

SKRIPSI

**MENGIKUTI MAYORITAS:
DATA MINING HISTORI PENCARIAN RUTE ANGKOT
DATA MINING**



JOVAN GUNAWAN

NPM: 2011730029

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2014**

UNDERGRADUATE THESIS

**FOLLOWING THE MAJORITY:
DATA MINING HISTORY SEARCHING ROUTE
DATA MINING**



JOVAN GUNAWAN

NPM: 2011730029

**DEPARTMENT OF INFORMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2014**

LEMBAR PENGESAHAN

**MENGIKUTI MAYORITAS:
DATA MINING HISTORI PENCARIAN RUTE ANGKOT
DATA MINING**

JOVAN GUNAWAN

NPM: 2011730029

Bandung, 14 September 2014

Menyetujui,

Pembimbing Tunggal

Pascal Alfadian, M.Com.

Ketua Tim Penguji

Anggota Tim Penguji

Plato

Euclid

Mengetahui,

Ketua Program Studi

Thomas Anung Basuki, Ph.D.

ABSTRAK

Data mining merupakan teknik untuk mengolah *data*, mencari suatu pola, kemudian dievaluasi untuk mendapatkan *knowledge*. Pola yang dihasilkan oleh *data mining* dapat digunakan untuk mencari suatu hal yang menarik dan unik. Dari pola tersebutlah, akan didapatkan suatu informasi yang sangat bermakna yang biasa disebut sebagai *knowledge*.

Pada penelitian ini, akan dilakukan analisis data penggunaan KIRI dengan menggunakan teori-teori *data mining*. Pada penelitian ini, diharapkan dapat ditemukan suatu pola yang menarik dan menghasilkan suatu ide atau gagasan yang baru. Maka dari itu, penelitian ini akan membutuhkan banyak *data* penggunaan KIRI, yang dapat diperoleh dari *log* histori KIRI.

Log tersebut memiliki 5 *field* untuk setiap *entry* sebagai berikut:

- *statisticId*, primary key dari *entry*
- *verifier*, mengidentifikasi sumber dari pencarian ini
- *timestamp*, waktu ketika pengguna KIRI mencari rute angkot
- *type*, tipe log, untuk penelitian ini selalu berisi FINDROUTE
- *additionalInfo*, mencatat koordinat awal, koordinat akhir, dan banyak rute yang ditemukan pada pencarian ini

Dengan menggunakan *data log* histori pencarian rute angkot dari KIRI, diharapkan bisa diperoleh pola yang menarik. Pola tersebut memiliki kemungkinan untuk digunakan kembali sebagai informasi tambahan pada perangkat lunak navigasi lain seperti KIRI.

Kata-kata kunci: Lintasan, Homotopy, Fréchet, Lintasan Median, Penyangga

Dipersembahkan untuk diri sendiri

KATA PENGANTAR

Bandung, September 2014

Penulis

DAFTAR ISI

| | |
|---|-------------|
| KATA PENGANTAR | xi |
| DAFTAR ISI | xiii |
| DAFTAR GAMBAR | xiv |
| DAFTAR TABEL | xv |
| 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Perumusan Masalah | 1 |
| 1.3 Tujuan | 1 |
| 1.4 Batasan Masalah | 2 |
| 1.5 Metode Penelitian | 2 |
| 1.6 Sistematika Pembahasan | 2 |
| 2 LANDASAN TEORI | 3 |
| 2.1 <i>Data Mining</i> | 3 |
| 2.1.1 Tahap <i>Data Mining</i> | 3 |
| 2.1.2 <i>Data Cleaning</i> | 3 |
| 2.1.3 <i>Data integration</i> | 5 |
| 2.1.4 <i>Data selection</i> | 5 |
| 2.1.5 <i>Data Transformation</i> | 5 |
| 2.1.6 <i>Data Mining</i> | 6 |
| 2.1.7 <i>Pattern Evaluation</i> | 6 |
| 2.1.8 <i>Knowledge Presentation</i> | 6 |
| 2.2 <i>Spatial and Spatiotemporal</i> | 6 |
| DAFTAR REFERENSI | 7 |

DAFTAR GAMBAR

| | |
|---|---|
| 2.1 Tahap <i>Data Mining</i> , Sumber Data Mining Concepts and Techniques | 4 |
|---|---|

DAFTAR TABEL

BAB 1

PENDAHULUAN

1.1 Latar Belakang

KIRI merupakan suatu organisasi yang memiliki tujuan untuk mengurangi pemanasan global, kemacetan, dan harga BBM tinggi. Untuk mencapai hal tersebut, KIRI memberikan solusi dengan mengajak Indonesia untuk menggunakan fasilitas transportasi publik. Maka dari itu, KIRI bergerak di bidang transportasi publik dengan memberikan petunjuk jalan menggunakan transportasi publik yang dapat diakses melalui website atau handphone. Hingga saat ini, KIRI sudah menyentuh 3 wilayah, yaitu Bandung, Jakarta, dan Surabaya.

Pertumbuhan teknologi hingga saat ini telah menghasilkan banyak sekali data-data, namun sering kali pemilik data hanya menggunakan data tersebut seperlunya saja. Jika dilihat lebih rinci, sebenarnya jika data tersebut diolah lebih lanjut, dapat menghasilkan sesuatu yang lebih. Perusahaan KIRI, telah menghasilkan banyak data log histori setiap bulan namun masih belum dilakukan tindakan pengolahan data. Salah satu cara mengolah data tersebut adalah dengan menggunakan teknik *data mining*. Dengan menggunakan teknik *data mining* akan mempermudah menganalisa masalah, pengambilan kesimpulan, bahkan mempermudah konsumen dalam menggunakan jasa.

Dengan menggunakan proses *data mining*, data log histori KIRI yang berisi informasi seorang pengguna berangkat dari suatu lokasi ke lokasi tertentu, dimungkinkan dapat menghasilkan suatu *pattern* yang menarik dan berguna baik untuk KIRI ataupun pemerintah.

1.2 Perumusan Masalah

Dengan mengacu pada uraian diskripsi diatas, maka permasalahan yang dibahas dan diteliti oleh penulis adalah

- Bagaimana cara mengolah pola yang diperoleh dari *data log histori* KIRI agar pola menjadi menarik dan bermakna?
- Bagaimana membuat perangkat lunak untuk melakukan data mining pada data log history?

1.3 Tujuan

Penelitian ini bertujuan untuk

- Mencari pola dan informasi yang menarik dari *log histori* KIRI

- Perangkat lunak dapat melakukan data mining dari *log histori* KIRI

1.4 Batasan Masalah

Penelitian *data mining* yang diatas akan ditentukan batasan masalah yang diteliti berupa :

- Penelitian ini dibatasi hanya pada permasalahan pada penerapan *data mining* pada *data log* KIRI
- Data log yang merupakan masukan akan dibatasi sebanyak 10000 buah data

1.5 Metode Penelitian

Berikut adalah Metode Penelitian yang digunakan :

- Melakukan studi literatur tentang algoritma-algoritma yang berkaitan dengan pemrosesan *data mining*
- Melakukan penelitian *data mining* yang diterapkan pada *log* KIRI
- Merancang dan mengimplementasikan algoritma untuk *data mining*
- Mengimplementasikan pembangkit pola *data mining*
- Melakukan pengujian dan eksperimen

1.6 Sistematika Pembahasan

Sistematika pembahasan dalam penelitian ini adalah: Bab 1: Pendahuluan, berisi latar belakang dari penelitian ini, rumusan masalah yang timbul, tujuan yang ingin dicapai, ruang lingkup atau batasan masalah dari penelitian ini, serta metode penelitian yang akan digunakan dan sistematika pembahasan dari penelitian ini. Bab 2: Landasan Teori, berisi dasar teori mengenai *data mining* dan *spatial and Spatiotemporal* Bab 3: Bab 4: Bab 5:

BAB 2

LANDASAN TEORI

2.1 *Data Mining*

2.1.1 Tahap *Data Mining*

Proses pada *data mining* dapat dibagi menjadi beberapa tahap seperti pada gambar 2.1

Terdapat 7 tahap pada *data Mining*:

- *Data cleaning*
- *Data integration*
- *Data selection*
- *Data transformation*
- *Data mining*
- *Pattern Evaluation*
- *Knowledge presentation*

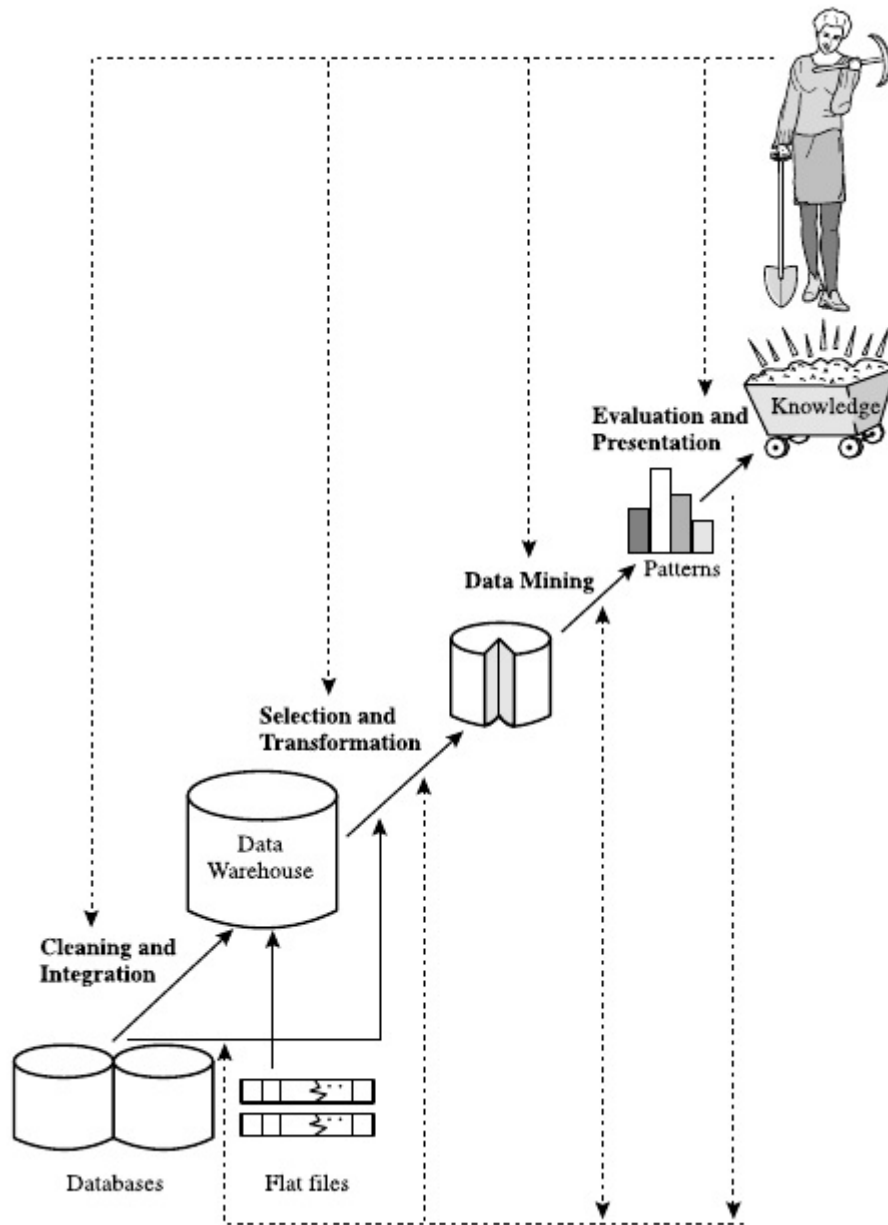
2.1.2 *Data Cleaning*

Data cleaning merupakan tahap *data mining* untuk menghilangkan *missing value* dan *noisy data*. Pada umumnya, *data* yang diperoleh dari *database* terdapat nilai yang tidak sempurna seperti nilai yang hilang, nilai yang tidak *valid* atau bahkan salah ketik. Nilai-nilai tersebut dapat diatasi dengan cara *smoothing techniques*. Atribut dari suatu *database* yang tidak relevan atau redundansi bisa diatasi dengan menghapus atribut tersebut.

Missing Values

Missing values tentu akan mengganggu proses *data mining* pada komputer dan dapat menghasilkan nilai akhir yang tidak sesuai. Terdapat beberapa teknik untuk mengatasi *missing values* yaitu

- Membuang tuple yang terdapat nilai yang hilang
- Mengisi nilai yang hilang secara manual
- Mengisi nilai yang hilang dengan menggunakan nilai konstan yang bersifat umum
- Menggunakan nilai rata-rata dari suatu atribut untuk mengisi nilai yang hilang



Gambar 2.1: Tahap *Data Mining*, Sumber Data Mining Concepts and Techniques

Noisy Data

2.1.3 Data integration

Data integration merupakan tahap menggabungkan data dari berbagai sumber. Sumber tersebut bisa termasuk beberapa *database*, *data cubes*, atau bahkan *flat data*. Tahap ini harus dilakukan secara teliti terutama ketika dalam memasangkan nilai-nilai yang berasal dari sumber yang berbeda. Pada tahap ini, perlu dilakukan identifikasi data apakah data tersebut dapat diturunkan atau tidak agar data yang diperoleh tidak terlalu besar. *Data integration* yang baik merupakan integrasi yang dapat memaksimalkan kecepatan dan meningkatkan akurasi dari proses *data mining*.

2.1.4 Data selection

2.1.5 Data Transformation

Data transformation merupakan tahap pengubahan data agar siap dilakukan proses *data mining*. *Data transformation* bisa melibatkan,

- *Smoothing*, proses untuk membuang *noise* seperti yang dilakukan pada tahap *data cleaning*
- *Aggregation*, proses mengganti nilai-nilai menjadi suatu nilai yang dapat mewakili nilai sebelumnya
- *Generalization*, proses dimana membuat suatu nilai yang bersifat khusus menjadi nilai yang bersifat umum
- *Normalization*, proses dimana suatu nilai dapat diubah skalanya menjadi nilai yang lebih kecil dan spesifik
- *Attribute construction*, proses membuat atribut baru yang berasal dari beberapa atribut untuk membantu proses data mining

—tanyakan!!!—>Proses *smoothing* merupakan proses yang sama seperti pada tahap *data cleaning*.<—tanyakan!!!—

Data Reduction

Proses *aggregation* dan *generalization* akan dilakukan dalam bentuk proses *data reduction* dan *Data Cube Aggregation*. *Data reduction* dilakukan untuk mendapatkan nilai yang representatif namun tetap menjaga keakuratan hasil *data mining*. Terdapat beberapa cara dalam mengimplementasikan *data reduction* yaitu

- *Data cube aggregation*
- *Data subset selection*
- *Dimensionality reduction*
- *Numerosity reduction*
- *Discretization and concept hierarchy generation*

Data Cube Aggregation

Data cube menyimpan semua informasi agregat dalam bentuk kubus, guna melakukan analisis. gambar 2.1 merupakan contoh dari *data cube*. Setiap sel, terdapat nilai data agregat yang saling berhubungan untuk menjadi data multidimensional point. *Data cube* dapat membuat proses *data mining* menjadi cepat untuk diakses dan membuat kesimpulan data.

Attribute Subset Selection

Attribute subset selection merupakan salah satu cara melakukan *data reduction* dengan menghilangkan atribut-atribut yang tidak relevan atau data yang redundansi. Hal ini dapat mempermudah pencarian pola dikarenakan banyak atribut yang muncul akan berkurang.

Dimensionality Reduction

Dimensionality Reduction merupakan metode pengurangan nilai secara random dengan cara melakukan konversi data. Jika data original dapat dibuat ulang dari data yang sudah dikompresi tanpa kehilangan informasi, maka akan dikatakan *lossless*, namun jika hanya mendapatkan data pendekatannya saja, akan disebut *lossy* [1].

Numerosity Reduction

Numerosity Reduction merupakan metode dimana data diganti atau ditentukan dengan cara parametrik atau nonparametrik.

Discretization and Concept Hierarchy Generation

lewat dulu

2.1.6 Data Mining

2.1.7 Pattern Evaluation

2.1.8 Knowledge Presentation

2.2 Spatial and Spatiotemporal

DAFTAR REFERENSI