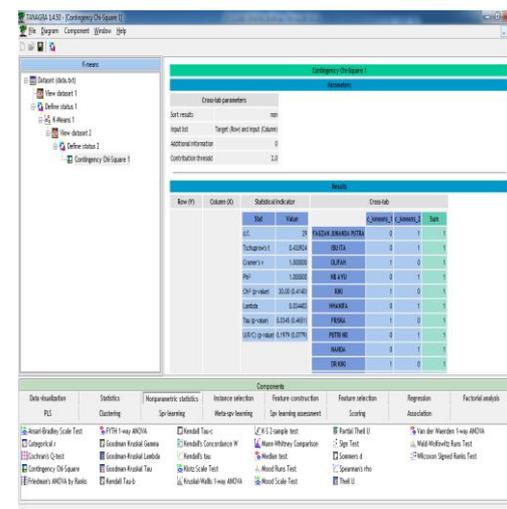


Gambar 5. Tampilan Parameter K-means



Gambar 3. Import Data MC Laundry

Untuk melihat isi data yang kita masukkan dapat dilakukan dengan cara drag dan drop *view dataset*, seperti pada gambar



Gambar 6. Tampilan Hasil K-Means Dengan Software Tanagra

Nama pelanggan	Jumlah transaksi	Total transaksi Cluster_Mean_1
FAKUSI	12	240000
IBU TIA	12	240000
CELEKHA	13	262500
BO AYO	20	480000
YESS	24	704000
HSANCHA	13	305500
FRIZKA	20	800000
PUTRI BO	11	437100
RAMON	12	249500
DR KIKI	11	600000
IBNU	12	180000
LORFI	23	506300
BOI	14	440100
RAMI	23	270000
RANI	11	553300
DESI	15	454300
EFFENDI	40	624000
DELA	14	370000
RAMON SETYAN	24	800000
RIZAFUSO	23	628000
INDRA	17	354000
SHALIM	19	445400
ALIN	13	330000
ASIT	14	293700
AFIN	23	714000
PRIDI	11	233000
BO BOY	14	340000
OSHO SAMON	11	274300
TRINDA	21	441000
BOY	23	544000

Gambar 7. Hasil Clustering

Knowledge discovery In Database adalah keseluruhan proses *non-trivial* untuk mencari dan mengidentifikasi pola (*pattern*) dalam data, dimana pola yang ditemukan bersifat sah, baru dapat bermanfaat dan dapat dimengerti (Ndaumanu R.I., Kusrini dan M. Rudyanto A., 2014).

Proses KDD secara garis besar (Nasari F. dan Surya D., 2015) dapat dijelaskan sebagai berikut:

1. *Data Selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses *Data mining* disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. *Pre-processing / Cleaning*

Sebelum proses *Data mining* dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi focus KDD. Proses pembersihan mencakup antara lain membuang *duplikasi data*, memeriksa data yang *inkosisten*, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (*tipografi*).

3. *Transformation*

Coding adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses *Data mining*. Proses *coding* dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *Data mining* sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Seperti halnya Susanto dan Sudiyatno (2014) meneliti tentang penerapan data mining untuk memprediksi prestasi belajar siswa berdasarkan status sosial ekonomi orang tua, motivasi, kedisiplinan siswa dan prestasi masa lalu dengan algoritma J48.

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses *Data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut *interpretation*. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya

Defenisi Data Mining

Data mining merupakan sebuah inti dari proses KDD, meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dapat didefinisikan sebagai pengorganisasian proses untuk

pengidentifikasi yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan kompleks. *Data mining* merupakan penyelesaian masalah dengan menganalisa data yang ada pada database, dimana data tersimpan secara elektronik dan pencariannya dilakukan otomatis seperti pada komputer (Arumawadu et. al., 2015).

Kemudian Harryanto dan Hansun (2017) menggunakan penerapan Algoritma C4.5 dalam mengidentifikasi calon pegawai suatu perusahaan dengan memprediksi dan mengklasifikasi terhadap calon yang potensial untuk masuk ke dalam perusahaan dengan cara membuat pohon keputusan berdasarkan data-data yang sudah ada. Dan Algoritma C4.5 juga digunakan untuk memprediksi masa studi mahasiswa pada perguruan tinggi (Haryati et. al., 2015).

SIMPULAN

Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut

1. Mengelompokkan data dengan algoritma K-Means dilakukan dengan cara menentukan jumlah *cluster*, hitung jarak terdekat dengan pusat *cluster*. Data dengan jarak terdekat menyatakan anggota dari *cluster* tersebut, dilakukan perhitungan kembali sampai data tidak berpindah pada *cluster* lain, untuk meminimalkan fungsi objektif.
2. Data pelanggan yang potensial didapatkan setelah perhitungan algoritma K-Means selesai, data dengan pusat *centroid* terbesar yang termasuk ke dalam pelanggan yang paling potensial.
3. Data pelanggan potensial didapatkan dengan cara menggunakan *tools Tanagra*, data yang diproses meliputi

data nama pelanggan, data jumlah transaksi dan data total transaksi. Data pelanggan potensial, didapatkan setelah menginput data pelanggan yang telah disimpan dalam *notepad* kemudian data tersebut *diimport* ke dalam *Tanagra*.

DAFTAR PUSTAKA

- Arumawadu, H. I., R. M., Kapila T. R., & Illangarathe, S. K. (2015). Mining Profitability of Telecommunication Customer Using K-Means Clustering. *Journal of Data Analysis and Information Processing*, 3(3), 63-71
- Harryanto, F. F., & Hangsung, S. (2017). Penerapan Algoritma C4.5 untuk Meprediksi Penerimaan Calon Pegawai Baru di PT. WISE. *JATISI*, 3(2), 95-103
- Haryati, S., Aji, S., & Eko, S. (2015). Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5. *Jurnal Media Infotama*, 11(2), 130-138
- Nasari, F., & Surya, D. (2015). Penerapan *K-Means Clustering* pada Data Penerimaan Mahasiswa Baru (Studi Kasus Universitas Potensi Utama). *Seminar Nasional Teknologi Informasi dan Multimedia*, 3(1), 73-78
- Ndaumanu, R. I., Kusrini, K., & Rudyanto, A. (2014). Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa dengan Metode K-Nearest Neighbor. *JATISI*, 1(1), 1-15
- Pralita, F. (2010). *Penerapan Jaringan Syaraf Tiruan untuk Memprediksi Jumlah Penumpang Bus dengan Metode Back Propagation*, Skripsi. Fakultas MIPA, Universitas Islam Indonesia

- Susanto, H., & Sudiyatno. (2014). *Data Mining* untuk Memprediksi Prestasi Siswa Berdasarkan Sosial Ekonomi, Motivasi, Kedisiplinan dan Prestasi Masa lalu. *Jurnal Pendidikan Vokasi*, 4(2), 222-231
- Widiarina, W., & Romi, S. W. (2015). Algoritma *Cluster* Dinamik untuk Optimasi Cluster pada Algoritma *K-Means* dalam Pemetaan Nasabah Potensial. *Journal of Intelligent Systems*, 1(1), 3