
Prediksi Target Sistem *Productivity Services* menggunakan Penerapan Algoritma C4.5 pada PT. Sanggar Sarana Baja Jakarta

Anis Rahmawati, Syifa Nur Rakhmah, Lusa
Indah Prahartiwi

Sistem Informasi, Fakultas Ilmu Komputer, Sekolah Tinggi Manajemen
Informatika dan Komputer Nusa Mandiri

Jl. Kaliabang No.8, Perwira, Kec. Bekasi Utara,
Kota Bks, Jawa Barat 17124

e-mail: anisra11162847@nusamandiri.ac.id, syifa.snk@nusamandiri.ac.id,
lusa.lip@nusamandiri.ac.id

Abstract

There are many ways that each service provider company does, especially services to win the competition, among others, by increasing service productivity targets. One service provider company that is committed to increasing service productivity targets is PT. Sanggar Sarana Baja. This study aims to predict service productivity system targets using the application of Algoritma C4.5 at PT. Sanggar Sarana Baja. The attributes of working time input in this study include area, performance, efficiency, and productivity. In this study, it was found that the results obtained came from several input attributes which resulted in a causal relationship in classifying the results of service productivity targets at PT. Sanggar Sarana Baja. This research is expected to help PT. Sanggar Sarana Baja in increasing customer satisfaction to retain customers and increase profits of PT. Sanggar Sarana Baja. Based on the classification results using the C4.5 Algorithm, it shows that the accuracy reaches 95.00%, which indicates that the C4.5 algorithm is suitable for measuring the target level at PT. Sanggar Sarana Baja.

Keywords: Accuracy, Validation, Decision Tree, Data mining, KDD, C4.5 Algorithm, Services Companies, Target Services Productivity Systems

Abstrak

Banyak cara yang dilakukan oleh masing - masing perusahaan penyedia jasa, khususnya servis untuk memenangkan persaingan, antara lain dengan meningkatkan target produktivitas servis. Salah satu perusahaan penyedia jasa servis yang berkomitmen dalam meningkatkan target produktivitas servis adalah PT. Sanggar Sarana Baja. Penelitian ini bertujuan untuk memperdiksi target sistem produktivitas servis menggunakan penerapan Algoritma C4.5 pada PT. Sanggar Sarana Baja. Atribut masukan waktu kerja dalam penelitian ini mencakup daerah, kinerja, efisiensi, dan produktivitas. Dalam penelitian ini, didapatkan bahwa hasil yang didapatkan berasal dari beberapa atribut masukan menghasilkan hubungan sebab -akibat dalam mengklasifikasikan hasil dari target produktivitas servis pada PT. Sanggar Sarana Baja. Penelitian ini diharapkan dapat membantu pihak PT. Sanggar Sarana Baja dalam meningkatkan kepuasan konsumen untuk mempertahankan pelanggan dan meningkatkan laba PT. Sanggar Sarana Baja tersebut. Berdasarkan hasil klasifikasi menggunakan Algoritma C4.5 menunjukkan bahwa diperoleh akurasi mencapai 95,00%, yang menunjukkan bahwa algoritma C4.5 cocok digunakan untuk mengukur tingkat target pada PT. Sanggar Sarana Baja.

Kata kunci: Akurasi, Validasi, Decision Tree, Data mining, KDD, Algoritma C4.5, Perusahaan Jasa, Target Sistem *Productivity Services*

1. Pendahuluan

Dalam usaha meningkatkan dunia bisnis, PT. Sanggar Sarana Baja berupa menjaga kepercayaan yang diberikan pelanggan dengan kualitas dari barang buat dan *services* di pesan. Setiap perusahaan mempunyai orientasi bisnis yang sama yaitu ingin menghasilkan keuntungan yang maksimal dengan meminimalkan biaya yang dikeluarkan. Keinginan perusahaan dapat terwujud dengan adanya pemilihan target sistem *productivity services* yang tepat. Karena itu target memiliki peranan yang sangat penting bagi perusahaan dalam jangka waktu kedepannya. Oleh karena itu, pemilihan target sebagai menyediakan *services* menjadi sangat penting dalam dunia bisnis ini. Pemilihan target dalam *services* yang baik memenuhi beberapa standart yaitu target untuk jangka menengah dan panjang dalam pengerjaan *services*nya.

Prediksi target *service* adalah sebuah data yang berupa dari *history* hasil *productivity services* yang ditahun lalu dijadikan target di tahun yang akan datang. Sistem pada *productivity services* adalah sebuah kumpulan data *timesheet* yang dimana sebagai suatu proses pekerjaan yang dilakukan di *services*. Pada PT. Sanggar Sarana Baja sendiri sudah menggunakan *timesheet*. *Timesheet* adalah sebuah pencatatan suatu pekerjaan yang sudah dikerjakan atau sedang dilakukan pekerjaan untuk mengetahui hasil dari proses pekerjaan tersebut. PT. Sanggar Sarana Baja adalah yang bergerak di bidang proses desain, manufaktur, fabrikasi baja, pemasok, instalasi, jasa, remanufaktur & refurbish dari produk, peralatan dan komponen atau *part* untuk pertambangan, minyak & gas dan industri umum lainnya, secara keseluruhan, bertekad untuk menjadi perusahaan yang diakui secara global.

Permasalahan pada PT. Sanggar Sarana Baja yaitu hanya pada saat waktu pengerjaan *services*. Pengerjaan *services* pada tepat waktu yang telah di sesuaikan dengan perjanjian yang ada, agar produksi berjalan dengan baik tidak ada kendala. Target yang telah di rencanakan di awal sesuai prediksi akhir. Selama ini proses target dilakukan dengan cara *history* dalam jangka waktu 6 bulan dan di liat dari tahun sebelumnya. Maka dari itu di butuhkan perancangan target baru untuk sistem pada *productivity services* agar mempermudah dalam perhitungan ketentuan secara tepat dan akurat.

Pada penelitian ini penulis akan menggunakan sistem pada *productivity services* yang nantinya kemudian akan di harapkan dapat membantu PT. Sanggar Sarana Baja dalam menentukan terget yang tepat untuk memenuhi semua kebutuhan waktu kerja. Maka hasil dari sistem pemilihan ini dapat memberikan pilihan yang akurat dalam menentukan target terbaik.

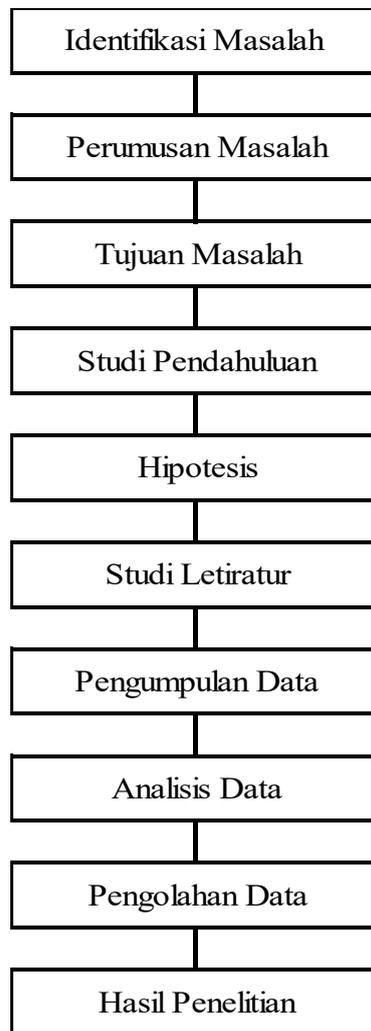
Untuk mengetahui target pada sistem pada *productivity services* diperlukan adanya prediksi. Prediksi target dapat menggunakan teknik *data mining*. *Data mining* bekerja dengan melakukan proses penggalian informasi dari data waktu kerja maupun dari basis data. [1]

Algoritma yang dapat digunakan untuk memprediksi atau mengklasifikasi suatu kejadian dengan pembentukan pohon keputusan antara lain algoritma C4.5, yang merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan oleh J. Ross Quinlan. Algoritma C4.5 merupakan kelompok Algoritma *decision tree*. Algoritma ini mempunyai input berupa *training samples dan samples*. *Training samples* merupakan data contoh yang digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan *samples* merupakan field-field data yang digunakan sebagai parameter dalam klasifikasi data. [2] dengan adanya Algoritma C4.5 untuk mengetahui prediksi di dalam permasalahan sistem yang belum ada.

Solusi dari penelitian untuk mempermudah PT. Sanggar Sarana Baja memberikan *services* yang memuaskan kepada *costumer* agar disediakan menjadi pelanggan setia untuk melakukan *services* dalam jangka waktu tidak tentu dan waktu pengerjaan sesuai dengan kontrak yang telah disepakati. Sedangkan penelitian ini mempunyai alternatif untuk menjaga *productivity* tidak di bawah 80% di seluruh area dalam periode tahun yang telah di rencanakan di awal tahun sebelumnya.

2. Metode Penelitian

Subjek Penelitian yang dilakukan adalah *Productivity services* sebuah pabrik remanufacturing yang terletak di Lingkungan Kawasan Industri Gedung TMT I, lantai 5, No 501 Jl. Cilandak KKO No. 1 Jakarta 12560. Sedangkan objek yang diamati adalah keseluruhan Target Sistem *Productivity Services*. Meliputi waktu kerja site setiap PT.SSB, rumus setiap PT.SSB, alur dan waktu produksi, mengujian menggunakan Algoritma C4.5, hasil *decision tree* dari proses awal hingga akhir untuk mengetahui targetnya. Langkah-langkah alur penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Pengolahan data dilakukan dengan tahapan-tahapan berikut:

1. Identifikasi Masalah

Dalam tahapan ini peneliti mencari permasalahan yang ada pada PT. Sanggar Sarana Baja yaitu hanya pada saat waktu pengerjaan *services*. Pengerjaan *services* pada tepat waktu yang telah di sesuaikan dengan perjanjian yang ada, agar produksi berjalan dengan baik tidak ada kendala. Target yang telah di rencanakan di awal sesuai prediksi akhir.

2. Merumuskan Masalah

Tahap ini merupakan kelanjutan dari penemuan masalah, yaitu merumuskan masalah berdasarkan masalah-masalah yang akan diteliti, tapi tetap dalam batasan ruang lingkup. Dengan demikian rumusan masalah pada penelitian ini adalah:

- Apabila terdapat kesalahan data pada salah satu *database* maka akan mengakibatkan kesalahan pada target sistem *productivity services*.
- Kesalahan dan ketidaktepatan target sistem *productivity services* yang terlalu besar akan menyebabkan kualitas *services* menurun dan dampak besarnya bisa menyebabkan kerugian.
- Masalah target sistem *productivity services* yang sering terjadi biasanya disebabkan karena ketidak-tepatnya perkiraan yang didapatkan.

3. Tujuan Penelitian

Menghasilkan sistem *productivity services* yang efektif khususnya pada PT. Sanggar Sarana Baja dan untuk memudahkan dalam proses *services* serta pencapaian laporan yang cepat, akurat, valid, efektif dan efisien.

4. Studi Pendahuluan

Yaitu dengan melihat literatur dari penelitian sebelumnya Hal ini dilakukan dengan tujuan untuk mengumpulkan informasi-informasi yang berkaitan dengan masalah yang akan diteliti pada PT. Sanggar Sarana Baja sehingga dapat diketahui keadaan dan kedudukan masalah tersebut baik secara teoritis maupun praktis.

5. Hipotesis

Hipotesis merupakan dugaan sementara yang akan dibuktikan kebenaran melalui penelitian di lapangan. Berdasarkan penjelasan di atas hipotesis dari penelitian ini adalah H_0 : tidak ada pengaruh rata-rata dalam menentukan budget dan orang dari jam setiap pekerjaan. H_1 : didapatkan ada pengaruh dari rata-rata 95% menentukan budget dan orang dari jam setiap pekerjaan. Perbulan persite itu 75%, sedangkan target waktu perbulan maksimal 85% dan minimal 65%.

6. Studi Literatur

Dilakukan dengan mempelajari dan memahami teori-teori yang digunakan, yaitu diantaranya Data Mining, Algoritma C4.5 dan metode pengumpulan data. Data-data tersebut dicari dengan cara mengumpulkan literatur, jurnal nasional, *browsing internet* dan bacaan-bacaan yang ada kaitannya dengan topik baik berupa *textbook* atau *paper*.

7. Pengumpulan Data

Tahap ini merupakan cara mengumpulkan data yang dilakukan dengan cara data primer dan Data sekunder.

8. Analisis Data

Sebuah proses untuk memeriksa dan mengevaluasi sistem *productivity services* dengan maksud untuk menemukan informasi yang bermanfaat sehingga dapat memberikan petunjuk bagi peneliti untuk mengambil keputusan terhadap penelitian.

9. Pengolahan Data

Mengenai pengolahan data, peneliti akan menggunakan Data Mining dan metode Algoritma C4.5. Di olah kembali dalam bentuk proses data mining yang terdiri dari *Data Selection* ke *Pre-processing* atau *Cleaning* di lihat datanya ada kosong, salah dalam penulisan salah di perbaiki, data ada yang di hilangkan jika tidak sesuai dengan kriteria, dan melakukan *Transformation ke dalam* normalisasi pada *Data Mining* ke proses penskalaan nilai atribut dari data sehingga bisa jatuh pada range tertentu. *Data mining* proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. *Interpretation/Evaluation* Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya atau tidak. Masuk ke dalam Data mining perumusan dalam Algoritma C4.5, sudah pengetahui hasil dari perhitungan Algoritma C4.5. Tools Terdiri No, *Site*, *Performace*, Efisiensi, dan *Productivity*.

10. Hasil Penelitian

Hasil yang diharapkan dari penelitian ini yaitu mendapatkan hasil yang akurat dan proporsional dalam sistem *productivity services* tersebut.

3. Hasil dan Analisis

3.1 Pengolahan Data

Dalam penelitian ini menggunakan metode Algoritma C4.5 untuk menentukan hasil penelitian dan pembahasan. Algoritma yang dapat digunakan untuk memprediksi atau mengklasifikasi suatu kejadian dengan pembentukan pohon keputusan antara lain algoritma C4.5, yang merupakan salah satu algoritma induksi pohon keputusan yang dikembangkan oleh J.Ross Quinlan

Algoritma C4.5 merupakan kelompok algoritma *decision tree*. Algoritma ini mempunyai input berupa *training samples* dan *samples*. *Training samples* merupakan data contoh yang digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang digunakan sebagai parameter dalam klasifikasi data.

Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam *data mining* adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap-tiap nilai
3. Bagi kasus dalam cabang.

Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Tabel 1 Data Asli Berdasarkan Ketentuan Persyaratan

No	Atribut		Syarat
1	Performance	Melebihi	124-256%
2		Mendekati	107-123%
3		Dibawah	41-102%
4	Efisiensi	Tinggi	87-100%
5		Sedang	78-85%
6		Rendah	43-77%
7	Productivity	Baik	85-199%
8		Buruk	32-82%
9			

Data ini menjelaskan dalam perhitungan manual menggunakan Algoritma C4.5 memilih data secara spesifik, dalam perhitungan terdiri dari 3 kategori untuk persyaratan yang terdiri dari:

1. *Performance* berjumlah 3 yaitu melebihi, mendekati, dan dibawah sesuai dengan persen.
2. *Efisiensi* berjumlah 3 yaitu tinggi, sedang, dan rendah sesuai dengan persen.
3. *Productivity* berjumlah 2 yaitu baik dan buruk sesuai dengan persen.

Tabel 2 Sampel Data Set

No	Site	Performance	Efisiensi	Productivity	Keterangan
1	Balikpapan	Mendekati	Tinggi	Baik	Tercapai
2	Balikpapan	Dibawah	Rendah	Buruk	Tidak Tercapai
3	Balikpapan	Mendekati	Rendah	Buruk	Tidak Tercapai
4	Balikpapan	Dibawah	Rendah	Buruk	Tidak Tercapai
5	Balikpapan	Dibawah	Sedang	Buruk	Tidak Tercapai
6	Balikpapan	Dibawah	Sedang	Buruk	Tidak Tercapai
7	Kassel	Mendekati	Rendah	Buruk	Tidak Tercapai
8	Kassel	Dibawah	Rendah	Buruk	Tidak Tercapai
9	Kassel	Mendekati	Rendah	Buruk	Tidak Tercapai
10	Kassel	Mendekati	Sedang	Baik	Tercapai
11	Kassel	Melebihi	Sedang	Baik	Tercapai
12	Kassel	Mendekati	Sedang	Baik	Tercapai
13	Grassberg	Dibawah	Tinggi	Baik	Tercapai
14	Grassberg	Dibawah	Tinggi	Baik	Tercapai
15	Grassberg	Melebihi	Tinggi	Baik	Tercapai
16	Grassberg	Mendekati	Tinggi	Baik	Tercapai
17	Grassberg	Dibawah	Tinggi	Baik	Tercapai
18	Grassberg	Dibawah	Tinggi	Baik	Tercapai
19	Others	Melebihi	Sedang	Baik	Tercapai
20	Others	Mendekati	Sedang	Baik	Tercapai
21	Others	Melebihi	Sedang	Baik	Tercapai
22	Others	Mendekati	Sedang	Baik	Tercapai
23	Others	Mendekati	Tinggi	Baik	Tercapai
24	Others	Mendekati	Tinggi	Baik	Tercapai
25	Sangatta	Dibawah	Sedang	Baik	Tidak Tercapai
26	Sangatta	Dibawah	Sedang	Buruk	Tidak Tercapai
27	Sangatta	Dibawah	Sedang	Buruk	Tidak Tercapai
28	Sangatta	Dibawah	Sedang	Baik	Tidak Tercapai
29	Sangatta	Dibawah	Tinggi	Baik	Tercapai
30	Sangatta	Dibawah	Sedang	Baik	Tidak Tercapai
31	Kuala Kencana	Dibawah	Rendah	Buruk	Tidak Tercapai
32	Kuala Kencana	Melebihi	Sedang	Baik	Tercapai
33	Kuala Kencana	Dibawah	Sedang	Buruk	Tidak Tercapai
34	Kuala Kencana	Melebihi	Sedang	Baik	Tercapai
35	Kuala Kencana	Dibawah	Tinggi	Baik	Tercapai
36	Kuala Kencana	Mendekati	Rendah	Buruk	Tidak Tercapai
37	Samarinda	Dibawah	Sedang	Buruk	Tidak Tercapai
38	Samarinda	Melebihi	Rendah	Baik	Tercapai
39	Samarinda	Melebihi	Tinggi	Baik	Tercapai
40	Samarinda	Melebihi	Sedang	Baik	Tercapai
41	Samarinda	Melebihi	Sedang	Baik	Tercapai
42	Samarinda	Melebihi	Sedang	Baik	Tercapai

Data sudah di pilih lalu kita olah berdasarkan kriteria. Dari *site*, *performance* dan efisiensi ke dalam bentuk 3 syarat dan *productivity* ke dalam bentuk 2 syarat. Keterangan adalah 3 data dalam persyaratan, misalkan salah satu dari 3 ini tidak tercapai. Sedangkan kategori melebihi, tinggi, dan baik itu adalah tercapai.

Information gain adalah atribut *selection measure* yang digunakan untuk memilih test attribut tiap node dalam *tree*. Pemilihan atribut sebagai simpul, baik simpul akar (*root*) atau simpul internal. didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Perhitungan nilai *gain* digunakan rumus sebagai berikut:

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

Keterangan:

- S : Himpunan kasus
 A : Atribut
 N : Jumlah partisi atribut A
 |S_i| : Jumlah kasus pada partisi ke-i
 |S| : Jumlah kasus dalam S

Entropy dapat dikatakan sebagai kebutuhan bit untuk menyatakan suatu kelas dan juga *entropy* digunakan untuk mengukur ketidaksihan S. Untuk menghitung *entropy* menggunakan rumus:

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

Keterangan:

- S : Himpunan kasus
 A : Fitur
 N : Jumlah Partisi S
 P_i : Proporsi dari S_i terhadap S

Hasil dari perhitungan *gain* dan *entropy* berdasarkan rumus:

1. *Entropy* total: Jumlah kasus = 42
 Tercapai = 25
 Tidak tercapai = 17
 $= \left(-\frac{25}{42} * \log_2 \left(\frac{25}{42} \right) + \left(-\frac{17}{42} * \log_2 \left(\frac{17}{42} \right) \right) = 0,974$
2. Salah satu contoh dari *performance entropy*:
 Jumlah kasus = 42
 Tercapai = 8
 Tidak tercapai = 4
Entropy total = 0,974
 $= \left(-\frac{8}{42} * \log_2 \left(\frac{8}{42} \right) + \left(-\frac{4}{42} * \log_2 \left(\frac{4}{42} \right) \right) = 0,918$
3. Salah satu contoh dari *performance gain*:
Entropy total = 0,974
 Jumlah kasus = 42
Performance sum = 11,12,19
Entropy
 $= (0,974) - \left(\left(\frac{11}{42} * 0 \right) - \left(\left(\frac{12}{42} * 0,918 \right) - \left(\left(\frac{19}{42} * 0,900 \right) \right) \right) = 0,30427089$

Tabel 3 Hasil Sampel Data Set

Jml Kasus	Tercapai	Tidak Tercapai	Entropy Total
42	25	17	0,974

Perhitungan nilai Entropy dan Gain

Analisis Atribut, Nilai, Banyaknya peristiwa, Entropy dan Gain

No	Atribut	Nilai	Sum(Nilai)	Sum(Tercapai)	Sum(Tidak Tercapai)	Entropy	Gain
1	Performance	Melembi	11	11	0	0	
		Mendekati	12	8	4	0,918	
		Dibawah	19	6	13	0,900	
							0,30427089
2	Efisiensi	Tinggi	12	12	0	0	
		Sedang	21	12	9	0,985	
		Rendah	9	1	8	0,503	
							-14,21911587
3	Productivity	Baik	28	25	0	0	
		Buruk	14	3	14	0,476	
							0,81492575

Hasil Dari perhitungan Gain dan Entropy berdasarkan rumus:

1. *Entropy* total:

Jumlah kasus = 42

Tercapai = 0

Tidak tercapai = 14

$$= \left(\left(-\frac{0}{42} \right) * \log_2 \left(\frac{0}{42} \right) + \left(-\frac{14}{42} \right) * \log_2 \left(\frac{14}{42} \right) \right) = 0$$

2. Salah satu contoh dari *performance entropy*:

Jumlah kasus = 42

Tercapai = 0

Tidak tercapai = 4

Entropy total = 0

$$= \left(\left(-\frac{0}{42} \right) * \log_2 \left(\frac{0}{42} \right) + \left(-\frac{4}{42} \right) * \log_2 \left(\frac{4}{42} \right) \right) = 0$$

3. Salah satu contoh dari *performance gain*:

Entropy total = 0

Jumlah kasus = 42

Performance sum = 4, 10

Entropy

$$= (0) - \left(\left(\frac{4}{42} \right) * 0 \right) - \left(\left(\frac{10}{42} \right) * 0 \right) = 0$$

Tabel 4 Sampel Data Set Node 1

No	Site	Performance	Efisiensi	Productivity	Keterangan
1	Balikpapan	Dibawah	Rendah	Buruk	Tidak Tercapai
2	Balikpapan	Mendekati	Rendah	Buruk	Tidak Tercapai
3	Balikpapan	Dibawah	Rendah	Buruk	Tidak Tercapai
4	Balikpapan	Dibawah	Sedang	Buruk	Tidak Tercapai
5	Balikpapan	Dibawah	Sedang	Buruk	Tidak Tercapai
6	Kassel	Mendekati	Rendah	Buruk	Tidak Tercapai
7	Kassel	Dibawah	Rendah	Buruk	Tidak Tercapai
8	Kassel	Mendekati	Rendah	Buruk	Tidak Tercapai
9	Sangatta	Dibawah	Sedang	Buruk	Tidak Tercapai
10	Sangatta	Dibawah	Sedang	Buruk	Tidak Tercapai
11	Kuala Kencana	Dibawah	Rendah	Buruk	Tidak Tercapai
12	Kuala Kencana	Dibawah	Sedang	Buruk	Tidak Tercapai
13	Kuala Kencana	Mendekati	Rendah	Buruk	Tidak Tercapai
14	Samarinda	Dibawah	Sedang	Buruk	Tidak Tercapai

Tabel 5 Hasil Sampel Data Set Node 1

Jml Kasus	Tercapai	Tidak Tercapai	Entropy Total
14	0	14	0

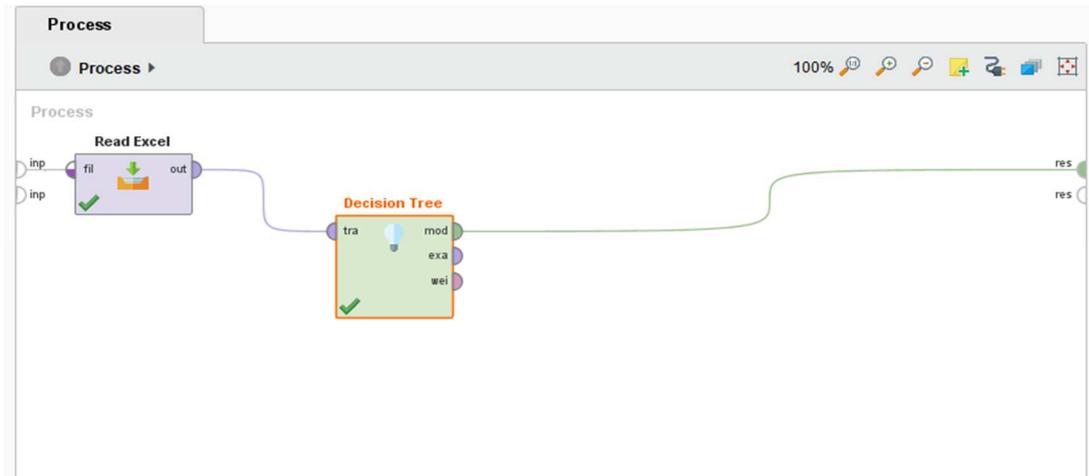
Perhitungan nilai Entropy dan Gain

Analisis Atribut, Nilai, Banyaknya peristiwa, Entropy dan Gain

No	Atribut	Nilai	Sum(Nilai)	Sum(Tercapai)	Sum(Tidak Tercapai)	Entropy	Gain
1	Performance	Mendekati	4	0	4	0,000	0,00000000
		Dibawah	10	0	19	0,000	
2	Efisiensi	Sedang	6	0	6	0,000	0,00000000
		Rendah	8	0	8	0,000	
3	Productivity	Buruk	14	0	14	0,000	0,00000000

Masuk kedalam aplikasi *Rapid Miner*

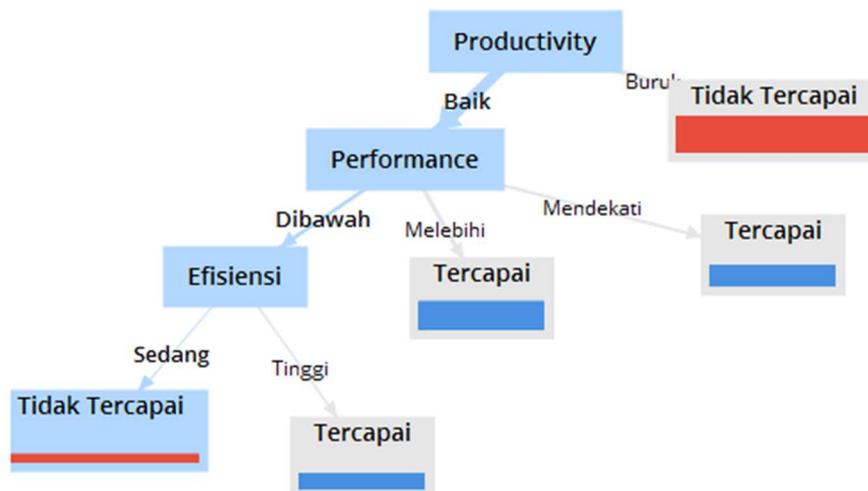
Langkah Pertama adalah kriteria terdapat dalamnya *site*, *performance*, efisiensi, dan *productivity* dan untuk melakukan proses *decision tree* terhadap data *test* melalui *tools* Rapidminer 9.6 yang dimulai dari proses data excel dan *decision tree* seperti gambar 2 berikut ini:



Gambar 2 Pembuatan *Decision Tree*

Hasil perhitungan *Decision Tree*:

Berdasarkan tabel 3 hasil perhitungan ke-1 contoh kasus, diperoleh nilai gain tertinggi adalah **productivity dengan nilai gain 0,974** maka yang menjadi simpul pertama atau simpul akar adalah *productivity*. Kemudian dengan melihat isi atribut *performance*, ketika isi atribut efisiensi ternyata sudah menjadi 2 keputusan yaitu tidak tercapai dan tercapai, maka akar *productivity* dengan isi atribut *performance* dan efisiensi, menjadi simpul daun/simpul keputusan.



Gambar 3 Hasil *Decision Tree*

Tree

```

Productivity = Baik
| Performance = Dibawah
| | Efisiensi = Sedang: Tidak Tercapai {Tercapai=0, Tidak Tercapai=3}
| | Efisiensi = Tinggi: Tercapai {Tercapai=6, Tidak Tercapai=0}
| Performance = Melebihi: Tercapai {Tercapai=11, Tidak Tercapai=0}
| Performance = Mendekati: Tercapai {Tercapai=8, Tidak Tercapai=0}
Productivity = Buruk: Tidak Tercapai {Tercapai=0, Tidak Tercapai=14}
  
```

Gambar 4 Deskripsi *Decision Tree*

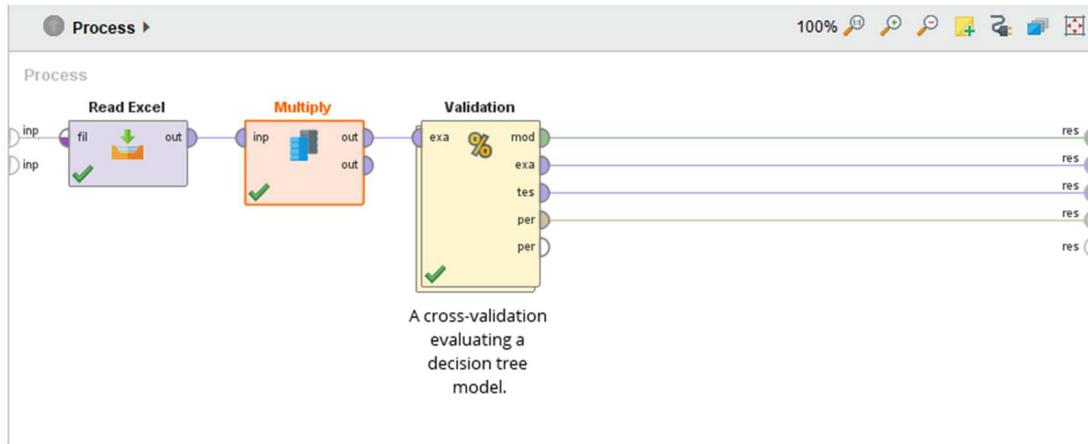
1. IF nilai Productivity = Baik
2. IF nilai Performance = Dibawah
3. IF nilai Efisiensi = Sedang, Tercapai = 0 and Tidak Tercapai = 3
4. IF nilai Efisiensi = Tinggi, Tercapai = 6 and Tidak tercapai = 0
5. IF nilai performance = Melebihi, tercapai = 11 and Tidak tercapai = 0
6. IF nilai Performance = Mendekati, Tercapai = 8 and tidak tercapai = 0
7. IF nilai Produktivity = Buruk, Tercapai = 0 and tidak tercapai 14

3.2 Validasi

Tabel 6 Hasil Data Validasi Menggunakan Excel

No	Performance	Efisiensi	Productivity	Keterangan	Jumlah
1	2	3	2	2	7
2	1	1	1	1	3
3	2	1	1	1	3
4	1	1	1	1	3
5	1	2	1	1	4
6	1	2	1	1	4
7	2	1	1	1	3
8	1	1	1	1	3
9	2	2	1	1	4
10	2	2	2	2	6
11	3	2	2	2	6
12	2	2	2	2	6
13	1	3	2	2	7
14	1	3	2	2	7
15	3	3	2	2	7
16	2	3	2	2	7
17	1	3	2	2	7
18	1	3	2	2	7
19	3	2	2	2	6
20	2	2	2	2	6
21	3	2	2	2	6
22	2	2	2	2	6
23	2	3	2	2	7
24	2	3	2	2	7
25	1	2	2	1	5
26	1	2	1	1	4
27	1	2	1	1	4
28	1	2	2	1	5
29	1	3	2	2	7
30	1	2	2	1	5
31	1	1	1	1	3
32	3	2	2	2	6
33	1	2	1	1	4
34	3	2	2	2	6
35	1	3	2	2	7
36	2	1	1	1	3
37	1	2	1	1	4
38	3	1	2	2	5
39	3	3	2	2	7
40	3	2	2	2	6
41	3	2	2	2	6
42	3	2	2	2	6
rHitung	#DIV/0!	0,87205750	0,89887047	0,89884231	
rTabel	0,51	0,51	0,51	0,51	
V/F	#DIV/0!	V	V	V	

Hasil dari validasi menggunakan Ms. Excel rHitung. Nilai dari *performance* dan jumlah dapat muncul ketika menggunakan nilai validasi 0,5, sedangkan di bawah 0,4 angka tidak muncul. Selanjutnya adalah melakukan proses akurasi terhadap data *test* melalui *tools* Rapidminer 9.6 yang dimulai dari proses data excel, *multiply* dan *cross validation* seperti gambar berikut ini:



Gambar 5 Pembuatan Akurasi

Jumlah prediksi tercapai (pT) sebanyak 25 data, prediksi tidak tercapai (fT) sebanyak 0 data, *true* tidak tercapai (Ttt) sebanyak 2 data dan *true* tidak tercapai (Ttt) sebanyak 15 data. *Accuracy* nilai pada target 95,00% tercapai dan *average* 95,24%

accuracy: 95.00% +/- 15.81% (micro average: 95.24%)

	true Tercapai	true Tidak Tercapai	class precision
pred. Tercapai	25	2	92.59%
pred. Tidak Tercapai	0	15	100.00%
class recall	100.00%	88.24%	

Gambar 6 Hasil Akurasi

PerformanceVector

```

PerformanceVector:
accuracy: 95.00% +/- 15.81% (micro average: 95.24%)
ConfusionMatrix:
True:  Tercapai      Tidak Tercapai
Tercapai:      25      2
Tidak Tercapai: 0      15
precision: 100.00% (positive class: Tidak Tercapai)
ConfusionMatrix:
True:  Tercapai      Tidak Tercapai
Tercapai:      25      2
Tidak Tercapai: 0      15
recall: 90.00% +/- 31.62% (micro average: 88.24%) (positive class: Tidak Tercapai)
ConfusionMatrix:
True:  Tercapai      Tidak Tercapai
Tercapai:      25      2
Tidak Tercapai: 0      15
AUC (optimistic): 1.000 +/- 0.000 (micro average: 1.000) (positive class: Tidak Tercapai)
AUC: 0.500 +/- 0.000 (micro average: 0.500) (positive class: Tidak Tercapai)
AUC (pessimistic): 0.900 +/- 0.316 (micro average: 0.900) (positive class: Tidak Tercapai)
    
```

Gambar 7 Deskripsi Akurasi

4. Kesimpulan

Berdasarkan penelitian sistem pada *productivity services* terhadap target di PT. Sanggar Sarana Baja Jakarta, dapat di simpulkan bahwa:

1. Permasalahan perancangan target sistem pada *productivity services* di PT. Sanggar Sarana Baja dapat menyelesaikan menggunakan proses *data mining*, yaitu Algoritma C4.5.
2. Didapatkan perancangan target untuk tahun 2021 di PT. Sanggar Sarana Baja sesuai dengan metode Algoritma C4.5 adalah 95%.

H_0 tidak ada pengaruh rata-rata dalam menentukan *budget* dan orang dari jam setiap pekerjaan sedangkan H_1 didapatkan ada pengaruh dari rata-rata 95% menentukan *budget* dan orang dari jam setiap pekerjaan. Perbulan persite adalah 75%, sedangkan target waktu perbulan maksimal 85% dan minimal 65%.

Referensi

- [1] Yulia and N. Azwanti, "Data Mining Prediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam Dengan Menggunakan Algoritma C4.5," *Semin. Nas. Ilmu Sos. dan Teknol.*, vol. 1, no. 1, pp. 175–180, 2018.
- [2] R. Novita, "Teknik Data Mining : Algoritma C 4 . 5," pp. 1–12, 2016.