



**PENERAPAN DATA MINING PADA RSUP DR.MOH HOESIN
SUMATERA SELATAN UNTUK PENGELOMPOKAN HASIL
DIAGNOSA PASIEN PENGGUNA ASURANSI KESEHATAN MISKIN
(ASKIN)**

PROPOSAL PENELITIAN

Diajukan guna melakukan penelitian skripsi

OLEH :

SANDRO ROMARIO.S

09142255

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS BINADARMA

2013

HALAMAN PENGESAHAN

**PENERAPAN DATA MINING PADA RSUP DR.MOH HOESIN
SUMATERA SELATAN UNTUK PENGELOMPOKAN HASIL
DIAGNOSA PASIEN PENGGUNA ASURANSI KESEHATAN MISKIN
(ASKIN)**

OLEH :

SANDRO ROMARIO.S

09142255

PROPOSAL PENELITIAN

Disusun sebagai salah satu syarat untuk melakukan penelitian

Disetujui,

Dosen Pembimbing I

Program Studi Teknik Informatika
Universitas Bina Darma,
Ketua program studi,

PH. Saksono, ST.MSc.PhD.

Syahril Rizal, S.T, M.M, M.Kom

Dosen Pembimbing II

Fatmasari, MM. M.Kom

KATA PENGANTAR

Puji Dan Syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa karena atas berkah, rahmat, dan karunia-Nya penulis dapat menyelesaikan proposal penelitian yang berjudul **“Penerapan Data Mining Pada RSUP DR.MOH HOESIN Sumatera Selatan Untuk Pengelompokan Hasil Diagnosa Pasien Pengguna Asuransi Kesehatan Miskin (ASKIN)”**

Dalam penulisan proposal penelitian ini penulis telah berusaha semaksimal mungkin untuk memberikan dan menyajikan yang terbaik. Tetapi penulis juga menyadari bahwa proposal penelitian ini masih jauh dari sempurna, hal ini dikarenakan terbatasnya pengetahuan yang dimiliki oleh penulis. Oleh karena itu, penulis mengharapkan saran dan kritik yang bersifat membangun untuk kesempurnaan skripsi ini.

Pada kesempatan yang baik ini, tidak lupa penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasehat, dan pemikiran dalam penyelesaian skripsi ini, terutama kepada :

1. Prof. Ir. Bochari Rachman, Msi, Rektor Universitas Bina Darma Palembang.
2. M. Izman Herdiansyah, ST.,MM.,Ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Bina Darma Palembang.
3. Syahril Rizal, S.T, M.M, M.Kom., selaku Kepala Program Studi Teknik Informatika.

4. PH. Saksono, ST.MSc.PhD.. selaku pembimbing I yang telah memberikan bimbingan penulisan proposal skripsi ini.
5. Fatmasari, MM. M.Kom selaku pembimbing II yang telah memberikan bimbingan dalam proposal skripsi ini.
6. Orang Tua, saudara-saudaraku, seluruh teman dan sahabat-sahabatku yang selalu memberikan dorongan dan masukan serta bantuan baik moril maupun materil yang tak ternilai harganya.

Semoga apa yang telah diberikan mereka kepada penulis, akan mendapat imbalan dari Tuhan Yang Maha Esa, Amin. Akhir kata semua kritik dan saran atas proposal ini akan penulis terima dengan senang hati, dan akan menjadi bahan pertimbangan bagi penulis untuk menyempurnakan proposal ini.

Palembang, Mei 2013

Penulis

ABSTRAK

RSUP Dr.Moh Hoesin merupakan Rumah Sakit Pemerintah Sumatera Selatan.Pihak RSUP Dr.Moh Hoesin bekerja sama dengan Pemerintah Provinsi Sumatera Selatan memberikan Askin. Dengan adanya layanan berupa askin tersebut, jumlah pasien pengguna askin akan meningkat pula. Oleh karena itu RSUP Dr.Moh Hoesin perlu menerapkan teknik data mining untuk menampilkan informasi pengelompokan pasien pengguna askin berdasarkan hubungan antara data jenis penyakit dengan data jumlah pasien pada RSUP Moh.Hoesin Palembang. Dengan adanya penerapan data mining ini diharapkan nantinya akan memberikan manfaat pada RSUP Dr.Moh Hoesin Palembang. Dapat memberikan informasi dan pengetahuan yang belum diketahui secara pasti dalam gudang data, sehingga menjadi informasi yang benar-benar berguna untuk dapat membantu mengelompokkan pengguna askin pada RSUP Moh.Hoesin Palembang. Metode penelitian ini berdasarkan *Knowledge Discovery in Databases (KDD)* yang merupakan keseluruhan proses konversi data mentah menjadi pengetahuan yang bermanfaat yang terdiri dari serangkaian tahap transformasi meliputi data *preprocessing* dan *postprocessing*. Berdasarkan uraian diatas maka dilakukan penelitian dengan judul **“Penerapan Data Mining Pada RSUP DR.MOH HOESIN Sumatera Selatan Untuk Pengelompokan Hasil Diagnosa Pasien Pengguna Asuransi Kesehatan Miskin (ASKIN)”** Dengan adanya Data mining menggunakan teknik clustering tersebut, diharapkan akan menghasilkan informasi berupa data diagnosa pasien pengguna askin, melakukan pencegahan terhadap penyakit yang sering diderita oleh pengguna askin di Sumatera Selatan.

Kata kunci : Askin,Data Mining, KDD, Clustering,

DAFTAR ISI

	Halaman
HALAMAN JUDUL	
HALAMAN PENGESAHAN.....	i
KATA PENGANTAR.....	ii
ABSTRAK.....	iv
DAFTAR ISI.....	v
I PENDAHULUAN	
1.1. Latar Belakang.....	1
1.2. Perumusan Masalah.....	3
1.3. Batasan Masalah.....	4
1.4. Tujuan dan Manfaat.....	4
1.4.1. Tujuan Penelitian.....	4
1.4.2. Manfaat Penelitian.....	4
II TINJAUAN PUSTAKA	
2.1. landasan teori.....	6
2.1.1 Data Mining.....	6
2.1.2 Fungsi Data Mining	7

2.1.3 Teknik Clustering	7
2.1.4 Algoritma CLHM	8
2.1.5 Analisa Cluster	9
2.1.6 Metode Knowledge Discovery In Databases (KD).....	10
2.1.7 Weka	13
2.2. Penelitian sebelumnya	14

III. METODOLOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian	18
3.2. Alat dan Bahan.....	18
3.3. Metode Penelitian.....	19
3.4. Metode Analisis Data	19

IV. JADWAL PENELITIAN

V . DAFTAR PUSTAKA

VI. LAMPIRAN

VII. JADWAL PENELITIAN

PROPOSAL SKRIPSI
PENERAPAN DATA MINING PADA RSUP DR.MOH HOESIN
SUMATERA SELATAN UNTUK PENGELOMPOKAN HASIL
DIAGNOSA PASIEN PENGGUNA ASURANSI KESEHATAN MISKIN
(ASKIN)

I. Pendahuluan

1.1 Latar belakang

Seiring dengan pesatnya perkembangan jaman dan majunya ilmu serta teknologi, maka pesat pula bermunculan instansi-instansi baik itu dalam dunia bisnis, kesehatan, asuransi, perusahaan, perbankan, ataupun instansi pemerintahan. Tidak terkecuali juga dengan instansi dalam yang bergerak dalam bidang dunia kesehatan. Seperti kita ketahui perkembangan teknologi dalam dunia termasuk Negara Indonesia, kemajuan dalam dunia kesehatan turut berperan penting dalam menekan angka jumlah kematian yang ada di dunia khususnya Negara Indonesia.

Oleh Karena itu kemajuan teknologi, mendorong setiap instansi-instansi dalam dunia kesehatan yaitu rumah sakit untuk meningkatkan mutu pelayanan terhadap pasien melalui cara melibatkan kemajuan teknologi dalam dunia kesehatan. maka pemerintah Indonesia saat ini termasuk pemerintah daerah yaitu pemerintah provinsi Sumatera Selatan segera mengambil tindakan berupa pemberian asuransi untuk rakyat miskin (askin) kepada rakyat yang miskin, agar pasien yang kurang mampu dapat berobat sebagaimana mestinya dan mendapatkan perawatan yang layak. Dengan adanya layanan berupa askin tersebut, maka dapat dipastikan jumlah pasien pengguna askin yang umumnya pasien yang kurang mampu akan meningkat pula. Dengan meningkatkannya

pasien pengguna askin, maka pihak RSUP Moh.Hoesin Palembang dapat melihat apa saja yang menjadi penyebab penyakit yang diderita pasien pengguna askin selama ini. sehingga nantinya akan sangat membantu pemerintah provinsi sumatera selatan dalam menemukan apa yang menjadi penyebabnya dengan tetap memperhatikan dari segala jenis faktornya.

Dimana nantinya pihak RSUP Moh.Hoesin Palembang dapat bekerja sama dengan pemerintah provinsi sumatera selatan untuk melakukan usaha pencegahan dalam bentuk pemeriksaan disertai penyuluhan ketempat-tempat tinggal masyarakat yang kurang mampu, yang dimana tempat tinggal mereka sangat rentan sekali terkena berbagai jenis penyakit. Untuk itulah diperlukan data untuk mengetahui hal tersebut, dan data tersebut didapat dari pihak RSUP Moh.Hoesin Palembang yang sebelumnya telah mengidentifikasi menjadi penyebab penyakit yang diderita oleh masyarakat kurang mampu yang menggunakan askin. Untuk mewujudkan hal tersebut, maka pihak RSUP Moh.Hoesin Palembang membutuhkan data mining untuk membantu kegiatan mereka nantinya yaitu berupa data-data yang akurat mengenai pokok permasalahan yang terjadi, dan memastikan agar apa yang dilakukan oleh RSUP Moh.Hoesin Palembang dan pemerintah provinsi sumatera selatan tepat sasaran dan efektif.

Data mining adalah solusi dalam dunia teknologi untuk mengatasi masalah yang dihadapi rumah sakit dalam memberikan informasi yang tepat dan akurat serta yang efisien kepada yang membutuhkan informasi yang tersebut, dimana informasi tersebut terdapat dalam media penyimpanan data yang memang khusus dipersiapkan oleh rumah sakit tersebut, terlebih lagi apabila rumah sakit

itu melayani pasien dalam jumlah banyak maka sudah pasti memerlukan media penyimpanan data dalam kapasitas yang besar dalam gudang data mereka.

Data mining dalam prosesnya menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat serta pengetahuan yang terkait dari berbagai database yang besar (Turban,dkk. 2005).

Dengan adanya masalah tersebut serta ada solusi untuk mengatasi keadaan seperti itu, maka penulis tertarik untuk melakukan penerapan data mining terhadap jumlah pasien yang menggunakan askin pada RSUP Moh.Hoesin Palembang. Untuk mengimplementasikan ketertarikan penulis terhadap masalah tersebut, maka penulis akan menuangkan penelitiannya dalam bentuk penulisan skripsi yang berjudul “ **Penerapan Data Mining Pada Rsup Dr.Moh Hoesin Sumatera Selatan Untuk Pengelompokkan Hasil Diagnosa Pengguna Asuransi Kesehatan Miskin (ASKIN)**”.

1.2 Perumusan Masalah

Berdasarkan penjelasan yang telah diuraikan diatas tadi, maka pokok permasalahan yang akan dirumuskan dalam penelitian ini adalah bagaimana menerapkan teknologi data mining dapat menghasilkan informasi mengenai pengelompokkan hasil diagnosa pasien yang menggunakan asuransi kesehatan miskin (askin).

1.3 Batasan Masalah

Agar penelitian ini terfokus pada pokok permasalahannya, maka penulis berinisiatif untuk membatasi permasalahan hanya pada :

- a. Penerapan data mining menggunakan teknik clustering dan menggunakan algoritma *decision tree* C4.5 mengikuti tahapan *knowledge Discovery in Databases (KDD)*.
- b. Informasi yang ditampilkan berupa pengelompokkan hasil diagnosa jenis penyakit yang sering diderita oleh pasien pengguna askin, berdasarkan hubungan data antara jenis diagnosa penyakit dengan jumlah pasien pengguna askin pada RSUP Moh.Hoesin Palembang.

1.4 Tujuan dan manfaat Penelitian

1.4.1 Tujuan Penelitian

Penelitian ini mempunyai tujuan untuk menerapkan teknik data mining untuk menampilkan informasi pengelompokkan pasien pengguna askin berdasarkan hubungan antara data jenis penyakit dengan data jumlah pasien pada RSUP Moh.Hoesin Palembang.

1.4.2 Manfaat Penelitian

Dengan adanya penerapan data mining ini diharapkan nantinya akan memberikan manfaat pada RSUP Dr.Moh Hoesin Palembang.

1. Dapat memberikan informasi –informasi serta pengetahuan yang selama ini mungkin belum diketahui secara pasti jumlah pengguna askin dalam gudang data sehingga menjadi informasi yang benar-benar berguna untuk

dapat membantu mengelompokkan pengguna askin pada RSUP Moh.Hoesin Palembang.

2. menjadi bahan pertimbangan oleh pihak instansi yang terkait untuk meningkatkan mutu pelayanan kesehatan kepada pengguna askin sehingga selaras dengan salah satu visi dan misi pemerintah daerah sumatera selatan yaitu berobat gratis tapi bermutu.

II. Tinjauan Pustaka

2.1 Landasan Teori

2.1.1 *Data Mining*

Data mining merupakan istilah yang sering dikatakan sebagai suatu cara untuk menguraikan serta mencari penemuan berupa pengetahuan didalam suatu *database*. *Data mining* adalah proses pemilihan atau “menambang” pengetahuan dari sekumpulan data dalam jumlah yang banyak (Han, Jiawei 2006).

Data mining juga sering disebut sebagai kegiatan mengeksplorasi dan menganalisis data dalam jumlah yang besar untuk menemukan *pattern* dan *rule* yang berarti (Berry, 2004). *Data mining* digunakan untuk mencari informasi bisnis berharga yang berharga dari basis data yang sangat besar, yang dipakai untuk memprediksi tren dan sifat-sifat bisnis serta menemukan pola-pola yang tidak diketahui sebelumnya.

Berdasarkan beberapa arti dan pengertian mengenai data mining, maka dapat diambil suatu kesimpulan mengenai *data mining*. Jadi *data mining* adalah proses pencarian informasi dan pengetahuan dari keseluruhan data yang tersembunyi, dengan menggunakan teknik yang sesuai dengan tempat yang “digali” sehingga nantinya akan menemukan suatu pola dari suatu data yang mungkin selama ini terabaikan dan dianggap sebagai suatu hal yang tak penting untuk kemudian diimplementasikan dalam bentuk grafik maupun diagram serta bentuk jenis lainnya yang mempunyai inti yang berisikan pengetahuan dan informasi yang berguna bagi penggunaannya.

2.1.2 Fungsi Data mining

Data mining bukan hanya pelengkap saja dalam suatu database, melainkan mempunyai fungsi yang penting untuk membantu penggunanya mendapatkan informasi yang berguna serta meningkatkan pengetahuan bagi sang pengguna itu sendiri dan dapat nantinya berguna untuk orang banyak. Pada dasarnya, *data mining* mempunyai empat fungsi dasar yaitu (Berson Dkk, 2000) :

1. Fungsi Klasifikasi (*classification*)

Data mining dapat digunakan untuk mengelompokkan data-data yang jumlahnya besar menjadi data-data yang lebih kecil.

2. Fungsi Segmentasi (*Segmentation*)

Disini data mining juga digunakan untuk melakukan segmentasi (pembagian) terhadap data berdasarkan karakteristik tertentu.

3. Fungsi Asosiasi (*Association*)

Disini data mining juga digunakan untuk mencari hubungan antara karakteristik tertentu .

4. Fungsi pengurutan (*Sequencing*)

Pada Fungsi ini, data mining digunakan untuk mengidentifikasi perubahan pola yang telah terjadi dalam jangka waktu yang tertentu.

2.1.3 Teknik Clustering

Teknik clustering melakukan pengelompokkan pada beberapa record, dan biasanya diberikan kepada pengguna akhir untuk memberikan gambaran tentang apa yang terjadi pada basis data.

Pada Data mining, teknik clustering berbeda dengan teknik association rule mining dan classification dimana kelas data telah ditentukan sebelumnya. Clustering melakukan pengelompokan data tanpa berdasarkan kelas data tertentu. Bahkan clustering dapat dipakai untuk memberikan label pada kelas data yang belum diketahui itu. Karena itu clustering sering digolongkan sebagai metode unsupervised learning. Prinsip clustering adalah memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas atau cluster. Clustering dapat dilakukan pada data yang memiliki atribut yang dipetakan sebagai ruang multi dimensi.

Banyak algoritma clustering memerlukan fungsi jarak untuk mengukur kemiripan antar data, selain itu juga diperlukan metode normalisasi bermacam atribut yang dimiliki data. Beberapa kategori algoritma clustering yang banyak dikenal adalah metode partisi dimana pemakai harus menentukan jumlah k partisi yang diinginkan lalu setiap data dites untuk dimasukkan pada salah satu partisi.

2.1.4 Algoritma CLHM (Centroid Linkage Hierarchical Method)

Algoritma centroid linkage merupakan salah satu proses untuk membuat pengklusteran yang didasarkan pada jarak centroidnya. Metode ini baik untuk untuk kasus data mining menggunakan teknik clustering dengan normal data set distribution. Akan tetapi metode ini tidak cocok untuk data yang mengandung outlier. dan untuk lebih memperjelas mengenai algoritma centroid linkage hierarchical method, berikut tahap kerja algoritma ini :

- a. Diasumsikan setiap data dianggap sebagai cluster. Kalau n = jumlah data dan c = jumlah cluster, berarti ada $c=n$.

- b. menghitung jarak antar cluster dengan Euclidian distance.
- c. Mencari 2 cluster yang mempunyai jarak antar cluster yang paling minimal dan digabungkan kedalam cluster baru (sehingga $c = c-1$).
- d. kembali ke langkah 3, dan di ulangi sampai dicapai cluster yang diinginkan

2.1.5 Analisis Cluster

Analisis cluster adalah teknik atau metode analisa multivariate (banyak variable) untuk mencari dan mengorganisir informasi tentang variabel tersebut sehingga secara relatif dapat dikelompokkan dalam bentuk homogen dalam sebuah cluster. Selain itu juga analisa cluster dapat diartikan sebagai pengorganisasian kumpulan pola kedalam cluster berdasarkan kesamaanya. Dalam cluster tersebut terdapat pola-pola yang terbentuk melalui kesamaan ciri/sifat yang terbentuk dari pola-pola cluster lainnya.

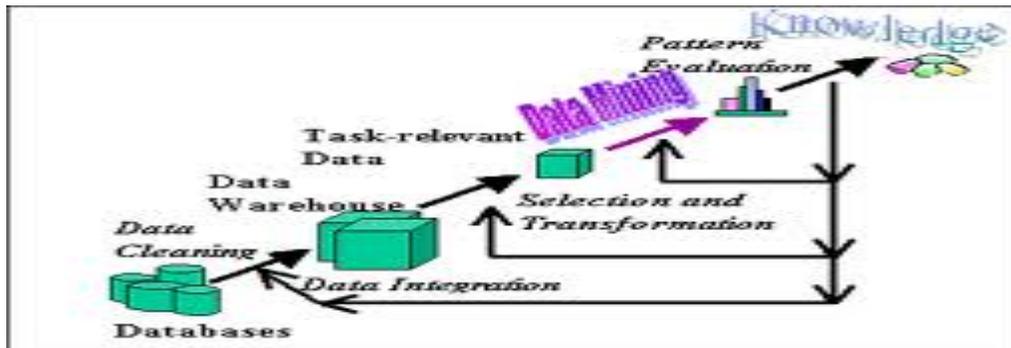
Fokus dari analisis cluster adalah membandingkan objek berdasarkan set variabel, hal inilah yang menyebabkan para ahli mendefinisikan set variabel sebagai tahap kritis dalam analisis cluster. Set variabel cluster adalah suatu set variabel yang merpresentasikan karakteristik yang dipakai objek-objek. Bedanya dengan analisis faktor adalah bahwa analisis cluster terfokus pada pengelompokan objek sedangkan analisis faktor terfokus pada kelompok variabel. Solusi analisis cluster bersifat tidak unik, anggota cluster untuk tiap penyelesaian/solusi tergantung pada beberapa elemen prosedur dan beberapa solusi yang berbeda dapat diperoleh dengan mengubah satu elemen atau lebih. Solusi cluster secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar

untuk menilai kesamaan. Penambahan atau pengurangan variabel-variabel yang relevan dapat mempengaruhi substansi hasil analisis cluster

2.1.6 Metode *Knowledge Discovery in databases (KDD)*

Knowledge Discovery in Databases (KDD) merupakan keseluruhan proses konversi data mentah menjadi pengetahuan yang bermanfaat yang terdiri dari serangkaian tahap transformasi meliputi data *preprocessing* dan *postprocessing*. Dimana data *preprocessing* merupakan langkah awal untuk mengubah data mentah menjadi format yang sesuai untuk tahap analisis selanjutnya. Data *preprocessing* kemungkinan akan membutuhkan waktu yang sangat lama, hal ini dikarenakan data yang mentah kemungkinan disimpan dengan format dan database yang berbeda. Sedangkan *Postprocessing* meliputi semua operasi yang harus dilakukan agar hasil data mining dapat diakses dan lebih mudah diinterpretasikan oleh para pihak analisis.

Memang sering kali terdengar bahwa *Knowledge Discovery in Databases* mempunyai ikatan yang sangat kuat dengan data mining, bahkan data mining mempunyai nama lain yaitu *Knowledge Discovery in Databases (KDD)*. Memang anggapan itu benar adanya, mengingat tujuan data mining itu sendiri. Namun dalam arti sebenarnya data mining dan *knowledge in databases* tidaklah sama karena data mining merupakan bagian proses dalam *Knowledge Discovery in Databases* itu sendiri (KDD), walaupun begitu memang data mining merupakan proses yang terpenting dalam proses *Knowledge Discovery in Databases (KDD)*. Dan hal itu bisa dilihat pada gambar 2.3 (Han & Kamber, 2001) untuk lebih menjelaskan proses *Knowledge Discovery in Databases*



Gambar 2.1 Proses tahapan-tahapan *Knowledge Discovery in Databases* (KDD)

Untuk lebih lebih mengetahui mengenai tahapan- tahapan proses pada knowledge in databases, berikut penjelasan mengenai tahapan-tahapan tersebut :

1. *Database*

Koleksi data yang saling berhubungan untuk dipergunakan secara bersama kemudian dirancang untuk memenuhi kebutuhan informasi organisasi

2. *Data Cleaning*

Pada umumnya, data yang diperoleh, baik dari database suatu perusahaan maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaanya bisa mengurangi mutu atau akurasi dari hasil data mining nantinya. Pembersihan data juga akan mempengaruhi performasi dari system data mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.

3. *Data integration*

Integrasi data dilakukan pada atribut-attribut yang mengidentifikasi *entitas-entitas* yang unik seperti atribut nama,nomor pegawai, tempat

lahir, agama dan lain sebagainya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil data yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

4. *Task relevan data*

Setelah semua sumber data digabung atau diintegrasikan menjadi satu keseluruhan *database*, maka tahap selanjutnya adalah melakukan task *relevant data*. Pada tahap ini melakukan relevansi atribut dari data yang *relevant* atau yang sesuai dengan target atau *output* yang akan dihasilkan.

5. *Data transformation*

Coding adalah proses 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 *database*.

6. *Data mining*

Data mining adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Tahap ini merupakan inti dari tahapan KDD yang dilakukan untuk menganalisis data yang telah dibersihkan.

7. *Pattern evaluation*

Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai dengan hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki *data mining* lain yang lebih sesuai, atau menerima hasilnya sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

8. *Knowledge*

Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak paham mengenai *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah itu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, *visualisasi* juga bisa membantu mengkomunikasikan hasil data mining.

2.1.7 Weka

Weka adalah aplikasi data mining open source yang berbasis java. Aplikasi ini dikembangkan pada tahun 1994 dan pertama kali oleh sebuah universitas di selandia baru yang bernama universitas Waikato. aplikasi weka ini mulai menjadi aplikasi *data mining open source* yang sangat terkenal pada awal perkembangannya. Hal itu dikarenakan aplikasi weka memiliki kelebihan yang tidak dimiliki aplikasi *data mining* lainnya yaitu pada aplikasi weka terdapat

banyak algoritma yang terdapat didalam aplikasi dan disertai juga *machine learning*, lalu juga dalam penggunaanya tidak terlalu rumit sehingga tidak menyulitkan penggunanya, dan ditambah dengan kelebihan lainnya bahwa algoritma-algoritma yang terdapat pada aplikasi weka selalu baru dan ter *update*, sehingga dengan beberapa kelebihan aplikasi weka tersebut digunakan banyak perusahaan dalam dunia bisnis untuk membantu dalam usaha bisnisnya, akademik pun juga tak mau ketinggalan untuk menggunakan aplikasi weka ini, serta instansi dalam bidang kesehatan yaitu rumah sakit juga saat ini juga menggunakan aplikasi weka.

Aplikasi weka merupakan *software* yang terdiri dari koleksi algoritma *machine learning* yang dapat digunakan untuk melakukan generalisasi atau formulasi dari sekumpulan data *sampling*. Inti dari kekuatan pada aplikasi weka terletak pada algoritma yang makin lengkap dan canggih, namun walaupun begitu canggihnya aplikasi weka tersebut, letak keberhasilan *data mining* tetap ditentukan oleh manusia itu sendiri sebagai penggunanya/*user*. Keberhasilan *data mining* itu berdasarkan pada pengumpulan data yang berkualitas tinggi, penggunaan model dan algoritma yang tepat. Sehingga secanggih serta sehebat apapun aplikasi *data mining*, tanpa kemampuan sang penggunanya untuk menerapkannya maka tidak akan menghasilkan data mining yang tepat dan bermanfaat, Sumber : <http://www.cs.waikato.ac.nz/ml/weka/>, diakses pada 22 april 2013.

2.2 Penelitian Sebelumnya

Penelitian sebelumnya berguna bagi penulis untuk dapat menjadi pedoman serta pegangan penelitian yang akan penulis lakukan sehingga nantinya dengan

adanya penelitian sebelumnya dapat membantu dan memudahkan penulis dalam melakukan penelitiannya sesuai dengan tema dan membuat system yang baru dan bermanfaat.

Menurut Lindawati dalam judul penelitiannya yang berjudul “ **Data Mining dengan teknik clustering dalam pengklasifikasian data mahasiswa studi kasus prediksi lama studi mahasiswa universitas bina nusantara**” Penggunaan teknik Data Mining (DM) clustering berbeda dengan teknik-teknik DM yang lainnya, seperti association rule mining dan classification yang memerlukan tahapan training dan evaluasi. Teknik ini menggunakan metode unsupervised learning yang berarti DM tidak perlu melakukan training terlebih dahulu tapi bisa langsung menggunakannya untuk pengelompokan. Teknik ini masih jarang digunakan dibanding dengan teknik-teknik DM yang lain.

Oleh karena itu, artikel ini berfokus pada pengembangan aplikasi DM dengan teknik clustering pada penelitian untuk mengklasifikasikan data prediksi lama studi mahasiswa di Universitas Bina Nusantara dengan menggunakan algoritma Self Organizing Maps dan pengujian keakuratan aplikasi DM dengan teknik clustering. Tahapan yang dilakukan dibagi menjadi tahapan analisa, perancangan, pengembangan dan pengujian aplikasi DM. Pada tahapan analisa dilakukan beberapa percobaan dalam mengklasifikasikan prediksi lama studi mahasiswa berdasarkan delapan atribut yang digunakan, yaitu: rata-rata Indeks Prestasi Kumulatif (rIpk), simpangan rata-rata Indeks Prestasi Kumulatif (srIpk), rata-rata jumlah SKS per Semester (rSksem), rata-rata jumlah SKS yang tidak lulus per semester (rSksemTL), jumlah SKS Kumulatif (skKum), jumlah SKS

yang akan diambil pada semester keempat (sksYad), jumlah SKS yang wajib diambil (sksMin) dan hak SKS yang dapat diambil pada semester lima dst (hakSks). Hasil dari tahapan analisa ini digunakan sebagai acuan pada tahapan perancangan dan pengembangan aplikasi DM. Selanjutnya dilakukan pengujian terhadap aplikasi DM yang telah dibuat untuk mengetahui keakuratan pengklasifikasian dari aplikasi DM tersebut. Evaluasi dilakukan melalui beberapa variasi pengujian dengan menggunakan paramter-parameter jumlah data, jumlah iterasi, learning rate, radius, neighbourhood function dan urutan data. Dari pengujian-pengujian yang dilakukan dapat diketahui bahwa rata-rata kesalahan hasil klasifikasi prediksi lama studi yang diperoleh relatif kecil, kurang dari atau maksimal 5%.

Kata Kunci: Data Mining, clustering, Self-Organizing Maps, Prediksi Lama Studi Mahasiswa

Menurut yulia purnama sari dalam judul penelitiannya yang berjudul **“Perancangan Aplikasi Sistem Data Mining Untuk Memprediksi Transaksi Gagal Serah Dana Dengan menggunakan Metode Clustering Pada Perusahaan PT.KPEI”**. Dalam penulisan penelitian ini dilatar belakangi oleh kebutuhan untuk menganalisis potensi kegagalan serah dana dari data yang jumlahnya besar. Tujuan pembuatan skripsi ini adalah menganalisis kebutuhan informasi yang diperlukan dalam proses analisis potensi akan kegagalan serah dana pada suatu transaksi jual beli saham, merancang sistem *data mining*, dan merancang aplikasi sistem *data mining* untuk mendukung proses analisis akan kegagalan serah dana pada suatu transaksi jual beli saham. Untuk memprediksi

potensi kegagalan serah dana digunakan teknik Data Mining. Teknik ini berguna untuk menghasilkan informasi tersembunyi dari data yang jumlahnya besar, sehingga memudahkan melakukan analisis terhadap potensi gagal serah dana, yang dapat menurunkan tingkat resiko kerugian dari PT. KPEI, sehingga penulis tertarik melakukan penelitian ini.

Metode yang digunakan adalah Clustering Kmean untuk memecahkan permasalahan yang ada dalam PT KPEI yaitu bagaimana menganalisis data agar dapat mengetahui transaksi mana yang dapat menyebabkan gagal serah dana lalu mengelompokannya berdasarkan *Receiving Failure*. Hasil pengelompokan akan menghasilkan pengetahuan yang berguna dalam pengambilan keputusan untuk jajaran top *level management* PT KPEI. Kesimpulan dari skripsi ini adalah dengan menggunakan aplikasi sistem *data mining* yang telah dirancang, pihak *management* dapat mengambil keputusan dengan lebih tepat sehingga menghindarkan kerugian untuk perusahaan.

Kata kunci : Data Mining, *Clustering*, Gagal Serah Dana

III METODOLOGI PENELITIAN

3.1 Waktu dan tempat penelitian

Penelitian ini dilakukan bertempat di RSUP DR.Moh Hoesin Palembang, dari bulan april 2013 sampai dengan juli 2013.

3.2 Alat dan bahan

1. Perangkat keras

Perangkat keras yang penulis gunakan dalam pembuatan program dan penulisan skripsi adalah :

- a. *Laptop*
- b. *Hardisk* dengan kapasitas 160 GB
- c. *Mouse*
- d. *Printer*
- e. Modem

2. Perangkat Lunak

Perangkat Lunak yang digunakan adalah sebagai berikut :

- a. Sistem operasi XP-3
- b. *PHP My admin* sebagai pengolah *database* awal
- c. Weka

3.3 Metode Penelitian

Dalam penelitian ini, penulis menggunakan metode *deskriptif* karena permasalahan yang sedang diteliti saat ini berdasarkan data-data yang bersifat fakta yang ada mengenai data jenis diagnosa penyakit dan data pasien pengguna askin pada RSUP Dr.Moh Hoesin Palembang.

3.4 Metode Pengumpulan Data

Metode untuk pengumpulan data yang dilakukan adalah metode pengumpulan data sekunder, yaitu dimana data-data yang menjadi bahan penelitian diperoleh langsung dari tempat penelitian yaitu RSUP Dr.Moh Hoesin Palembang , dimana nantinya data-data tersebut lah yang akan menjadi bahan penelitian untuk kami teliti yang diperoleh secara langsung dari pihak IT RSUP Dr.Moh Hoesin Palembang.

3.5 Metode Analisis Data

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan proses tahapan *knowledge discovery in databases (KDD)* yang terdiri dari *Database, Data Cleaning, Data integration, Task relevan data, Data transformation, Data mining, Pattern evolution, knowledge* :

1. *Database*

Koleksi data yang saling berhubungan untuk dipergunakan secara bersama kemudian dirancang untuk memenuhi kebutuhan informasi organisasi

2. *Data Cleaning*

Pada umumnya, data yang diperoleh, baik dari database suatu perusahaan maupun eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau hanya sekedar salah ketik. Data-data yang tidak *relevan* itu juga lebih baik dibuang karena keberadaanya bisa mengurangi mutu atau akurasi dari hasil *data mining* nantinya. Pembersihan data juga akan mempengaruhi performansi dari system data

mining karena data yang akan ditangani akan berkurang jumlah dan kompleksitasnya.

3. *Data integration*

Integrasi data dilakukan pada atribut-attribut yang mengidentifikasi *entitas-entitas* yang unik seperti atribut nama, nomor pegawai, tempat lahir, agama dan lain sebagainya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil data yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

4. *Task relevan data*

Setelah semua sumber data digabung atau diintegrasikan menjadi satu keseluruhan *database*, maka tahap selanjutnya adalah melakukan task *relevant data*. Pada tahap ini melakukan *relevansi* atribut dari data yang *relevant* atau yang sesuai dengan target atau output yang akan dihasilkan.

5. *Data transformation*

Coding adalah proses 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 *database*.

6. *Data mining*

Data mining adalah proses mengeksplorasi dan menganalisa data dalam jumlah yang besar yang bertujuan untuk menemukan suatu pola atau informasi yang menarik dari data yang tersimpan dalam jumlah yang besar dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau

algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Tahap ini merupakan inti dari tahapan KDD yang dilakukan untuk menganalisis data yang telah dibersihkan.

7. *Pattern evaluation*

Dalam tahap ini, merupakan hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah *hipotesa* yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai dengan hipotesa, ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki data mining lain yang lebih sesuai, atau menerima hasilnya sebagai suatu hasil yang diluar dugaan yang mungkin bermanfaat.

8. *Knowledge*

Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak paham mengenai data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah itu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining*.

DAFTAR PUSTAKA

Han J & Kamber. 2001. *Data Mining: Concepts and Techniques*. Simon Fraser University. USA: Morgan Kaufman Publisher

Han J & Kamber. 2006. *Data Mining: Concepts and Techniques second edition*. Simon Fraser University. USA: Morgan Kaufman Publisher

Michael berry & Gordon S. Linoff. 2004. *Data Mining Techniques : For marketing, sales, and customer relationship management*. John Willey & Sons, Inc.

Berson, alex dkk. (2000). Building data mining application for CRM. Mc Graw – Hill. United states of America.

[http:// www.blog.uin-malang.ac.id](http://www.blog.uin-malang.ac.id)

Machine learning group at the university of Waikato. (2013), standar machine learning techniques into a softaware “workbench” called weka. Diakses 22 april 2013 dari, <http://www.cs.waikato.ac.nz/ml/index.html>

LAMPIRAN

1. Data Awal

RSUP Dr. Mohammad Hoesin
Jl. Jend. Sudirman KM.3,5
Palembang
Telp. (0711) 364086

LAPORAN JUMLAH DIAGNOSA PASIEN

Periode : 01-JAN-10 s/d 01-JAN-12

Pasiien : ASKIN

NO.	KODE	NAMA DIAGNOSA	JAN	FEB	MAR	APRIL	MEI	JUNI	JULI	AGU	SEP	OKTO	NOV	DES	TOTAL
1	J00	ACUTE NASOPHARYNGITIS (COMMON COLD)	0	0	0	2	1	1	0	0	0	0	2	0	8
2	R25	ABNORMAL INVOLUNTARY MOVEMENTS	0	1	0	0	0	0	0	0	0	0	0	0	1
3	E82.1	ABSCESS	0	0	0	0	0	2	2	0	0	0	0	0	4
4	D69.9	ACQUIRED HAEMOLYTIC ANAEMIA, UNSPECIFIED	0	0	0	0	0	0	0	0	0	0	1	0	1
5	J21.9	ACUTE BRONCHIOLITIS => UNSPECIFIED	0	0	0	0	0	0	0	2	0	0	0	2	4
6	J20	ACUTE BRONCHITIS	1	0	0	1	1	0	0	0	0	0	0	0	3
7	I24.9	ACUTE ISCHAEMIC HEART DISEASE, UNSPECIFIED	1	0	0	0	0	0	0	0	0	0	0	0	1
8	J04.0	ACUTE LARYNGITIS (NOS) (oedematous) (subglottic) (suppurative) (ulcerative)	0	0	0	0	0	0	0	1	0	0	0	0	1
9	N00	ACUTE NEPHRITIC SYNDROME	3	4	1	0	1	1	0	2	4	2	1	0	19
10	J05.0	ACUTE OBSTRUCTIVE LARYNGITIS (CROUP) (NOS)	0	0	0	0	0	2	0	0	2	0	0	0	4
11	J02	ACUTE PHARYNGITIS	0	0	0	0	0	0	0	0	0	0	1	0	1
12	R18	ASCITES	0	0	0	0	0	0	0	0	0	0	1	2	3
13	R09.0	ASPHYXIA	0	0	0	0	0	0	1	0	2	0	0	0	3
14	J45	ASTHMA	0	1	0	0	0	1	4	6	0	0	0	0	12
15	J45.9	ASTHMA UNSPECIFIED	0	1	2	3	3	2	2	0	2	2	8	4	27
16	I25.1	ATHEROSCLEROTIC HEART DISEASE (CAD)	0	0	0	0	0	0	3	0	0	0	0	0	3
17	I23.1	ATRIAL SEPTAL DEFECT AS CURRENT COMPLICATION FOLLOWING ACUTE MYOCARDIAL INFARCTION	0	0	0	0	0	1	0	0	0	0	0	2	3
18	Z49.0	ATTENTION TO SURGICAL DRESSINGS & SUTURES	0	0	0	0	0	2	0	0	0	0	0	0	2
19	R80	Abnormal findings on diagnostic imaging of central nervous system	0	0	0	0	0	2	0	0	0	0	0	0	2
20	B15	Acute Hepatitis A	0	0	2	2	9	2	4	1	1	0	0	0	21
21	B15.9	Acute Hepatitis A => Hepatitis Virus A	0	1	0	0	0	0	1	1	0	0	0	0	3
22	A08.0	Acute Amoebic Dysentery	0	0	0	0	0	0	0	0	0	1	0	0	1
23	C82.0	Acute Myeloid Leukaemia	0	0	0	0	0	0	0	0	0	0	0	1	1
24	C81.0	Acute lymphoblastic leukaemia	14	5	8	3	17	11	9	12	17	16	14	19	145
25	J98.0	Acute respiratory failure	0	0	0	0	0	0	1	0	0	0	0	0	1
26	J03.9	Acute tonsillitis, unspecified (follicular) (gangrenous) (infective) (ulcerative)	0	0	0	0	0	0	2	0	0	0	0	0	2
27	J08.9	Acute upper respiratory infection,unspecified (disease, acute) (infecton NOS)	0	0	0	1	0	0	0	0	0	0	0	0	1
28	D64.9	Anaemia,unspecified	23	9	21	17	18	16	21	21	17	23	24	29	239
29	A82.2	Arenaviral Haemorrhagic Fever => Lassa Fever	0	0	0	0	0	0	0	0	0	1	0	0	1
30	H35.0	BACKGROUND RETINOPATHY & RETINAL VASCULAR	0	0	0	0	0	0	0	0	0	0	2	0	2

2. Konsep Data

Kode Diagnosa	Nama Diagnosa	Tahun	Jumlah
J00	<i>ACUTE NASOPHARYNGITIS (COMMON COLD)</i>	2010	6
R25	<i>ABNORMAL INVOLUNTARY MOVEMENTS</i>	2010	1
E32.1	<i>ABSCESS</i>	2010	4
D59.9	<i>ACQUIRED HAEMOLYTIC ANAEMIA UNSPECIFIED</i>	2010	1
J21.9	<i>ACUTE BROCHIOLITIS => UNSPECEFIED</i>	2010	4
J20	<i>ACUTE BRONCHITIS</i>	2010	3
I24.9	<i>ACUTE ISCHAEMIC HEART DISEASE, UNSPECIFIED</i>	2010	1
J04.0	<i>ACUTE LARYNGITIS (NOS)(OEDEMETUS) (SUGBOTIC)</i>	2010	1
N00	<i>ACUTE NEPHIRITIC SYNDROME</i>	2010	19
J05.0	<i>ACUTE OBSTRUCTIVE LARYNGITIS (CROUP) (NOS)</i>	2010	4
J02	<i>ACUTE PHARYNGITIS</i>	2010	1
R18	<i>ASCITES</i>	2010	3

JADWAL PENELITIAN

No	Kegiatan	April				Mei				Juni				Juli				1			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1			
1	Penyusunan Proposal	■	■	■	■																
2	Ujian Proposal					■															
	Analisis Data Mining																				
	Penerapan Data mining						■	■													
	Analisa							■	■	■											
	Penyusunan Data									■	■	■	■								
	Rancangan													■	■	■	■				
	Dokumentasi																	■	■	■	■
	Transisi																				
	Produksi																				
3	Ujian Komprehenship																				
4	Perbaikan Skripsi																				
5	Bimbingan	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Keterangan :

■ : Yang Dilakukan

□ : Tidak Dilakukan