

PERBANDINGAN SOFTWARE DATA MINING BERBAYAR DAN OPEN SOURCE

Dwi Nurul Huda

Mahasiswa STMIK Amikom Yogyakarta

E-mail : dwi.nurulhuda@yahoo.com



ABSTRAK

Data mining merupakan proses penggalian informasi tersembunyi dari suatu data yang berukuran besar. Pola-pola yang ditemukan dalam data mining dapat menjadi suatu informasi yang bermanfaat untuk digunakan selanjutnya seperti pada sistem pendukung keputusan. Kasus data mining masih menjadi kasus yang menarik untuk diteliti oleh para peneliti. Banyaknya software yang tersedia baik yang bersifat open source maupun berbayar yang digunakan dalam membantu dalam pengolahan data mining membuat peneliti dapat dengan mudah memilih software yang sesuai dengan kebutuhan kasus yang sedang diteliti dan kemampuan pengeluaran dari peneliti itu sendiri. Pada penulisan paper ini penulis hendak membahas perbandingan antara software berbayar dan software open source dimana penulis mengambil satu buah sample berbayar yaitu IBM SPSS Modeler dan satu buah sample yang bersifat open source yaitu WEKA. Kedua software tersebut banyak digunakan dalam membantu menyelesaikan proses data mining.

Kata Kunci: data mining, WEKA, IBM SPSS Modeler

1. PENDAHULUAN

Dunia ini memiliki data yang sangat besar tetapi informasi yang diperoleh dari data yang berjumlah besar tersebut masih sedikit. Data mining merupakan evolusi dari suatu teknologi informasi. Pada zaman primitif suatu data yang mengalami penumpukan dibiarkan begitu saja sebab dianggap tidak memiliki informasi rahasia didalamnya.



Gambar 1. Dunia Memiliki data yang kaya tetapi miskin akan informasi yang diperoleh

Belum adanya teknologi yang membantu dalam menggali informasi baru yang berasal dari penumpukan data tersebut karena data masih disimpan dalam bentuk kertas/arsip. Namun, seiring dengan perkembangan teknologi informasi yang semakin berevolusi diiringi dengan permasalahan yang semakin rumit maka data tersebut dapat terakumulasi, penggalian data terhadap data yang telah menumpuk tersebut perlu dilakukan.

Pemanfaatan data mining sendiri dapat dipandang secara komersial bisnis dan keilmuan. Pada komersial bisnis misalnya pengolahan data penjualan pada suatu swalayan, dimana proses penjualan dilakukan setiap hari yang mengakibatkan data terakumulasi setiap hari dan berakibat memiliki volume besar. Penggunaan data mining dapat menangani masalah penumpukan data yaitu dengan penggalian informasi yang dapat bermanfaat

dari data tersebut yang akan berguna dalam membantu menjawab permasalahan dalam suatu bisnis. Sedangkan dipandang secara keilmuan data mining dapat digunakan dalam menangkap, menganalisis dan menyimpan data yang berukuran besar yang dapat digunakan dalam menjawab permasalahan yang berhubungan dengan keilmuan.

Adapun proses pengolahan data mining dilakukan menggunakan *software* pembantu data mining. *Software* data mining ini terdiri dari 2 jenis pilihan yaitu *software* berbayar dan *software open source*. Contoh *software* berbayar antara lain adalah IBM SPSS Modeler, MATLAB, CART, KXEN, Angoss, ADAPA, Bayes Server dan masih banyak lainnya, sedangkan *software open source* antara lain adalah WEKA, Rapid Miner, AR Miner, AR Tools, KNIME, R, Orange, Rattle GUI, UNIMA, Cluto, Tanagra dan lain sebagainya.

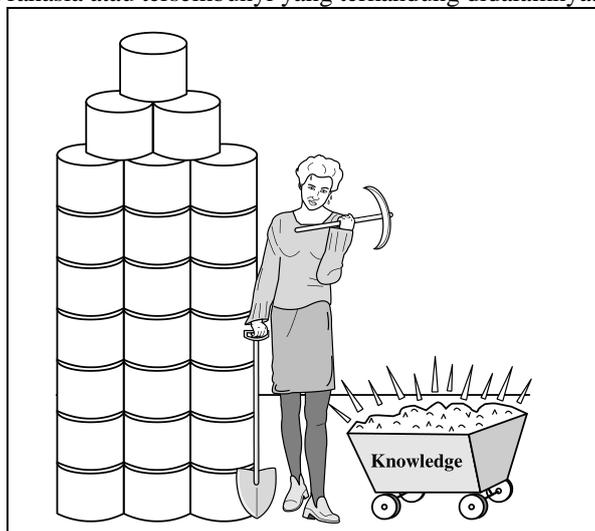
Pengguna *software* data mining dapat memilih *software* yang dikehendaki seperti *software* berbayar atau *open source* disesuaikan dengan kebutuhannya yang bersangkutan, juga pendekatan yang dilakukan dalam menyelesaikan kasus data mining seperti *clustering*, *classification*, regresi, visualisasi data atau lainnya. Masing-masing *software* dapat saja memiliki fitur yang berbeda-beda dalam melakukan pendekatan.

Perbandingan mengenai *software* berbayar ataupun *open source* merupakan hal menarik untuk dibahas, sebab terkadang pengguna yang memiliki *cost* yang lebih cenderung memilih *software* berbayar karena biasanya *software* berbayar bebas *bugs* serta menyajikan layanan yang lebih memuaskan dibanding *open source*, sedangkan *software open source* biasanya masih membutuhkan pengembangan-pengembangan sebab masih terdapat *bugs* didalamnya, sehingga layanan yang ditawarkan kurang memuaskan. Namun, seluruhnya kembali kepada pengguna *software* sebab jika dilihat dari sisi *cost* tentu saja lebih menguntungkan *open source*.

2. TINJAUAN UMUM

2.1 Data mining

Data mining merupakan proses penggalian informasi atau pengetahuan dengan cara menemukan pola-pola tersembunyi dari suatu data yang berukuran besar. Data mining dapat memenuhi kebutuhan akan informasi tersembunyi atau rahasia yang dapat digunakan dalam proses pengambilan keputusan suatu organisasi. Jika berbicara tentang data, maka tidak akan ada habisnya, sebab dalam dunia ini terdapat banyak data yang tidak terbatas sebab makin hari makin bertumpuk data tersebut. Namun, informasi yang diperoleh dari data yang dimiliki tersebut masih minim diantaranya disebabkan tidak mampunya seseorang dalam mengolah data yang dimiliki tersebut, sehingga informasi yang terkandung didalam data tersebut akan sulit untuk ditemukan. Penumpukan data yang semakin hari semakin banyak, menyebabkan proses penggalian data dibutuhkan dalam proses penggalian informasi rahasia atau tersembunyi yang terkandung didalamnya.

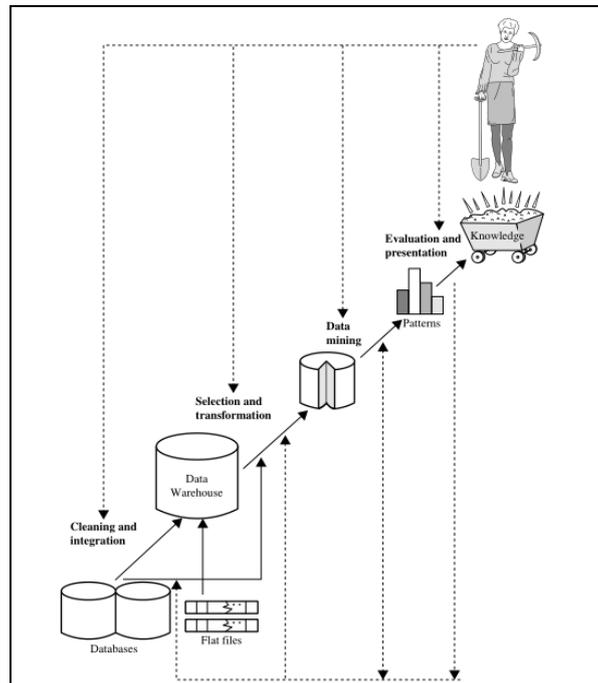


Gambar 2. Penggalian pengetahuan berdasarkan pola yang menarik dari suatu data

Penemuan *knowledge* dari suatu data diperoleh dengan melakukan beberapa tahapan yaitu :

1. *Data cleaning*, dengan menghilangkan *noise* dan ketidakconsistenan data
2. *Data integration*, dengan menggabungkan beberapa sumber data yang sekiranya dapat digabungkan
3. *Data selection*, dengan mengambil data yang relevan dari *database* untuk dianalisis
4. *Data transformation*, yaitu dengan melakukan transformasi data untuk dilakukan proses selanjutnya yaitu penggalian data
5. *Data mining*, pemilihan metode untuk mengekstrak pola data
6. *Evaluasi pola*, mengevaluasi pola yang menarik yang mewakili suatu pengetahuan/informasi

- Presentasi pengetahuan, dimana representasi dan teknik pengetahuan tadi akan dipresentasikan untuk selanjutnya dapat digunakan untuk pengguna



Gambar 3. Tahapan penemuan pengetahuan

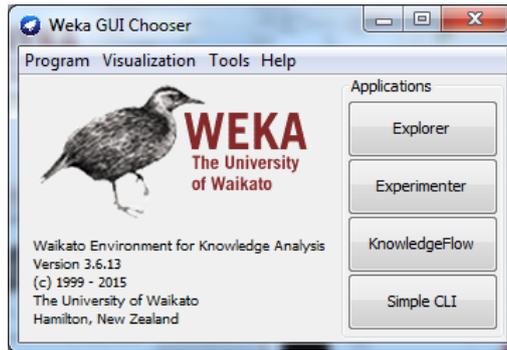
Menurut Witten dkk (2011) “*Data mining is about solving problems by analyzing data already present in databases*” artinya bahwa ketika berbicara tentang data mining berarti berbicara tentang pemecahan masalah dengan cara menganalisis suatu data yang berada dalam *database*. Pemecahan masalah disini artinya adalah menemukan pola tersembunyi dari suatu *database*, sehingga dapat membantu dalam pemecahan suatu masalah. Sebagai contoh misalkan untuk melihat loyalitas pelanggan terhadap suatu produk berkenaan dengan pasar yang semakin bersaing satu sama lainnya. Pada kasus ini akan dianalisis pola perilaku pelanggan yang berubah-ubah terhadap loyalitas penggunaan suatu produk berdasarkan karakteristik masing-masing pelanggan ketika tetap menggunakan produk tersebut atau memilih berpaling kepada produk lainnya. Hal ini dapat dilihat berdasarkan profil pelanggan serta pilihan produk pelanggan

Data mining menurut pendapat (Sulianta dan Juju, 2010) adalah metode yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada *database*. Sedangkan menurut (Turban dkk, 2007) data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Berdasarkan pendapat tadi, maka dapat diketahui bahwa data mining sendiri terdiri dari berbagai macam disiplin ilmu dari statistik, matematika hingga kecerdasan buatan

2.2 WEKA

Salah satu *software open source* yang digunakan dalam membantu proses data mining adalah WEKA. WEKA merupakan singkatan dari *Waikato Environment for Knowledge Analysis*. WEKA merupakan *software open source* yang dikembangkan oleh Universitas Waikato yang berada di New Zealand. WEKA sendiri ditulis menggunakan Java dan di distribusikan dibawah lisensi GNU Public. WEKA memiliki berbagai macam *function* didalamnya yang dapat digunakan dalam membantu dalam melakukan proses yang berkaitan dengan data mining. Selain itu WEKA dapat berjalan di berbagai *platform* dari windows, linux, mac dan lainnya.

WEKA mampu menyelesaikan masalah dalam data mining terutama klasifikasi. WEKA memiliki GUI yang bersifat mudah untuk dipelajari, sehingga cocok untuk digunakan/diterapkan oleh banyak tingkatan pengguna. Adapun fasilitas yang disediakan oleh WEKA antara lain adalah mampu menyelesaikan *pre-processing* data, klasifikasi, *clustering*, *regresi*, asosiasi dan visualisasi. Pengguna dapat melakukan *preprocessing* data tanpa menuliskan kode program samasekali. Hal ini membuat pengguna dapat lebih berfokus kepada algoritma tanpa memikirkan kode untuk hasil evaluasi.



Gambar 4. Tampilan GUI WEKA

Beberapa penelitian menggunakan WEKA sebagai *software* pembantu berkaitan dengan proses data mining, diantaranya adalah :

Tabel 1. Tabel penelitian terdahulu menggunakan *software* WEKA

No	Penelitian Terdahulu
1	<p>Judul : <i>Using Apriori with WEKA for Frequent Pattern Mining</i></p> <p>Penulis : Paresh Tanna dan Dr. Yogesh Ghodasara</p> <p>Sumber : <i>International Journal of Engineering Trends and Technology (IJETT)</i> – Volume 12 Number 3 - Jun 2014</p> <p>Keterangan : penelitian ini membahas tentang pengaplikasian WEKA dengan menggunakan algoritma apriori. Dimana algoritma apriori merupakan algoritma yang paling simple pada data mining. Penggunaan algoritma ini memanfaatkan <i>software</i> WEKA dengan merujuk pada aturan asosiasi pada data mining</p>
2	<p>Judul : <i>Data Mining: An experimental approach with WEKA on UCI Dataset</i></p> <p>Penulis : Ajay Kumar dan Indranath Chatterjee</p> <p>Sumber : <i>International Journal of Computer Applications (0975 – 8887). Volume 138 – No.13, March 2016</i></p> <p>Keterangan : penelitian ini membahas tentang penyajian klasifikasi dasar dan algoritma <i>clustering</i>. Perbandingan pohon keputusan, <i>bayes</i>, <i>KNN</i>, <i>K-Means</i>, <i>partional clustering</i>, <i>hierarchy clustering</i> dan <i>fuzzy clusterin</i>. Penelitian ini memanfaatkan ketersediaan <i>IRISH dataset</i> dari UCI pusat data.</p>
3	<p>Judul : <i>Review on Meta Classification Algorithms using WEKA</i></p> <p>Penulis : Rausheen Bal dan Sangeeta Sharma</p> <p>Sumber : <i>International Journal of Engineering Trends and Technology (IJETT)</i> – Volume 35 Number 1 - Mei 2016.</p> <p>Keterangan : penelitian ini membahas tentang penggunaan WEKA sebagai</p>

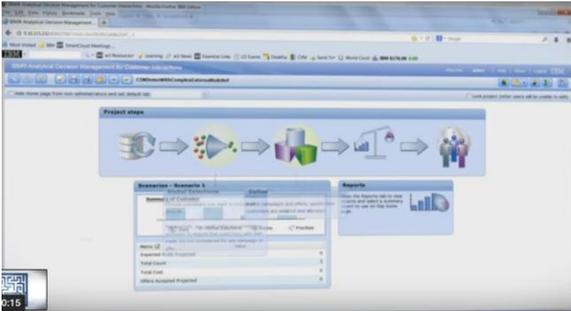
	software klasifikasi terbaik. Dan membahas 3 buah algoritma yaitu <i>Forest Algoritma</i> , <i>REP-Tree</i> , <i>Bagging Classifier</i>
--	---

Saat ini versi terakhir WEKA berada pada WEKA 3.6, dimana terdapat algoritma klasifikasi baru yang termasuk didalamnya, antara lain :

1. *Bayesian logistic regression*, merupakan metode untuk mengkategorikan *text*, dengan both Gaussian and Laplace priors.
2. *Best-first decision tree*, membangun pohon keputusan menggunakan strategi pencarian *best-first*
3. *Decision table naive Bayes hybrid*, merupakan pembelajaran hibrid dengan menggabungkan *decision tables* dan *bayes*
4. *Discriminative multinomial naive Bayes*, pengklasifikasian *bayes* dengan pembelajaran paramete diskriminatif untuk mengkategorisasikan teks
5. *Functional trees*, merupakan pohon keputusan dengan *oblique split* dan fungsi linier
6. *Gaussian processes*, mengimplementasikan metode Gaussian untuk regresi
7. *Simple CART*, merupakan pohon keputusan yang mengimplementasikan minimal biaya
8. *Variants of AODE (Averaged One-Dependence Estimators)* dengan subsumption resolution (AODEsr) and *weighted AODE (WAODE)*.
9. *Wrapper classifiers*, memungkinkan algoritma LibSVM dan LibLINEAR digunakan di WEKA

2.3 IBM SPSS Modeler

IBM SPSS Modeler merupakan *software* yang berasal dari IBM yang digunakan untuk membangun *predictive models* atau memprediksi tren. IBM SPSS Modeler memiliki antarmuka yang memungkinkan pengguna untuk memanfaatkan algoritma statistik dan data mining tanpa melakukan pemrograman.



Gambar 5. Tampilan IBM SPSS Modeler

Beberapa jenis layanan IBM SPSS Modeler yang dapat dipilih antara lain :

1. IBM SPSS Modeler Gold, membangun dan menyebarkan model prediksi secara langsung kedalam proses bisnis, sehingga dapat menghasilkan keputusan yang tepat. Terdapat pula IBM SPSS Modeler Cloud, dengan layanan bulanan berbasis web
2. IBM SPSS Modeler Premium, mampu menganalisis teks, analisis entitas dan analisis jaringan sosial bersama dengan pemodelan dan persiapan teknik otomatis
3. IBM SPSS Modeler Professional, mampu mengungkap pola tersembunyi dalam data terstruktur. IBM SPSS Modeler Professional mampu memberikan algoritma yang canggih , manipulasi data dan pemodelan otomatis untuk membangun model prediksi
4. IBM SPSS Modeler Personal, mampu merancang dan membangun model *predictive*, dapat membantu memecahkan masalah bisnis dengan mengungkapkan pola dan tren dalam data terstruktur.

IBM SPSS Modeler merupakan *software* berbayar, maka pengguna software harus rela mengeluarkan biaya lebih. Jumlah biaya yang dikeluarkan sendiri bervariasi tergantung pilihan layanan yang diinginkan. Makin lengkap dan makin banyak layanan yang ditawarkan maka akan semakin mahal harganya.

Beberapa penelitian terdahulu menggunakan software IBM SPSS Modeler

Tabel 2. Tabel penelitian terdahulu menggunakan *software* IBM SPSS Modeler

No	Penelitian Terdahulu
1	Judul : <i>A Survey and Compare the Performance of IBM SPSS Modeler and</i>

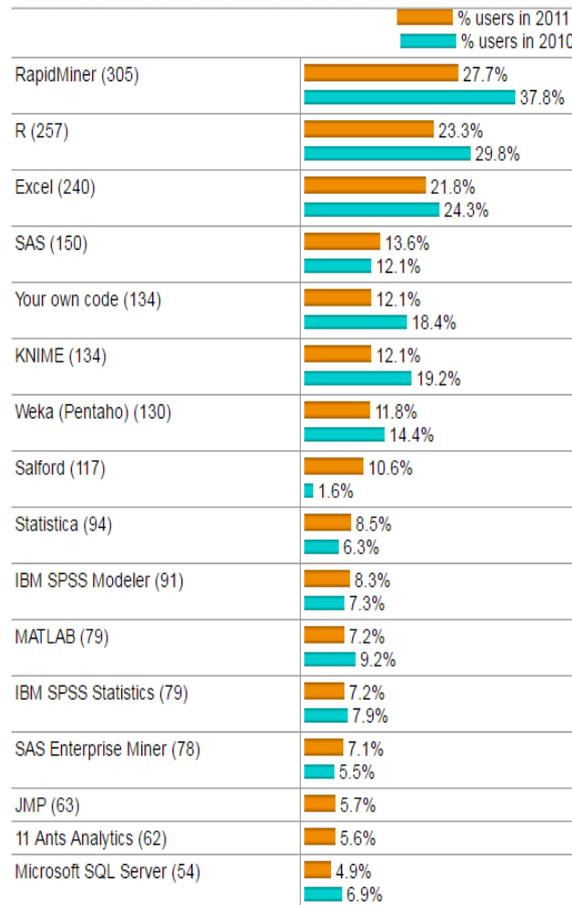
	<p><i>Rapid Miner Software for Predicting Liver Disease by Using Various Data Mining Algorithms</i></p> <p>Penulis : Moloud Abdar</p> <p>Sumber : <i>Cumhuriyet Science Journal (CSJ)</i> – Volume 36 Number 3 - Jun 2015</p> <p>Keterangan : penelitian ini membandingkan <i>software</i> IBM SPSS Modeler dengan rapidminer untuk kasus catatan pasien liver. Diperoleh hasil bahwa tingkat akurasi yang dimiliki oleh algoritma rapid miner memiliki persentasi yang lebih sedikit daripada IBM SPSS Modeler, tetapi lebih akurat sebab rincian yang diberikan oleh rapidminer lebih lengkap dibanding dengan IBM SPSS Modeler.</p>
2	<p>Judul : <i>Disaster Prediction System Using IBM SPSS Data Mining Tool</i></p> <p>Penulis : B.Renuka Devi , Dr.K.NageswaraRao , Dr.S.PallamSetty , Dr.M.NagabhushanaRao</p> <p>Sumber : <i>International Journal of Engineering Trends and Technology (IJETT)</i> Volume 4 – Ags 2013</p> <p>Keterangan : penelitian ini membahas tentang penggunaan algoritma C5.0, Neural Net dan QUEST dalam memprediksi bencana. Hasilnya algoritma C5.0 terbukti lebih efisien dibanding lainnya.</p>

3. PEMBAHASAN

3.1 Analisis Jumlah Pengguna *Software* Data Mining

Sebelumnya diketahui bahwa berdasarkan polling perbandingan penggunaan *software* data mining beberapa tahun yang lalu. Perbandingan pengguna *software* ini membandingkan pengguna *software* yang menggunakan *software* data mining

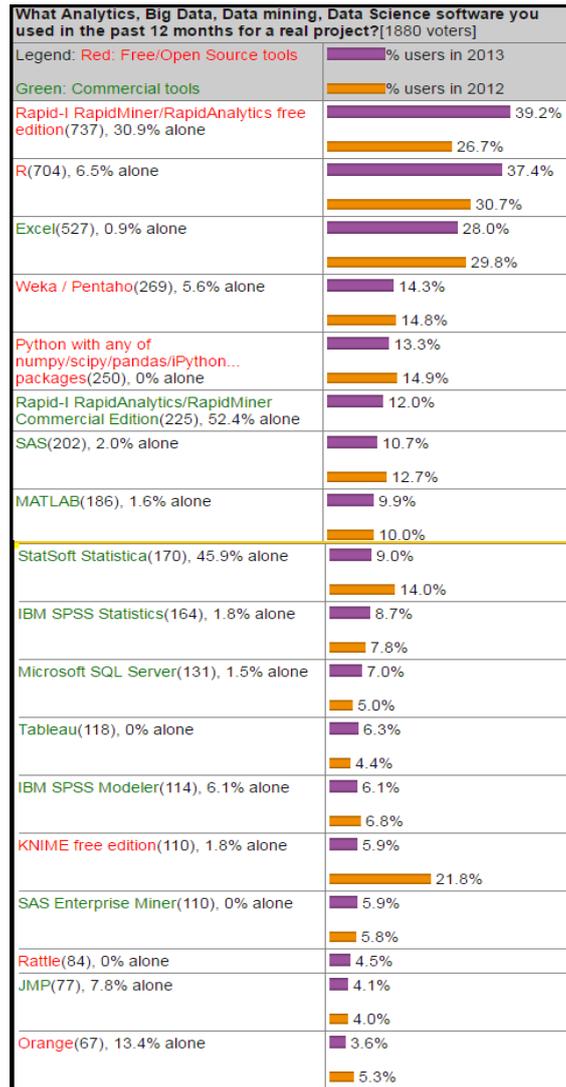
yang berbayar dan bersifat *open source* berdasarkan urutan 15 besar. Dimana dari data tersebut diketahui bahwa jumlah pengguna WEKA menurun sedangkan IBM SPSS Modeler meningkat dibanding tahun sebelumnya.



Gambar 6. Perbandingan pengguna *software* data mining antara tahun 2010 dan 2011

Kemudian dua tahun kemudian dilakukan perbandingan kembali *software* data mining yang digunakan oleh pengguna *software* data mining. Adapun pengguna yang melakukan vote lebih banyak dari tahun sebelumnya. Namun, pada gambar 7 dibawah ini hanya ditampilkan sekitar 18 *software* data mining saja. Untuk lebih lengkapnya dapat diakses melalui laman <http://www.kdnuggets.com/polls/2013/analytics-big-data-mining-data-science-software.html>.

Adapun *software* yang dibandingkan masih dengan *software* yang sama yaitu WEKA dan IBM SPSS Modeler. Terdapat perbedaan penurunan pengguna untuk kedua *software* tersebut dari tahun sebelumnya. Tetapi jika dilihat dari segi pemilihan *software*, lebih banyak pengguna yang melakukan vote yang menggunakan *software* WEKA di bandingkan dengan IBM SPSS Modeler. Pada tahun 2013 tersebut jumlah pengguna yang melakukan vote lebih meningkat dibanding 2 tahun sebelumnya. Walaupun begitu vote ini bukanlah menjadi satu-satu hasil penentu bahwa *software* data mining WEKA lebih baik dibandingkan dengan *software* data mining IBM SPSS Modeler.



Gambar7.Perbandingan pengguna *software* data mining antara tahun 2012 dan 2013

3.2 Analisis Perbandingan *Software* Data Mining WEKA dan IBM SPSS Modeler

Berikut ini akan di jelaskan perbandingan *software* data mining WEKA dan IBM SPSS Modeler termasuk dari segi kelebihan dan kekurangan dari masing-masing *software* tersebut.

Tabel 3.Tabel Perbandingan *Software* Berbayar (IBM SPSS Modeler)dan *Open Source*(WEKA) Data Mining

No	Vendor	Di publikasikan oleh University Of Waikato New Zeland	SPSS Inc an IBM Company
1	tool Name	WEKA 3.6 merupakan versi terakhir WEKA	IBM SPSS Modeler terakhir versi 17
2	License	Open source software GNU General Public License (GPL)	Commercial software /Berbayar
3	Pendekatan data mining	Machine learning, preposesing data mining,	Classification Discovery, Cluster Discovery, Regression

		<i>clustering, classification, regresi, association rules, attribute selection, experiments, workflow, visualization</i>	<i>Discovery, Association Discovery, Text Mining, Outlier Discovery, Data Visualisation, Discovery Visualisation, Sequence Analysis, Web Analytics, Social Network Analysis</i>
4	Algoritma	Tergantung pendekatan, contohnya apriori, filtered associator, generalized sequential pattern, tertius, LVQ, SOM (model algoritma lain dapat http://weka.class.algos.sourceforge.net/)	Auto classifier, pohon keputusan, logistic, SVM, time series, auto clustering, K-means, anomali detection, apriori, carma, sequence (model algoritma lain dapat mengunjungi http://knowledgeproviders.com/wp-content/uploads/2013/08/IBM-SPSS-Modeler-Professional.pdf)
5	Format data	*.ARFF Singkatan dari <i>Attribute-Relation FileFormat</i>	*.ODBC dan seluruh ekstensi database yang masuk ke dalam vendor IBM
6	Kelebihan	<ol style="list-style-type: none"> 1. <i>Software WEKA berlisensi open source, sehingga tidak berbayar</i> 2. Dapat berjalan diberbagai macam platform, sehingga bersifat portable 3. Memiliki berbagai jenis teknik dalam preprocessing data dan juga pemodelan data 4. Memiliki GUI yang mudah dipahami 	<ol style="list-style-type: none"> 1. Membantu membangun, mengevaluasi dan menyebarkan <i>predictive model</i>, model <i>clustering</i> dan model asosiasi 2. Membantu dalam menganalisis banyak data secara bersamaan 3. Membantu dalam mengambil keputusan secara cepat 4. Mendukung akses <i>database</i> menggunakan ODBC dan produk lainnya yang termasuk kedalam vendor IBM

		oleh pengguna terutama pengguna awam	
		5. Mampu diimplementasikan menggunakan bahasa pemrograman Java	
6	Kekurangan	<ol style="list-style-type: none"> 1. Tidak dapat memasukkan data yang banyak untuk diproses, sehingga jika <i>database</i> yang dimasukan melebihi kemampuan tampung WEKA akan terjadi <i>overloading</i> data hingga <i>software</i> WEKA dapat <i>close/keluar</i> secara otomatis 2. Hanya mampu mendukung akses menggunakan database SQL 	<ol style="list-style-type: none"> 1. IBM SPSS Modeler ini merupakan <i>software</i> data mining berbayar, sehingga harus siap mengeluarkan cost sesuai dengan layanan yang diinginkan 2. Untuk penggunaan akademis yang menggunakan pendekatan bootstrap, <i>software</i> ini tertinggal bila dibandingkan dengan R atau SAS 3. Grafis bawaan <i>software</i> masih kurang baik terkadang terdapat menu yang dirasa kurang 4. Sering ada masalah kompatibilitas dengan versi sebelumnya

KESIMPULAN

Berdasarkan perbandingan sebelumnya diketahui bahwa pengguna *software* cenderung memilih menggunakan *software* WEKA dibanding dengan IBM SPSS Modeler. Hal ini dimungkinkan karena *software* WEKA yang memiliki lisensi GNU GPL atau tidak bersifat komersil. Selain itu tidak adanya tingkatan layanan yang ditawarkan oleh WEKA menjadi salah satu pilihan pengguna menggunakan *software* WEKA, sehingga layanan yang diberikan kepada seluruh pengguna *software* akan sama. Sedangkan IBM SPSS Modeler yang selain berbayar, layanan yang diberikan oleh vendor sesuai dengan biaya yang dikeluarkan seperti IBM SPSS Gold Cloud yang merupakan layanan berlangganan perbulan dapat menjadi pilihan layanan yang paling baik sebab layanan yang diberikan lebih banyak dibanding dengan IBM SPSS Modeler jenis lainnya. Selain itu kekurangan yang dimiliki IBM SPSS Modeler lebih banyak dibanding dengan WEKA apabila dilihat dari Tabel 3 diatas.

PUSTAKA

- Abdar, Moloud. 2015. *A Survey and Compare the Performance of IBM SPSS Modeler and Rapid Miner Software for Predicting Liver Disease by Using Various Data Mining Algorithms*. Cumhuriyet Science Journal (CSJ), Volume. 36 Number. 3
- Devi, B.Renuka., Rao, K.Nageswara., Setty, S.Pallam dan Rao, M. Nagabhushana. 2013. *Disaster Prediction System Using IBM SPSS Data Mining Tool*. International Journal of Engineering Trends and Technology (IJETT) Volume 4
- Han, Jiawei., Kamber, Michelin., dan Jian, Pei. (2012). *Data Mining Concepts And Techniques*. Morgan Kaufman..Elsevier. United States. 2012
- Kumar, Ajay.,Chatterjee, Indranath . 2016. *Data Mining: An experimental approach with WEKA on UCI Dataset*. International Journal of Computer Applications Volume 138 Number 13
- Mehmed, kantardzic. (2011). *Data Mining : Concepts, Models, Methods, And Algorithms*. Institute of Electrical and Electronics Engineers. Canada. 2011
- Rausheen.,Sharma, Sangeeta. 2016.*Review on Meta Classification Algorithms using WEKA*.International Journal of Engineering Trends and Technology (IJETT) Volume 35 Number 1
- Sulianta, Feri., Juju, Dominikus. *Data Mining : Meramalkan Bisnis Perusahaan*. PT. Elex Media Komputindo. Jakarta. 2010
- Tanna, Paresh., Ghosara, Yogesh. 2014. *Using Apriori with WEKA for Frequent Pattern Mining*.International Journal of Engineering Trends and Technology (IJETT) Volume 12 Number 3
- Turban, Efraim., Aronson, Jay E, Liang, Ting-Peng dan Sharda, Ramesh. (2007). *Decision Support And Bussiness Intellegence Systems (8th Edition)*. Pretince Hall. New Jersey. 2007
- Witten, Ian. H., Frank, Eibe., dan Hall, Mark A. (2011). *Data Mining Practical Machine Learning Tools And Techniques*.Morgan Kaufman.Elsevier. United States. 2011
- Algorithm WEKA*. Diakses pada 4 Juni 2016 dari <http://weka.classalgos.sourceforge.net/>
- Algorithm IBM SPSS Modeler*. Diakses pada 4 Juni 2016 dari <http://knowledgeproviders.com/wp-content/uploads/2013/08/IBM-SPSS-Modeler-Professional.pdf>
- Data Mining Tool, Software And Applications*. Diakses pada 2 Juni 2016 dari <http://www.the-datamine.com/Software/DataMiningSoftware>
- Data Mining/ Analytic Tools used Polling 2011*. Diakses pada 4 June 2016 dari <http://www.kdnuggets.com/polls/2011/tools-analytics-data-mining.html>.
- Data Mining/ Analytic Tools used Polling 2013*. Diakses pada 4 June 2016 dari <http://www.kdnuggets.com/polls/2013/analytics-big-data-mining-data-science-software.html>
- SPSS Modeler*. Diakses pada 4 June 2016 dari <http://www-03.ibm.com/software/products/en/spss-modeler>
- What are the advantages and disadvantages of using The SPSS*. Diakses pada 4 June 2016 dari <https://uk.answers.yahoo.com/question/index?qid=20111208160641AAKg2PV>