**KNOWLEDGE DISCOVERY IN DATABASES (KDD)**

**UNTUK MENENTUKAN SEGMENTASI PELANGGAN**

****

**Oleh :**

**Agil Furqaan Nur Rayyan**

**09031281621047**

**JURUSAN SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

**Pendahuluan**

*Knowledge Discovery in Databases* (KDD) adalah proses menemukan pengetahuan yang berguna atau bermanfaat dari kumpulan data. Teknik penambangan data yang banyak digunakan ini adalah proses yang meliputi persiapan dan pemilihan data, pembersihan data, menggabungkan pengetahuan sebelumnya pada set data dan menafsirkan solusi akurat dari hasil yang diamati. KDD adalah proses otomatis mengubah data mentah menjadi informasi yang berguna dimana sistem komputer menyaring dan menyortir data untuk mencari pola atau memprediksi tren. Hal ini umumnya dianggap sebagai ekstraksi non-trivial dari informasi yang implisit, sebelumnya tidak diketahui, dan informasi yang berpotensi berguna dari data.

KDD mengacu pada proses non-trivial dalam mengidentifikasi pola yang valid, baru, berpotensi berguna, dan dapat dipahami dalam data Gambar 1 (diadaptasi dari Fayyad et al.) menunjukkan bahwa proses KDD memiliki sifat iteratif dan interaktif. Hasil yang diperoleh dalam proses ditingkatkan secara bertahap, kemungkinan mempertimbangkan kembali keputusan sebelumnya dalam proses. Langkah-langkah dari proses KDD termasuk:

1. Pemilihan data yang relevan dengan tugas analisis
2. Preprocessing data, termasuk tugas-tugas seperti pembersihan data-data dan integrasi data
3. Transformasi data menjadi bentuk yang sesuai untuk penambangan .
4. Aplikasi algoritma Data Mining untuk ekstraksi pola .
5. Interpretasi / evaluasi pola yang dihasilkan sehingga dapat mengidentifikasi pola-pola yang mewakili pengetahuan nyata, berdasarkan beberapa langkah-langkah menarik.



Gambar 1. Langkah-langkah proses KDD

**Pembahasan**

*Datasets* adalah kumpulan data. Paling umum satu set data sesuai dengan isi dari tabel database tunggal, atau matriks data statistik tunggal, di mana setiap kolom tabel mewakili variabel tertentu, dan setiap baris sesuai dengan anggota tertentu dari kumpulan data yang bersangkutan. *Datasets* mencantumkan nilai untuk masing-masing variabel, seperti tinggi dan berat suatu objek, untuk setiap anggota kumpulan data. Setiap nilai dikenal sebagai datum. Kumpulan data dapat terdiri dari data untuk satu atau lebih anggota, sesuai dengan jumlah baris.

 Pada tulisan kali ini, dataset didapat dari situs data.world ( https://data.world/lpetrocelli/czech-financial-dataset-real-anonymized-transactions) **1999 Czech Financial Dataset - Real Anonymized Transactions**. Dimana data tersebut terdiri dari 1,056,320 baris dan 10 kolom.

**Analisis Jurnal**

**Jurnal 1**

**Data Mining on Customer Segmentation: A Review**

*Menurut jurnal ini, data mining* digunakan untuk mengekstrak informasi tersembunyi dari sejumlah besar data. Jurnal tersebut juga menunjukkan tentang perbandingan berbagai teknik berdasarkan segmentasi citra dan big data. Pendekatan ini penting untuk industri yang sangat disesuaikan yang menggunakan sejumlah besar data. Namun masih ada beberapa masalah yang belum mempertimbangkan penggunaan data besar dengan segmentasi pelanggan serta penggunaan evolusi diferensial diabaikan dalam literatur yang ada untuk mengklasifikasikan data. Sehingga dalam waktu dekat akan menggunakan evolusi diferensial dan juga akan melindungi perangkat lunak untuk hasil yang premature.

Pendekatan yang berbeda untuk segmentasi pelanggan dalam penggalian data adalah: pengelompokan dan penemuan subkelompok. Karena beberapa keterbatasan dan ruang lingkup teknik pengelompokan, itu mengarah pada perbaikan lebih lanjut dalam metodologi dalam *data mining*.

**Jurnal 2**

**The New Approach in Market Segmentation by Using RFM Model**

Jurnal ini menunjukkan pendekatan baru membagi target pasar yang luas ke dalam *subset* konsumen yang memiliki kebutuhan umum, minat, dan prioritas, dan kemudian merancang dan menerapkan strategi untuk menargetkan mereka untuk mencapai maksimalisasi keuntungan. tujuan penelitian ini ada dua; Pertama, adalah menggunakan data historis (seperti barang yang dibeli dan biaya moneter asosiatif), model yang diusulkan mengidentifikasi segmen pelanggan berdasarkan *Firefly* *Algorithm* (FA), dan Kedua, adalah identifikasi segmen yang paling menguntungkan menurut model RFM ( *Recency*, *Frequency* dan *Monetary*).

**Jurnal 3**

**RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies**

Jurnal ini tidak hanya menggunakan model RFM dalam menentukan segmentasi pelanggan tetapi juga menggunakan model CLV (*Customer* *Lifetime* *Value*) dimana keterkaitan RFM dan CLV adalah Perilaku pembelian sebelumnya sering diringkas oleh RFM: waktu sejak pembelian terakhir (R = rekonsiliasi), jumlah pembelian sebelumnya (F = frekuensi), dan jumlah total yang dibelanjakan (M = nilai uang). Karena variabel ini secara luas dikenal untuk pelanggan yang sudah ada, mereka sering dimasukkan dalam model CLV dan digunakan untuk membuat perkiraan tingkat pelanggan.

**Jurnal 4**

**Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)**

Pada jurnal ini pengelompokan data menggunakan metode K-means untuk melakukan pengelompokan dengan model *Recency*, *Frequency*, *Monetary* (RFM), Hasil dari model RFM yakni variabel *Recency*, *Frequency*, dan *Monetary* dapat mendeskripsikan profil atau karakteristik setiap segmen. Hal tersebut dapat dibuktikan dengan nilai *Recency* untuk mengetahui rentang waktu transaksi terakhir pelanggan, nilai *Frequency* untuk mengetahui jumlah transaksi pelanggan, dan *Monetary* untuk mengetahui besaran rupiah yang dikeluarkan pelanggan untuk mendapatkan layanan di Belle Crown Malang. Kemudian Metode *K-Means* dapat menjadi suatu pilihan dalam menyelesaikan masalah segmentasi pelanggan. Akan tetapi, hasil pencarian nilai k dengan metode *Elbow* yang telah dilakukan secara *intuitive* sebelum proses *clustering* belum sesuai dengan hasil uji performa yakni nilai jumlah cluster pada metode *Elbow* sebanyak 3, sedangkan hasil uji performa sebanyak 2 dan 5. Dari perbedaan hasil tersebut, maka dilakukan *clustering* ulang untuk *cluster* sebanyak 2.

**Jurnal 5**

**Opinion Mining: Issues and Challenges (A survey)**

Menurut jurnal ini, Pendapat opini sangat penting baik bagi individu maupun perusahaan. Individu mungkin ingin melihat pendapat pelanggan lain tentang suatu produk untuk menganalisanya sebelum membelinya. Perusahaan ingin menganalisis umpan balik pelanggan tentang produk mereka untuk membuat keputusan di masa depan. Jadi, menganalisis pendapat pelanggan dan tanggapan mereka adalah penting. Penambangan digunakan pada ulasan produk yang tersedia di berbagai blog, forum web, dan situs ulasan produk untuk mengevaluasi pendapat pelanggan. Dengan begitu, pelanggan baru dapat menemukan pandangan orang lain tentang suatu produk dan dapat memutuskan produk mana yang akan dibeli dengan bantuan pendapat pelanggan yang sudah menggunakan produk tersebut. Selain itu perbandingan fitur produk yang sama oleh vendor yang berbeda dilakukan. Dengan cara ini perusahaan dapat fokus pada peningkatan fitur produk mereka yang tidak populer di kalangan pelanggan. Ini mengarah untuk mengatasi persyaratan intelijen pemasaran dan pembandingan produk dalam industri produksi. Dalam makalah ini kami melakukan survei terhadap makalah dan akan merangkum isu dan tantangan penambangan opini yang mempengaruhi hasil penambangan opini.

**Daftar Pustaka**

Iman Makhani, M. M. ohammadian. (2016). RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies. *International Academic Journal of Accounting and Financial Management*, *3*(6), 21–35.

Kaur, E. R., Kiranbir, E., & Assistant, K. (2017). Data Mining on Customer Segmentation: A Review. *International Journal of Advanced Research in Computer Science*, *8*(5), 857–861.

Roshan, H., & Afsharinezhad, M. (2017). The New Approach in Market Segmentation by Using RFM Model, *4*(4), 259–267. https://doi.org/10.22105/jarie.2017.91297.1011

Savitri, A. D., Bachtiar, F. A., & Setiawan, N. Y. (2018). Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan ( Studi Kasus : Belle Crown Malang ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(9), 2957–2966.

Seerat, B. (2012). Opinion Mining : Issues and Challenges ( A survey ), *49*(9), 42–51.