

IMPLEMENTASI ALGORITMA C5.0 PADA KELULUSAN PESERTA UJIAN KEMAHIRAN BERBAHASA INDONESIA (UKBI) PADA BALAI BAHASA SUMATERA UTARA

Mauhati Pardede¹, Efori Buulolo², Eferoni Ndruru³

Program Studi Teknologi Informatika STMIK Budi Darma, Medan, Indonesia
Email: ¹mauhati.pardede02@gmail.com, ²buuloloeffori21@gmail.com, ³eferonindruru@gmail.com

Abstrak

Berkembangnya era modern membuat banyak orang semakin terpengaruh dalam menggunakan bahasa asing di kehidupan sehari-hari. Hampir setiap hari kalangan masyarakat terlebih generasi di zaman sekarang ini lebih memahami dalam penggunaan bahasa asing dari pada bahasa yang berasal dari negerinya sendiri. Salah satu usaha memperkuat kedudukan Bahasa Indonesia dilakukan oleh Badan Balai Bahasa dengan mengembangkan alat uji kemahiran berbahasa Indonesia yang disebut dengan Uji Kemahiran Berbahasa Indonesia (UKBI). Uji Kemahiran Berbahasa Indonesia (UKBI) dinilai berdasarkan beberapa faktor kemampuan peserta yaitu mendengarkan, meresponkan kaidah, membaca, menulis dan berbicara. Selama ini peserta UKBI pada Balai Bahasa Sumatera Utara sebagian ada yang tidak lulus atau gagal, salah satu faktor penyebabnya adalah penilaian tidak terpenuhi. Untuk mengatasi permasalahan diatas diperlukan sebuah metode. Adapun teknik yang dapat dimanfaatkan dalam proses mengolah data adalah klasifikasi dengan algoritma C5.0. Algoritma C5.0 adalah salah satu algoritma klasifikasi yang diterapkan pada teknik *decision tree* yang diproses menggunakan *information gain*. Dalam memilih atribut untuk pemecah obyek dalam beberapa kelas harus dipilih atribut yang menghasilkan *information gain* paling besar atribut dengan nilai *information gain* tertinggi akan dipilih sebagai parent bagi node selanjutnya.

Kata kunci: Uji Kemahiran Berbahasa Indonesia, Algoritma C5.0

Abstract

The development of the modern era has made many people more influenced in using foreign languages in their daily lives. Almost every day, the people especially the generation in this era understand more about the use of foreign languages than the language from their own country. One of the efforts to strengthen the position of Indonesian is carried out by the Balai Bahasa by developing an Indonesian language proficiency test tool called the Indonesian Language Proficiency Test (UKBI). The Indonesian Language Proficiency Test (UKBI) was assessed based on several factors of the participant's ability to listen, respond to the rules, read, write and speak. So far some of the UKBI participants at the North Sumatra Language Center have failed or failed, one of the contributing factors is that the assessment was not fulfilled. To overcome the problems above, a method is needed. The technique that can be utilized in the process of processing data is the classification with the C5.0 algorithm. C5.0 algorithm is a classification algorithm applied to decision tree techniques that are processed using information gain. In selecting attributes for object breakers in several classes, the attribute that produces the greatest information gain with the highest information gain value will be chosen as the parent of the next node.

Keywords: Indonesian Language Proficiency Test, C5.0 Algorithm

1. PENDAHULUAN

Kelulusan adalah kegiatan paling akhir dari manajemen seorang peserta didik, kelulusan juga dapat diartikan dengan pernyataan dari lembaga pendidikan tentang telah diselesaikannya program pendidikan yang harus diikuti peserta. Kelulusan dinilai berdasarkan kompetensi lulusan dan proses kelulusan. Kompetensi lulusan ialah kemampuan minimal yang harus dicapai oleh peserta setelah mengikuti pendidikan pada jenjang atau satuan pendidikan tertentu. Proses kelulusan adalah kegiatan paling akhir dari manajemen peserta didik.[1]

Balai Bahasa Sumatera Utara adalah lembaga yang memiliki visi mewujudkan insane berkarakter dan jati diri bangsa melalui bahasa dan sastra Indonesia yang memegang peran penting dalam memperkuat kedudukan Bahasa Indonesia di tingkat nasional, mengembangkan penggunaan Bahasa Indonesia di tingkat regional, dan memartabatkan Bahasa Indonesia di tingkat Internasional. Salah satu usaha memperkuat kedudukan Bahasa Indonesia dilakukan oleh Badan Balai Bahasa dengan mengembangkan alat uji kemahiran berbahasa Indonesia yang disebut dengan Uji Kemahiran Berbahasa Indonesia (UKBI).[5]

Uji Kemahiran Berbahasa Indonesia (UKBI) dinilai berdasarkan beberapa faktor kemampuan peserta yaitu mendengarkan, meresponkan kaidah, membaca, menulis dan berbicara. Selama ini peserta UKBI pada Balai Bahasa Sumatera Utara sebagian ada yang tidak lulus atau gagal, salah satu factor penyebabnya adalah penilaian tidak terpenuhi, dan begitu juga pada kelulusan peserta UKBI. Selama ini Balai Bahasa Sumatera Utara belum memanfaatkan data-data kelulusan peserta UKBI sebelumnya untuk menganalisis factor penyebab lulus atau gagalnya peserta UKBI .[5]

Untuk mengatasi permasalahan diatas diperlukan sebuah metode. Adapun teknik yang dapat dimanfaatkan dalam proses mengolah data adalah klasifikasi dengan algoritma C5.0. Algoritma C5.0 adalah salah satu algoritma klasifikasi yang diterapkan pada teknik *decision tree* yang diproses menggunakan *information gain*. Dalam memilih atribut untuk pemecah obyek dalam beberapa kelas harus dipilih atribut yang menghasilkan *information gain* paling besar atribut dengan nilai *information gain* tertinggi akan dipilih sebagai parent bagi node selanjutnya. Dengan algoritma C5.0

diharapkan menghasilkan analisis yang akurasi tinggi dan keputusan yang tepat berdasarkan atribut-atribut yang telah dianalisis untuk sebuah pengklasifikasian kelulusan peserta ujian.[2]

2. TEORITIS

2.1 Implementasi

Dalam kenyataannya, implementasi merupakan proses untuk melaksanakan ide, program atau seperangkat aktivitas baru dengan harapan orang lain dapat menerima dan melakukan perubahan. Dalam konteks implementasi kurikulum pendekatan-pendekatan yang telah dikemukakan diatas memberikan tekanan pada proses, suatu aktivitas yang digunakan untuk mentransfer ide atau gagasan, program yang dituangkan dalam bentuk kurikulum desain (tertulis) agar dilaksanakan sesuai dengan desain tersebut. Masing-masing pendekatan itu mencerminkan tingkat pelaksanaan yang berbeda[3].

2.2 Data Mining

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisa dan mengekstrak pengetahuan secara otomatis. Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan diwaktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan lainnya[3][4].

2.3 Algoritma C5.0

Algoritma C5.0 adalah salah satu algoritma klasifikasi data mining yang khususnya diterapkan pada teknik *decision tree*. C5.0 merupakan penyempurnaan algoritma sebelumnya yang dibentuk oleh *Ross Quinlan* pada tahun 1987, yaitu ID3 dan C4.5. Dalam algoritma ini pemilihan atribut yang akan diproses menggunakan *information gain*. Dalam memilih atribut untuk pemecah objek dalam beberapa kelas harus dipilih atribut yang menghasilkan *information gain* paling besar. Atribut dengan nilai *information gain* tertinggi akan dipilih sebagai parent bagi *node* selanjutnya[2].

Berikut ini merupakan rumus untuk perhitungan nilai *entropy*, seperti pada persamaan satu berikut:

$$Entropy(S) = \sum_{i=0}^n - p_i * \log_2 p_i \quad (1)$$

Keterangan :

S : Himpunan kasus

n : Jumlah Partisi S

p_i : Jumlah kasus pada partisi ke – I

Untuk mendapatkan nilai *gain* selanjutnya digunakan formula dibawah ini :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan :

S : Himpunan kasus

n : Jumlah partisi atribut A

$|S_i|$: Jumlah kasus pada partisi ke-i

$|S|$: Jumlah kasus dalam S[2]

Sebuah pohon keputusan adalah *classifier* yang direpresentasikan sebagai struktur pohon, di mana masing-masing simpul adalah simpul daun. *Klasifikasi* yang berlaku untuk semua kasus yang mencapai daun atau *node non-leaf*, beberapa tes dilakukan pada nilai atribut tunggal, dengan satu cabang dan sub-pohon untuk setiap kemungkinan hasil tes. *Node* dalam pohon keputusan melibatkan pengujian atribut tertentu. Biasanya, tes pada *node* membandingkan nilai atribut dengan konstan. Namun, beberapa pohon membandingkan dua atribut satu sama lain, atau menggunakan beberapa fungsi dari satu atau lebih atribut. *Node* daun memberikan *klasifikasi* yang berlaku untuk semua kasus yang mencapai daun, atau satu set *klasifikasi*, atau distribusi probabilitas atas semua klasifikasi yang mungkin[2].

3. ANALISA DAN PEMBAHASAN

3.1 Analisa Masalah

Analisa masalah dapat dikatakan sebagai suatu proses untuk menemukan penyebab masalah. Sebagaimana telah diuraikan pada latar belakang masalah bahwa Balai Bahasa Sumatera Utara yang bergerak dalam mengembangkan alat uji kemahiran berbahasa Indonesia (UKBI) masih ada sebagian peserta yang tidak lulus atau gagal yang menjadi salah satu faktor penyebabnya adalah penilaian tidak terpenuhi. Uji Kemahiran Berbahasa Indonesia (UKBI) dinilai berdasarkan beberapa faktor kemampuan peserta yaitu mendengarkan, meresponkan kaidah, membaca, menulis dan berbicara.

Dengan menggunakan klasifikasi Algoritma C5.0 dapat membantu Balai Bahasa dalam pengklasifikasian kelulusan peserta ujian. Proses mencari data yang diperoleh dari hasil studi lapangan seperti wawancara dan pengamatan yang dilakukan dan hasilnya akan membentuk pohon keputusan.

Tabel 1. Data Peserta Ujian Kemahiran Berbahasa Indonesia (UKBI)

Nama	Seksi I Mendengarkan	Seksi II Merespon Kaidah	Seksi III Membaca	Seksi IV Menulis	Seksi V Berbicara	Peringkat	Keterangan
Niken Ayu	420	512	475	350	460	V.Semenjana	Tidak Lulus
Eka Nianti	520	650	625	485	425	IV.Madya	Lulus
Lidia Lestari	560	710	685	510	490	III.Unggul	Lulus
Seri Wahyuni	680	720	545	680	580	II.Sangat Unggul	Lulus
Nasution							
Dwi Ayu Utami	252	310	320	400	260	VII.Terbatas	Tidak Lulus
Nasution							
Mai Nanda	253	420	535	470	320	VI.Marginal	Tidak Lulus
Wiwik Hastuti	600	540	520	490	515	V.Madya	Lulus
Roni Kesuma	320	415	510	300	290	VI.Marginal	Tidak Lulus
Dini	260	330	410	385	200	VII.Terbatas	Tidak Lulus
Intan Rahmaya	400	500	600	700	300	IV.Madya	Lulus
Sari Andini	310	290	450	520	360	VII.Terbatas	Tidak Lulus
Melin Nadya Gusti	795	780	690	685	685	I.Istimewa	Lulus
Khairunisa	220	325	280	455	520	VI.Marginal	Tidak Lulus
Sudarman Gulo	410	280	370	425	495	VI.Marginal	Tidak Lulus
Fitria Dinda	690	585	680	505	635	III.Unggul	Lulus
Eka Meilina	680	675	750	660	790	II.Sangat Unggul	Lulus
Indah Ariska	520	395	400	650	530	V.Semenjana	Tidak Lulus
Josafat	690	595	670	490	560	III. Unggul	Lulus
Tampubolon							
Parta Monica	615	425	575	520	485	IV.Madya	Lulus
Novitasari	465	395	210	430	475	VI.Marginal	Tidak Lulus
Krisna Wati	265	320	360	400	295	IV.Marginal	Tidak Lulus
Dewi Nurhalizar	500	730	495	560	470	IV.Madya	Lulus
Hardi Kurniawan	225	240	315	280	390	VII.Terbatas	Tidak Lulus
Zefri	600	515	495	465	580	IV.Madya	Lulus
Ade Wahyu	595	480	395	625	415	IV.Madya	Lulus

3.2 Pengumpulan Data

Dataset yang dilakukan dalam penelitian ini adalah menggunakan dataset peserta-peserta ujian kemahiran berbahasa Indonesia di Balai Bahasa Sumatera Utara data yang digunakan dalam penelitian ini adalah data mendengarkan, merespon kaidah, membaca, menulis dan berbicara.

3.3 Pembahasan

Untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut- atribut yang ada. Untuk mendapatkan nilai *gain*, terlebih dahulu mencari nilai *entropy*. Dengan menggunakan dua persamaan berikut, seperti yang telah dijelaskan pada bab sebelumnya maka akan didapatkan nilai *entropy* dan *gain* yang digunakan sebagai akar dalam membuat pohon keputusan.

Berikut ini merupakan rumus untuk perhitungan nilai *entropy*, seperti pada persamaan satu berikut:

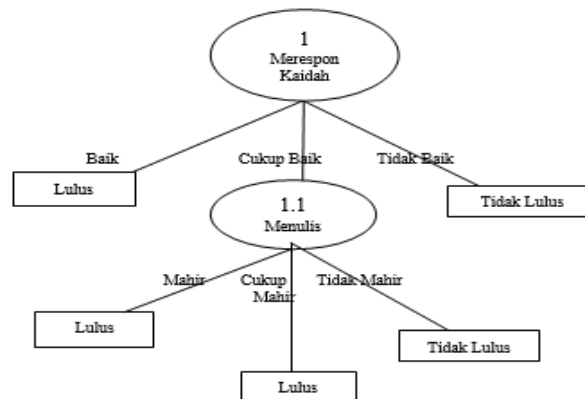
Setelah didapatkan nilai *Entropy* dan *Gain* dari sampel data yang dimiliki, berikut ini adalah rekapitulasi perhitungan nilai *Entropy* dan *Gain*nya.

Tabel 2. Rekapitulasi Hasil

NODE	Keterangan	Jumlah Kasus	Lulus	Tidak Lulus	Entropy	Gain
1	TOTAL	25	13	12	0.998845536	
	Mendengarkan					0.38985248075
	Jelas	6	5	1	0.65002242162	
	Cukup Jelas	8	6	2	0.81127812445	
	Tidak Jelas	11	1	10	0.4394969869	
						0.83317715802
	Merespon Kaidah					
	Baik	6	6	0	0	
	Cukup Baik	7	6	1	0.59167277851	
	Tidak Baik	12	0	12	0	
						0.45448553286
	Membaca					
	Lancar	6	6	0	0	

	Cukup Lancar	9	5	4	0.9910760598	
	Tidak Lancar	10	1	9	0.46893155402	
Menulis	Mahir	6	5	1	0.65002242162	0.47088574273
	Cukup Mahir	11	8	3	0.84535093654	
	Tidak Mahir	8	0	8	0	
Berbicara	Lancar	3	3	0	0	0.16642690856
	Cukup Lancar	12	6	6	1	
	Tidak Lancar	10	3	7	0.8810465686	
	Lancar					

Adapun pohon keputusan dari node 1 adalah sebagai berikut:



Gambar 3. Hasil akhir dari Pohon Keputusan

Maka basis pengetahuan atau *rule* yang terbentuk yaitu:

1. Jika Merеспон Kaidah = Baik maka Hasil = Lulus
2. Jika Merеспон Kaidah = Cukup Baik dan Menulis = Mahir maka Hasil = Lulus
3. Jika Merеспон Kaidah = Cukup Baik dan Menulis = Cukup Mahir maka Hasil = Lulus
4. Jika Merеспон Kaidah = Cukup Baik dan Menulis = Tidak Mahir maka Hasil = Tidak Lulus
5. Jika Merеспон Kaidah = Tidak Baik maka Hasil = Tidak Lulus

4. IMPLEMENTASI

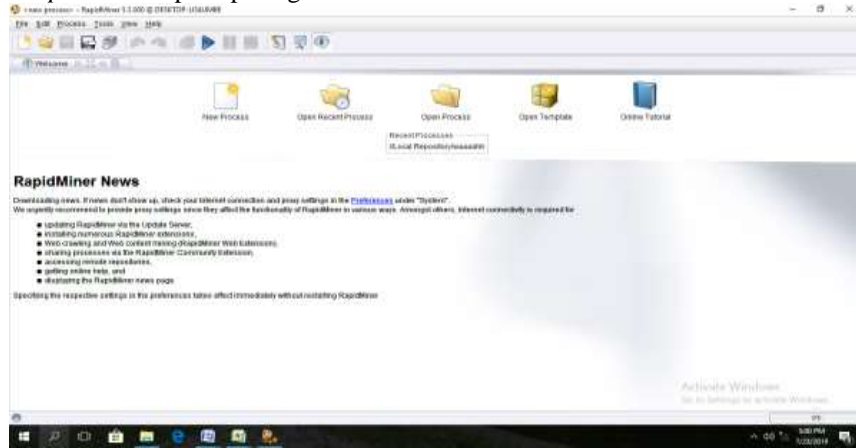
Implementasi merupakan suatu tindakan untuk mewujudkan terlaksananya rencana dalam mencapai tujuan tertentu. Implementasi sistem meliputi spesifikasi kebutuhan perangkat keras (*hardware*), perangkat lunak (*software*) dan hasil pengujian. Data diperoleh dari hasil suatu pengamatan dalam bentuk angka dengan fakta dan disajikan sebagai bahan dasar untuk mendapatkan suatu informasi yang bermanfaat setelah data tersebut diolah. Data yang akan digunakan dalam pengolahan merupakan data praproses dari data bencana alam yang telah dilakukan dan dibuat dalam *Microsoft Excel* dengan atribut yang sudah ditentukan.[6]



Gambar 2. Tampilan data Praproses dalam *Microsoft Excel*

Data tersebut disimpan pada lembar kerja *Ms.Excel* yang menjadi database penyimpanan data yang siap dilakukan pengujian pada *tools rapidminer*, dengan *save as type* adalah *Excel Workbook* dengan format ekstensi (*.xlsx). kemudian dikoneksikan ke *rapidminer* dengan cara :

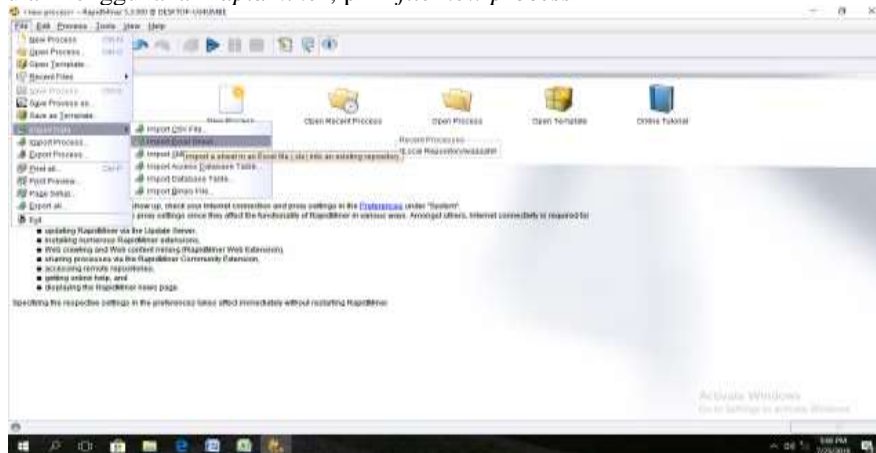
1. Buka *tools rapidminer* pada laptop atau komputer yang telah terinstal *rapidminer*
2. Tampilan awal *rapidminer* seperti pada gambar dibawah ini :



Gambar 3. Tampilan Awal *Rapidminer*

Pada gambar 5.2 merupakan tampilan awal dari *rapidminer* sebelum mulai mencari dan mengkoneksikan database ke *rapidminer*.

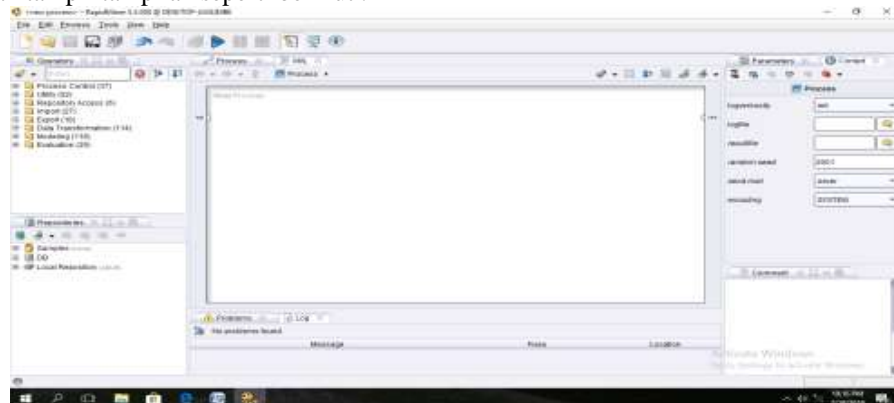
3. Untuk memulai menggunakan *rapidminer*, pilih *file-new process*



Gambar 4. Isi dari Menu File

Gambar 5.3 dipilih *file-new process* karena akan dikoneksikan *database* yang baru dibuat atau yang belum tersimpan pada *rapidminer*.

4. Kemudian akan tampil tampilan seperti berikut :



Gambar 5. Tampilan *New Process*

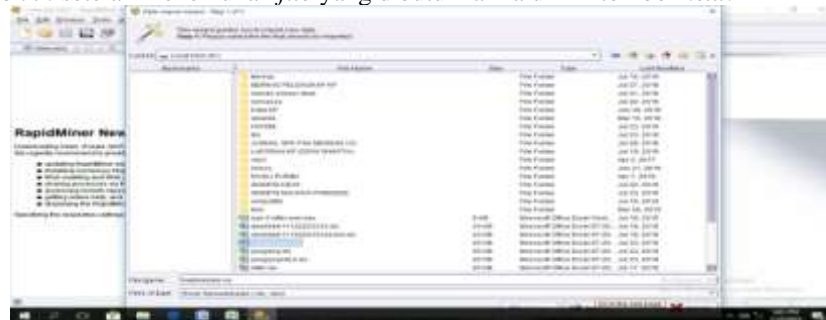
Pada gambar 5.4 menunjukkan untuk melakukan proses analisis data mining atau sebagai lingkungan kerja.

5. Setelah itu klik *file* pilih *Import Data- Import Excel Sheet*, seperti gambar berikut :



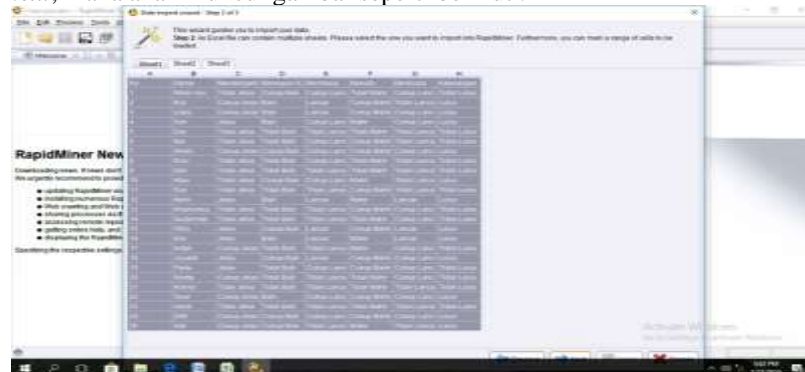
Gambar 6. Pilihan untuk membuka data yang akan diolah

6. Kemudian akan muncul tampilan seperti *data import wizard step 1 of 5*, setelah itu cari *file excel* dengan klik pada bagian *look in* setelah menemukan *file* yang dibutuhkan lalu klik tombol *next*.



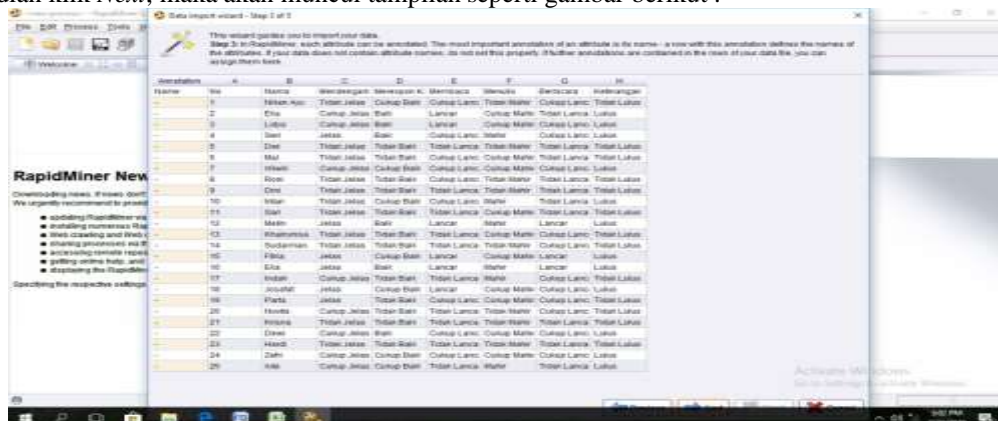
Gambar 7. Data Import Wizard Step 1 of 5

7. Pada gambar 7 merupakan langkah pertama yang dilakukan untuk *import data*. Kemudian klik *Next*, maka akan muncul gambar seperti berikut :



Gambar 8. Data Import Wizard Step 2 of 5

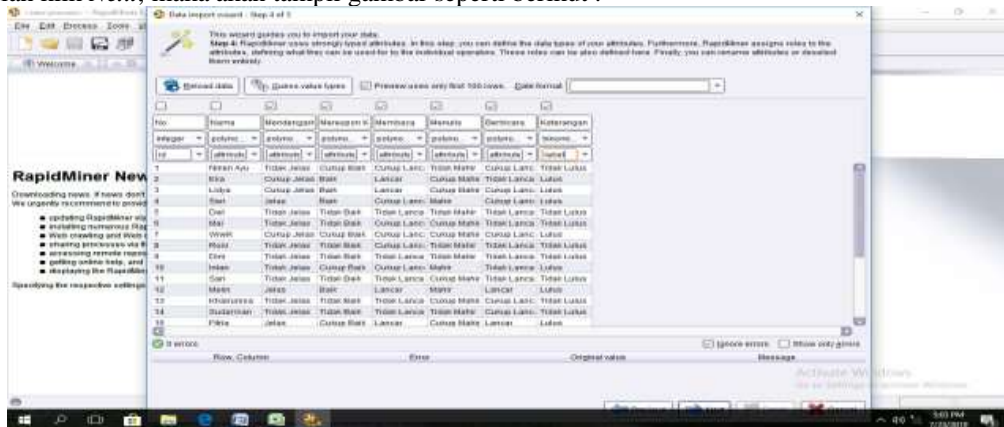
8. Pada gambar 8 diatas menunjukkan untuk memilih sheet yang akan dimasukkan. Kemudian klik *Next*, maka akan muncul tampilan seperti gambar berikut :



Gambar 9. Data Import Wizard Step 3 of 5

Pada gambar 9. menunjukkan untuk memberikan anotasi, karena data tidak memiliki nama *attribute*, maka tidak perlu dilakukan apa-apa pada step ini.

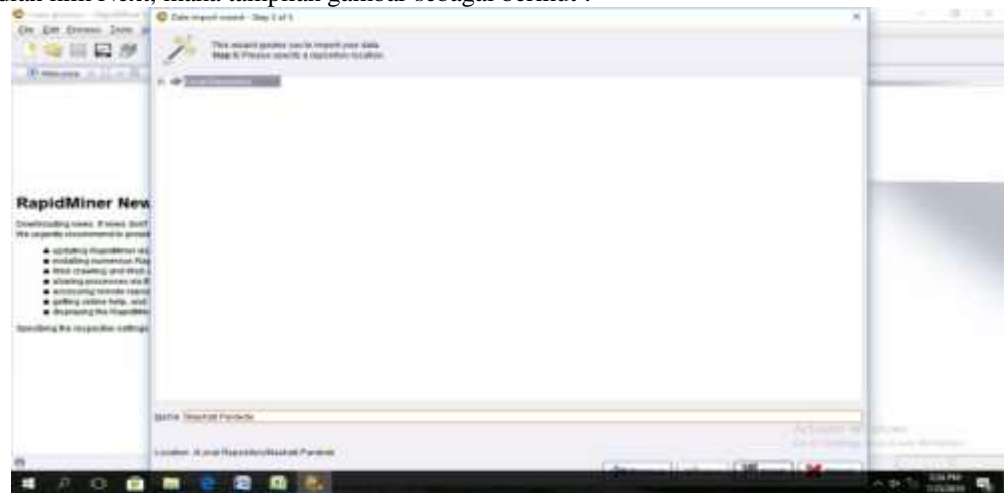
9. Kemudian klik *Next*, maka akan tampil gambar seperti berikut :



Gambar 10. Data Import Wizard Step 4 of 5

Pada gambar 10. menunjukkan untuk memberikan tipe data pada masing-masing atribut. Pada atribut “No” hilangkan ceklis, lalu pada atribut “name” pada kolom atribut ubah menjadi “Id” dan pada atribut “Keterangan” kolom atribut ubah jadi “Label”.

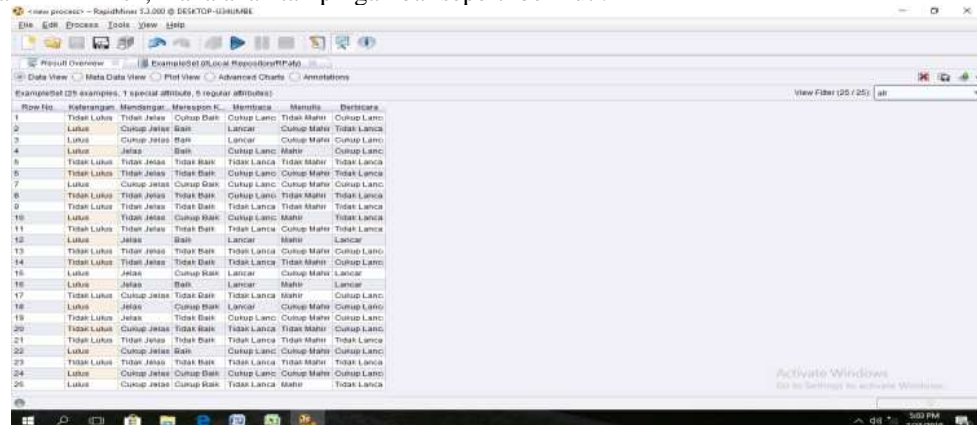
10. Kemudian klik *Next*, maka tampilan gambar sebagai berikut :



Gambar 11. Data Import Wizard Step 5 of 5

Gambar 5.10 adalah untuk memasukkan *database* kedalam repository seta memberikan nama repository.

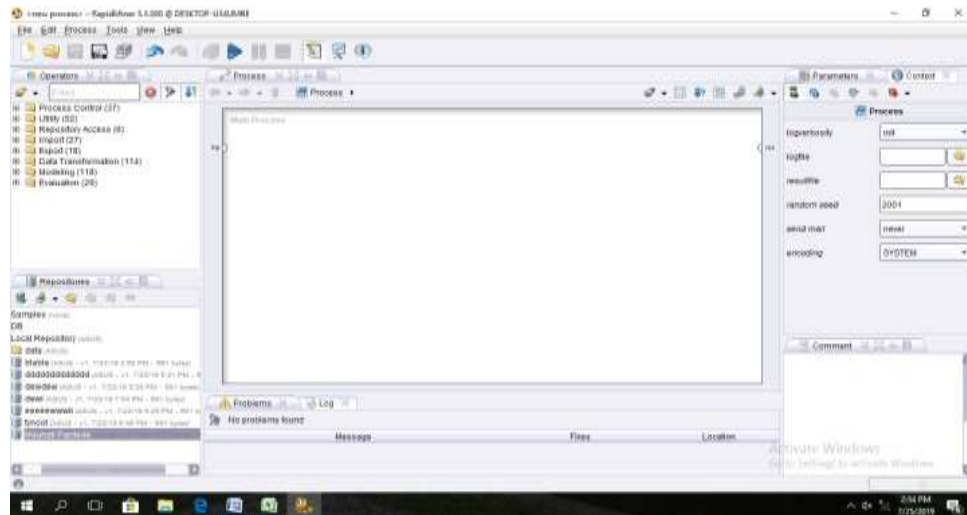
11. Kemudian klik *Finish*, maka akan tampil gambar seperti berikut :



Gambar 12. Resul Overview

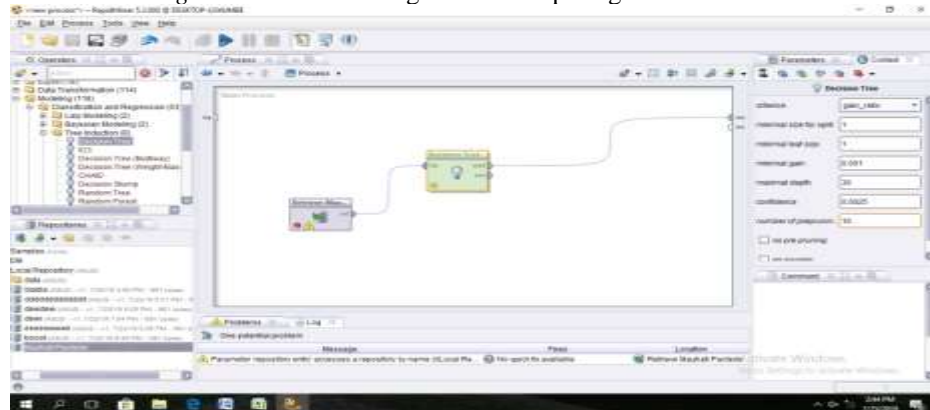
Gambar 12. adalah evaluasi terhadap dataset yang kita *import* kedalam repository.

12. Klik *Switch To Design Perspective*, kemudian akan muncul tampilan untuk melakukan proses analisis data mining atau sebagai lingkungan kerja seperti gambar dibawah ini.



Gambar 13. Main Process

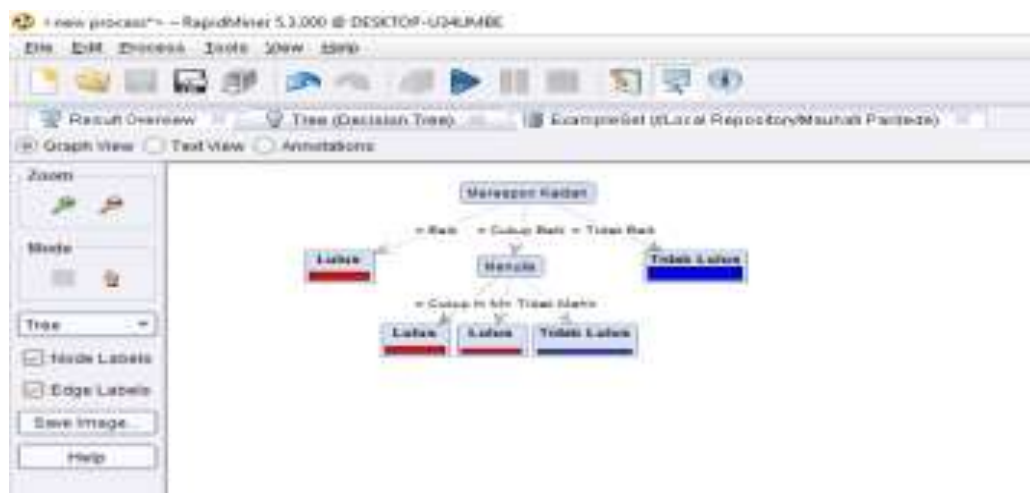
13. Klik *Switch To Design Perspective*, kemudian akan muncul tampilan untuk melakukan proses analisis data mining atau sebagai lingkungan kerja, setelah itu cari hasil *import repository* yang disimpan tadi pada bagian *repository*, lalu lakukan *drag drop* atau tarik dan lepaskan *repository* kedalam *main process*, lalu klik satu kali operator *decision tree*, maka pada bagian parameters biarkan semuanya ke *default*, namun pada *column minimal leaf size* dan pada *column minimal gain* ubah sesuai keinginan kita. Seperti gambar berikut ini :



Gambar 14. Mengubah Nilai *Gain* dan *Leaf Size*

Pada gambar 14. adalah menentukan minimal *gain* yang berfungsi untuk menghasilkan simpul pohon keputusan, dan *minimal leaf size* untuk menghasilkan pohon sebanyak *leaf size* yang di *input*.

14. Kemudian klik *Ikona Play*, maka akan muncul tampilan *result overview, tree (decision tree)* seperti gambar berikut :



Gambar 15. *Decision Tree* (Pohon Keputusan)

Pada gambar 15. merupakan *role decision tree* atau pohon keputusan terhadap data peserta Ujian Kemahiran Berbahasa Indonesia(UKBI).

Berdasarkan hasil pengujian pada *tools rapidminer* untuk mencari aturan (*role*) pohon keputusan (*decision tree*) dari sampel dataset diatas menunjukkan bahwa yang menjadi pohon keputusan dalam data peserta ujian Uji Kemahiran Berbahasa Indonesia pada Balai Bahasa Sumatera Utara adalah “Merespon Kaidah”. Karena atribut merespon kaidah yang merupakan faktor utama dalam menentukan peserta lulus pada Balai Bahasa Sumatera Utara.

5. KESIMPULAN

Berdasarkan uraian pada bab-bab sebelumnya, maka penulis mengambil keputusan sebagai berikut.

1. Faktor – faktor peserta Uji Kemahiran Berbahasa Indonesia (UKBI) pada Balai Bahasa Sumatera Utara dapat menghasilkan pohon keputusan
2. Algoritma C5.0 dapat diterapkan pada kelulusan
3. Adapun hasil dari klasifikasi algoritma C5.0 Jika merespon kaidah baik hasil lulus dan jika merespon kaidah tidak baik maka hasil tidak lulus.

6. REFERENCES

- [1] K. S. AliGusman Halawa, efori buulolo, “Implementasi algoritma c 4.5 untuk memprediksi kelulusan sisi di gereja bnkb teladan medan,” vol. 13, no. 2339–210X, 2018.
- [2] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, “Implementasi Algoritma C5 . 0 Dalam Klasifikasi Pendapatan Masyarakat (Studi Kasus : Kelurahan Mesjid Kecamatan Medan Kota),” Inf. dan Teknol. Ilm., vol. 13, pp. 192–198, 2018.
- [3] K. dan E. T. Luthfi, Algoritma Data Mining. Yogyakarta: CV.ANDI OFFSET, 2009.
- [4] D. H. Kamagi and S. Hansun, “Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa,” J. Ultim., vol. 6, no. 1, pp. 15–20, 2018.
- [5] N. E. W. Wijang Iswara Mukti, Andayani, “Pengajaran BIPA dan TES UKBI dalam Upaya Menjaga Eksistensi Bahasa Indonesia di Era Masyarakat Ekonomi ASEAN,” vol. 911–916, 2017.
- [6] https://www.academia.edu/7712860/Belajar_Data_Mining_dengan_RapidMiner, “Belajar_Data_Mining_dengan_RapidMiner.”.