KLASIFIKASI DATA NILAI MAHASISWA DENGAN MENGGUNAKAN METODE *DECISION TREE.*

Henrico[1)] dan Elis Sondang Dasawaty[2)]
[1)]Alumni Program Studi Sistem Informasi
[2)]Staf Pengajar Program Studi Sistem Informasi
Institut Bisnis dan Informatika Kwik Kian Gie
Jl.Yos Sudarso Kav.87. Sunter Jakarta Utara 14350
ellis.tampubolon@kwikkiangie.ac.id

ABSTRACT

In the era of globalization where technology is becoming more advanced so is necessary for an institutions or organizations to take further strategic steps, quick data processing capabilities, and produce information quickly and accurately depending on technological reliability inside those organizations. Kwik Kian Gie School of Business is one of those institutions that tries to appeal itself, and securing it's positions more realizing the need for data mining for further use.

The theory used are the meaning of data, database, data mining, system, information system, and decision tree.

Data mining can be used by numerous and compatible software, during the making the following software used are : Rapidminer, PHP, Notepad++, dan xampp.

The result is a program which works in a similar way to a decision tree which can give specification, or easiness during the data searching which less time consuming and gave accurate results of the wanted report .

In conclusion with the existence of data mining even the existing data is uncountable, all of those data inside the organization can be easily searched, also with the use of data mining we can reduce the chances of human error.

Key Word : Data Mining, Decision Tree

1. PENDAHULUAN

Seiring dengan berkembangnya teknologi dan kemajuan zaman di berbagai bidang maka teknologi menjdai bagian tolak ukur dalam majunya suatu proses dan berkembangnya suatu usaha dalam bidang ekonomi, politik, budaya , social, maupun dalam bidang pendidikan. Teknologi akhirnya juga menjadi sebuah bagian dalam organisasi atau usaha bisa sebagai alat bersaing dengan organisasi lain maupun juga sebagai penunjang dalam aktivitas dalam berorganisasi. Pemakaian teknologi diharapkan dapat digunakan untuk membantu menyelesaikan pekerjaan dengan lebih cepat dan lebih efisien, selain itu teknologi juga dapat digunakan sebagai alat bantu dalam mengambil keputusan karena bersifat real time dan dapat memberi informasi yang akurat. Karena itu penggunaan teknologi merupakan bagian yang sangat penting dalam penunjang keberhasilan dalam mencapai tujuan organisasi.

Untuk mendukung dalam mengambil keputusan yang tepat, maka salah satu cara yang digunakan dalam organisasi yang bersangkutan adalah dengan menggunakan *Decision Tree*. *Decision Tree* mendukung pengambilan keputusan manajemen dengan mengumpulkan dan mengorganisir data untuk kebutuhan analisis dan laporan bagi pihak manajemen sehingga pihak manajemen dapat melihat gambaran umum dari data dalam jumlah yang sangat besar dengan lebih mudah.

Perkembangan pendidikan saat ini membawa persaingan yang sangat ketat sehingga membuat institusi saling bersaing untuk meningkatkan kualitas pendidikan serta meningkatkan sarana dan prasarana yang dimiliki. Salah satu institusi yang bersaing ini adalah Universitas Kwik Kian Gie School of Business.

Diharapkan kedepannya Kwik Kian Gie School of Business mampu mengikuti perkembangan zaman yang cepat berubah seiring dengan permintaannya. Meskipun sistem Kwik Kian Gie Schol of Business sudah terkomputerisasi dan memiliki *database* dalam untuk menyimpan banyaknya data jumlah transaksi per-hari menyebabkan universitas mengalami kesulitan dalam menangani data dalam jumlah yang banyak dan terus bertambah.

Decision Tree akan diterapkan sehingga mendukung pengambilan data mahasiswa sehingga menjadi sangat mudah dalam menyortir data.

Selama ini kesulitan dalam menghasilkan laporan yang cepat dan akurat yang berujung pada lambatnya dan kurang efektifnya pengambilan keputusan. Karena itu pihak Manajemen Kwik Kian Gie School of Business memutuskan untuk menggunakan *Data Mining* dalam menunjang pihak manajemen dan eksekutif dalam pengambilan keputusan strategis.

2. LANDASAN/KERANGKA PEMIKIRAN

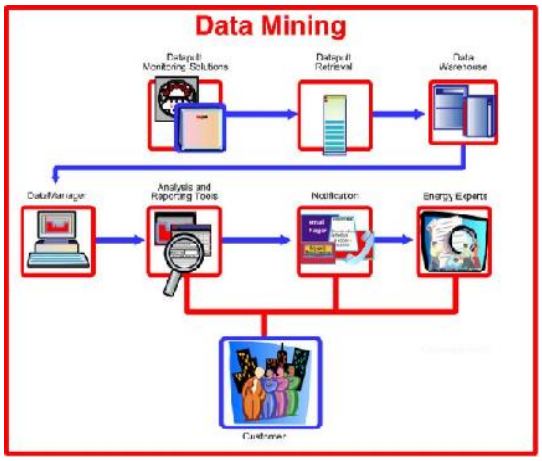
2.1.Data Mining

Menurut Turban et al (2005), *data mining* merupakan proses semi otomatik yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam database besar.

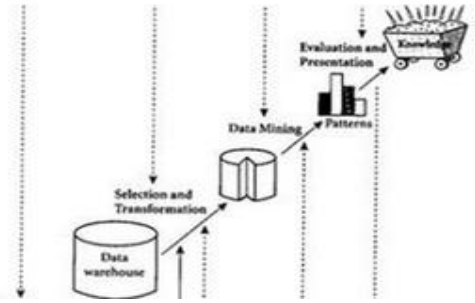
Menurut Pramudiono (2006), *data mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.

Data mining adalah ilmu untuk menggali informasi dari data set yang besar dan *database* yang digunakan dalam proses mengelola data yang mencakup analisa dan penentuan kesalahan (Ghencea & Immo, 2010). *Data mining* merupakan teknik untuk mengekstrak pengetahuan dari informasi. Proses ini mempunyai peran penting dalam menentukan estimasi standar (Titan, Tudor, & Loana, 2007). Sebuah sistem data mining pada khususnya dapat melakukan satu tugas atau secara umum mampu melakukan banyak tugas *data mining*.

Data mining adalah analisis data yang berjumlah besar untuk mencari pola dan aturan yang bisa digunakan untuk membantu pengambilan keputusan dan memprediksi behavior masa depan (Laudon, 2014).



Gambar 2.1 ilustrasi Data Mining
Sumber : www.c3dmw.com



Gambar 2.2 Tugas/Tahap Data Mining
Sumber : www. stumbleupon.com

2.2.OLAP (*Online Analytical Processing*)

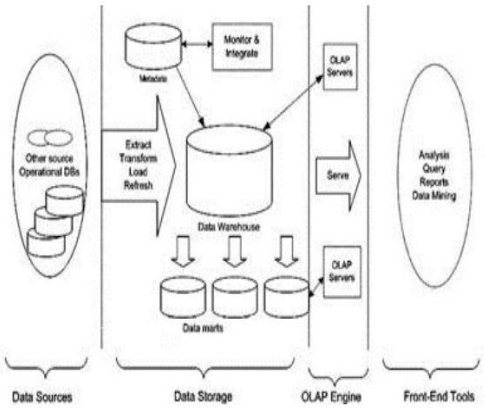
Menurut Hoffer et al (2010 : 569), OLAP adalah pengunaan sekumpulan alat grafikal yang menyediakan kepada *user* sebuah tampilan multi-dimensional dari data *user* dan memampukan *user* untuk menganalisa data menggunakan teknik penyajian yang sederhana.

Menurut Connolly dan Begg (2005: 1205), OLAP adalah suatu istilah yang mendeskripsikan suatu teknologi yang menggunakan multi-dimensional *view* dari agregasi data untuk menyediakan akses cepat ke informasi strategis untuk tujuan analisa lebih lanjut. OLAP tools berdasarkan konsep database multi dimensi dan memperbolehkan pengguna ahli untuk

menganalisa data memakai pandangan kompleks dan multi dimensi.

Codd memformulasikan 12 aturan sebagai basis untuk memilih tools OLAP:

- *Multi-dimensional conceptual view*
- *Transparency*
- *Accessibility*
- *Consistent reporting performance*
- *Client-server architecture*
- *Generic dimensionality*
- *Dynamic sparse matrix handling*
- *Multi-user support*
- *Unrestricted cross-dimensional operations*
- *Intuitive data manipulation*
- *Flexible reporting*
- *Unlimited dimensions and aggregation levels*



Gambar 2.8 OLAP

Sumber : bimentalist.com

2.3. Sistem Basis Data

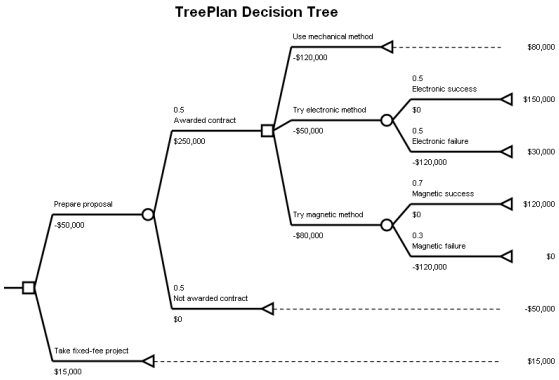
Sistem basis Data menurut Kendall & Kendall (2005 : 704) adalah software yang mengorganisir data di dalam database menyediakan gudang informasi, organisasi, dan kapasitas pengambilan kembali. Menurut Carlos Coronel (2013 : 7) Sistem basis data adalah kumpulan program yang memanage struktur *database* dan mengontrol akses ke data yang

2.4. RapidMiner

RapidMiner adalah sebuah lingkungan untuk *machine learning, text mining, data mining,* dan *predictive analysis*. *RapidMiner* merupakan *open source* berlisensi AGPL versi 3, Dibentuk pada tahun 2001 oleh Ralf KlinkenBerg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence unit* dari *University of Dortmund.* Di host oleh Source Forge sejak 2004. Peringkat satu sebagai tools untuk data mining projek nyata pada poll oleh KDnuggets sebuah Koran data mining, pada tahun 2010-2011. Menyediakan prosedur data mning dan machine learning termasuk ETL, data preprocessing, visualisasi, modeling, dan evaluasi. Proses *data mining* tersusun atas operator-operator yang *nestable*, dideskripsikan dengan XML, dan dibuat dengan GUI. Ditulis dengan bahasa pemrograman Java. Diintegrasikan dengan *proyek data mining* Weka dan statistika R.

2.6. Decision Tree

Decision Tree (Pohon Keputusan) adalah pohon dimana setiap cabangnya menunjukkan pilihan diantara sejumlah alternatif pilihan yang ada, dan setiap daunnya menunjukkan keputusan yang dipilih. *Decision tree* biasa digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. *Decision tree* dimulai dengan sebuah *root node* (titik awal) yang dipakai oleh user untuk mengambil tindakan. Dari *node root* ini,user memecahnya sesuai dengan algoritma *decision tree.* Hasil akhirnya adalah sebuah *decision tree* dengan setiap cabangnya menunjukkan kemungkinan sekenario dari keputusan yang diambil serta hasilnya. Algoritma yang dapat dipakai dalam pembentukan decision tree adalah ID3, C4.5, CART.



Gambar 2.6 Decision Tree

Sumber: treeplan.com

Tabel dimensi dinormalisasi untuk mengurangi redudansi data (duplikasi), sehingga struktur tabelnya akan lebih ramping. Dengan pengelompokan ini, data akan lebih mudah dibaca dan membantu pengembang aplikasi untuk menata desain antarmuka sistem dan *filtering* data. Struktur ini akan menghemat kapasitas *storage*, namun waktu eksekusi data akan lebih lama

mengingat jumlah tabel dimensi yang direlasikan lebih banyak dan membutuhkan tambahan relasi *foreign key*. Kueri yang terbentuk lebih kompleks, yang mengakibatkan kinerja kueri menurun. Pada penerapan yang lebih umum, tabel dimensi tidak diturunkan dengan lebih banyak tabel dimensi lain dan pengaturan UI atau pengelompokan data diatur secara *hard-coded* di kode program aplikasinya.

Pada decision tree terdapat 3 jenis node, yaitu:

a. ***Root Node***, merupakan node paling atas, pada node ini tidak ada input dan bisa tidak mempunyai output atau mempunyai output lebih dari satu.

b. ***Internal Node*** , merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal dua.

c. ***Leaf node*** atau terminal node , merupakan node akhir, pada node ini hanya terdapat satu input dan tidak mempunyai output.

2.5.Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti di bidang kecerdasan buatan bernama J.Rose Quinlan pada akhir tahun 1970-an. Algortima C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun.

Secara umum, algoritma C4.5 untuk membangun sebuah pohon keputusan adalah sebagai berikut:

a) itung jumlah data berdasar anggota atribut dengan syarat tertentu, untuk proses pertama syarat masih kosong.
b) ilih atribut sebagai Node.
c) Buat cabang untuk tiap-tiap anggota dari Node.
c) eriksa apakah nilai entropy ada yang bernilai nol. Jika ada tentukan daun yang terbentuk. Jika seluruh nilai adalah nol, maka proses pun akan berhenti.
d). Jika ada anggota Node yang memiliki nilai entropy lebih dari nol, ulang lagi proses dari awal dengan Node sebagai syarat sampai semua anggota dari Node bernilai nol.

Node adalah atribut yang mempunyai nilai gain tertinggi dari atribut-atribut yang ada. Rumus untuk menghitung gain adalah sebagai berikut:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \left(\sum_{i=1}^{n} \frac{Ai}{S} * \text{Entropy}(Ai) \right)$$

Keterangan :

S = kasus, A = Atribut, n = jumlah partisi atribut A, Ai = jumlah kasus pada partisi ke-i, Sigma = jumlah kasus

3.METODOLOGI PENELITIAN

3.1 Metode pengumpulan data

Metode pengumpulan data yang penulis gunakan adalah dengan penggunaan wawancara dengan pihak ICT dan melakukan observasi langsung terhadap hal-hal yang dibutuhkan. Untuk proses ETL *(Extract, Transform,Load)* data menggunakan RapidMiner dalam bentuk *decision tree.* Penggambaran yang dilakukan adalah dengan menggunakan model sistem:

1) Data Flow diagram
2) Entity Relationship Diagram

Untuk pembuatan *user interfacenya* digunakan program PHP dalam pembuatannya sehingga dapat menyaksikan langsung dalam prosesnya.

H

3.2.Teknik Pengumpulan Data

Proses pengumpulan data dilakukan dengan cara:

1.Wawancara Tak Terstruktur

Pertama dilakukan wawancara terhadap sistem yang berjalan. Lalu menulis apa yang menjadi kendala. Penulis juga menanyakan laporan apa yang diinginkan, tentunya dengan harapan bisa membantu dalam pengambilan keputusan.

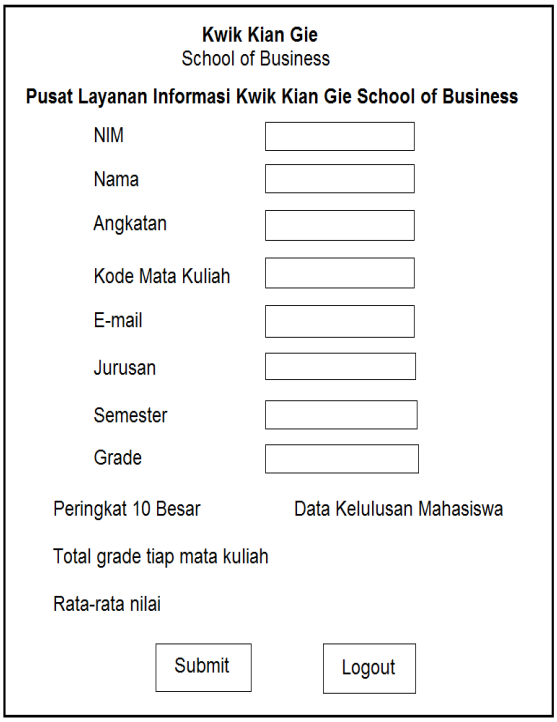
2.Observasi Tidak Langsung

Penulis juga melakukan observasi terhadap data-data yang diperlukan. Tetapi penulis tidak terlalu terlibat dalam melakukan pengambilan data sehingga penulis hanya mengamati apa yang dibutuhkan penulis.

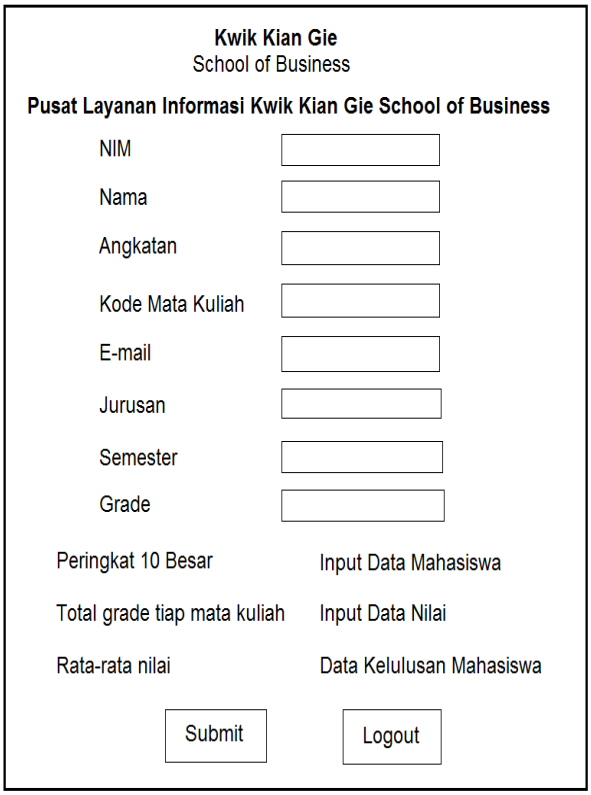
4.HASIL PENELITIAN DAN PEMBAHASAN

4.1. Rancangan User Interface

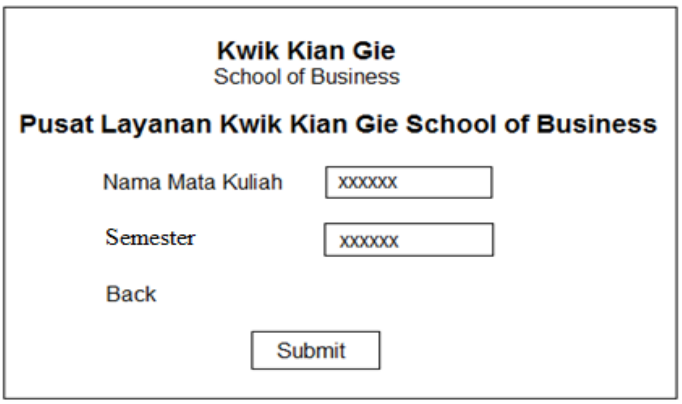
Gambar 4.1 Rancangan Form Login



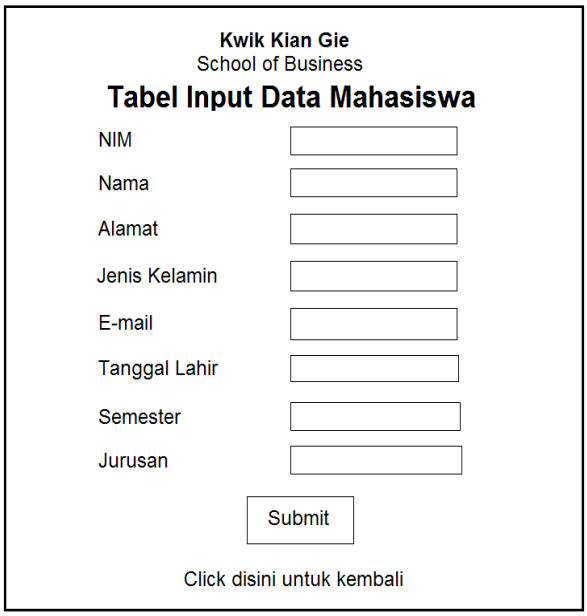
Gambar 4.2 Rancangan Form Mencari Nilai Untuk User



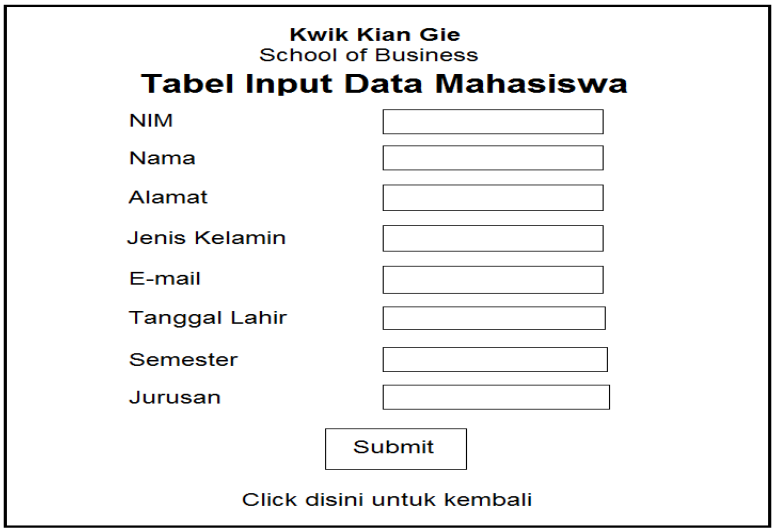
Gambar 4.3. Rancangan Form Mencari Nilai Untuk Admin



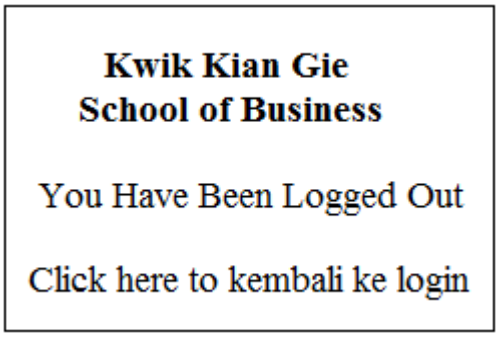
Gambar 4.4. Rancangan Form Untuk Mencari Mahasiswa yang Berperingkat 10 Besar



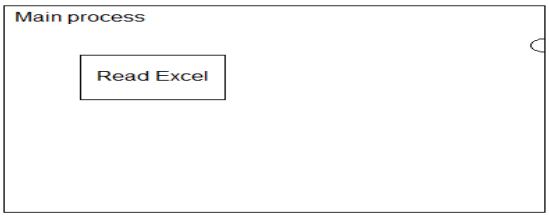
Gambar 4.5. Rancangan Form Untuk Menginput Data Mahasiswa



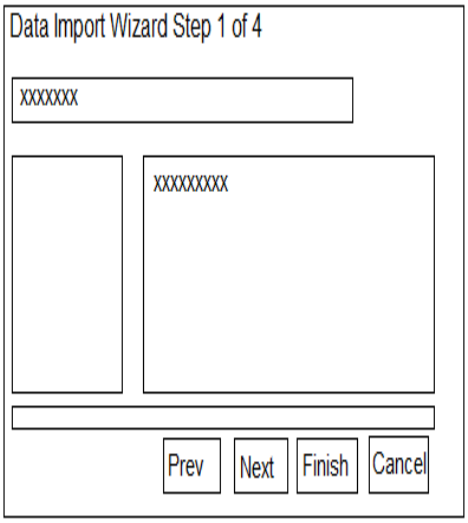
Gambar 4.6. Rancangan Form Untuk Menginput Nilai Mahasiswa



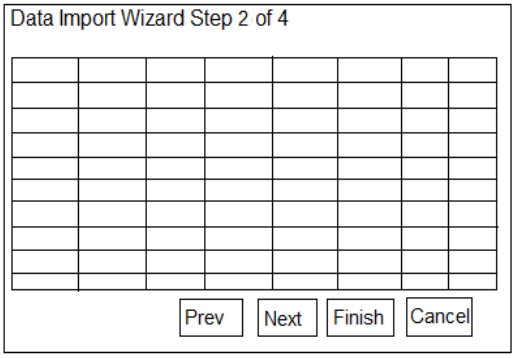
Gambar 4.7 *Form* Logout



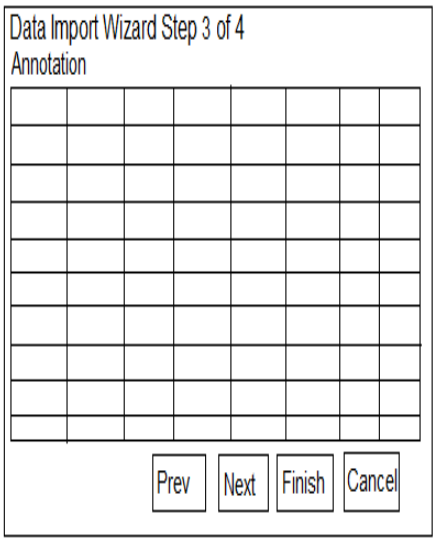
Gambar 4.8. Interface Rapidminer dengan ReadExcel



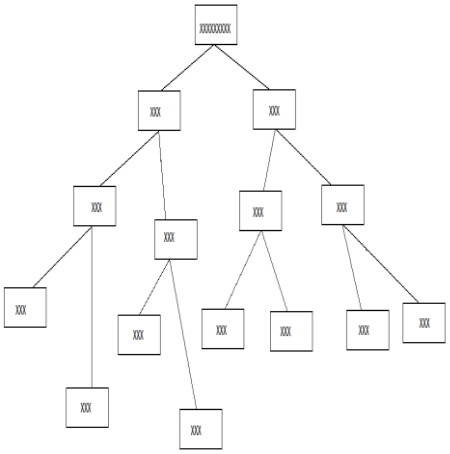
Gambar 4.9. Langkah Pertama Import Data dari Excel



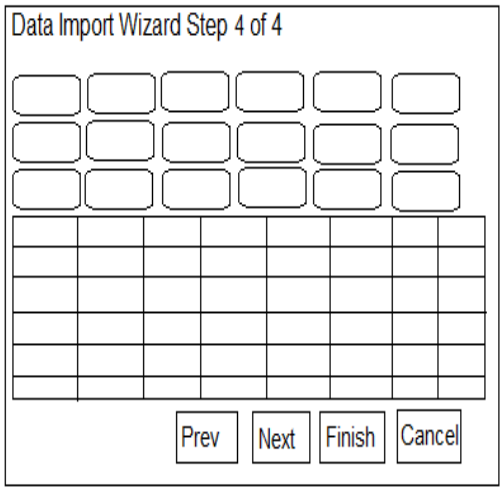
Gambar 4.10. Langkah Kedua Import Data dari Excel



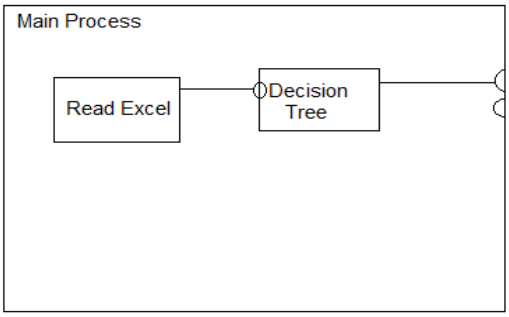
Gambar 4.11 Langkah Ketiga Import Data dari Excel



Gambar 4.14. Hasil Decision Tree



Gambar 4.12 Langkah Keempat Import Data dari Excel



Gambar 4.13. Read Excel yang Telah Digabung dengan Rapidminer

5. SIMPULAN

Simpulan

Dengan adanya proses pembuatan klasifikasi data nilai mahasiswa ini, maka dapat disimpulkan antara lain :

1. Mempermudah pencarian data yang diinginkan sehingga juga memermudah dalam pengambilan keputusan.
2. Membantu dalam mengintegrasikan data dan mencari data tentang nilai mahasiswa dengan mudah.
3. Melalui proses ini, proses yang dibuat dapat mempercepat proses penghasilan laporan dan meningkatkan ketepatan informasi yang dihasilkan.

6.DAFTAR REFEENSI

[1]. Connolly, Thomas and Carolyn Begg (2005), *Database System: A Pratical Approach to Design, Implementation, and Management*, 4th Edition, Harlow: Addison Wesley.

[2]. Coronel, Carlos (2013). Database *Systems Design, Implementation, and Management.*Kanada: Penerbit Joe Sabatino.

[3]. Hoffer, A. Jeffrey, Mary B. Prescott, Fred R. McFadden (2005*), Modern Database Management* , Edisi ke-7, New Jersey : Pearson Prentice Hall.

[4]. Laudon, C. Kenneth and Jane Laudon (2014),*Management Information Systems : Managing the Digital Firm*, Edisi ke-13, Global Edition.

[5]. McLeod, Jr. Raymond and George P. Schell (2007), *Management Information Systems*, Edisi ke -10, New Jersey : Pearson International Edition.
[6]. Neuman, W.Lawrence (2000), *Qualitative and Quantitative Approaches*, Edisi ke-4, Pearson Education Company.
[7]. Ramakhrisnan, Raghu dan Johannes Gehrke (2004), *Sistem Manajemen Database*, Edisi ke-3, McGraw Hill Education.
[8]. Riordan, M.Rebecca (2005), *Designing Effective Database System*, Pearson Education, Inc.
[9]. Welling, Luke and Laura Thomson (2009), *PHP and MYSQL Web Development*, Edisi ke-4, Pearson Education, Inc.
[10]. Yank, Kevin (2004), *Build Your Own Database Driven Website Using PHP & MYSQL*, Sitepoint.