

# Perbandingan Hasil Penggunaan Metode Decision Tree dan Random Tree Pada Data Training Aplikasi Pencarian Tukang

Teguh Prasandy<sup>1</sup>, Kartika Nurkhasanah<sup>2</sup>, Meity Puspita Sari<sup>3</sup>, Tria Rahmatul Fazry<sup>4</sup>

<sup>1</sup> Information System, BINUS Online Learning, Bina Nusantara University, Jakarta, Indonesia, 11480  
teguh.prasandy@binus.edu

<sup>2</sup> Information System, BINUS Online Learning, Bina Nusantara University, Jakarta, Indonesia, 11480  
tikafarhika@gmail.com

<sup>3</sup> Information System, BINUS Online Learning, Bina Nusantara University, Jakarta, Indonesia, 11480  
meity.psari05@gmail.com

<sup>4</sup> Information System, BINUS Online Learning, Bina Nusantara University, Jakarta, Indonesia, 11480  
triarhm@gmail.com

Diterima 28 Mei 2019

Disetujui 20 Desember 2019

**Abstract**—Pembuatan Aplikasi pencarian tukang dengan memanfaatkan pohon keputusan yang sesuai dengan pelayanan yang diminta oleh pelanggan dan sesuai dengan posisi tukang, dengan memanfaatkan *data training* dari data transaksi dari transaksi yang telah ada sebagai data awal dalam pembuatan pohon keputusan dan menggunakan *data mining* sebagai data analisis. Tujuan penelitian ini adalah mengetahui perbandingan hasil *tree* antara *decision tree* dengan *random tree* dan merekomendasikan metode *tree* yang sesuai dengan aplikasi pencarian tukang. *Decision tree* memiliki parameter dan lebih kemungkinan lebih banyak dibandingkan *random tree* serta memiliki value 67% untuk *max depth* sebesar 20, memiliki value 77% untuk *confidence* sebesar 0,25 dan memiliki 71% untuk *minimal gain* 0,1. Sedangkan untuk *random tree* tidak memiliki value tersebut, sehingga *decision tree* lebih direkomendasikan untuk aplikasi pencarian tukang.

**Keyword**—Metode, Decision Tree, Random Tree, Data Training, Aplikasi Pencarian Tukang

## I. PENDAHULUAN

Paper ini merupakan pengembangan dari skripsi mahasiswa dengan judul Desain Aplikasi “Helptask” Pada Modul Pemesanan Helper Di Pt Horison Selaras Indonesia. Data saat ini mudah sekali mendapatkannya melalui berbagai macam sumber seperti dari penelitian, dari database aplikasi baik aplikasi berbasis *desktop* maupun berbasis *web*, dari *social media*, dari transaksi yang terjadi pada aplikasi-aplikasi *android* dsb. Semua data tersebut jumlahnya melimpah, pertambahannya sangat cepat, volumenya makin membesar (konsep big data). Sehingga data merupakan sumber daya baru yang bisa digunakan dalam meraih keuntungan finansial, promosi barang ke konsumen, hingga mengetahui kebiasaan-kebiasaan

yang dilakukan konsumen. Salah satu contoh perusahaan berbasis aplikasi yang memanfaatkan data tersebut adalah Go-Jek[1], diantara untuk meningkatkan strategi fitur mana yang diprioritaskan dan produk apa yang akan dibuat, hal ini merupakan konsep dari *Decision Support System*. Go-Jek menyimpan datanya dalam *repository* yang dimilikinya untuk menemukan pengetahuan, biasa disebut sebagai *Knowledge Database Discovery* (KDD). KDD merupakan bagian dari data mining [2].

Dengan adanya data yang melimpah tersebut maka untuk mendapatkan *data training* yang akan digunakan sebagai dasar pengambilan keputusan menjadi sangat mudah. *Data training* adalah seperangkat data awal yang digunakan untuk membantu program dalam ekstraksi data untuk mendapatkan informasi dan algoritma pembentukan keputusan[3][4]. *Data training* tersebut biasanya dianalisis untuk menghasilkan pengetahuan baru dan informasi yang lebih bermanfaat menggunakan *data mining*.

*Data mining* adalah proses menemukan korelasi, pola, dan *tren* baru yang bermakna memilah-milah sejumlah besar data yang disimpan dalam *repository*, menggunakan pengenalan pola teknologi serta teknik statistik dan matematika[5]. Pengertian lainnya data mining didefinisikan sebagai proses menganalisis basis data besar, biasanya gudang data atau internet, untuk menemukan informasi baru, pola dan perilaku tersembunyi [6]. Sedangkan menurut [7] *data mining* merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Dari ketiga pengertian diatas dapat

disimpulkan bahwa *data mining* adalah proses analisis untuk menemukan korelasi, pola, tren, dan perilaku tersembunyi serta informasi baru dari data yang tersimpan dalam *repository* perusahaan menggunakan teknik matematika dan statistika.

Perkembangan aplikasi pelayanan berbasis *mobile* yang semakin meningkat menyebabkan adanya keinginan untuk membuat aplikasi yang memberikan pelayanan pada masyarakat seperti tukang servis *air conditioner* (AC), tukang listrik (*electrician*), tukang bersih-bersih plafon (*Ceiling Service*), tukang cat (*Room Painter*), tukang pipa (*Plumber*) dan layanan *laundry* yang tersedia dalam aplikasi yang dinamakan sebagai aplikasi pencarian tukang. Aplikasi ini memanfaatkan metode *decision tree* dalam mencari tukang yang sesuai dengan pelayanan yang diminta oleh pelanggan yang sesuai dengan posisi tukang, dengan memanfaatkan *data training* dari data transaksi dari transaksi yang telah ada sebagai data awal dalam pembuatan pohon keputusan dan menggunakan *data mining* sebagai analisis data.

Di dalam *data mining* terdapat beberapa macam metode yang bisa digunakan untuk pengambilan keputusan diantaranya adalah *decision tree* dan *random tree*. Maka diperlukan penelitian yang membandingkan dan merekomendasikan metode mana yang digunakan apakah *decision tree* atau *random tree*? Sehingga aplikasi tukang dapat menggunakan salah satu dari metode tersebut.

Dari uraian diatas maka judul penelitian ini adalah Perbandingan Hasil Penggunaan Algoritma *Decision Tree* dan *Random Tree* Pada *Data Training* Aplikasi Pencarian Tukang. Sedangkan tujuan penelitian ini adalah mengetahui perbandingan hasil tree antara *decision tree* dengan *random tree* dan merekomendasikan metode yang sesuai dengan aplikasi pencarian tukang.

*Decision Tree* adalah pohon seperti kumpulan *node* yang dimaksudkan untuk membuat keputusan tentang nilai afiliasi ke kelas atau perkiraan nilai target numerik [8]. Setiap *node* mewakili aturan pemisahan untuk satu Atribut tertentu. Untuk klasifikasi aturan ini memisahkan nilai-nilai milik kelas yang berbeda, untuk regresi itu memisahkan mereka untuk mengurangi kesalahan secara optimal untuk kriteria parameter yang dipilih[9].

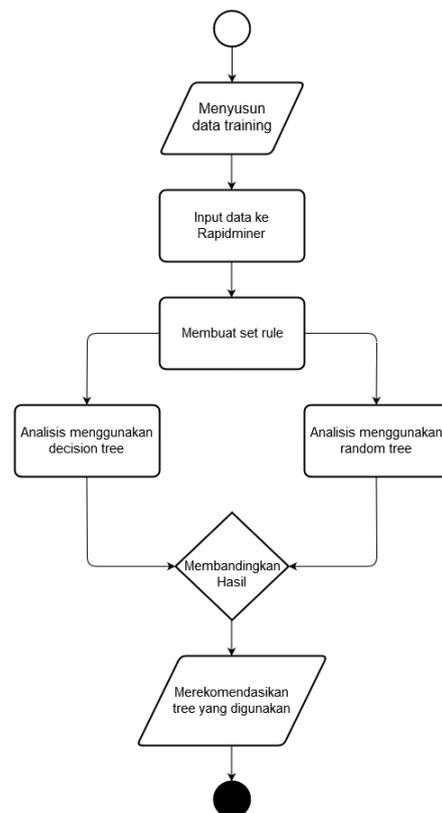
*Random Tree* berfungsi persis seperti operator *Decision Tree* dengan satu pengecualian: untuk setiap pemisahan hanya subset atribut acak yang tersedia. Representasi data sebagai *Tree* memiliki keunggulan dibandingkan dengan pendekatan lain yang bermakna dan mudah diinterpretasikan. Tujuannya adalah untuk membuat model klasifikasi yang memprediksi nilai label berdasarkan beberapa atribut input dari *ExampleSet*. Setiap simpul interior pohon sesuai dengan salah satu atribut input. Jumlah tepi sebuah simpul interior sama dengan jumlah nilai yang

mungkin dari atribut input yang sesuai. Setiap simpul daun mewakili nilai label yang diberi nilai atribut input yang diwakili oleh jalur dari akar ke daun[10].

Untuk lebih mudah dalam proses pembuatan *tree* digunakan *rapidminer*. *Rapidminer* adalah aplikasi populer gratis yang digunakan dalam data mining dengan memiliki tampilan grafis *user interface*, terdapat banyak algoritma, dan berbagai fitur visualisasi data[2] sedangkan menurut [11] *RapidMiner*, platform analitik prediktif *open source* # 1 di industri, mengganggu industri dengan memberdayakan perusahaan untuk memasukkan analitik prediktif dalam setiap proses bisnis — menutup lingkaran antara wawasan dan tindakan. Solusi *RapidMiner* yang mudah membuat analitik prediktif secepat kilat untuk analisis modern saat ini, secara radikal mengurangi waktu untuk menggali peluang dan risiko. *RapidMiner* menghadirkan keahlian mengubah permainan dari komunitas analitik prediktif terbesar di dunia. Maka *rapidminer* adalah aplikasi #1 dalam digunakan dalam data mining yang dapat membuat analitik prediktif secepat kilat dan menghasilkan tampilan grafis *user interface*, terdapat banyak algoritma, dan berbagai fitur visualisasi data.

## II. METODE PENELITIAN

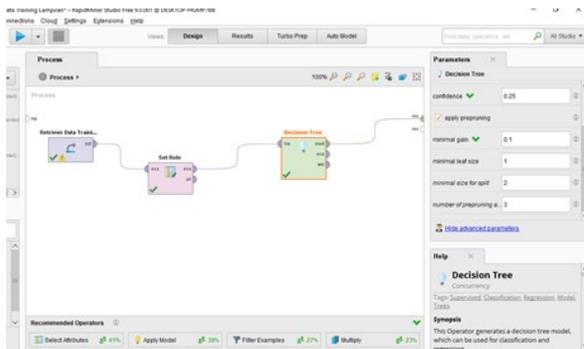
Metode yang digunakan dalam penelitian ini dijabarkan pada gambar 1.



Gambar 1. Flowchart Penelitian

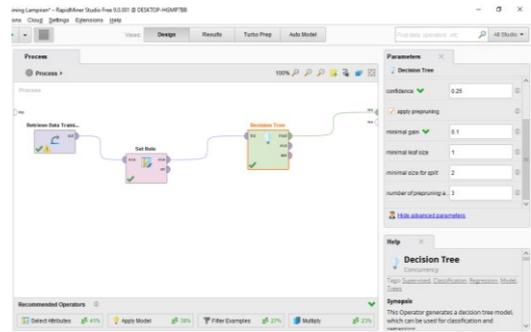
Penjelasan gambar 1, sebagai berikut:

1. Menyusun data training dari transaksi aplikasi tukang sebelumnya.
2. Memasukkan data ke aplikasi *Rapidminer*
3. Membuat *set rule* dengan ketentuan:
  - a. *Service order* (label)
  - b. *City helper* (regular)
  - c. *Service helper* (regular)
  - d. *Available* (prediksi)
4. Membuat *decision tree* menggunakan data training menggunakan analisis *Decision Tree* dapat dilihat pada gambar 1, dengan menggunakan parameter:
  - a. *Criteria gain ratio*
  - b. *Max depth* 20
  - c. *Confidence* 0,25
  - d. *Minimal gain* 0,1
  - e. *Minimal leaf size* 1
  - f. *Minimal size for split* 2
  - g. *Number prepruning* 3



Gambar 2. *Design Decision Tree* pada *rapidminer*

5. Membuat *decision tree* menggunakan data training menggunakan analisis *random tree* dapat dilihat pada gambar 2, dengan menggunakan parameter:
  - a. *Criteria gain ratio*
  - b. *Max depth* 20
  - c. *Confidence* 0,25
  - d. *Minimal gain* 0,1
  - e. *Minimal size for split* 2
  - f. *Number of preprun* 3



Gambar 3. *Design Random Tree* pada *rapidminer*

6. Membandingkan hasil *tree* dengan melihat *node tree* yang mewakili seluruh transaksi.
7. Merekomendasikan *tree* yang akan digunakan oleh aplikasi pencarian tukang.
8. *Data training* yang digunakan

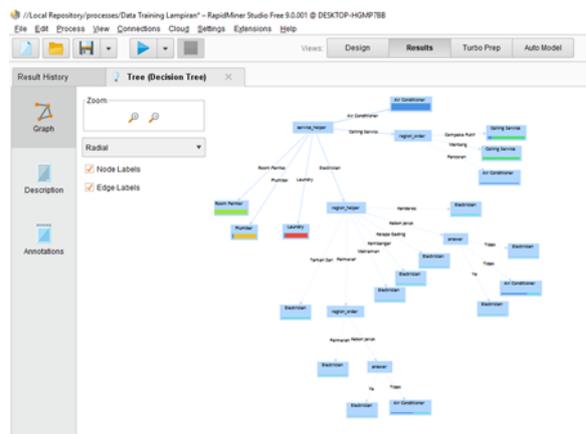
Tabel 1. Data Training

service_order	service_helper	region_order	region_helper	city_order	city_helper	available	answer
Air Conditioner	Air Conditioner	Kebon Jeruk	Kebon Jeruk	Jakarta Barat	Jakarta Barat	Ya	Ya
Air Conditioner	Air Conditioner	Kebon Jeruk	Palmerah	Jakarta Barat	Jakarta Barat	Ya	Ya
Electrician	Electrician	Kebon Jeruk	Kebon Jeruk	Jakarta Barat	Jakarta Barat	Ya	Ya
Electrician	Electrician	Kebon Jeruk	Palmerah	Jakarta Barat	Jakarta Barat	Ya	Ya
Ceiling Service	Ceiling Service	Cempaka Putih	Cempaka Putih	Jakarta Pusat	Jakarta Pusat	Ya	Ya
Ceiling Service	Ceiling Service	Cempaka Putih	Menteng	Jakarta Pusat	Jakarta Pusat	Ya	Ya
Room Painter	Room Painter	Pancoran	Pancoran	Jakarta Selatan	Jakarta Selatan	Ya	Ya
Plumber	Plumber	Matraman	Matraman	Jakarta Timur	Jakarta Timur	Ya	Ya
Plumber	Plumber	Matraman	Jatinegara	Jakarta Timur	Jakarta Timur	Ya	Ya
Laundry	Laundry	Kelapa Gading	Kelapa Gading	Jakarta Utara	Jakarta Utara	Ya	Ya

### III. HASIL DAN PEMBAHASAN

#### A. Hasil

Hasil analisis data training menggunakan rapidminer menggunakan metode *decision tree* dapat dilihat pada gambar 3.

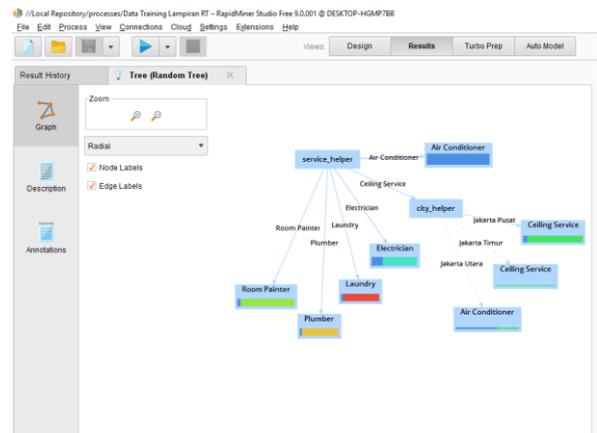


Gambar 4. Hasil Analisis menggunakan *Decision Tree*

Dari gambar diatas dapat dijelaskan sebagai berikut:

1. Jika *service helper* dengan pelayanan *air conditioner*, *room painter*, *plumber*, *laundry availability* iya yang berarti tersedia.
2. Jika *service helper* dibagi berdasarkan *region order* pada daerah menteng dan cempaka putih dengan pelayanan *ceiling service*, *availability* iya, hal ini berarti tukang tersedia pada daerah tersebut. Jika *region order* berada di daerah pancoran dengan pelayanan *air conditioner*, *availability* iya, berarti tukang tersedia di daerah tersebut.
3. Jika *service helper* dibagi berdasarkan *region helper* taman sari, matraman, Kembangan, kelapa gading, dengan *service electrician* maka *availability* iya, hal ini berarti tukang tersedia.
4. Jika *service helper* dibagi berdasarkan *region helper* pada daerah kebon jeruk dengan pelayanan jawaban iya dan pelayanan *electrician* maka tukang selalu bersedia melaksanakan pelayanan, tetapi ada sebagian tukang yang tidak bersedia melaksanakan, sedangkan dengan jawaban tidak berarti tukang menolak memberikan pelayanan.
5. Jika *service helper* dibagi berdasarkan *region helper* dan *region order* pada daerah palmerah dengan pelayanan *electrician*, sedangkan pada daerah kebon jeruk dengan melihat jawaban iya maka pelayanan yang diberikan berupa *electrician*, tetapi pelayanan *air conditioner* pasti ditolak.

Sedangkan dengan menggunakan *random tree* dapat dilihat pada gambar 4.



Gambar 5. Hasil Analisis menggunakan *Random Tree*

Dari gambar diatas dapat dijelaskan sebagai berikut:

1. Jika *service helper* dengan pelayanan *room painter*, *plumber*, *laundry*, *electrician* dan *air conditioner* maka *availability* iya, berarti tukang tersedia.
2. Jika *service helper* dengan pelayanan *ceiling service* dengan *city helper* Jakarta pusat dan Jakarta timur maka *availability* iya, maka tukang tersedia, sedangkan pada Jakarta utara tidak ada pelayanan dikarenakan adanya ketidaksinkronan data *service helper* dengan *service order*.

#### B. Pembahasan

Dengan melihat hasil analisis menggunakan rapidminer dalam sekilas saja *random tree* lebih sederhana dibandingkan *decision tree*. Pada *random tree* hanya terdapat dua kemungkinan yaitu (1). *Service helper* dan layanannya, (2). *Service helper* dengan pelayanan *ceiling service*, kemudian melihat *city helper* pada Jakarta Jakarta pusat dan timur. Sedangkan hasil analisis menggunakan rapidminer pada *decision tree* terdapat 5 kemungkinan, sehingga *decision tree* memiliki lebih banyak kemungkinan dan lebih terperinci. Dengan melihat lebih dalam lagi ke dalam penyusunan parameter *decision tree* memiliki value 67% untuk *max depth* sebesar 20, memiliki value 77% untuk *confidence* sebesar 0,25 dan memiliki 71% untuk *minimal gain* 0,1. Sedangkan untuk *random tree* tidak memiliki value tersebut.

Dengan kelengkapan parameter dan perincian tiap *tree* berarti metode *decision tree* lebih baik dibandingkan *random tree*, sehingga rekomendasi untuk aplikasi pencarian tukang adalah *decision tree*.

## IV. SIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah *decision tree* memiliki parameter dan lebih kemungkinan lebih banyak dibandingkan *random tree*, sehingga *decision tree* lebih direkomendasikan untuk aplikasi pencarian tukang.

## DAFTAR PUSTAKA

- [1] Kumparan Tech, "Big Data Bikin Go-Jek Bisa Ketahui Kebiasaan Pengguna - kumparan.com," 2017. [Online]. Available: <https://kumparan.com/@kumparantech/big-data-bikin-go-jek-bisa-ketahui-kebiasaan-pengguna>. [Accessed: 01-May-2019].
- [2] E. Turban, R. Sharda, D. Delen, J. E. Aronson, T.-P. Liang, and T.-P. Liang, *Business Intelligence and Analytics*, vol. 10, no. 4. Person, 2014.
- [3] R. A. Anggraini, G. Widagdo, A. S. Budi, and M. Qomaruddin, "Penerapan Data Mining Classification untuk Data Blogger Menggunakan Metode Naïve Bayes," *J. Sist. dan Teknol. Inf.*, vol. 7, no. 1, p. 47, 2019.
- [4] R. Adiputra, N. Made, and S. Iswari, "Personality Based Lipstick Color Recommender System using K-Nearest Neighbors Algorithm," vol. VI, no. 1, pp. 55–59, 2019.
- [5] Daniel T. Larose, *DISCOVERING KNOWLEDGE IN DATA An Introduction to Data Mining*, 1st ed. Hoboken, New Jersey: John Wiley & Sons, Inc.
- [6] A. R. Al-azmi, "Data , Text , and Web Mining for Business Intelligence : a Survey," *Int. J. Data Min. Knowl. Manag. Process.*, vol. 3, no. 2, pp. 1–21, 2013.
- [7] V. S. Moertini, "Data Mining Sebagai Solusi Bisnis," *Integral*, vol. 7, no. 1, pp. 44–56, 2002.
- [8] S. H. David Hartanto Kamagi, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Tek. Inform.*, vol. VI, no. 1, pp. 15–20, 2014.
- [9] Rapidminer, "Decision Tree - RapidMiner Documentation." [Online]. Available: [https://docs.rapidminer.com/8.1/studio/operators/modeling/predictive/trees/parallel\\_decision\\_tree.html](https://docs.rapidminer.com/8.1/studio/operators/modeling/predictive/trees/parallel_decision_tree.html). [Accessed: 28-May-2019].
- [10] Rapidminer, "Random Tree - RapidMiner Documentation." [Online]. Available: [https://docs.rapidminer.com/8.1/studio/operators/modeling/predictive/trees/random\\_tree.html](https://docs.rapidminer.com/8.1/studio/operators/modeling/predictive/trees/random_tree.html). [Accessed: 28-May-2019].
- [11] about rapidminer, "RapidMiner Embraces its Community and Open Source Culture." [Online]. Available: <https://rapidminer.com/news/rapidminerembracesitscommunity/>. [Accessed: 28-May-2019].