# Penerapan Knowledge Discovery in Database (KDD) di sebuah Toko Handphone

****

**Disusun :**

**Oleh : Indah Irawati Pardede**

**NIM : 09031181621130**

**Kelas : SI REG 5B**

**SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

1. **Informasi tentang Knowladge Discovery In Database (KDD )**
* KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interprestasi dan visualisasi dari pola-pola sejumlah kumpulan data.
* Knowledge discovery in databases (KDD) 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.



**2. Tahapan Proses KDD**

**1. Data Selection**

* Menciptakan himpunan data target , pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan.
* 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**

* Pemprosesan pendahuluan dan pembersihan data merupakan operasi dasar seperti penghapusan noise dilakukan.
* Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD.
* Proses cleaning mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).
* Dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.

**3. Transformation**

* Pencarian fitur-fitur yang berguna untuk mempresentasikan data bergantung kepada goal yang ingin dicapai.
* Merupakan proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data

**4. Data mining**

* Pemilihan tugas data mining; pemilihan goal dari proses KDD misalnya klasifikasi, regresi, clustering, dll.
* Pemilihan algoritma data mining untuk pencarian (searching)
* Proses Data mining yaitu 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.

**5. Interpretation/ Evaluation**

* Penerjemahan pola-pola yang dihasilkan dari data mining.
* Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah mimengerti oleh pihak yang berkepentingan.
* Tahap ini merupakan bagian dari proses KDD yang mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesa yang ada sebelumnya.

### Metode

Banyak teknik dan metode yang ada untuk melakukan berbagai jenis tugas data mining. Metode ini dikelompokkan dalam 3 paradigma utama data mining: Predictive Modeling, Discovery, dan Deviation Detection.

#### Predictive Modeling

Aplikasi Predictive Modeling menghasilkan klasifikasi atau prediksi. Tujuan dari predictive modeling adalah menemukan pola yang melibatkan variabel untuk memprediksi dan mengklasifikasi perilaku masa depan dari sebuah entitas. Ada dua tipe masalah yang diselesaikan oleh predictive modeling: klasifikasi dan regresi.

Klasifikasi melibatkan model pembelajaran yang memetakan (atau mengklasifikasi) data contoh ke dalam satu atau beberapa kelas yang telah didefinisikan. Sebagai contoh, bank dapat menggunakan skema klasifikasi untuk menentukan pengajuan pinjaman yang akan disetujui. Teknik klasifikasi meliputi Naive Bayesian, neural networks, dan decision trees.

Regresi melibatkan model yang memetakan data contoh ke prediksi real-valued. Teknik regresi meliputi neural networks dan decision (atau regression) trees.

#### Discovery

Aplikasi discovery adalah pendekatan eksploratoris untuk analisis data. Aplikasi discoverymenggunakan teknik yang menganalisis data set yang besar untuk menemukan association rules(atau pola), atau menemukan kluster dari sampel yang dapat dikelompokan. Hasil dari metode discovery umumnya dimaksudkan untuk pengguna. Walau begitu, hasilnya juga dapat diaplikasikan ke metode data mining yang lain

#### Deviation Detection

Deviation Detection melakukan deteksi anomali secara otomatis. Tujuannya untuk mengidentifikasi kebiasaan suatu entitas dan menetapkan sejumlah norm melalui pattern discovery. Sampel yang berdeviasi dari norm lalu diidentifikasi sebagai tidak biasa. teknik Deviation Detection melalui visualisasi melalui parallel coordinates, scatterplots, dan surface plots.



### Aplikasi

Aplikasi dari KDD dalam bisnis antara lain:

* 1. **Market segmentation** – Mengidentifikasi karakteristik umum dari pelanggan yang membeli produk yang sama dari perusahaan
	2. **Customer churn** – Memprediksi pelanggan yang mungkin meninggalkan perusahaan untuk beralih ke kompetitor
	3. **Fraud detection** – Mengidentifikasi transaksi yang mungkin merupakan transaksi curang.
	4. **Direct marketing** – Memprediksi prospek yang harus dimasukkan dalam mailing-listuntuk mendapatkan respon maksimum
	5. **Interactive marketing** – Memprediksi hal-hal yang menarik bagi setiap individu yang mengakses website.
	6. **Market basket analysis** – Memahami produk atau layanan yang biasanya dibeli bersamaan.
	7. **Trend analysis** – Mengungkap perbedaan antara tipikal pelanggan bulan ini dan bulan lalu dan prediksi tipikal pelanggan bulan depan.

# Penerapan Knowledge Discovery in Database (KDD) di sebuah Toko Handphone

**Contoh kasus yang saya ambil adalah transaksi yang terjadi di toko Handphone**

Data sampelnya berupa accessories handphone

|  |  |
| --- | --- |
| **NO.** | **JENIS ITEM** |
| 1 | CASING |
| 2 | BATERAI |
| 3 | MEMORY |
| 4 | HEADPHONE |
| 5 | CHARGER |
| 6 | SIM CARD |
| 7 | ANTI-GORES |
| 8 | KONDOM (HP) |

**Sample Data Penjualan Toko Handphone**

|  |  |
| --- | --- |
| No. Transaksi | Items |
| 1 | Headphone, memory, baterai |
| 2 | Casing, kondom (hp) |
| 3 | Sim card, anti-gores, casing |
| 4 | Headphone, charger, baterai |
| 5 | Memory |
| 6 | Baterai, charger,  headphone |
| 7 | Kondom (hp), casing, sim card, charger |
| 8 | Charger |
| 9 | Anti-gores, memory, headphone, sim card |
| 10 | Headphone, sim card, kondom (hp) |
| 11 | Charger, casing, memory |
| 12 | Anti-gores, kondom (hp), memory |
| 13 | Baterai, headphone, sim card |
| 14 | Kondom (hp) |
| 15 | Anti-gores |
| 16 | Sim card, charger |
| 17 | Headphone, casing |
| 18 | Headphone |
| 19 | Kondom (hp), anti-gores |
| 20 | Memory, charger, sim card |



**Pemilihan Data (Data Selection)**

|  |  |
| --- | --- |
| No. Transaksi | Items |
| 1 | Headphone, memory, baterai |
| 2 | Casing, kondom (hp) |
| 3 | Sim card, anti-gores, casing |
| 4 | Headphone, charger, baterai |
| 5 | Memory |
| 6 | Baterai, charger,  headphone |
| 7 | Kondom (hp), casing, sim card, charger |
| 8 | Charger |
| 9 | Anti-gores, memory, headphone, sim card |
| 10 | Headphone, sim card, kondom (hp) |

Dalam data selection kita harus mencari lebih dulu yang disebut "frequent item" (item sering), yaitu item yang dibeli oleh pengunjung sebanyak 2 kali atau lebih.

|  |  |
| --- | --- |
| Items | Frequent Items |
| Casing | 3 |
| Baterai | 3 |
| Memory | 3 |
| Headphone | 4 |
| Sim card | 3 |
| Charger | 3 |
| Anti-gores | 2 |
| Kondom (hp) | 3 |



Frequent item ini juga mencakup untuk pembelian secara bersama-sama, sehingga didapatkan juga frequent items seperti berikut:

|  |  |
| --- | --- |
| Items | Frequent |
| Casing dan sim card | 2 |
| Casing dan kondom (hp) | 2 |
| Baterai dan headphone | 3 |
| Baterai dan charger | 2 |
| Memory dan headphone | 2 |
| Headphone dan charger | 2 |
| Headphone dan sim card | 2 |
| Sim card dan anti-gores | 2 |
| Sim card dan kondom (hp) | 2 |

Nilai *Support*item barang menunjukkan frekuensi item barang yang muncul terhadap jumlah transaksi yang terjadi. Untuk menghitung kemungkinan dari association rule dapat dilakukan dengan membuat perhitungan nilai confidence (kepercayaan).

= Support

= Confidence

Dari data penjualan di atas association rule sebagai berikut:

|  |  |  |
| --- | --- | --- |
| Association | Support(%) | Confidence(%) |
| Casing  =>  Sim card | 20 | 66,67 |
| Casing  =>  Kondom (hp) | 20 | 66,67 |
| Baterai  =>  Headphone | 30 | 100 |
| Baterai  =>  Charger | 20 | 66,67 |
| Memory  =>  Headphone | 20 | 66,67 |
| Headphone  =>  Charger | 20 | 50 |
| Headphone  => Sim card | 20 | 50 |
| Sim card  => Anti-gores | 20 | 66,67 |
| Sim card  => Kondom (hp) | 20 | 66,67 |

Dari association rule di atas , didapatkan pada rule ke-2 yaitu, pembeli yang membeli Baterai akan membeli juga Headphone. Tingkat kekuatan, keterkaitan pembelian Baterai dengan Headphone adalah 100% (*Conf%*). Nilai *confidence* adalah hasil bagi support (Baterai dan Headphone) dengan support (Baterai) dikali 100%. Perhitungan nilai *confidence*= (3/3)\*100% = 100%.

Association Rule ini dapat dituliskan sebagai berikut:

{ Baterai }  =>  { Headphone}

Pengetahuan yang diperoleh dari hasil analisis di atas, pemilik toko mengambil keputusan (strategi) agar dapat menarik konsumen baru bergabung atau membeli yaitu :

1.      Memberikan bonus Headphone kepada pengunjung yang melakukan pembelian Charger.

2.      Memberikan bonus Headphone dan Sim card kepada pengunjung yang melakukan pembelian chrager dan baterai

3.      Memberikan bonus Kondom (hp) dan Anti-gores kepada pengunjung yang melakukan pembelian headphone dan baterai

Dengan demikian maka Didapatkan suatu data berupa jenis accessories handphone  yang paling sering dibeli secara bersamaan, yaitu :

 { Baterai }  =>  { Headphone}

Analisis :

 Dalam penerapan knowledge discovery in database (kdd) di sebuah toko handphone kita dapat mengetahui karakteristik umum dari pelanggan yang membeli produk yang sama dari toko handphone dan dapat mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. Data akan melalui proses selection dimana ada pemilihan himpunan data, atau memfokuskan pada subset variabel atau sampel data, dimana penemuan (discovery) akan dilakukan. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, data selection yang disebut "frequent item" (item sering), yaitu item yang dibeli oleh pengunjung. Frequent item mencakup untuk pembelian secara bersama-sama, sehingga kita dapat menentukan nilai *Support*item barang yang menunjukkan frekuensi item barang yang muncul terhadap jumlah transaksi yang terjadi. Untuk menghitung kemungkinan dari association rule dapat dilakukan dengan membuat perhitungan nilai confidence (kepercayaan). Dengan demikian maka didapatkan suatu data berupa jenis accessories handphone  yang paling sering dibeli secara bersamaan yaitu baterai dan headphone.

**Referensi** :

**Introduction to Data Mining and Knowledge Discovery Third Edition**
Penulis : Herbert A. Edelstein
Publisher : Two Crows Corporation (October 8, 1999)

Penulis : Kurt Thearling
**Judul : An Introduction to Data Mining, Discovering hidden value in your data warehouse**
<http://www.thearling.com/text/dmwhite/dmwhite.htm>

Publisher : Wikipedia Foundation
Judul : Data Mining
[**http://en.wikipedia.org/wiki/Data\_mining**](http://en.wikipedia.org/wiki/Data_mining)

Publisher : ALG (Automated Learning Group), NCSA (National Center for Supercomputing Applications)
Judul : Data Mining
[**http://alg.ncsa.uiuc.edu/tools/docs/d2k/manual/dataMining.html**](http://alg.ncsa.uiuc.edu/tools/docs/d2k/manual/dataMining.html)

Penulis : Jiawei Han, Micheline Kamber
Judul : **Data Mining: Concept and Techniques**
Publisher : Academic Press Sand Diego, CA, 2001

 <https://shabri-prayogi.blogspot.com/2012/05/penerapan-knowledge-discovery-in.html> diakses tanggal 18 November 2018 pukul : 22.12

<http://www.sigitprabowo.id/2013/04/data-mining-tahap-tahapan-knowladge.html?m=1> diakses tanggal 18 November 2018 pukul : 22.12

<https://dnial.wordpress.com/2007/04/27/kdd-dan-data-mining/> diakses tanggal 18 November 2018 pukul : 22.12

<http://edocs.ilkom.unsri.ac.id/id/eprint/2862>