

Model Aturan dalam Menentukan Prestasi Nilai Siswa di SMK GKPS 1 Raya Menggunakan Algoritma C4.5

Rofiana Simanullang^{*1}, Dedy Hartama², Poningsih³, Iin Parlina⁴, Muhammad R. Lubis⁵

^{1,2,3,4,5}Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar, Indonesia
Email: ¹rofianamanullang27@gmail.com, ²dedyhartama@yahoo.com,
³poningsih@amiktunasbangsa.ac.id, ⁴iin@amiktunasbangsa.ac.id,
⁵ridwanlubis@amiktunasbangsa.ac.id

Abstrak

Data nilai siswa merupakan suatu data penting bagi pihak departemen, maupun pada pihak sekolah karena perlu untuk melihat bagaimana perkembangan nilai siswa-siswi di SMK GKPS 1 Raya dikemudian hari. Data nilai siswa pun semakin bertambah bila semakin tahun berganti tahun, dan data tersebut dapat memberi informasi yang berguna bila diolah dengan baik. Maka dari itu, dalam penelitian ini penulis akan memanfaatkan 202 data nilai siswa yang diperoleh dari pihak sekolah SMK GKPS 1 Raya yang dikelola menggunakan data mining untuk mendapatkan suatu informasi klasifikasi perkembangan nilai siswa dan menentukan prestasi siswa. Dalam metode ini, algoritma yang digunakan yaitu algoritma C 4.5 decision tree yang didukung dengan Software RapidMiner, Kriteria yang digunakan adalah seperti NISN, Nama Siswa, Nilai Rata-rata dan Nilai Kehadiran yang di input ke input kedalam Microsoft Excel 2007 dan dilakukan transformasi dari nilai angka ke nilai huruf yang dimana Jika nilai >90 maka nilainya A, 80 – 89 = B, 70 – 79 = C, dan <60 = D. Dengan menggunakan metode ini dapat menjadi salah satu alat untuk dapat membantu pihak sekolah dalam melihat perkembangan nilai siswa, sehingga hasil yang di dapat dalam metode ini dapat menentukan nilai yang berprestasi dan yang tidak berprestasi serta dapat memberikan rekomendasi untuk pihak sekolah semakin memperbaiki sistem pembelajaran yang berlaku untuk kedepannya.

Kata kunci: Algoritma C4.5, Data Mining, Decision Tree, Nilai Siswa

Abstract

Student score data is an important data for the department, as well as for the school because it is necessary to see how the development of student scores at SMK GKPS 1 Raya in the future. Data on student scores also increases as the years go by, and the data can provide useful information if processed properly. Therefore, in this study the author will utilize 202 student score data obtained from the SMK GKPS 1 Raya school which is managed using data mining to obtain an information on the classification of student grade developments and determine student achievement. In this method, the algorithm used is the C 4.5 decision tree algorithm which is supported by RapidMiner Software. The criteria used are NISN, Student Name, Average Value and Attendance Value which are inputted into Microsoft Excel 2007 and transformed from the value numbers to letter grades where if the value is >90 then the value is A, 80 – 89 = B, 70 – 79 = C, and <60 = D. Using this method can be a tool to help the school in seeing the progress of the grades students, so that the results obtained in this method can determine the value of achievers and those who do not excel and can provide recommendations for the school to further improve the learning system that applies to the future.

Keywords: C4.5 Algorithm, Data Mining, Decision Tree, Student Score

1. PENDAHULUAN

Prestasi nilai siswa merupakan tingkat kemampuan siswa dalam menerima suatu pelajaran yang menyangkut ilmu pengetahuan, keterampilan dan sikap, sebagai hasil pengalaman siswa dalam interaksi dengan lingkungannya. Sehingga mutu dan kualitas dari suatu institusi pendidikan atau pihak departemen salah satunya dilihat dari prestasi yang dicapai oleh lulusannya. Bagi pihak sekolah SMK GKPS 1 Raya hingga saat ini masih sering menjadi masalah bagaimana cara untuk mengeksplorasi karakteristik dari siswa hingga dapat menentukan prestasi nilai siswa tersebut, belum dapat dilakukan

secara mudah dan sederhana sehingga masih dilakukan dengan melihat prestasi beberapa siswa dengan cara manual dan belum melakukan klasifikasi secara keseluruhan.

Perkembangan prestasi nilai siswa selalu dilakukan penilaian secara berkala, yaitu yang dapat berbentuk ujian, pelaksanaan tugas, dan pengamatan oleh guru. Untuk pengamatan guru terhadap siswa biasanya diselenggarakan melalui ujian tengah semester, ujian akhir semester, hingga dari nilai kehadiran siswa.

Tingkat prestasi atau mutu dari siswa merupakan barometer dalam menentukan perkembangan akreditasi dari pihak departemen. Dalam hal ini perlu untuk dilakukan membuat model aturan dalam penentuan prestasi nilai siswa apakah terjadi perkembangan.

Sehingga penelitian ini mengaplikasikan teknik data mining dengan algoritma C4.5 dalam membuat suatu model aturan untuk menentukan prestasi nilai siswa berdasarkan mata pelajaran berbasis kompetensi dan data yang mendukung peningkatan prestasi siswa. Model aturan yang diperoleh untuk mengklasifikasikan prestasi siswa yang terdiri dari Sangat Baik, Baik, Cukup, dan Kurang.

Dalam penelitian ini, penulis memberikan suatu solusi untuk membuat suatu model aturan dalam menentukan prestasi nilai siswa dengan menggunakan algoritma C4.5 untuk melakukan penentuan prestasi siswa, karena algoritma C4.5 merupakan salah satu metode pada Decision Tree atau Pohon keputusan yang sering banyak digunakan untuk melakukan suatu prediksi terhadap suatu permasalahan. Penelitian ini diharapkan dapat memberikan kontribusi bagi pihak departemen pendidikan khususnya pada SMK GKPS 1 Raya. Walaupun telah banyak peneliti yang telah melakukan penelitian berkaitan dengan prestasi nilai siswa, namun masih banyak yang belum mengetahui apa yang menjadi faktor berkembang tidaknya suatu prestasi nilai siswa tersebut.

Adapun beberapa penelitian terdahulu terkait dengan penggunaan algoritma C4.5 yang berkaitan dengan membuat suatu model aturan dalam menentukan prestasi nilai salah satunya yaitu: penelitian yang dilakukan oleh [1] tentang model aturan keterhubungan data mahasiswa menggunakan algoritma C4.5 Untuk Meningkatkan Indeks Prestasi. Beliau menyimpulkan bahwa Dalam studi kasus pada Akademi Manajemen Informatika Komputer (AMIK) Tunas Bangsa Pematangsiantar bahwa sebagian besar mahasiswa yang memiliki predikat buruk apabila faktor rata-rata ekonomi = 3 dan ratamkk bernilai E dan predikat dengan pujian apabila jika ratamkk = A dan ratamkb = A kemudian rata-rata dukungan orang tua = 4.8.

Penelitian [2] telah melakukan survei, data mining dalam bidang pendidikan antara tahun 1995 sampai 2005, mereka menyimpulkan bahwa data mining yang berhubungan dengan pendidikan sangat baik untuk diteliti terutama di bidang *e-learning*, multimedia, artificial intelligent dan web database.

Penelitian [3] menggunakan teknik data mining yang digunakan untuk membangun prototipe Penilaian Kinerja Monitoring System (PAMS) untuk mengevaluasi kinerja mahasiswa.

Penelitian [4] penerapan algoritma C4.5 untuk klasifikasi predikat keberhasilan mahasiswa di amik tunas bangsa.

2. TINJAUAN PUSTAKA

2.1. Pengertian Algoritma C4.5

Algoritma C4.5 merupakan algoritma terkuat untuk proses pembuatan decision tree jika dibandingkan dengan algoritma lain seperti ID3, C5.0, dan CART [5]. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Sebuah model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih homogen dengan memperhatikan pada variabel tujuannya. Sebuah pohon keputusan mungkin dibangun dengan seksama secara manual atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi.

Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan probability dari tiap-tiap record terhadap kategori-kategori tersebut atau untuk mengklasifikasi record dengan mengelompokkannya dalam satu kelas. Pohon keputusan juga

dapat digunakan untuk mengestimasi nilai dari variabel continue meskipun ada beberapa teknik yang lebih sesuai untuk kasus ini. Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5.

2.2. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi [6].

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [7]. Pengertian lain mengenai Data mining, Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Data mining adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya. Data mining digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut Knowledge Discovery Databases (KDD) yaitu tahapan yang dilakukan dalam menggali pengetahuan dari sekumpulan data. Proses penggalian informasi tersembunyi dalam suatu basis data yang besar sering menggunakan istilah data mining dan knowledge discovery in databases (KDD) [8].

2.3. Decision Tree

Decision Tree merupakan suatu metode klasifikasi dan prediksi yang sangat terkenal yang sering digunakan. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami, juga dapat diekspresikan dalam bentuk bahasa basis data seperti Structured Query Language untuk mencari record pada kategori tertentu.

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun juga mudah untuk dipahami, sehingga Decision Tree ini merupakan metode klasifikasi yang paling populer digunakan [9].

Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, pohon keputusan sangat bagus sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain [10].

2.4. RapidMiner

RapidMiner adalah aplikasi data mining berbasis open-source yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data mining seperti untuk loading data, transformasi data, pemodelan data, dan metode visualisasi data [11].

Dengan menggunakan RapidMiner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. RapidMiner dikhususkan untuk penggunaan data mining. Model yang disediakan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction, Neural Network dan lain-lain. Banyak metode yang disediakan oleh RapidMiner mulai dari klasifikasi, klustering, asosiasi dan lain-lain.

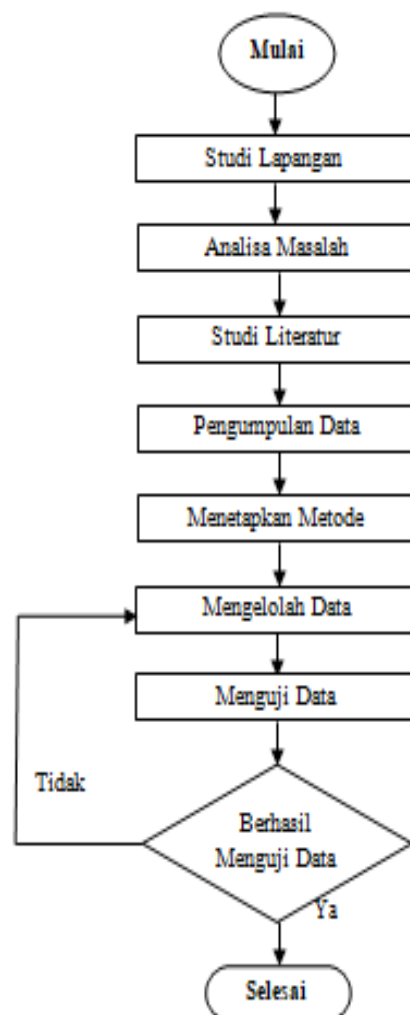
3. METODE PENELITIAN

Metodologi penelitian merupakan suatu prosedur atau langkah yang dilakukan oleh peneliti dalam rangka untuk mengumpulkan suatu data atau informasi. Metodologi penelitian memiliki gambaran rancangan penelitian yang meliputi antara lain yaitu mulai dari cara mengumpulkan data, menganalisa data.

3.1. Rancangan Penelitian

Pada Rancangan penelitian ini yang pertama kali dilakukan penulis yaitu dengan melakukan observasi atau pengamatan untuk mempelajari klasifikasi data mata pelajaran. selanjutnya dilakukan eksperimen data dengan menggunakan perangkat lunak yaitu RapidMiner yang merupakan software open source untuk membuat model aturan data yang diambil dari database pendidikan SMK GKPS 1 Raya. Yang dimana hasil dari eksperimen data ini nantinya yang akan menjadi suatu pengembang ilmu pengetahuan yang dapat menjadi solusi bagi pemecahan masalah yang terjadi pada pihak sekolah SMK GKPS 1 Raya.

Dalam melakukan klasifikasi data nilai siswa SMK GKPS 1 Raya penulis melakukan beberapa tahapan atau prosedur untuk rancangan penelitian, Berikut rancangan penelitian yang disajikan dalam rancangan flowchart, dapat di tunjukkan oleh Gambar 1.



Gambar 1. Rancangan Penelitian

Keterangan dari gambar rancangan penelitian di atas adalah :

1. Studi Lapangan
Merupakan suatu cara untuk memperoleh data dengan melakukan observasi langsung terhadap obyek yang akan diteliti untuk memperoleh informasi mengenai permasalahan. Penulis melakukan studi lapangan tepatnya di SMK GKPS 1 Raya.
2. Analisa Masalah
Masalah yang terkait dalam model aturan penentuan prestasi nilai siswa di SMK GKPS 1 Raya agar dapat membantu pihak sekolah pada bagian dalam mengambil keputusan untuk menentukan

perkembangan prestasi siswa serta Sebagai model untuk klasifikasi predikat akhir ujian semester siswa SMK GKPS 1 Raya

3. Studi Literatur

Mulai dari pengertiannya penulis akan memulai dengan mencari data-data yang dibutuhkan yang berhubungan dengan topik atau judul yang diangkat. Adapun kebutuhan itu dimulai dengan uji coba materi dimana penulis mengumpulkan referensi-referensi yang didapat dari berbagai sumber, jurnal, dokumentasi internet dan pustaka.

4. Pengumpulan Data

Ditahap ini Pengumpulan data dapat dilakukan dalam berbagai cara, ada beberapa cara dalam pengumpulan data yaitu dengan Interview, Kuisioner dan juga observasi. dan pada tahap ini penulis melakukan pengumpulan data dengan cara Observasi (Pengamatan) yang dimana penulis memperoleh data langsung dari pihak sekolah yaitu SMK GKPS 1 Raya.

5. Menetapkan Metode Data Mining

Dalam penetapan metode ini penulis mengklasifikasikan prestasi nilai siswa SMK GKPS 1 Raya menggunakan algoritma C4.5

6. Pengolah Data

Dalam mengelola data penulis menggunakan decision tree C4.5 untuk pengolahan data yang telah diteliti, Pada tahap ini sangat menentukan keberhasilan dari penelitian karena mulai dari dataset, atribut dan variabel sangat diperhatikan.

7. Menguji Data

Tahap pengujian data dilakukan dengan menggunakan perangkat lunak yaitu Rapid Miner sehingga dapat diperoleh kategori keputusan yang sesuai dengan metode decision tree C4.5. Data pengujian yang digunakan merupakan data yang diperoleh dari pihak sekolah yaitu SMK GKPS 1 Raya yang datanya terdiri dari 202 orang. Pengujian juga merupakan tahapan yang sangat menentukan apakah pengujian yang dibuat telah layak digunakan atau belum. Jika pengujian belum menghasilkan luaran yang sesuai dengan kebutuhan maka tahapan penelitian kembali ke pengolahan data.

3.2. Analisa Data

Dalam penelitian ini data yang digunakan akan diolah dari hasil observasi kedalam excel yang dimana penulis mengumpulkan data dari pihak sekolah SMK GKPS 1 Raya, data yang dikumpulkan ialah sebanyak 202 orang/siswa yang dimana dalam hasil observasi tersebut terdapat kriteria yang berisi seperti:

4. NISN
5. Nama Siswa
6. Rata MN
7. Rata MK
8. Rata MPK
9. Kehadiran

Selanjutnya hasil dari data yang sudah diperoleh dari pihak sekolah SMK GKPS 1 Raya dan di input ke Excel lalu dilakukan transformasi atau perubahan dari nilai berupa angka menjadi sebuah grade yang berupa nilai huruf dan kemudian akan diolah kembali menggunakan RapidMiner dengan operator Performance yang berfungsi untuk validasi dan reabilitas data untuk mencari keakuratan data. Data yang akurat akan dilakukan pengolahan data untuk mencari hasil dari masalah penelitian dengan menggunakan perangkat lunak yaitu RapidMiner dan diperoleh keputusan dari hasil. Dan ketentuan nilai siswa berupa angka diubah menjadi nilai huruf atau grade ditunjukkan oleh Tabel 1.

Tabel 1. Ketentuan Nilai Siswa

No	Nilai Angka	Nilai Huruf	Keterangan
1	90 – 100	A	Sangat Baik
2	80 – 89	B	Baik
3	70 – 79	C	Cukup
4	< 69	D	Kurang

Dan data yang digunakan sebanyak 202 data nilai siswa yang diperoleh dari pihak sekolah SMK GKPS 1 Raya, dapat ditunjukkan pada Tabel 2

Tabel 2. Data Penelitian

No	NIS/NISN	Nama Siswa	Rata MN	Rata MK	Rata MPK	Kehadiran	Rata-rata	Grade
1	0038959811	Ade Putra Damanik	90	94	90	95	92	A
2	0010922233	Aga Kencana Sinaga	80	89	90	90	87	B
3	0043835111	Age Risnauli Sipayung	90	90	94	85	90	B
4	0038405126	Agnes Kristin Purba	81	85	81	90	84	B
5	0022669079	Aldivan Maholtra Zega	90	90	90	87	89	B
6	0034957534	Aldrianson Sembiring	85	87	78	80	83	B
7	0028388355	Arjun Saragih	83	85	84	83	84	B
8	0016418798	Arta Naully Girsang	83	86	82	85	84	B
9	0012748247	Asked Aritionang	70	77	80	80	77	C
10	0028157879	Ayu Lestari Sidauruk	80	88	90	82	85	B
11	0022849079	Ayu Monicha Dabukke	82	84	86	80	83	B
12	0035533802	Benni Darfin Purba	69	78	85	89	80	B
13	0037099040	Betty Stevani	89	90	78	90	87	B
14	0037870769	Bina Ofta Mariana Purba	92	84	88	70	84	B
15	0046158874	Budi Harianto Saragih	81	81	79	75	79	C
16	0025630361	Chesya Alfionita Malau Christian Stevani	84	86	80	80	83	B
17	0028193715	Sihaloho Chyndrella Classya	80	81	81	81	81	B
18	0048216407	Hiskia Sidauruk	81	81	82	82	82	B
19	0034713206	Cinta Lestari Cintya Putri Lestari	82	83	81	81	82	B
20	0028758046	Saragih	81	82	80	82	81	B
21	0036306215	Citditiro Saragih	82	82	81	90	84	B
22	0028887801	Citra Dewita Ambarita	90	92	89	89	90	A
23	0033308082	Clara Purba	89	92	85	82	87	B
24	0041431016	Clara Sartika Br Sitepu	82	83	83	80	82	B
25	0015585141	Cyndy Claudia P Sinaga	80	85	83	84	83	B
26	0028372372	Dandy Alexander Malau Dani Endes Wari	84	83	80	87	84	B
27	0038818249	Saragih Dani Kristian Frananda	87	83	82	83	84	B
28	0046629294	Purba Daniel Oktripandi	83	80	82	85	83	B
29	0028157959	Nasution	85	82	83	90	85	B
30	0041433272	Daniel Roy Hend Purba	90	82	89	87	87	B
31	0048433061	Daniel Sinaga	87	83	92	90	88	B
32	0028234359	Dasmon Tuahta Munthe Daud Manambos	90	90	80	80	85	B
33	0036429401	Harianto Purba Daud Parsaulian	86	92	84	82	86	B
34	0027347344	Damanik	82	81	80	85	82	B
35	0023488295	Dedi Syahputra Sitopu	85	84	79	84	83	B
36	0019000129	Deniarta Sianturi Desi Putri Ramayana	84	80	85	85	84	B
37	0028680390	Sinaga	85	81	80	80	82	B
38	0027239794	Desry Kristiani Purba	80	82	81	70	78	C
39	0022171499	Devi Esrayani Sinaga Devi Milayanti	82	82	88	81	83	B
40	0034957888	Sembiring	81	81	78	80	80	B
41	0033552793	Devi Vita Lora Saragih	80	82	84	81	82	B
42	0029320130	Dewi Fransiska Saragih	81	82	82	89	84	B
43	0041631841	Dewi Yupita	89	89	87	76	85	B
44	9996987526	Domi Julpria Damanik	76	82	81	83	81	B
45	0031286814	Dons Mikhael Purba	83	80	86	70	80	C
46	0012664713	Dory Safitri Saragih	83	84	90	80	84	B
47	0026969746	Echa Feronika Girsang	80	87	78	82	82	B

48	0040752058	Edi Saputra Tarigan	82	83	88	82	84	B
49	0015973961	Egia Pranata Damanik	82	85	79	83	82	B
50	0045917784	Egla Sery Redina						
		Haloho	83	90	80	89	86	B
		Eka Soniananda Purba						
51	0037881819	Dasuha	89	87	81	92	87	B
52	0028157788	Elista Purba	92	88	91	81	88	B
53	0028837531	Elma Nofrida Sinaga	81	86	83	84	84	B
54	0023785463	Elsa Oktavia Siahaan	84	82	82	83	83	B
55	0045877061	Elysabeth Cry Aruan	83	85	82	80	83	B
56	0039273054	Enjelia Purba	80	84	83	84	83	B
57	0011067744	Eplin Suhendra Haloho	84	92	92	87	89	B
		Erina Yemima Chesya						
58	0033989169	Damanik	87	81	83	83	84	B
59	0046253967	Ernasari Sinurat	83	84	85	85	84	B
60	0045837866	Esra Yesika Tondang	85	80	83	90	85	B
61	0016294537	Ester Maruliana Purba	90	90	91	87	90	B
		Eva Kristina Hidayat						
62	0010208832	Lingga	87	82	80	85	84	B
63	0031220907	Eva Meliana .S	85	81	82	70	80	C
64	0040591950	Exa Neria Purba	86	82	82	87	84	B
		Exsaudina Lorencia						
65	0022999587	Sinurat	82	82	83	85	83	B
66	0020567911	Fabio Perdana Putra	85	89	89	86	87	B
67	0046357247	Fani Yuliana Sipayung	84	82	80	77	81	B
		Felix Jontri Wilson						
68	0037930182	Tondang	78	80	81	88	82	B
69	0041311002	Fernandes Hutapea	90	84	89	84	87	B
70	0036961324	Fika Aglesia Purba	84	87	85	78	84	B
71	0027035066	Fitri Kristiana Malau	81	83	83	90	84	B
72	0048135274	Fitriani Purba	86	85	83	84	85	B
73	0036275952	Floranita Sitanggang	81	90	80	81	83	B
74	0001806546	Frandoni Gulo	81	87	82	86	84	B
75	0040691115	Frans Dwi Putra Marbun	83	85	82	81	83	B
		Frans Toni Syaputra						
76	0028349794	Sumbayak	82	86	83	81	83	B
		Fransiska Desri Nataria						
77	0028158033	Damanik	82	82	83	83	83	B
78	0017280286	Frenky Arianto Purba	78	80	80	82	80	B
79	0026938412	Frensi Enjelina Saragih	92	84	86	82	86	B
80	0023338881	Fridayani Purba	83	85	82	80	83	B
81	0036560042	Friskila Hani Saragih	85	80	83	100	87	B
82	0001943336	Geopani Sitanggang	83	82	84	83	83	B
83	0021466549	Gokmandika Sinaga	83	81	85	85	84	B
		Gokmaria Mahdalena						
84	0030042486	Purba	80	80	90	83	83	B
		Gom Gom Pardomuan						
85	0002165995	Sinurat	82	81	84	83	83	B
86	0022999595	Heldi Gratia Sinaga	82	89	81	80	83	B
87	0049553040	Helma Friskila Silitonga	83	85	86	82	84	B
88	0034770785	Helpina Saragih	89	83	81	82	84	B
89	0028399522	Heni Erfita Saragih	92	83	81	83	85	B
90	0015589365	Herbin Manurung	67	69	67	90	73	C
91	0030143400	Herd Arra Purba	77	75	82	75	77	C
92	0034957711	Hizkia Putri Butar Butar	80	82	82	90	84	B
93	0036771819	Hot Imanuel Girsang	81	83	83	95	86	B
94	0030418039	Hot Starni Girsang	82	83	92	85	86	B
95	9991666779	Hotmauli Purba	76	89	76	75	79	C
96	0011067642	Hotnai Sari Sinaga	81	92	85	82	85	B
		Hotni Uhur Idahonta						
97	0028430491	Purba	82	81	79	82	81	B
		Ica Bella Chintia						
98	0038735907	Sumbayak	82	84	85	81	83	B
99	0023081457	Ice Hariani Sinaga	89	80	80	82	83	B

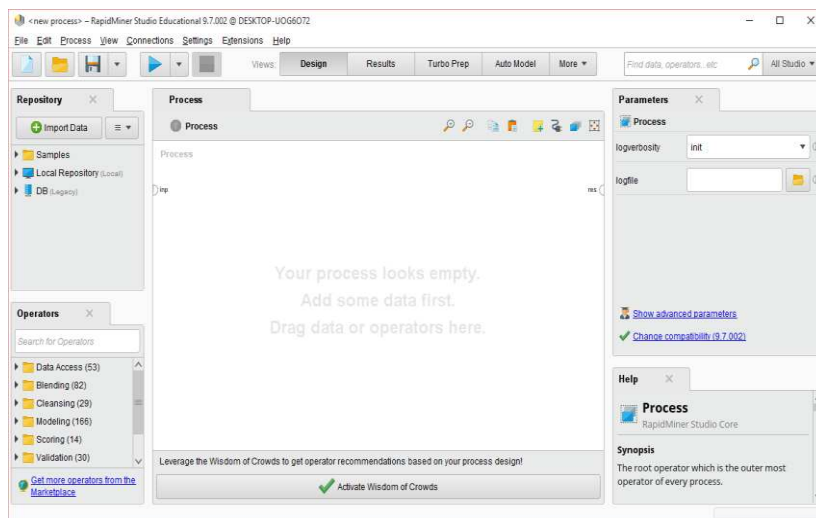
100	0042706672	Icha Angriani Simarmata	82	81	81	82	82	B
101	0030433148	Ida Monalisa	80	82	88	89	85	B
		Ijonli Hamonangan						
102	0022287418	Saragih	69	70	78	82	75	C
103	0010922279	Ilasari Rambei	87	81	84	85	84	B
104	0044969306	Imanuel Saragih	83	82	82	84	83	B
105	0032077021	Imarhaenis Nainggolan	85	82	87	87	85	B
		Imron Frederik						
106	0029628565	Verodomes Sumbayak	90	88	80	83	85	B
107	0042918274	Indah Ramayana Sinaga	65	82	77	78	76	C
108	0039902456	Inra Frendi Saragih	85	80	90	90	86	B
109	0026783152	Iswar Rulianto Purba	86	84	78	87	84	B
		Iyolanda Switami						
110	0024242589	Sidauruk	82	87	88	88	86	B
111	0028282434	Iyut Novita Sani Saragih	85	90	79	86	85	B
		Jeki Putra Antonius						
112	0034770945	Saragih	84	85	80	82	83	B
113	0039011565	Jekky Siputra Purba	68	78	75	85	77	C
		Jelita Ayu Jintia Huta						
114	0046109075	Gaol	85	87	82	84	85	B
		Jelita Erva Maya						
115	0032987240	Sihombing	80	85	81	92	85	B
		Jenny Konirlir Br						
116	0025702420	Tarigan	82	86	80	81	82	B
117	0032446361	Jesindo Saragih	81	82	81	84	82	B
118	0047534749	Jessi Aprianto Saragih	89	90	95	80	89	B
119	0009587037	Jhon Agustian Sinaga	81	84	83	90	85	B
120	0039746913	Jhon Kristiando Saragih	89	85	83	82	85	B
121	0038058575	Jhon Wari Pintu Batu	85	90	80	81	84	B
122	0019236204	Johana Sayang Tamba	83	84	82	82	83	B
123	0014708838	Joly Oktar Situmorang	83	81	82	82	82	B
		Jordan Marchell						
124	0042739474	Simanihiuruk	77	78	75	78	77	C
125	0039635712	Joshua Parulian Sitopu	82	81	89	90	86	B
126	0028157932	Joy Lasro Purba	82	81	92	87	86	B
127	0035086675	Juby Lenda Purba	83	83	82	85	83	B
128	0022288178	Jui Melpa Sinaga	89	82	79	86	84	B
129	0017265579	Juli Yanti Saragih	92	82	80	90	86	B
130	0047380645	Juliana Br Sianturi	82	83	82	88	84	B
131	0040499356	Julmiati Enjelia Purba	79	84	83	84	83	B
132	0028430379	Jultri Purba	80	87	83	78	82	B
133	0015761285	Junaedi Sinaga	82	83	80	90	84	B
134	0025948420	Juner Permanda Sinaga	83	85	90	84	86	B
135	0035533448	Juwisrel Onesis Saragih	68	80	69	81	75	C
136	0046158871	Juwita Mayasari Sinaga	80	87	85	86	85	B
		Keisya Adelia						
137	0027181431	Simarmata	79	85	86	81	83	B
138	0038514355	Kevin Flado Sinaga	88	86	87	81	86	B
		Khristin Meysia. P.						
139	0045877050	Sihaloho	78	82	80	83	81	B
		Kiranda Yaptopian						
140	0046109612	Simarmata	90	92	97	82	90	A
141	0046728121	Krisna Girsang	84	84	80	82	83	B
		Kristin Putri Enjeli						
142	0021948612	Saragih	87	85	81	92	86	B
143	0020271531	Kristina Simarmata	90	90	80	92	88	B
144	0036416013	Lasta Sri Ayu Saragih	89	84	79	83	84	B
145	0028158005	Leni Marlina Saragih	86	81	90	100	89	B
146	0044992906	Lewisa Dewanti Turnip	87	86	86	83	86	B
147	0011840458	Lia Damai Silaban	80	81	88	100	87	B
148	3003247118	Lorina Natalia Sipahutar	80	81	87	80	82	B
149	0022673298	Luis Weldi Saragih	80	83	83	82	82	B
150	0044338428	Maju Wando Sinaga	69	75	77	90	78	C
151	0042429829	Marel Andrew	80	82	82	83	82	B

152	0028157927	Nathanael Simbolon May Dasni Sinaga May Hendra Priadi	79	83	83	90	84	B
153	0042634688	Sinaga	90	92	83	92	89	B
154	0022673090	Meilani Lida Siahaan	86	83	85	81	84	B
155	0011067597	Melda Nova Sari Purba	88	85	83	84	85	B
156	0027553123	Meli Teresya Purba	87	83	83	90	86	B
157	0044859302	Melvi Damanik	83	83	82	90	85	B
158	0034770947	Mesrani Tinneki Saragih	82	82	79	100	86	B
159	0022673250	Messya Duita Purba	77	79	80	75	78	C
160	0041239082	Mira Monica Saragih	83	80	85	100	87	B
161	0026665121	Monica Sazy Saragih Nanci Ade Lestari	84	82	86	82	84	B
162	0037504187	Sinaga	87	83	82	82	84	B
163	0027322091	Naomi Amrita Siringo- Ringo	75	78	85	79	79	C
164	0038735913	Natalia Klaska Saragih	85	80	84	82	83	B
165	0022404878	Natasya Sinaga	90	92	96	80	90	B
166	0025355677	Nehemia Br Manjorang	87	88	90	84	87	B
167	0048250213	Nia Lestari Sinaga	85	78	84	87	84	B
168	0020944071	Nofia Sinaga	75	78	81	75	77	C
169	0019854408	Noprich Helty Damanik	82	84	86	85	84	B
170	0028430318	Nora Setty Purba Nosela Ifilly Padhomu	85	87	81	90	86	B
171	0024367595	Sinaga	89	90	90	87	89	B
172	0022673297	Nova Hotvita Saragih Noventi Ervinawati	85	89	88	88	88	B
173	0038951462	Sijabat	90	86	87	86	87	B
174	0028758054	Novia Saragih Novita Pasmawati	84	87	83	82	84	B
175	0028026188	Saragih	81	80	82	70	78	C
176	0028157885	Novri Mayasari Purba	86	80	82	84	83	B
177	0048092167	Nurini Dela Saragih	81	85	83	92	85	B
178	0022072941	Nurmalem Putri Saragih	81	84	84	81	83	B
179	0033281708	Oktri Verawati Saragih	75	78	87	70	78	C
180	0028430334	Osde Riahni Sitopu Pankresia Krisnalia	82	88	83	80	83	B
181	0036177126	Purba	82	90	83	80	84	B
182	0001548225	Parnadion Simbolon	80	82	83	82	82	B
183	0030877918	Pebryanta Br Barus	80	82	80	81	81	B
184	0014958539	Pesta Parsaulian Sinaga	88	90	89	82	87	B
185	0046357237	Pison Manurung	77	84	70	82	78	C
186	0023248692	Priska Nasution	86	87	85	89	87	B
187	0025355689	Putra Baik Saragi M	83	83	82	83	83	B
188	0027597097	Putrayado Saragih	72	68	69	72	70	C
189	0026293480	Putri Kristiani Damanik	88	90	80	88	87	B
190	0042179625	Rafita Ulina Saragih Rajes Andriansen	83	87	79	83	83	B
191	0037496222	Tarigan	83	85	82	100	88	B
192	0042637125	Rama Indah Purba Rano Alpencus	80	86	80	80	82	B
193	0038877664	Sipayung	70	82	75	70	74	C
194	0046956797	Rebeka Sinaga	80	85	82	80	82	B
195	0034401716	Refa Fransiska Purba	85	84	83	85	84	B
196	0028157795	Resi Prasanta Purba	82	85	84	90	85	B
197	0031489632	Reski Kristasya Purba	88	90	87	88	88	B
198	0023622426	Restiana Sipayung Reynaldo Pratama	80	75	75	78	77	C
199	0026368213	Sipayung	79	79	80	100	85	B
200	9991767526	Riadi Purba	82	86	80	85	83	B
201	0003967730	Riana Hotmaida Saragih	80	81	79	80	80	B
202	0030211602	Ribka Febrianti	79	81	70	78	77	C

3.3. Alat Analisa Data

Dalam alat analisis data penelitian ini penulis menggunakan aplikasi perangkat lunak RapidMiner. RapidMiner merupakan Software untuk pengolahan Data Mining. RapidMiner (Yale) adalah “perangkat lunak open source atau perangkat lunak yang bersifat terbuka untuk knowledge discovery dan data mining”, RapidMiner mempunyai lebih kurang 400 prosedur (operator) data mining, termasuk operator untuk masukan, output, data preprocessing dan visualisasi. Operator meta secara otomatis mengoptimalkan desain eksperimen dan pengguna tidak memerlukan waktu yang panjang untuk menentukan langkah dan parameter yang lebih panjang.

RapidMiner Studio Versi 9.7 digunakan pada studi ini untuk menampilkan kelompok siswa pada kumpulan data dan memperlihatkan matriks presentasi yang tersebar dari kelompok data. Tampilan RapidMiner Studio Versi 9.7 dapat ditunjukkan oleh Gambar 2.



Gambar 2. Tampilan RapidMiner Studio versi 9.7

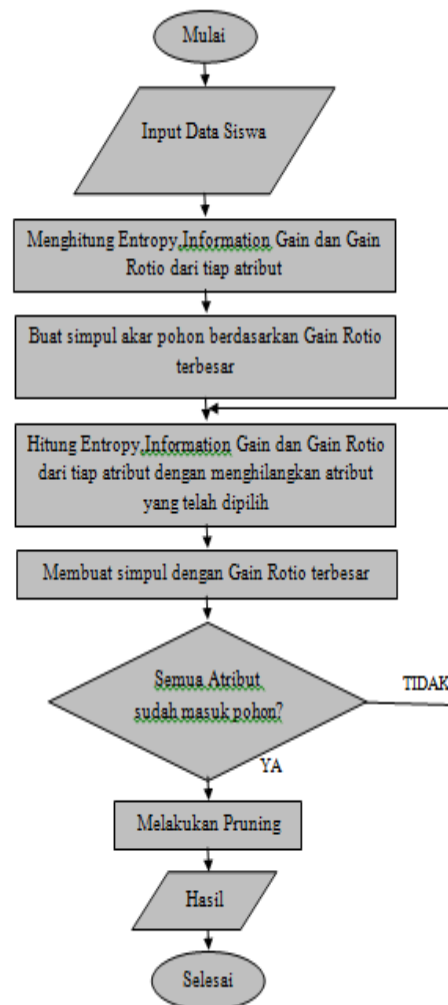
Adapun penjelasan atau keterangan dari antarmuka di RapidMiner Studio Versi 9.7 antara lain sebagai berikut:

1. Operators View: Yang berisi operator-operator yang dapat dimanfaatkan pada proses analisis. Pada operator View ini terdapat beberapa kelompok operator yaitu yang terdiri dari Data Access, Blending, Cleansing, Modeling, Scoring, Validation, Utility, dan Extensions.
2. Repository View: Merupakan komponen utama dalam Design Perspective selain Operators View. Pada Repository digunakan untuk mengelolah dan menata proses analisis, dapat digunakan sebagai sumber data dan yang berkaitan dengan metadata.
3. Process View: Merupakan suatu yang menunjukkan langkah-langkah tertentu dalam proses analisis dan sebagai penghubung langkah-langkah tersebut.
4. Parameter View: Parameter View memiliki toolbar sendiri sama seperti View-view yang lain. Dalam beberapa operator Rapidminer membutuhkan satu atau lebih parameter agar dapat diindikasikan sebagai fungsionalitas yang benar.
5. Help & Comment View: Operator ini berfungsi untuk menunjukkan penjelasan mengenai operator yang dipilih seperti ketika memilih operator view atau Process view, maka operator ini akan menunjukkan penjelasan mengenai operator ini.
6. Problem & Log View: Merupakan komponen yang sangat penting dan merupakan sumber bantuan bagi pengguna selama merancang proses analisis. Setiap peringatan dan pesan kesalahan jelas ditunjukkan dalam problem view.

Pada penelitian ini penulis menggunakan Operator Decision Tree yang ada pada software RapidMiner sebagai operator utama dalam mengerjakan penelitian ini.

3.4. Instrumen Penelitian

Pada penelitian ini, penulis melakukan dataset file dalam bentuk spreadsheet file excel 2007 (xls). Transformasi data ini digunakan sebagai masukan untuk perangkat lunak yang akan digunakan dalam penelitian ini. Dalam pengolahan data transformasi, penulis menggunakan perangkat lunak rapidminer versi 9.7 Instrumen penelitian seperti Gambar 3 dapat dijelaskan bahwa data penelitian berasal dari database pendidikan SMK GKPS 1 Raya.

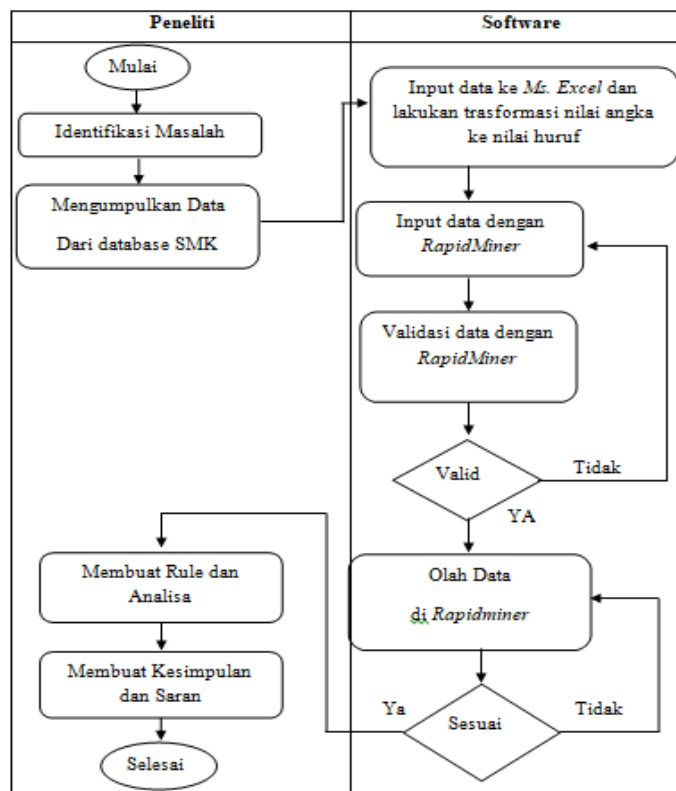


Gambar 3. Proses Percobaan

3.5. Diagram Aktivitas Kerja Penelitian

Berikut ini alur kerja yang akan dilakukan pada penelitian ini yang digambar dalam diagram aktifitas pada Gambar 4.

Pada Gambar 4 merupakan diagram aktivitas yang dilakukan penelitian pada sistem yang digunakan, penulis mengidentifikasi masalah dan tujuan penelitian yang dilakukan, menggumpulkan data, menggunakan data yang diberikan pihak SMK GKPS 1 Raya, kemudian memvalidasi data dengan menggunakan RapidMiner Studio mengukur target eror. Hasil dari informasi yang diberikan aplikasi RapidMiner Studio maka penulis membuat pola decision tree dan analisa yang telah dilakukan perhitungan C4.5.



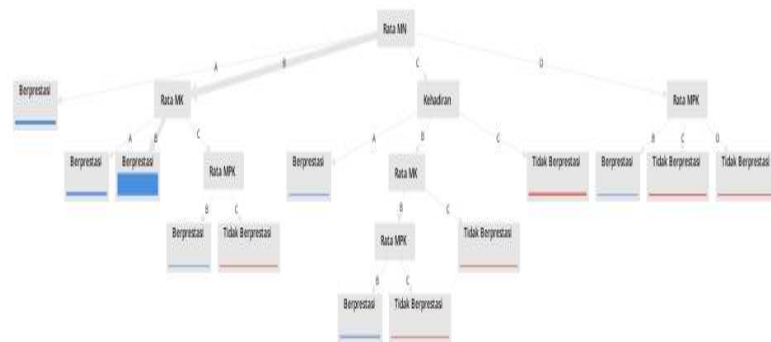
Gambar 4. Diagram Aktivitas Kerja Penelitian

3.6. Pemodelan Metode Decision Tree

Pada pohon keputusan dalam aturan keputusan (Decision rule) merupakan metodologi data mining yang banyak diterapkan sebagai solusi untuk klasifikasi. Pohon keputusan merupakan suatu metode klasifikasi yang menggunakan struktur pohon, yang dimana setiap node merepresentasikan atribut dan cabangnya merepresentasikan nilai dari atribut, dan sedangkan daunnya digunakan untuk merepresentasikan kelas. Node teratas dari Decision Tree ini disebut dengan root.

Dalam pembentukan pohon keputusan terdiri dari beberapa tahapan, antara lain sebagai berikut :

1. **Konstruksi Pohon**
Yang diawali dengan pembentukan akar (terletak paling atas) yang kemudian data dibagi berdasarkan atribut-atribut yang cocok untuk dijadikan daun.
2. **Pemangkasan Pohon (Tree Pruning)**
Untuk mengidentifikasi dan membuang cabang yang tidak diperlukan pada pohon yang telah dibentuk. Tree Pruning dilakukan untuk menyederhanakan tree sehingga akurasi dapat bertambah. Pruning ada dua pendekatan, yaitu:
 - a. **Pre-pruning**, yaitu menghentikan pembangunan suatu subtree lebih awal (yaitu dengan memutuskan untuk tidak lebih jauh mempartisi data training). Saat seketika berhenti, maka node berubah menjadi leaf (node akhir). Node akhir ini menjadi kelas yang paling sering muncul di antara subset sampel.
 - b. **Post-pruning**, yaitu menyederhanakan tree dengan cara membuang beberapa cabang subtree setelah tree selesai dibangun. Node yang jarang dipotong akan menjadi leaf (node akhir) dengan kelas yang paling sering muncul.
3. **Pembentukan Aturan Keputusan**
Membuat aturan keputusan dari pohon yang telah dibentuk. Aturan tersebut dapat berupa bentuk If-then diturunkan dari pohon keputusan dengan melakukan penelusuran dari akar sampai ke daun.
Pada model Decision tree dengan model pohon keputusan yang akan digunakan dapat dilihat pada Gambar 5 berikut.



Gambar 5. Model *Decision Tree*

Pada Gambar 5 Model *Decision Tree* diatas dapat dijelaskan bahwa nilai rata-rata nilai MN (Muatan Nasional) memiliki pengaruh paling besar dalam data penelitian skripsi ini, setelah dilakukan percobaan dengan menggunakan RapidMiner Studio 9.7 bahwa atribut Muatan Nasional berada pada node paling atas atau disebut sebagai akar. Berikut ini Rule yang diterjemahkan kedalam bahasa sederhana dapat ditunjukkan pada Tabel 3.

Tabel 3. Rule

Rule	Keterangan Rule	Keterangan
1.	Jika nilai rata-rata MN = A	Berprestasi
2.	Jika nilai rata-rata MN = B dan nilai rata-rata MK = A	Berprestasi
3.	Jika nilai rata-rata MN = B dan nilai rata-rata MK = B	Berprestasi
4.	Jika nilai rata-rata MN = B, nilai rata-rata MK = C dan nilai rata-rata MPK = B	Berprestasi
5.	Jika nilai rata-rata MN = B, nilai rata-rata MK = C dan nilai rata-rata MPK = C	Tidak Berprestasi
6.	Jika nilai rata-rata MN = C, nilai kehadiran = A	Berprestasi
7.	Jika nilai rata-rata MN = C, nilai kehadiran = B, nilai rata-rata MK = B, nilai rata-rata MPK = B	Berprestasi
8.	Jika nilai rata-rata MN = C, nilai kehadiran = B, nilai rata-rata MK = B, nilai rata-rata MPK = C	Tidak Berprestasi
9.	Jika nilai rata-rata MN = C, nilai Kehadiran = B dan nilai rata-rata MK = C	Tidak Berprestasi
10.	Jika nilai rata-rata MN = C, dan nilai Kehadiran = C	Tidak Berprestasi
11.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = B	Berprestasi
12.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = C	Tidak Berprestasi
13.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = D	Tidak Berprestasi

4. HASIL DAN PEMBAHASAN

Pada pembahasan sebelumnya telah dijelaskan untuk hasil dari penelitian sesuai dengan pernyataan yang diajukan pada awalnya. Penelitian ini dilaksanakan pada suatu data nilai siswa SMK GKPS 1 Raya yang dimana data yang digunakan penulis sebanyak 202 data siswa. Dan data siswa tersebut bersifat nominal yang dimana terdiri dari NISN, Nama, Nilai Rata-rata Muatan Nasional, Nilai Rata-rata Muatan Kewilayahan, Nilai Rata-rata Muatan Peminatan Kejuruan dan Nilai Kehadiran. Yang kemudian data tersebut di input kedalam Microsoft Excel 2007 dan dilakukan transformasi dari nilai angka ke nilai huruf. Yang dimana Jika nilai >90 maka nilainya A, $80 - 89 = B$, $70 - 79 = C$, dan $<60 = D$.

Data yang diperoleh penulis digunakan sebagai contoh sumber input untuk membuat suatu model aturan dalam menentukan prestasi nilai siswa di SMK GKPS 1 Raya menggunakan algoritma C4.5 yang didukung dengan Software RapidMiner. Pada percobaan ini penulis menampilkan gambar hasil dari pemetaan model aturan dalam menentukan prestasi nilai siswa menggunakan algoritma C4.5. Hasil dari pohon keputusan ini akan digunakan untuk membuat suatu model aturan yang akan dipilih dalam mengambil keputusan pada bagian pihak sekolah. Maka dari itu akan dijelaskan bagaimana proses data mining yang telah di aplikasikan menggunakan RapidMinerstudio9.7.

4.1. Pemodelan Algoritma C4.5

Berikut ini langkah-langkah perhitungan dalam algoritma C4.5 pada penyelesaian kasus dalam menentukan prestasi nilai siswa yang dimana akan dibagi menjadi suatu label “Iya” (Berprestasi) dan “Tidak” (Tidak Berprestasi). Dalam perhitungan algoritma C4.5 berikut tahapan yang akan dilakukan :

1. Perhitungan Mencari Nilai Entropy

Pada proses pertama perhitungan algoritma C4.5 yaitu menentukan nilai entropy. Untuk langkah pertama yang harus dilakukan ialah dengan menentukan entropy total kasus terlebih dahulu. Adapun rumus untuk mencari nilai Entropy dari kasus data nilai siswa tersebut dapat ditunjukkan pada persamaan (1) berikut.

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i * \text{Log}_2 p_i \quad (1)$$

Keterangan:

S : Himpunan Kasus

n : Jumlah Partisi S

p_i : Proporsi dari S_i terhadap S

Maka:

Entropy (Total) =

$$\left(-\frac{\text{sum}(ya)}{\text{Total}} * \text{Log}_2 \left(\frac{\text{sum}(ya)}{\text{Total}} \right) \right) \left(-\frac{\text{sum}(Tidak)}{\text{Total}} * \text{Log}_2 \left(\frac{\text{sum}(Tidak)}{\text{Total}} \right) \right)$$

Maka:

$$\text{Entropy (Total)} = \left(-\frac{183}{202} * \text{Log}_2 \left(\frac{183}{202} \right) \right) + \left(-\frac{19}{202} * \text{Log}_2 \left(\frac{19}{202} \right) \right)$$

Tabel 4. Perhitungan *Entropy* Total Kasus

Total Kasus	Jumlah (Berprestasi)	Jumlah (Tidak Berprestasi)	Entropy
202	183	19	0,641538569

Kemudian lanjut menghitung nilai *Entropy* dari semua kasus yang terbagi berdasarkan atribut “MN” (Muatan Nasional), “MK” (Muatan Kewilayahan), “MPK” (Muatan Peminatan Kejuruan), dan “Kehadiran”. Seperti berikut ini:

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i * \text{Log}_2 p_i$$

$$\text{Entropy [MN-A]} = \left(-\frac{18}{18} * \text{Log}_2 \left(\frac{18}{18} \right) \right) + \left(-\frac{0}{18} * \text{Log}_2 \left(\frac{0}{18} \right) \right)$$

$$= 0$$

$$\text{Entropy [MN-B]} = \left(-\frac{156}{157} * \text{Log}_2 \left(\frac{156}{157} \right) \right) + \left(-\frac{1}{157} * \text{Log}_2 \left(\frac{1}{157} \right) \right)$$

$$= 0,092925105$$

$$\text{Entropy [MN-C]} = \left(-\frac{8}{20} * \text{Log}_2 \left(\frac{8}{20} \right) \right) + \left(-\frac{12}{20} * \text{Log}_2 \left(\frac{12}{20} \right) \right)$$

$$= 1,235336214$$

$$\text{Entropy [MN-D]} = \left(-\frac{1}{7} * \text{Log}_2 \left(\frac{1}{7} \right) \right) + \left(-\frac{6}{7} * \text{Log}_2 \left(\frac{6}{7} \right) \right)$$

$$= 2,596926295$$

$$\begin{aligned}
Entropy [MK-A] &= \left(-\frac{25}{25} * \text{Log}_2\left(\frac{25}{25}\right)\right) + \left(-\frac{0}{25} * \text{Log}_2\left(\frac{0}{25}\right)\right) \\
&= 0 \\
Entropy [MK-B] &= \left(-\frac{155}{161} * \text{Log}_2\left(\frac{155}{161}\right)\right) + \left(-\frac{6}{161} * \text{Log}_2\left(\frac{6}{161}\right)\right) \\
&= 0,178909821 \\
Entropy [MK-C] &= \left(-\frac{3}{14} * \text{Log}_2\left(\frac{3}{14}\right)\right) + \left(-\frac{11}{14} * \text{Log}_2\left(\frac{11}{14}\right)\right) \\
&= 2,019533784 \\
Entropy [MK-D] &= \left(-\frac{0}{2} * \text{Log}_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} * \text{Log}_2\left(\frac{2}{2}\right)\right) \\
&= 0 \\
Entropy [MPK-A] &= \left(-\frac{22}{22} * \text{Log}_2\left(\frac{22}{22}\right)\right) + \left(-\frac{0}{22} * \text{Log}_2\left(\frac{0}{22}\right)\right) \\
&= 0 \\
Entropy [MPK-B] &= \left(-\frac{146}{152} * \text{Log}_2\left(\frac{146}{152}\right)\right) + \left(-\frac{6}{152} * \text{Log}_2\left(\frac{6}{152}\right)\right) \\
&= 0,368128817 \\
Entropy [MPK-C] &= \left(-\frac{15}{25} * \text{Log}_2\left(\frac{15}{25}\right)\right) + \left(-\frac{10}{25} * \text{Log}_2\left(\frac{10}{25}\right)\right) \\
&= 1,057542476 \\
Entropy [MPK-D] &= \left(-\frac{0}{3} * \text{Log}_2\left(\frac{0}{3}\right)\right) + \left(-\frac{3}{3} * \text{Log}_2\left(\frac{3}{3}\right)\right) \\
&= 0 \\
Entropy [Kehadiran-A] &= \left(-\frac{34}{36} * \text{Log}_2\left(\frac{34}{36}\right)\right) + \left(-\frac{2}{36} * \text{Log}_2\left(\frac{2}{36}\right)\right) \\
&= 0,463325 \\
Entropy [Kehadiran-B] &= \left(-\frac{139}{144} * \text{Log}_2\left(\frac{139}{144}\right)\right) + \left(-\frac{5}{144} * \text{Log}_2\left(\frac{5}{144}\right)\right) \\
&= 0,336666452 \\
Entropy [Kehadiran-C] &= \left(-\frac{10}{22} * \text{Log}_2\left(\frac{10}{22}\right)\right) + \left(-\frac{12}{22} * \text{Log}_2\left(\frac{12}{22}\right)\right) \\
&= 0,953966311 \\
Entropy [Kehadiran-D] &= \left(-\frac{0}{0} * \text{Log}_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \text{Log}_2\left(\frac{0}{0}\right)\right) \\
&= 0
\end{aligned}$$

2. Perhitungan Mencari Nilai Gain

Setelah mendapat hasil *Entropy* dari semua atribut maka langkah selanjutnya yang akan dilakukan ialah menghitung nilai *Gain* dari setiap atribut.

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

Gain (Total, MN)

$$\begin{aligned}
&= 0,641538569 - \left(\frac{18}{202} * 0\right) + \left(\frac{157}{202} * 0,092925105\right) + \\
&\quad \left(\frac{20}{202} * 1,235336214\right) + \left(\frac{7}{202} * 2,596926295\right) \\
&= 0,357011589
\end{aligned}$$

Gain (Total, MK)

$$\begin{aligned}
&= 0,641538569 - \left(\frac{25}{202} * 0\right) + \left(\frac{161}{202} * 0,178909821\right) + \\
&\quad \left(\frac{14}{202} * 2,019533784\right) + \left(\frac{2}{202} * 0\right) \\
&= 0,358974439
\end{aligned}$$

Gain (Total, MPK)

$$\begin{aligned}
&= 0,641538569 - \left(\frac{22}{202} * 0\right) + \left(\frac{152}{202} * 0,368128817\right) +
\end{aligned}$$

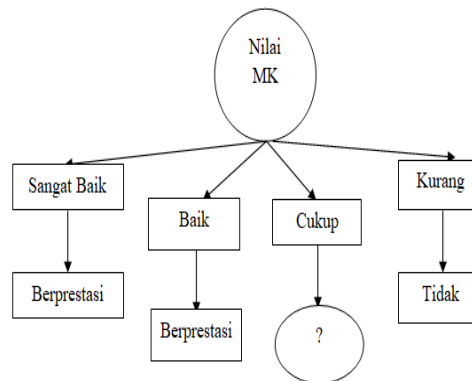
$$\begin{aligned} & \left(\frac{25}{202} * 1,057542476 \right) + \left(\frac{3}{202} * 0 \right) \\ & = 0,233646776 \\ \text{Gain (Total, Kehadiran)} \\ & = 0,641538569 - \left(\frac{36}{202} * 0,463325 \right) + \left(\frac{144}{202} * 0,336666452 \right) + \\ & \quad \left(\frac{22}{202} * 0,953966311 \right) + \left(\frac{0}{202} * 0 \right) \\ & = 0,215068628 \end{aligned}$$

Berikut perhitungan diatas ditunjukkan pada tabel 5 yang dikerjakan di *Excel* antara lain sebagai berikut:

Tabel 5. Tabel Hasil Perhitungan 1

Atribut	Jumlah Kasus	Berprestasi	Tidak Berprestasi	Entropy	Gain
	(S)	(S1)	(S2)		
Total	202	194	8	0,641538569	
Rata MN	A 18	18	0	0	0,357011589
	B 157	156	1	0,092925105	
	C 20	8	12	1,235336214	
	D 7	1	6	2,596926295	
Rata MK	A 25	25	0	0	0,358974439
	B 161	155	6	0,178909821	
	C 14	3	11	2,019533784	
	D 2	0	2	0	
Rata MPK	A 22	22	0	0	0,233646776
	B 152	146	6	0,368128817	
	C 25	15	10	1,057542476	
	D 3	0	3	0	
Kehadiran	A 36	34	2	0,463325	0,215068628
	B 144	139	5	0,336666452	
	C 22	10	12	0,953966311	
	D 0	0	0	0	

Pada Tabel 5 diatas dapat dilihat bahwa atribut Nilai Rata Muatan Kewilayahan (MK) memiliki Gain tertinggi yaitu 0,358974439, maka nilai rata Muatan Kewilayahan (MK) yang menjadi node akar. Rata MK memiliki 4 nilai yaitu A (Sangat Baik), B (Baik), C (Cukup), D (Kurang). Pada nilai Sangat Baik, Baik dan Kurang (A, B dan D) telah mengklasifikasi kasus menjadi 1 yaitu keputusan “Berprestasi” dan “Tidak Berprestasi” sedangkan untuk nilai Cukup (C) masih diperlukan perhitungan lagi karena masih memiliki hasil antara “Berprestasi” dan “Tidak”, perhitungan tersebut dilakukan untuk menentukan node akar selanjutnya, maka dapat digambarkan pohon keputusan dari tabel diatas dapat ditunjukkan pada Gambar 6.



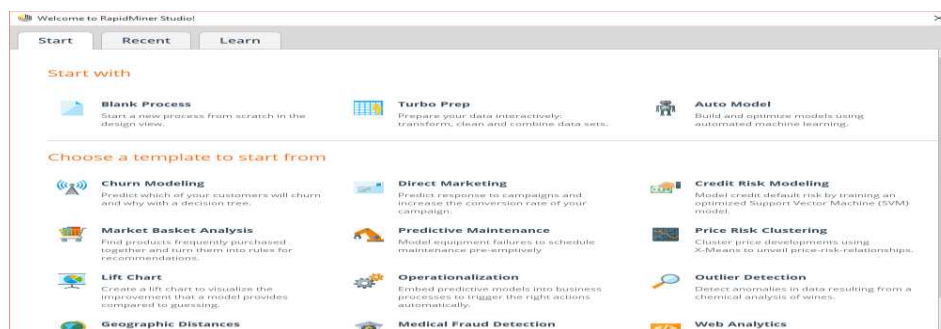
Gambar 6. Pohon Keputusan 1

Pada gambar 5 diatas dapat dilihat bahwa nilai rata Muatan Kewilayahan digunakan sebagai node akar. Nilai atribut Sangat Baik, nilai Baik telah terklasifikasi yaitu “Berprestasi” sedangkan nilai atribut yang bernilai Kurang terklasifikasi yaitu dengan atribut “Tidak Berprestasi” dan nilai atribut yang bernilai Cukup belum terklasifikasi. Dalam perhitungan selanjutnya yaitu untuk menghitung node akar yang bernilai Cukup perhitungannya dilakukan dengan cara yang sama seperti perhitungan pertama yang telah dilakukan.

4.2. Hasil Percobaan Algoritma C4.5 Menggunakan *RapidMiner Studio 9.7*

RapidMiner Studio 9.7 merupakan *software tool open source* untuk *data mining*. *RapidMiner Studio 9.7* juga menyediakan beberapa langkah dan *machine learning* termasuk ETL (*Extraction, Transformation, Loading*). Data siswa yang telah diperoleh dan sudah dianalisis kemudian diolah menggunakan aplikasi *RapidMiner Studio 9.7*. Untuk dapat mengetahui klasifikasi siswa yang berprestasi atau tidak. Berikut ini langkah-langkah percobaan algoritma C4.5 pada *RapidMiner Studio 9.7*:

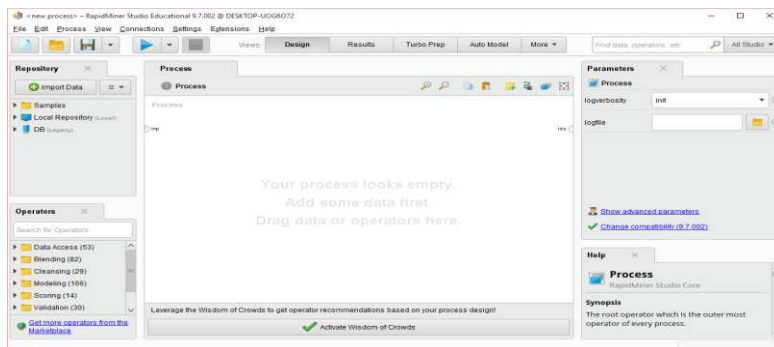
1. Buka aplikasi *RapidMiner Studio 9.7*, halaman depan *RapidMiner Studio 9.7* dapat ditunjukkan oleh Gambar 7.



Gambar 7. Tampilan *Rapidminer Studio* Versi 9.7

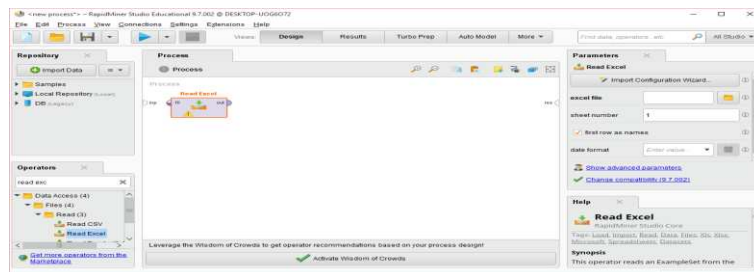
Tampilan menu dari Gambar 7 Penjelasannya adalah sebagai berikut :

- Start, adalah menu digunakan untuk menunjukkan bagian-bagian aksi yang akan kalian gunakan untuk prosesnya.
 - Recent, Berfungsi untuk menampilkan proses-proses yang sudah dilakukan sebelumnya.
 - Learn, Berfungsi untuk menampilkan pembelajaran lebih detail mengenai penggunaan *RapidMiner* yang sudah disediakan oleh aplikasi tersebut.
2. Untuk membuat proses baru pilih start, pada start Pilih New Process sehingga tampilan New Process yang dimaksud dapat di tunjukkan oleh Gambar 8.



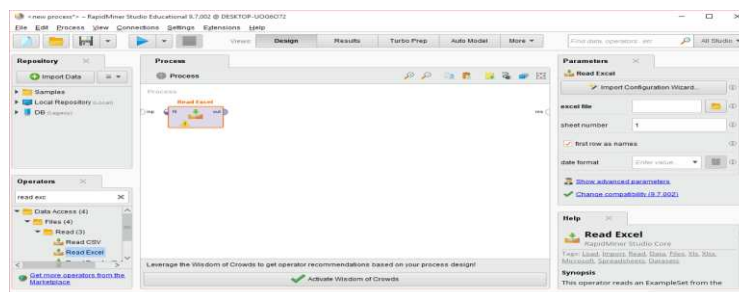
Gambar 8. Tampilan *New Process*

3. Pilih Operators lalu cari Read Excel kemudian drag and drop ke panel Process, dapat ditunjukkan oleh Gambar 9.



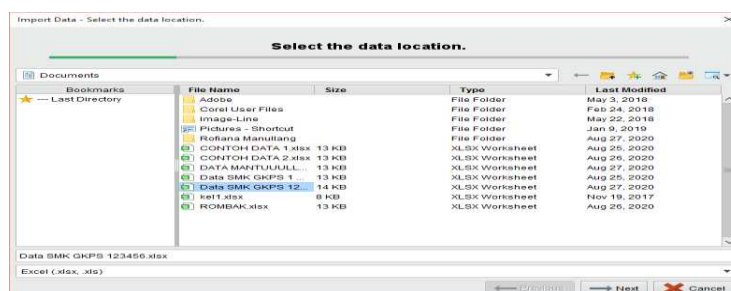
Gambar 9. Pemilihan Atribut

4. *Import Data*



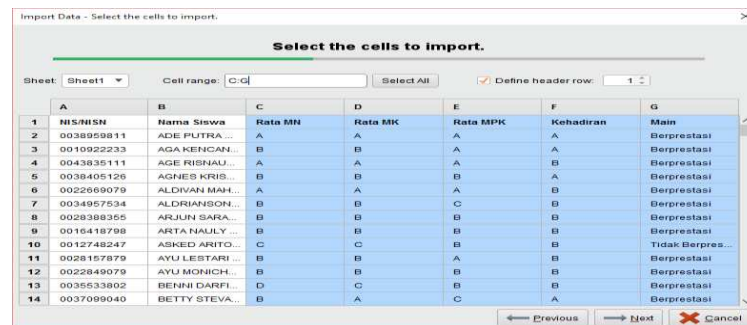
Gambar 10. *Importing Data*

Langkah selanjutnya yaitu melakukan *import data* dengan cara klik *import Configuration Wizard* pada panel *Parameters*, kemudian pilih file yang ingin di *import*. Dapat ditunjukkan oleh Gambar 11.



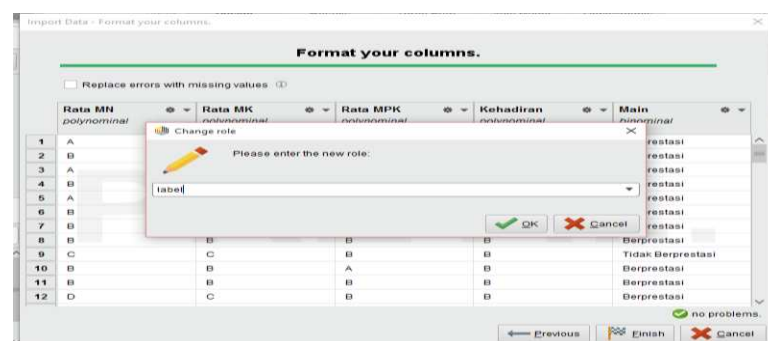
Gambar 11. Pemilihan Data yang Ingin di *Import*

Setelah data dipilih maka kemudian klik *next*, lalu seleksi data apa saja yang akan digunakan, pada penelitian ini data yang digunakan hanya nilai rata Muatan Nasional, Muatan Kewilayahan, Muatan Peminatan Kejuruan, Kehadiran dan *Main* atau Keterangan. Dapat ditunjukkan oleh Gambar 12.



Gambar 12. Pemilihan Atribut/Data Cleaning

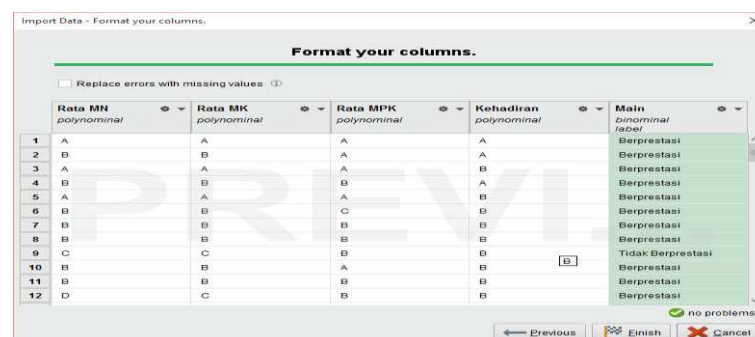
Kemudian klik *next* dan tetapkan type *Polynomial* Menjadi *Binominal* dan *rule* pada *Main* menjadi label. Dapat di tunjukkan oleh Gambar 13.



Gambar 13. Menentukan Role

Untuk langkah selanjutnya yaitu menetapkan atribut “Main” sebagai variabel keputusan. Dan sedangkan untuk variabel penentu diantaranya:

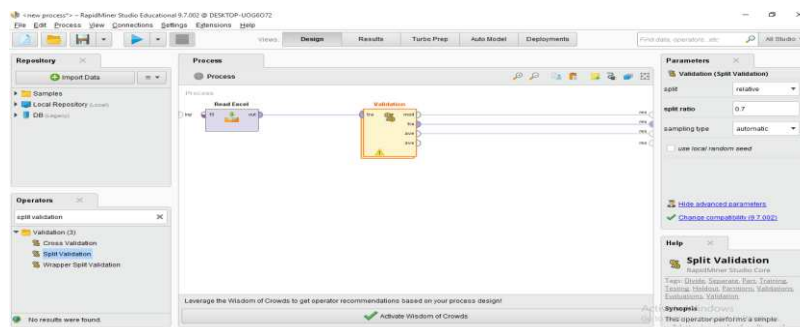
Nilai rata-rata MN (Muatan Nasional), Nilai rata-rata MK (Muatan Kewilayahan), Nilai rata-rata MPK (Muatan Peminatan Kejuruan), dan Nilai Kehadiran. Dapat ditunjukkan oleh Gambar 14.



Gambar 14. Data Selection

Pada Gambar 14 diatas menunjukkan atribut Main telah berhasil diubah menjadi label. Label tersebut menandakan bahwa atribut Main adalah atribut keputusan. Setelah selesai menetapkan label maka klik *Finish* dan proses *Importing Data* telah selesai.

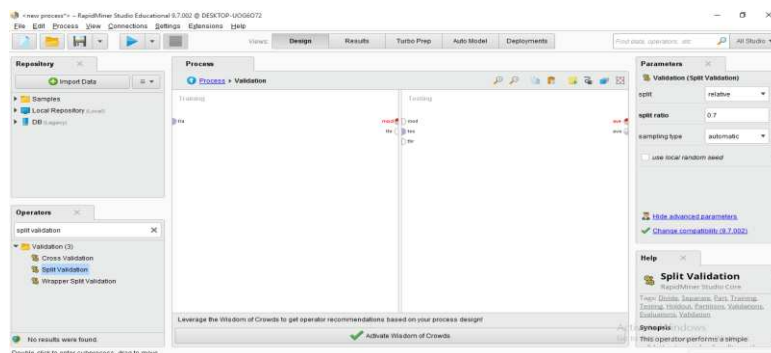
5. Menghubungkan *port read excel* dengan *port validation*



Gambar 15. Penghubungan *Port Read Excel* Dengan *Port Validation* Vata

Langkah selanjutnya yaitu untuk melakukan *testing* dengan *drag and drop split validation* pada menu *operators* kedalam panel *process*. Setelah *split validation* muncul maka selanjutnya lakukan penghubungan antara *read excel* dengan *split validation* seperti gambar 14 diatas.

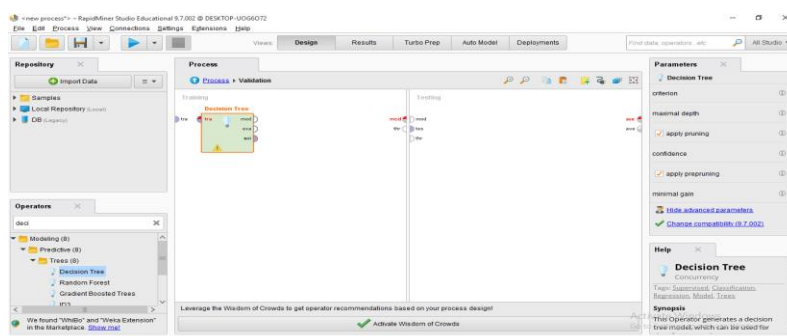
6. Tampilan panel *process validation*



Gambar 16. Panel *Process Validation*

Lalu kemudian *double click* pada *validation* untuk membuka tampilan sub *process training* dan *testing*.

7. Tampilan *operators decision tree* pada tabel *training*



Gambar 17. Tampilan *Operators Decision Tree* Pada *Training*

Pada tabel *Training* lakukan *drag and drop* algoritma yang akan digunakan yaitu *Decision tree*.

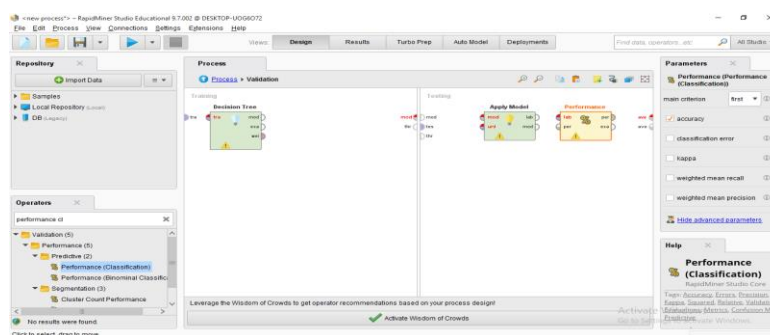
8. Tampilan *Operators apply model* pada tabel *testing*

Selanjutnya pada tabel *testing* lakukan *drag and drop Apply Model*, *Apply model* ini fungsinya untuk mempelajari informasi *ExampleSet* yang telah dilatih dan digunakan untuk prediksi menggunakan model ini.



Gambar 18. Tampilan *operators apply model* pada tabel *testing*

9. Tampilan *operators performance* pada tabel *testing*

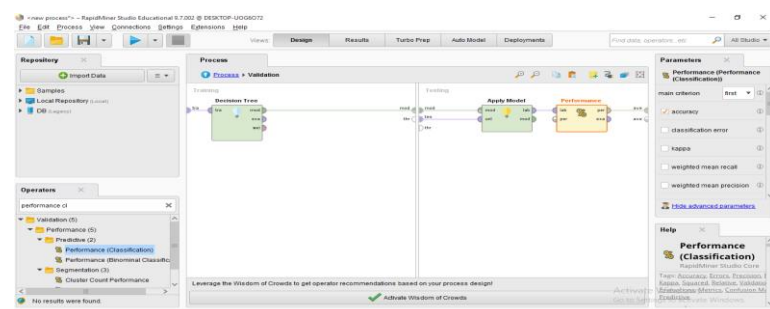


Gambar 19. Tampilan *operators performance* tabel *testing*

Lalu lakukan *Drag and drop performance classification* pada *table testing*. *Performance* digunakan untuk evaluasi statistic dari kinerja klasifikasi dan memberikan daftar nilai criteria dari kinerja klasifikasi tersebut.

10. Menghubungkan *port decision tree*, *apply model* dan *port performance*

Selanjutnya lakukan penghubung antara *port decision tree*, *apply model* dan *port performance* dapat ditunjukkan oleh gambar 19.

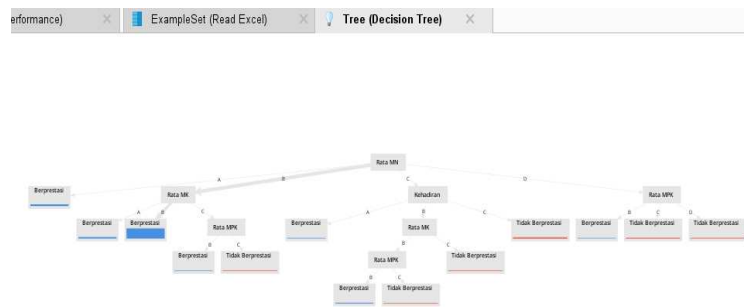


Gambar 20. Tampilan Menghubungkan *Port Decision Tree*, *Apply Model* Dan *Port performance*

Lalu kemudian langkah selanjutnya yaitu klik *Run* pada toolbar untuk menampilkan hasil.

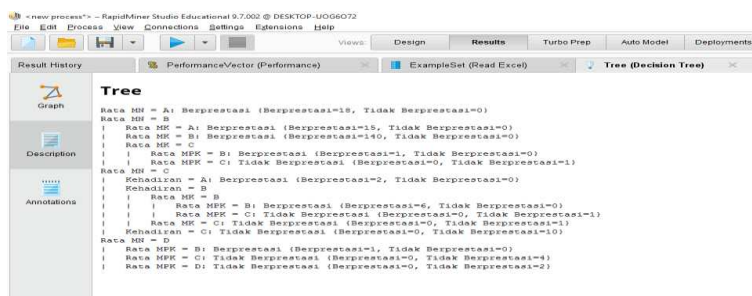
11. Hasil Percobaan *Decision Tree*

Setelah dilakukan perhitungan dan pengujian data pada masing-masing atribut dengan algoritma C4.5, maka akan didapatkan pola pohon keputusan akhir



Gambar 21. Hasil Percobaan *Decision Tree*

12. Tampilan Deskripsi *Decision Tree*



Gambar 22. Deskripsi *Decision Tree*

Gambar 22 menunjukkan hasil deskripsi secara lengkap dari pohon keputusan (*Decision Tree*) yang telah terbentuk dengan menggunakan algoritma C4.5. Dari hasil deskripsi juga menunjukkan bahwa penggunaan *data mining* algoritma C4.5 baik digunakan untuk proses menggali data (*Data mining process*) untuk menarik beberapa kesimpulan yang divisualisasikan dengan pohon keputusan (*Decision Tree*). Berikut *rules* yang dihasilkan dari pohon keputusan dapat di tunjukkan pada Tabel 6.

Tabel 6. Rule Hasil Pohon Keputusan

Rule	Keterangan Rule	Keterangan
1.	Jika nilai rata-rata MN = A	Berprestasi
2.	Jika nilai rata-rata MN = B dan nilai rata-rata MK = A	Berprestasi
3.	Jika nilai rata-rata MN = B dan nilai rata-rata MK = B	Berprestasi
4.	Jika nilai rata-rata MN = B, nilai rata-rata MK = C dan nilai rata-rata MPK = B	Berprestasi
5.	Jika nilai rata-rata MN = B, nilai rata-rata MK = C dan nilai rata-rata MPK = C	Tidak Berprestasi
6.	Jika nilai rata-rata MN = C, nilai kehadiran = A	Berprestasi
7.	Jika nilai rata-rata MN = C, nilai kehadiran = B, nilai rata-rata MK = B, nilai rata-rata MPK = B	Berprestasi
8.	Jika nilai rata-rata MN = C, nilai kehadiran = B, nilai rata-rata MK = B, nilai rata-rata MPK = C	Tidak Berprestasi
9.	Jika nilai rata-rata MN = C, nilai Kehadiran = B dan nilai rata-rata MK = C	Tidak Berprestasi
10.	Jika nilai rata-rata MN = C, dan nilai Kehadiran = C	Tidak Berprestasi
11.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = B	Berprestasi
12.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = C	Tidak Berprestasi
13.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = D	Tidak Berprestasi

Untuk lebih jelasnya dapat dilihat pada model aturan berbentuk teks dapat di tunjukkan oleh gambar 23.

```
Rata MN = A: Berprestasi {Berprestasi=18, Tidak Berprestasi=0}
Rata MN = B
| Rata MK = A: Berprestasi {Berprestasi=15, Tidak Berprestasi=0}
| Rata MK = B: Berprestasi {Berprestasi=140, Tidak Berprestasi=0}
| Rata MK = C
| | Rata MPK = B: Berprestasi {Berprestasi=1, Tidak Berprestasi=0}
| | Rata MPK = C: Tidak Berprestasi {Berprestasi=0, Tidak Berprestasi=1}
Rata MN = C
| Kehadiran = A: Berprestasi {Berprestasi=2, Tidak Berprestasi=0}
| Kehadiran = B
| | Rata MK = B
| | | Rata MPK = B: Berprestasi {Berprestasi=6, Tidak Berprestasi=0}
| | | Rata MPK = C: Tidak Berprestasi {Berprestasi=0, Tidak Berprestasi=1}
| | Rata MK = C: Tidak Berprestasi {Berprestasi=0, Tidak Berprestasi=1}
| Kehadiran = C: Tidak Berprestasi {Berprestasi=0, Tidak Berprestasi=10}
Rata MN = D
| Rata MPK = B: Berprestasi {Berprestasi=1, Tidak Berprestasi=0}
| Rata MPK = C: Tidak Berprestasi {Berprestasi=0, Tidak Berprestasi=4}
| Rata MPK = D: Tidak Berprestasi {Berprestasi=0, Tidak Berprestasi=2}
```

Gambar 23. Model Aturan Text *Decision Tree*

Dari Gambar 23 di atas dapat disimpulkan bahwa faktor yang menentukan prestasi nilai siswa berada pada *node* pertama adalah Nilai rata-rata MN (Muatan Nasional), *node* kedua Nilai rata-rata MK (Muatan Kewilayahan), *node* ketiga Nilai rata-rata MPK (Muatan Peminatan Kejuruan) dan *node* ke empat adalah Kehadiran. Untuk melihat *rule* dengan *gain* rasio predikat dapat ditunjukkan pada tabel 7 dengan ketentuan Berprestasi dan TB (Tidak Berprestasi).

Tabel 7. Keterangan *Rule Text* Dengan *Gain* Rasio

Rule	Keterangan <i>Rule</i>	Predikat <i>Gain</i> Rasio	
		Berprestasi	T.B
1.	Jika nilai rata-rata MN = A maka berprestasi	18	0
2.	Jika nilai rata-rata MN = B dan nilai rata-rata MK = A maka Berprestasi	15	0
3.	Jika nilai rata-rata MN = B dan nilai rata-rata MK = B maka Berprestasi	140	0
4.	Jika nilai rata-rata MN = B, nilai rata-rata MK = C dan nilai rata-rata MPK = B maka Berprestasi	1	0
5.	Jika nilai rata-rata MN = B, nilai rata-rata MK = C dan nilai rata-rata MPK = C maka Tidak Berprestasi	0	1
6.	Jika nilai rata-rata MN = C, nilai kehadiran = A maka Berprestasi	2	0
7.	Jika nilai rata-rata MN = C, nilai kehadiran = B, nilai rata-rata MK = B, nilai rata-rata MPK = B maka Berprestasi	6	0
8.	Jika nilai rata-rata MN = C, nilai kehadiran = B, nilai rata-rata MK = B, nilai rata-rata MPK = C maka Tidak Berprestasi	0	1
9.	Jika nilai rata-rata MN = C, nilai Kehadiran = B dan nilai rata-rata MK = C maka Tidak Berprestasi	0	1
10.	Jika nilai rata-rata MN = C, dan nilai Kehadiran = C maka Tidak Berprestasi	0	10
11.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = B maka Berprestasi	1	0
12.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = C maka Tidak Berprestasi	0	4
13.	Jika nilai rata-rata MN = D, dan nilai rata-rata MPK = D maka Tidak Berprestasi	0	2

Dari Tabel 7 di atas dapat disimpulkan bahwa siswa yang memiliki nilai berprestasi ada sejumlah 183 siswa dan jumlah siswa yang tidak berprestasi ada sebanyak 19 siswa.

5. KESIMPULAN

Kesimpulan dari penelitian ini diperoleh suatu model aturan yang dapat memperlihatkan aturan dalam menentukan prestasi nilai siswa SMK GKPS 1 Raya menggunakan algoritma C4.5 dan dalam studi kasus pada SMK GKPS 1 Raya bahwa sebagian siswa yang tidak memiliki prestasi Jika nilai rata-rata MN = B, nilai rata-rata MK = C dan nilai rata-rata MPK = C dan apabila nilai rata-rata Muatan Nasional, Muatan Kewilayahan, Muatan Peminatan Kejuruan dan Kehadiran berada pada nilai <69 atau sama dengan bernilai D serta algoritma C4.5 dianggap algoritma yang sangat membantu dalam melakukan klasifikasi data karena karakteristik data klasifikasi didapatkan dengan jelas seperti pada bentuk struktur pohon keputusan maupun model aturan, sehingga memudahkan pengguna dalam melakukan penggalan informasi terhadap yang bersangkutan.

DAFTAR PUSTAKA

- [1] M. I. Prestasi, "Universitas Sumatera Utara," 2011.
- [2] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, 2007, doi: 10.1016/j.eswa.2006.04.005.
- [3] E. N. Ogor, "Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques Department of Natural Sciences Turks & Caicos Islands Community College Visualization and Articulation," *Fourth Congr. Electron. Robot. Automot. Mech.*, pp. 354–359, 2007, doi: 10.1109/CERMA.2007.78.
- [4] Y. S. Luvia, A. P. Windarto, S. Solikhun, and D. Hartama, "Penerapan Algoritma C4.5 Untuk Klasifikasi Predikat Keberhasilan Mahasiswa Di Amik Tunas Bangsa," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 1, no. 1, p. 75, 2017, doi: 10.30645/jurasik.v1i1.12.
- [5] F. F. Harryanto and S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *Tek. Inform. Dan Sist. Inf.*, vol. 3, no. 2, pp. 95–103, 2017, [Online]. Available: <http://jurnal.mdp.ac.id/index.php/jatisi/article/view/71>.
- [6] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- [7] W. Julianto, R. Yunitarini, and M. K. Sophan, "Algoritma C4.5 Untuk Penilaian Kinerja Karyawan," *Scan*, vol. Vo. IX, no. No. 2, pp. 33–39, 2014.
- [8] A. K-medoids, R. Ulina, B. Barus, I. Gunawan, and B. E. Damanik, "Pengelompokan Data Penjualan Mie Berdasarkan Bulan Dengan Menggunakan," vol. 1, no. 2, 2021.
- [9] A. Muzakir, "Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) 2014 Yogyakarta, 15 November 2014 ISSN: 1979-911X," *Snast*, no. November, pp. 211–216, 2014.
- [10] 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, 2014, doi: 10.31937/ti.v6i1.327.
- [11] R. Nofitri and N. Irawati, "Integrasi Metode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 6, no. 1, pp. 35–42, 2019, doi: 10.33330/jurteks.v6i1.393.