# IMPLEMENTASI MODEL *DECISION TREE* MENGGUNAKAN ALGORTIMA *ID3* UNTUK MENETAPKAN STRATEGI PEMASARAN UNIVERSITAS BINA DARMA PALEMBANG

**Putri Inda Sari[(1)], Muhammad Nasir[(2)]**
Fakultas Ilmu Komputer, Universitas Bina Darma
Email: 151420164@student.binadarma.ac.id[(1)], nasir@binadarma.ac.id[(2)]

**ABSTRAK**

Universitas Bina Darma Palembang merupakan Universitas swasta terbaik di Sumatera Selatan, oleh sebab itu minat calon mahasiswa semakin tinggi terhadap Universitas. Dengan demikian pihak Universitas harus bisa mempertahankan dan menciptakan citra yang baik di masyarakat. Konsep pohon keputusan merupakan salah satu konsep teori graf yang paling penting, algoritma *ID3* memperlihatkan pemakaian pohon keputusan untuh memudahkan pengambilan keputusan strategi yang harus dipertahankan. *Tools RapidMiner* menggunakan algoritma *ID3* berhasil mendapatkan keputusan menentukan lokasi paling sedikit peminat pada tahun 2017 terdapat di Lahat, Banyuasin dan Pali dengan hasil akurasi kota asal tahun 2017 sebesar 87.39%, Tahun 2017 memperlihatkan lokasi paling sedikit peminatan tersebut berada pada Ogan Ilir, Lahat dan Musi Banyuasin dengan hasil akurasi sebesar 85.77%, Tahun 2018 memperlihatkan lokasi peminat paling sedikit berada di wilayah Prabumulih, Pagar Alam dan Musi Banyuasin dengan hasil akurasi sebesar 86.27%. Sedangkan untuk sumber informasi yang paling diminati oleh mahasiswa ataupun masyarakat luas pada umumnya adalah bersumber dari internet dengan rata-rata hasil akurasi sebesar 83.12%. Hasil keputusan tersebut diharapkan dapat dilakukan lebih pemasaran pada lokasi sedikit peminat tersebut.

**Kata Kunci:** *Tools RapidMiner*, Algoritma *ID*, Strategi Pemasaran.

## 1. PENDAHULUAN

### 1.1 Latar Belakang

Setiap perguruan tinggi swasta akan menghadapi persaingan yang semakin ketat. Kondisi ini yang menuntut setiap perguruan tinggi untuk selalu memperhatikan dan memenuhi kebutuhan, keinginan konsumen yaitu calon mahasiswa dan orang tuanya secara lebih memuaskan dibanding perguruan tinggi lain.

Dalam perencanaan kegiatan promosi ini guna meningkatkan minat calon mahasiswa untuk mendaftar di Universitas Bina Darma, Unit Pemasaran Universitas Bina Darma membutuhkan masukan guna meningkatkan minat lulusan Sekolah Menengah Atas (SMA) atau setara yang ada di kota Palembang khususnya dan di Sumatera Selatan pada umumnya untuk mejadi Mahasiswa Universitas Bina Darma, hal ini dikarenakan kegiatan promosi yang terdiri dari kegiatan pemasangan iklan, penyebaran brosur dan promosi dengan mengundang sekolah masih belum menyebar luas kedaerah yang tingkat peminatnya masih rendah sehingga masyarakat yang ada didaerah atau dipedesaan masih belum banyak yang mengetahui tentang Universitas Bina Darma. Dengan demikian, Unit Pemasaran Universitas Bina Darma Palembang perlu untuk

menggali atau mencari informasi dari data pendaftaran tahun-tahun sebelumnya sebagai bahan evaluasi dan analisa untuk mendukung strategi promosi tahun berikutnya.

Untuk pencapaian target bisa dikelompokkan sumber data PMB dari tahun sebelumya sehingga dapat dibuat suatu metode yang tepat untuk mempercepat pencapaian target. Dalam pengelolaan data tersebut dibutuhkan sebuah metode yang bisa digunakan untuk menggali informasi yang tersembunyi dari sebuah data. Metode tersebut adalah *data mining*. *Data mining* akan melakukan proses analisis data untuk menemukan pola atau aturan tersembunyi dalam lingkup himpunan data tersebut. Pada studi kasus ini, analisis *data mining* dilakukan dengan metode pohon keputusan yang mengunakan algoritma *Intrative Dichotomizes 3*. Pemanfaatan struktur pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan.

Berdasarkan latar belakang masalah diatas, maka penulis mengambil judul "**Implementasi Model *Decision Tree* Menggunakan Algoritma ID3 untuk Menetapkan Strategi Pemasaran di Universitas Bina Darma**".

## 1.2 Rumusan Masalah

Permasalahan yang diangkat adalah bagaimana cara menetapkan strategi pemasaran dengan mengidentifikasi minat dari data calon mahasiswa baru yang akan diterima pada Universitas Bina Darma?.

## 1.3 Batasan Masalah

Agar pembahasan tidak menyimpang dari permasalahan dan lebih terarah dari yang telah direncanakan maka penulis membuat batasan permasalahan, yaitu:

a. Penelitian ini hanya menentukan dan memprediksi daerah mana yang paling banyak minatnya terhadap Universitas Bina Darma Palembang.
b. Pada penelitian ini menggunakan model *Decision Tree* dengan algoritma ID3 *(Iterative Dichotomizes 3).*
c. Data yang digunaan dalam penelitian ini adalah data penerimaan mahasiswa baru selama 2 tahun terakhir yaitu pada tahun 2017 dan tahun 2018.

## 1.4 Tujuan dan Manfaat Penelitian

### 1.4.1 Tujuan Penelitian

Tujuan penelitian ini adalah membuat suatu pohon keputusan yang dapat digunkan untuk menetapkan strategi pemasaran yang paling efektif guna untuk pencapaian target jumlah mahasiswa baru di Universitas Bina Darma.

### 1.4.2 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Dapat memberikan solusi untuk menetapkan strategi pemasaran dari data penerimaan mahasiswa baru.
2. Hasil penelitian ini bisa dijadikan tolak ukur untuk kinerja pencapaian target penerimaan mahasiswa baru yang akan datang.

## 1.5 Metodologi Penelitian

### 1.5.1 Waktu dan Tempat Penelitian

Lokasi penelitian ini dilakukan pada Universitas Bina Darma Palembang di Jl. Jendral Ahmad Yani No. 3, 9/10 Ulu, Seberang Ulu I, Kota Palembang, Sumatera Selatan 30264. Lamanya waktu yang dibutuhkan untuk menyelesaikan penelitian ini yaitu selama 6 (Enam) bulan yang dimulai pada bulan September 2018 s/d Februari 2019.

### 1.5.2 Data Penelitian

Berdasarkan sumbernya, data penelitian dapat dikelompokkan dalam dua jenis yaitu data primer dan data sekunder.

a. Data primer merupakan data yang didapat / dikumpulkan oleh peneliti dari tempat penelitian yaitu Universitas Bina Darma Palembang. Data yang digunakan adalah

data penerimaan mahasiswa baru pada tahun 2017 dan tahun 2018 di Universitas Bina Darma Palembang.

b. Data Sekunder adalah data yang diperoleh selain dari Universitas Bina Darma Palembang atau dikumpulkan dari berbagai sumber yang telah ada. Data sekunder dapat diperoleh dari berbagai sumber seperti buku, laporan, jurnal, atau situs internet.

### 1.5.3 Metode Penelitian

Penelitian yang digunakan penulis dalam menyusun proposal skripsi ini menggunakan metode deskriptif, yaitu suatu metode dalam meneliti suatu objek, suatu metode yang mengemukakan masalah dengan mengumpulkan data-data yeng bertujuan untuk mendapatkan gambaran yang jelas mengenai suatu keadaan dengan cara menyajikan, mengumpulkan dan menganalisis data tersebut sehingga menjadi informasi baru yang dapat digunakan untuk menganalisa dan mengambil kesimpulan mengenai masalah yang sedang diteliti.

### 1.5.4 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini adalah:

1. Studi pustaka *(literature)*
   Studi literatur merupakan pengumpulan data dengan membaca buku-buku referensi, jurnal-jurnal serta sumber-sumber lainya yang terkait dengan penelitian ini.
2. Metode Wawancara (*Interview)*
   Merupakan teknik pengambilan data di mana peneliti langsung melakukan wawancara kepada unit pemasaran Universitas Bina Darma Palembang untuk menggali informasi dari data penerimaan mahasiswa baru.
3. Melakukan Observasi pada data PMB
   Observasi dilakukan dengan mengadakan pengamatan langsung terhadap objek yang diteliti dalam penelitian ini pengamatan yang dilakukan dengan melihat langsung data PMB dari dua tahun terakhir untuk mempelajari data calon mahasiswa baru.

### 1.5.5 Teknik Analisis

Adapun untuk menganalisis data dalam penerapan data mining ini menggunakan proses data mining atau *Knowledge Discovert in Database* (KDD), yaitu:

a. Data *Cleaning* (Pembersih Data)
b. Data *Integration* (Integrasi Data)
c. Data *Selection* (Seleksi Data)
d. Data *Transforation* (Transformasi Data)
e. Data Mining
f. *Pattern Evaluation* (Evaluasi Pola)
g. *Knowledge Presentation* (Presentasi Pengetahuan)

## 2. Tinjauan Pustaka

### 2.1 Data Mining

Data mining, sering juga disebut *Knowledge Discovery in Database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari data mining ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan(Vulandari, 2017).

Jadi dapat penulis simpulkan bahwa, *data mining* atau sering dikenal dengan *knowledge discovery in database* adalah proses untuk mengekstrak data untuk menemukan pola-pola yang penting atau menarik yang belum diketahui secara manual dari basis data yang besar untuk mendapatkan sebuah keputusan.

### 2.2 Proses atau tahapan *Knowledge Discovery in Database* (KDD)

Menurut Indrajani, S.Kom., MM. (2014:322) dalam bukunya berjudul *Database Systems Case Study All In One* menjelaskan bahwa "Fase-fase dimulai dari data mentah yang berakhir

dengan pengetahuan atau informasi yang telah diolah, yang didapatkan sebagai hasil dari tahapan-tahapan berikut:

a. Data C*leaning*
   Ini adalah sebuah fase dimana data-data tidak lengkap, mengandung error dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih dan relevan dapat digunakan untuk diproses ulang untuk penggalian pengetahuan (*discovery knowledge*).
b. Data *Integration*
   Pada tahap ini terjadi integrasi data, dimana sumber-sumber data yang berulang (*multiple* data), *file-file* yang berulang (*multiple file*), dapat dikombinasikan dan digabungkan ke dalam suatu sumber.
c. Data S*election*
   Pada langkah ini, data yang relevan terhadap analisis dapat dipilih dan diterima dari koleksi data yang ada.
d. Data *Transformation*
   Juga dikenal sebagai data *consolidation.* Pada tahap ini, dimana data-data yang telah terpilih, ditransformasikan ke dalam bentuk-bentuk yang cocok untuk prosedur penggali (mining *porcedure*) dengan cara melakukan normalisasi dan agregasi data.
e. Data Mining
   Tahap ini adalah tahap yang paling penting, dengan menggunakan teknik-teknik yang diaplikasikan untuk mengekstrak pola-pola potensial yang berguna.
f. *Pattern Evaluation*
   Pada tahap ini, pola-pola menarik dengan jelas merepresentasikan pengetahuan telah diidentifikasi berdasarkan measure yang telah diberikan.
g. *Knowledge Presentation*
   Ini merupakan tahap terakhir, dimana pengetahuan yang telah ditemukan secara visual ditampilkan kepada user. Tahap ini menggunakan teknik visualisasi untuk membantu user dalam mengerti dan menginterpretasikan hasil dari data mining.

## 2.3 Algoritma *ID3*

Algoritma ID3 merupakan singkatan dari *Intrative Dichotomiser 3* atau *Induction of Decision Tree*. Algoritma ID3 dicetuskan pertama kali pada akhir decade 70-an oleh Ross Quinian. Algoritma ID3 bertujuan untuk medapatkan pohon keputusan yang terbaik. Algoritma ini dapat menghasikan decision tree sederhana yang terbaik, yang konsisten dari sekumpulan data. Mekanisme konstruksi dilakukan secara *top-down* dengan memilih atribut mana yang harus diperiksa dapat *root* dari decision tree(Vulandari, 2017).

## 2.4 Persamaan Metode ID3

Metode pohon keputusan memiliki beberapa kelebihan bila dibanding dengan metode lain diantaranya: (1) Mengubah daerah pengambilan keputusan menjadi lebih spesifik. (2) Mengeliminasi perhitungan yang diperlukan, data pelatihan diuji berdasarkan kriteria atau kelas tertentu. (3) Fleksibel untuk memilih fitur dari internal node yang berbeda, fitur yang terpilih akan membedakan masing-masing kriteria. Fleksibelitas pohon kepurusan ini meningkatkan kualitas keputusan yang dihasilkan jika dibandingkan ketika menggunakan metode perhitungan satu tahap lebih konvensional. (4) Metode pohon keputusan dapat menghindari munculnya distribusi tinggi ataupun parameter tertentu dari distribusi analisis multivariat dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap internal node tanpa banyak mengurangi kualitas keputusan yang dihasilkan. (5) Mudah dipahami dan ditafsirkan. (6) Mempertimbangkan detail-detail kecil yang mungkin dilewatkan. (7) Mengehmat waktu, setelah terbentuk akan sangat mudah diimplementasikan. Sedangkan pohon keputusan memiliki beberapa kekurangan diataranya: (1) Terjadi overlap terutama ketika kelas-kelas dan kriteria yang digunakan jumlahnya sangat banyak,

hal tersebut juga menyebabkan meningkatnya waktu pengambilan keputusan dan jumlah memori yang diperlukan. (2) Pengakumulasian jumlah error dari setiap tingkat dalam sebuah pohon keputusan yang besar. (3) Kesulitan dalam merancang pohon keputusan yang optimal. (4) Hasil kualitas keputusan yang didapatkan dari metode pohon keputusan sangat tergantung pada bagaimana pohon tersebut dirancang. (Sumihar, 2015)

## 2.5 Rumus Algoritma ID3

Menurut Retno Tri Vulandari, S.Si.,M.Si. (2017:25) dalam bukunya berjudul Data Mining Teori dan Aplikasi *Rapidminer* menjelaskan bahwa "untuk memilih atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada. Dapat dilakukan dengan rumus *gain* berikut:

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|Sv|}{|S|} Entropy(S)$$

Sedangkan perhitungan nilai *entropy* dapat dilihat pada rumus *entropy* berikut:

$$Entropy(S) = \sum_{i=1}^{c} -P_i \log 2P_i$$

Catatan :

*Gain* (S,A) : *Information gain* dari sebuah atribut A pada koleksi S.
S : Himpunan Kasus.
A : Atribut.
n : Jumlah Partisi Atribut A.
|Si| : Jumlah kasus pada Partisi ke-i.
|S| : Jumlah kasus dalam S.
Pi : Proporsi dari Si terhadap S.

## 2.6 Strategi Pemasaran

Pengertian strategi pemasaran adalah Pendekatan pokok yang akan digunakan oleh unit bisnis dalam mencapai sasaran yang telah ditetapkan terlebih dahulu di dalamnya tercantum keputusan-keputusan pokok mengenai target pasar, penempatan produk di pasar, bauran pemasaran dan tingkat biaya pemasaran yang diperlukan(Nurcholifah, 2014).

Sedangkan menurut Philip Kotler strategi pemasaran merupakan pola pikir yang digunakan untuk mencapai tujuan pemasaran pada suatu perusahaan, bisa mengenai strategi spesifik untuk pasar sasaran ,penetapan posisi, bauran pemasaran (*marketing mix*), dan besarnya sebuah pengeluaran pemasaran.

Jadi penulis simpulkan bahwa strategi pemasaran adalah pola pikir dari suatu perusahaan untuk meningkatkan atau upaya memasarkan suatu produk baik itu barang maupun jasa.

## 2.7 Sinkronisasi Data

Sinkronisasi data dilakukan dengan menghubungkan dua aplikasi *server web service*. Langkah awal proses tersebut dilakukan sesuai dengan permintaan layanan yang dilakukan oleh *client* yang dikirim melalui http dalam bentuk html (PHP) yang selanjutnya akan dikelola oleh *web server*, data yang diperoleh akan dipergunakan beberapa modul, terdapat dua buah dokumen yang akan diproses yaitu dokumen yang berisi informasi skema basis data dan dokumen yang berisi data itu sendiri (Nasir, 2012)

Jadi dapat dikesimpulkan bahwa sinkronisasi data dilakukan dengan menggunakan dua aplikasi yang akan dipergunakan dengan beberapa modul yang dilakukan sesuai permintaan layanan dalam menyelesaikan masalah dengan penggunaan data berupa data yang berisis informasi skema basis data dan data itu sendiri yang akan diproses lebih lanjut.

## 2.8 Sistem Pendukung Keputusan

Sistem pendukung keputusan adalah sistem yang berbasis komputer yang fleksibel, interaktif dan dapat diadaptasi yang dikembangkan untuk mendukung solusi untuk masalah manajemen spesifik yang tidak terstruktur, sistem pendukung keputusan juga dapat diartikan sebagai sistem informasi berbasis komputer yang menggabungkan model dan data guna menyelesaikan masalah semistruktur dan beberapa masalah tidak terstuktur dengan keterlibatan pengguna secara luas. (Nasir, 2018)

Jadi dapat penulis ambil kesimpulan bahwa sistem pendukung keputusan merupakan sistem informasi berbasis komputer yang digunakan untuk menyelesaikan masalah dengan cara menggabungkan model dan data.

### 2.9 *RapidMiner*

Menurut situs resminya https://rapidminer.com/ menjelaskan bahwa "RapidMiner is a software platform for data science teams that unites data prep, machine learning, and predictive model deployment". Penulis terjemahkan yaitu "*RapidMiner* adalah *platform* perangkat lunak untuk tim ilmu data yang menyatukan persiapan data, pembelajaran mesin dan penyebaran model prediktif".

*RapidMiner* adalah *platform* untuk pengolahan data mining yang digunakan untuk analisis teks, mengekstrak pola-pola dari data set yang besar dan mengkombinasikannya dengan metode statistika, kecerdasan buatan, dan *database* serta telah terdapat prosedur data mining dan *machine learning,*didalamnya termasuk ETL (*Extraction, Transformation, oading*), data *preprocessing,* visualisasi, modeling dan evaluasi didalamnya.

## 3. ANALISIS DATA MINING

Data awal yang didapat dari pengambilan data pada Universitas Bina Darma sebelum dilakukannya *preprocessing* data, dengan data keseluruhan yang dapat diakses kembali Universitas Bina Darma untuk tahun 2017 sebanyak 2.001 data Penerimaan Mahasiswa Baru, dan tahun 2018 sebanyak 2.100 data Penerimaan Mahasiswa Baru, jadi jumlah seluruh data yang digunakan pada penelitian ini sebanyak 4.101 data Penerimaan Mahasiswa Baru. Adapun sumber data awal yang digunakan dalam penelitian ini adalah dalam bentuk *excel* (.xlsx). Data awal yang didapat akan ditambahkan *field* baru sebagai *training* untuk membantu penggalian data dalam memprediksi daerah atau kota asal mana yang banyak minat kuliah di Universitas Bina Darma. *Field traning* yang di tambahkan pada penelitian ini yaitu dengan *record* minat dan tidak minat. *Field* baru tersebut didapatkan dari data *survey* yang telah dilakukan sebelumnya kepada 20 mahasiswa dari setiap fakultasnya dengan hasil *survey* yang dapat dilihat pada bagian lampiran sehingga mendapatkan *rule* yang diinginkan.

### 3.1 Data *Cleaning*

Data *Cleaning* adalah sebuah fase dimana data-data tidak lengkap, mengandung error, dan tidak konsisten dibuang dari koleksi data, sehingga data yang telah bersih dan relevan dapat digunakan untuk diproses ulang untuk penggalian pengetahuan (*Discovery Knowledge*). Proses pembersihan data kosong adalah dengan cara menggunakan *operation* pada *tools rapidminer* yaitu *filter example* dengan memasukkan nama *field* yang mengandung data kosong dan setelah dipilih *field* yang mengandung data kosong, data tersebut akan dihapuskan dari sumber data. Tahapan dari proses pembersihan data kosong adalah sebagai berikut:

1. Memilih *operation filter example*
2. Memilih semua *field* yang mungkin mengandung data kosong
3. Memilih *operation replace missing values*
4. Memilih *Average, run*

Setelah dilakukan pembersihan data dari mulai menghapus *record* yang kosong serta menggantikan nilai yang tidak konsisten atau hilang maka didaptkan hasil untuk tahun 2017 sebelum dilakukan pembersihan data sebanyak 2002 berubah menjadi sebanyak 1563 data setelah dilakukan pembersihan data, sedangkan untuk tahun 2018 sebelumnya berjumlah sebanyak 2099 berubah menjadi sebanyak 1847 data.

### 3.2 Data *Integration*

Pada tahapan data mining selanjutnya setelah dilakukan pembersihan data adalah integrasi data dimana data yang berulang, file yang berulang akan digabungkan dan dihapuskan salah satunya. Tahap *integration* data berfungsi untuk menghilangkan data yang ganda dengan tujuan agar tidak terjadi kesalahan dalam perhitungan menentukan daerah atau kota asal mana yang paling banyak minat terhadap Universitas Bina Darma dari data Penerimaan mahasiswa baru. Tahap *integration* dilakukan dengan cara menggunakan *operation remove duplicate* pada *tools*

*rapidminer*. Setelah dilakukan penghilangan data yang duplikat atau ganda maka didapatkan hasil data yang telah relevan dari sebelumnya setelah dilakukan pembersihan data untuk tahun 2017 sebanyak 1563 data setelah dilakukan penghilangan duplikat maka didapatkan data yang relevan sebanyak 1563 data itu berarti data tidak berubah atau data sudah bersih, sedangkan untuk tahun 2018 sebelumnya setelah dilakukan *cleaning* data berjumlah 1847 data setelah dilakukan *integration* data berkurang menjadi 1839 data, dengan selisih data yang dihapuskan sebanyak 8 data.

### 3.3 Data *Selection*

Tahapan data mining selanjurnya adalah *selection* data yaitu memilih *field* dari koleksi data yang ada untuk digunakan dalam penggalian data pada proses data mining selanjutnya untuk mendapatkan *knowledge* yang diinginkan. *Selection* data berfungsi untuk memilih *field* yang sesuai kebutuhan untuk mendapatkan *knowledge*, pada penelitian saat ini *field* data yang sesuai untuk mendapatkan hasil yang diinginkan dan sesuai dengan prosesnya adalah no barcode, jenis barang dan jenis pengiriman. Adapun tahapan dari data *selection* adalah sebagai berikut:

1. Pilih *operation select attributes.*
2. Pada bagian parameters pilih *subset* di bagian *attribute filter type.*
3. Pada bagian *attribute* pilih *field* yang diinginkan.
4. Klik *apply, run*

hasil dari proses seleksi data yang telah dilakukan sebelumnya dengan *field* data yang dipilih sesuai kebutuhan dalam penelitian ini menggunakan *field* no tes, kota asal, training, dan info.

### 3.4 Data *Transformation*

Data *Transformation* juga dikanal sebagai data *consolidation*. Pada tahap ini, dimana data-data yang telah terpilih, ditransformasikan kedalam bentuk-bentuk atau tipe data yang cocok untuk dilakukan penggalian data pada tahap selanjutnya dengan cara melakukan normalisasi data. *Transformation* data dilakukan dilakukan dengan cara normalisasi data menggunakan *tools rapidminer* dengan mencari *operation normalize*. Adapun tahapan dari transformasi data adalah sebagai berikut:

1. Pilih *normalize* pada bagian *search operation*.
2. Klik *run*
3. Akan keluar hasil atau *result* bentuk data setelah ditransformasi sesuai format *tools rapidminer* untuk dilakukan penggalian data pada tahap selanjutnya .
4. Setelah dilakukan normalisasi data pada sumber data yang ada maka didapatlah hasil data yang siap dilakukan penggalian data pada tahap data mining selanjutnya.

Hasil dari normalisasi data dengan tidak ada perubahan karena tidak terdapat bilangan *numeric* dan data siap dimining.

## 4. HASIL DAN PEMBAHASAN

*Data mining* atau sering dikenal dengan *knowledge discovery in database* adalah proses untuk mengekstrak data untuk menemukan pola-pola yang penting atau menarik yang belum diketahui secara manual dari basis data yang besar untuk mendapatkan sebuah keputusan.

### 4.1 Perhitungan Manual

perhitungan manual dari algoritma *ID3* dengan menggunakan dataset training sebanyak 100 *record* dengan diambil acak dari data PMB tahun 2017, tahun 2018 dan gabungan untuk memprediksi tahun 2019 dengan diambil dari tahun 2017 sebanyak 50 *record* dan dari tahun 2018 sebanyak 50 *record*, proses pengambilan data tersebut dengan diambil urutan 2-101 pada data awal 2017 dan 2018, sedangkan untuk tahun 2019 diambil dari urutan 2-51 dari tahun 2017 dan begitupun tahun 2018 untuk memprediksi tahun 2019. Dalam perhitungan ini untuk memilih atribut sebagai akar maka dilakukan perhitungan nilai *entropy* terlebih dahulu dengan menggunakan rumus dan perhitungan sebagai berikut:

$$Entropy(S) = \sum_{i=1}^{c} -P_i \log 2P_i$$

Keterangan:
S : Himpunan Kasus.
Pi : Proporsi dari Si terhadap S.

Sedangkan untuk memilih atribut sebagai akar didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada, *gain* dapat dilakukan dengan rumus sebagai berikut:

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|Sv|}{|S|} Entropy(S)$$

Keterangan:
*Gain* (S,A) : *Information gain* dari sebuah atribut A pada koleksi S.
S : Himpunan Kasus.
A : Atribut.
|Si| : Jumlah kasus pada Partisi ke-i.
|S| : Jumlah kasus dalam S.

Adapun perhitungan manual dilakukan pada data 2017, 2018 dan prediksi tahun 2019 dengan 2 pengetahuan yang akan diketahui yaitu sumber informasi yang paling diminati dan lokasi terbanyak serta terdikit diminati. Berikut merupakan contoh perhitungan manual prediksi tahun 2019 dengan pengetahuan kota asal dan sumber informasi.

### 4.1.1 Perhitungan Manual Kota Asal

Perhitungan data kota asal untuk prediksi tahun 2019 menggunakan 50 *record* tahun 2017 dan 50 *record* tahun 2018 yang akan dijadikan *dataset* dalam menentukan lokasi peminat tersebut, adapun contoh *dataset* tersebut dapat dilihat pada tabel 4.1 dibawah ini.

Tabel 4.1 *Dataset* Kota Asal Prediksi 2019

| No. Tes | Kota Asal | Training |
|---|---|---|
| **20170002** | KOTA PALEMBANG | MINAT |
| **20170005** | KOTA PALEMBANG | TIDAK MINAT |
| **20170020** | KOTA BANDAR LAMPUNG | TIDAK MINAT |
| **20170023** | KOTA SURABAYA | MINAT |
| **20170039** | KABUPATEN LAHAT | MINAT |
| **20170042** | KABUPATEN EMPAT LAWANG | TIDAK MINAT |
| **20170045** | KOTA PALEMBANG | MINAT |
| **20170059** | KOTA PALEMBANG | MINAT |
| **20170062** | KABUPATEN BANYUASIN | MINAT |
| **20170065** | KOTA PALEMBANG | MINAT |
| **20170082** | KABUPATEN OKI | TIDAK MINAT |
| **20170102** | KOTA PALEMBANG | MINAT |
| **20170105** | KOTA PALEMBANG | MINAT |
| **20170121** | KABUPATEN OGAN ILIR | MINAT |
| **20170123** | KOTA PALEMBANG | MINAT |
| **20170126** | KOTA RIAU | TIDAK MINAT |
| **20170140** | KABUPATEN LAHAT | MINAT |
| **20170143** | KOTA PALEMBANG | MINAT |
| **20170146** | KOTA PALEMBANG | MINAT |
| **20170163** | KABUPATEN OKI | MINAT |

Dari *dataset* tersebut akan dibuat sebuah pohon keputusan untuk menentukan lokasi peminat dilihat dari kota dan kabupatennya dengan jumlah data yang digunakan adalah sebagai berikut.

Tabel 4.2 Hasil Penjumlahan Kota Asal Prediksi 2019

| Kota Asal | S | Minat | Tidak Minat |
|---|---|---|---|
| Banyuasin | 10 | 9 | 1 |
| Empat Lawang | 1 | 1 | 0 |
| Lahat | 7 | 7 | 0 |
| Muara Enim | 9 | 8 | 1 |
| Musi Banyuasin | 8 | 3 | 5 |
| Musi Rawas | 1 | 1 | 0 |
| Ogan Ilir | 6 | 5 | 1 |
| OKI | 10 | 8 | 2 |
| OKU | 1 | 1 | 0 |
| OKU Selatan | 1 | 0 | 1 |
| OKU Timur | 3 | 1 | 2 |
| Bandar Lampung | 2 | 0 | 2 |
| Bengkulu | 1 | 0 | 1 |
| Jakarta Barat | 1 | 1 | 0 |
| Lubuk Linggau | 1 | 0 | 1 |
| Palembang | 33 | 32 | 1 |
| Prabumulih | 3 | 2 | 1 |
| Riau | 1 | 0 | 1 |
| Surabaya | 1 | 1 | 0 |
| Total | 100 | 79 | 21 |

Setelah dilakukan penjumlahan diatas maka dilakukan perhitungan dengan rumus algoritma *ID3* untuk menghitung *entropy* dengan hasil perhitungan sebagai berikut.

$$Entropy(Total) = (-\frac{79}{100} x \log_2\left(\frac{79}{100}\right)) + (-\frac{21}{100} x \log_2(\frac{21}{100})) = 0.7413$$

$$Entropy(Banyuasin) = (-\frac{9}{10} x \log_2\left(\frac{9}{10}\right)) + (-\frac{1}{10} x \log_2(\frac{1}{10})) = 0.4671$$

$$Entropy(Muara\ Enim) = (-\frac{8}{9} x \log_2\left(\frac{8}{9}\right)) + (-\frac{1}{9} x \log_2(\frac{1}{9})) = 0.4811$$

$$Entropy(Musi\ Banyuasin) = (-\frac{3}{8} x \log_2\left(\frac{3}{8}\right)) + (-\frac{5}{8} x \log_2(\frac{5}{8})) = 0.953$$

$$Entropy(Ogan\ Ilir) = (-\frac{5}{6} x \log_2\left(\frac{5}{6}\right)) + (-\frac{1}{6} x \log_2(\frac{1}{6})) = 0.6382$$

$$Entropy(OKI) = (-\frac{8}{10} x \log_2\left(\frac{8}{10}\right)) + (-\frac{2}{10} x \log_2(\frac{2}{10})) = 0.721$$

$$Entropy(OKU\ T) = (-\frac{1}{3} x \log_2\left(\frac{1}{3}\right)) + (-\frac{2}{3} x \log_2(\frac{2}{3})) = 0.9066$$

$$Entropy(Palembang) = (-\frac{32}{33} x \log_2\left(\frac{32}{33}\right)) + (-\frac{1}{33} x \log_2(\frac{1}{33})) = 0.20738$$

$$Entropy(Prabumulih) = (-\frac{2}{3} x \log_2\left(\frac{2}{3}\right)) + (-\frac{1}{3} x \log_2(\frac{1}{3})) = 0.9066$$

Sedangkan perhitungan nilai *gain* dengan rumus *ID3* dapat dilihat sebagai berikut.

$$Gain(Total, K) = 0.7413 - \left(\frac{10}{100} x 0.4671\right) + \left(\frac{9}{100} x 0.4811\right) + \left(\frac{8}{100} x 0.953\right) + \left(\frac{6}{100} x 0.6382\right) + \left(\frac{10}{100} x 0.721\right) + \left(\frac{3}{100} x 0.9066\right) + \left(\frac{33}{100} x 0.20738\right) + \left(\frac{3}{100} x 0.9066\right) = 0.3423$$

Hasil perhitungan yang telah dilakukan diatas dapat dilihat sebagai berikut.

Tabel 4.3 Hasil Perhitungan Kota Asal Prediksi 2019

| | S | Minat | Tidak Minat | *Entropy* | *Gain* |
|---|---|---|---|---|---|
| Kota Asal | | | | | 0.3423 |
| Banyuasin | 10 | 9 | 1 | 0.4671 | |
| Empat Lawang | 1 | 1 | 0 | 0 | |
| Lahat | 7 | 7 | 0 | 0 | |
| Muara Enim | 9 | 8 | 1 | 0.4811 | |
| Musi Banyuasin | 8 | 3 | 5 | 0.953 | |
| Musi Rawas | 1 | 1 | 0 | 0 | |
| Ogan Ilir | 6 | 5 | 1 | 0.6382 | |
| OKI | 10 | 8 | 2 | 0.721 | |
| OKU | 1 | 1 | 0 | 0 | |
| OKU Selatan | 1 | 0 | 1 | 0 | |
| OKU Timur | 3 | 1 | 2 | 0.9066 | |
| Bandar Lampung | 2 | 0 | 2 | 0 | |
| Bengkulu | 1 | 0 | 1 | 0 | |
| Jakarta Barat | 1 | 1 | 0 | 0 | |
| Lubuk Linggau | 1 | 0 | 1 | 0 | |
| Palembang | 33 | 32 | 1 | 0.20738 | |
| Prabumulih | 3 | 2 | 1 | 0.9066 | |
| Riau | 1 | 0 | 1 | 0 | |
| Surabaya | 1 | 1 | 0 | 0 | |
| Total | 100 | 79 | 21 | 0.7413 | |

Dari perhitungan tersebut dapat diketahui bahwa banyak kota asal yang telah memiliki keputusan dengan melihat dari nilai *entropy* 0 yang tidak diikuti angka lainnya, sedangkan angka 0 yang diikuti angka lainnya dan nilai 1 memiliki penurunan yang berisikan minat dan tidak minat. Dari hasil perhitungan tersebut juga dapat diketahui 3 lokasi sebagai paling banyak peminat yaitu terdapat di Palembang dengan data sebanyak 32 peminat, Banyuasin sebanyak 9 peminat dan OKI sebanyak 8 peminat.

### 4.1.2 Perhitungan Manual Sumber Informasi

Pada perhitungan info dalam memprediksi sumber informasi paling diminati tersebut pada tahun 2019 dilakukan dengan menggunakan 100 *record* yang akan dijadikan sebagai *dataset* dalam memprediksi sumber informasi dengan data 100 *record* tersebut didapatkan secara acak dari 50 data tahun 2017 dan 50 data dari tahun 2018. Adapun contoh *dataset* yang digunakan adalah sebagai berikut.

Tabel 4.4 *Dataset* Info Prediksi 2019

| No. Tes | Info | Training |
|---|---|---|
| **20170002** | INTERNET | MINAT |
| **20170005** | SEKOLAH | TIDAK MINAT |
| **20170020** | BROSUR | TIDAK MINAT |
| **20170023** | INTERNET | MINAT |
| **20170039** | BROSUR | MINAT |
| **20170042** | BROSUR | TIDAK MINAT |
| **20170045** | INTERNET | MINAT |
| **20170059** | INTERNET | MINAT |
| **20170062** | SEKOLAH | MINAT |
| **20170065** | KELUARGA | MINAT |
| **20170082** | KELUARGA | TIDAK MINAT |
| **20170102** | INTERNET | MINAT |

| | | |
|---|---|---|
| **20170105** | BROSUR | MINAT |
| **20170121** | KELUARGA | MINAT |
| **20170123** | KELUARGA | MINAT |
| **20170126** | BROSUR | TIDAK MINAT |
| **20170140** | INTERNET | MINAT |
| **20170143** | INTERNET | MINAT |
| **20170146** | BROSUR | MINAT |
| **20170163** | KELUARGA | MINAT |

Dari data tersebut dilakukan penjumlahan nilai data dari *dataset* yang digunakan dengan melihat dari *field* info, adapun hasil penjumlahan data tersebut dapat dilihat pada tabel 4.5 dibawah ini.

Tabel 4.5 Hasil Penjumlahan Info Prediksi 2019

| Info | S | Minat | Tidak Minat |
|---|---|---|---|
| Brosur | 18 | 12 | 6 |
| Internet | 29 | 25 | 4 |
| Keluarga | 35 | 29 | 6 |
| Sekolah | 7 | 4 | 3 |
| Spanduk | 1 | 1 | 0 |
| Teman | 10 | 8 | 2 |
| Total | 100 | 79 | 21 |

Setelah dilakukan penjumlahan diatas maka dilakukan perhitungan dengan rumus algoritma *ID3* untuk menghitung *entropy* atau menentukan adanya penurunan atau tidak, dengan hasil perhitungan sebagai berikut.

$$Entropy(Total) = (-\frac{79}{100} x \log_2\left(\frac{79}{100}\right)) + (-\frac{21}{100} x \log_2(\frac{21}{100})) = 0.7413$$

$$Entropy(Brosur) = (-\frac{12}{18} x \log_2\left(\frac{12}{18}\right)) + (-\frac{4}{18} x \log_2(\frac{4}{18})) = 0.9095$$

$$Entropy(Internet) = (-\frac{25}{29} x \log_2\left(\frac{25}{29}\right)) + (-\frac{4}{29} x \log_2(\frac{4}{29})) = 0.5771$$

$$Entropy(Keluarga) = (-\frac{29}{35} x \log_2\left(\frac{29}{35}\right)) + (-\frac{6}{35} x \log_2(\frac{6}{35})) = 0.6631$$

$$Entropy(Teman) = (-\frac{8}{10} x \log_2\left(\frac{8}{10}\right)) + (-\frac{2}{10} x \log_2(\frac{2}{10})) = 0.721$$

Sedangkan perhitungan nilai *gain* atau nilai akar dengan rumus *ID3* dapat dilihat sebagai berikut.

$$Gain(Total, K) = 0.7413 - \left(\frac{18}{100} x 0.9095\right) + \left(\frac{29}{100} x 0.5771\right) + \left(\frac{35}{100} x 0.6631\right) + \left(\frac{7}{100} x 0.9789\right) + \left(\frac{10}{100} x 0.721\right) = 0.0522$$

Dari perhitungan yang telah dilakukan diatas maka didapatkan hasil *entropy* dan *gain* yang dapat dilihat pada tabel 4.6 berikut ini.

Tabel 4.6 Hasil Perhitungan Info Prediksi 2019

| | S | Minat | Tidak Minat | *Entropy* | *Gain* |
|---|---|---|---|---|---|
| Info | | | | | 0.0522 |
| Brosur | 18 | 12 | 6 | 0.9095 | |
| Internet | 29 | 25 | 4 | 0.5771 | |
| Keluarga | 35 | 29 | 6 | 0.6631 | |
| Sekolah | 7 | 4 | 3 | 0.9789 | |
| Spanduk | 1 | 1 | 0 | 0 | |

| | | | | | |
|---|---|---|---|---|---|
| Teman | 10 | 8 | 2 | 0.721 | |
| Total | 100 | 79 | 21 | 0.7413 | |

Dari perhitungan pada tabel 4.18 diatas dapat diketahui bahwa terdapat beberapa nilai yang telah memiliki keputusan dan banyak yang memiliki penurunan, dan dari tabel tersebut juga dapat diketahui bahwa sumber informasi yang paling diminati adalah bersumber dari keluarga dengan jumlah minat 29 data.

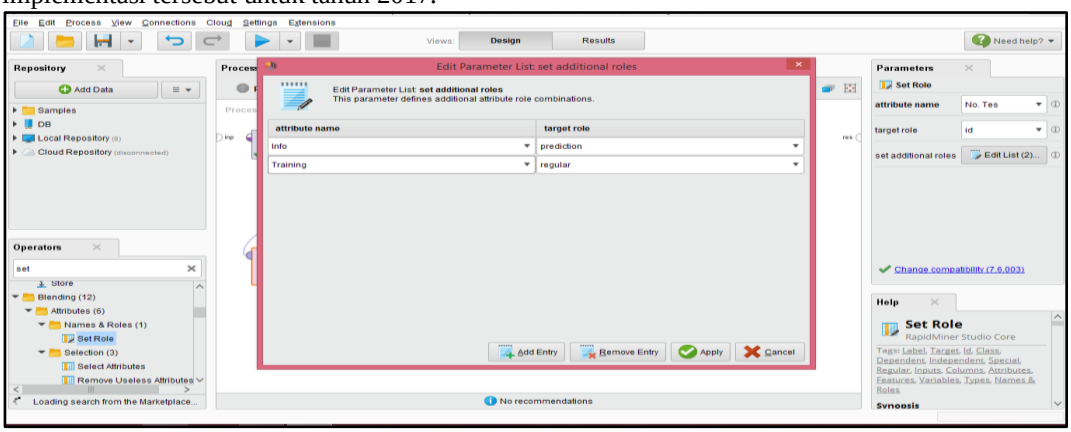
## 4.2 Penerapan *RapidMiner*

Implementasi algortima *ID3* menggunakan *tools rapidminer* dengan tahapan sebagai berikut:

1. Menambahkan *Operator set rule* dengan memberi label.
2. Menambakan *Operator split* data.
3. Menuliskan *edit parameters*
4. Menambakan algoritma *ID3.*
5. Menambakan *Operator apply model*.
6. Hasil sumber informasi paling diminati.

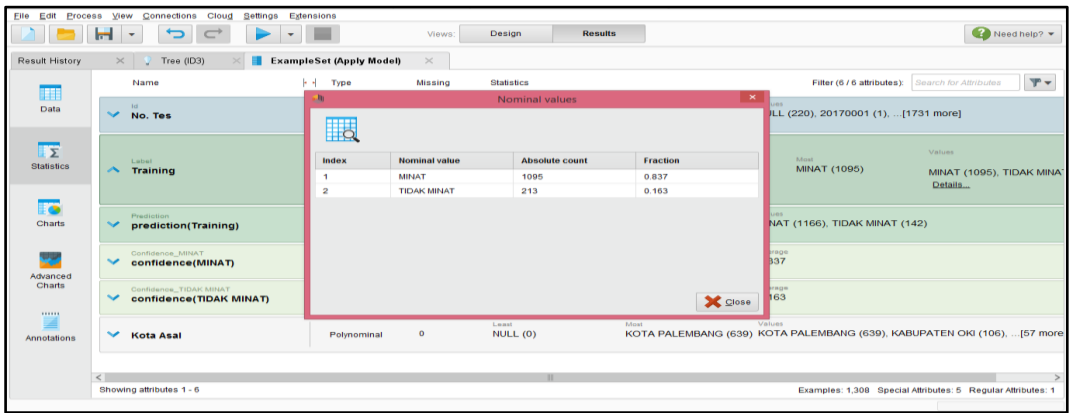
### 4.2.1 *RapidMiner* Kota Asal

Seperti yang telah dijelaskan sebelumnya untuk tahapan implementasi algoritma *ID3* menggunakan *tools rapidminer*, pada tahap ini adalah memperlihatkan gambar dari proses implementasi tersebut untuk tahun 2017.



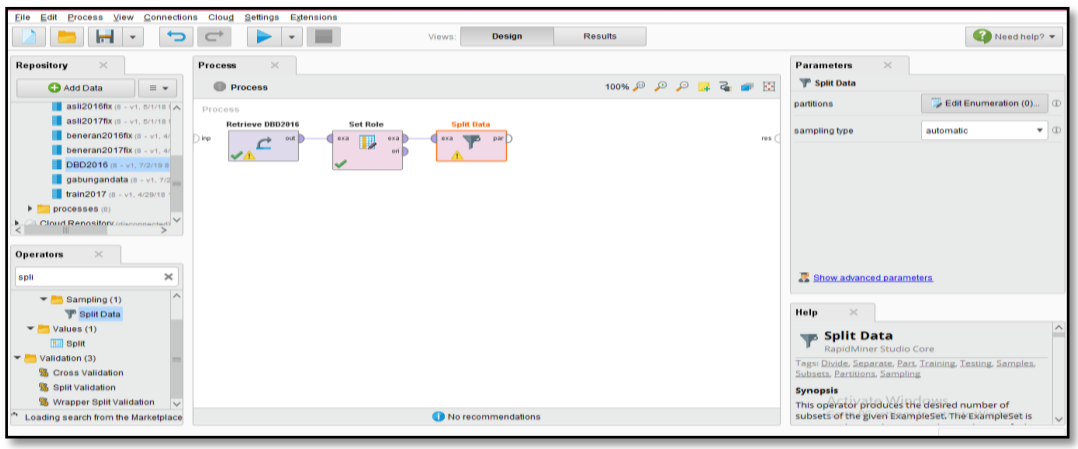

Gambar 4.1 *setrole*

Pada gambar diatas *setrole* merupakan proses untuk menambahkan sebuah atribut khusus. Pada penelitian ini *field* data yang di set adalah no tes sebagai id karena berperan sebagai *primary key*, untuk *edit parameter* dilakukan *edit list* dengan *attribute name* Info dengan *targetrole* sebagai *prediction* dan *training* sebagai label.



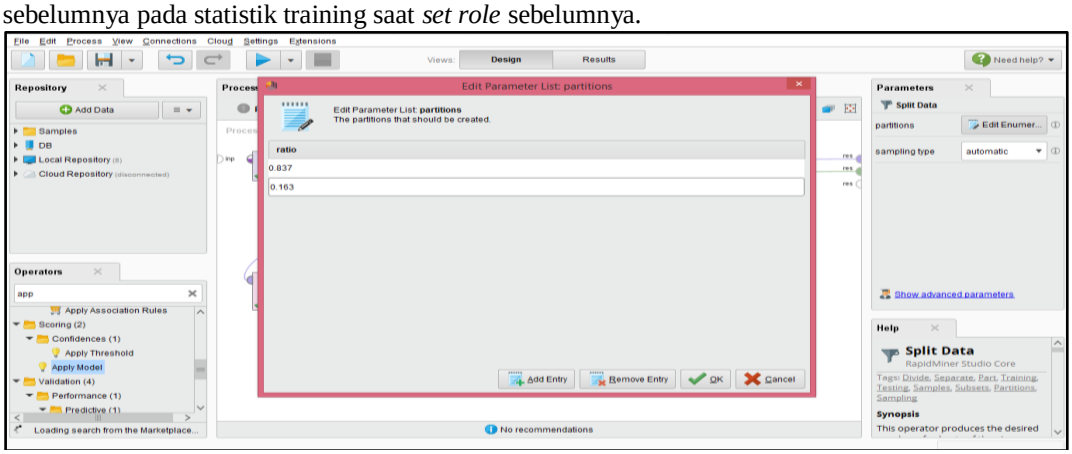

Gambar 4.2 statistik training

Pada gambar diatas terdapat *nominal values*, yang perlu dicatat atau yang sangat penting adalah *fraction* kota asal untuk pembagian *class* perhitungan *rapidminer*.



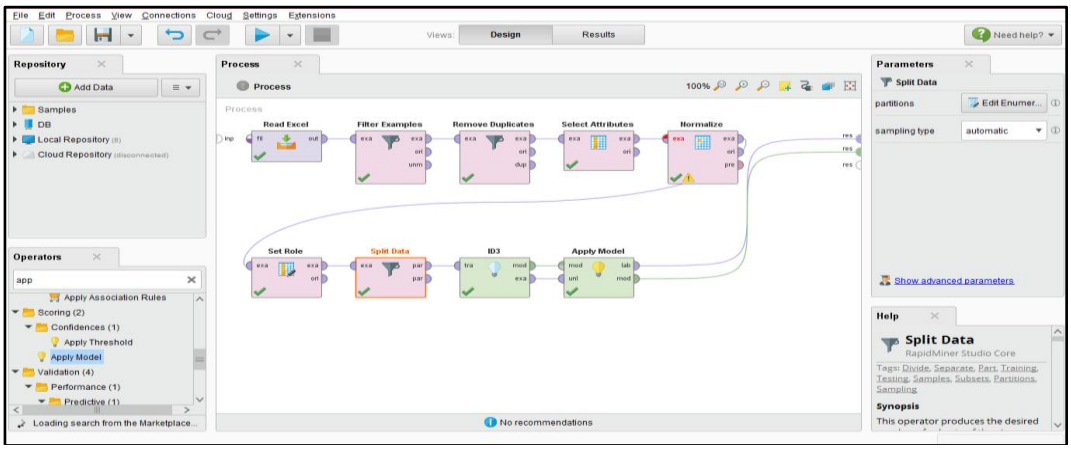

Gambar 4.3 *split* data

Pada proses *split* data memilih *edit parameters* dengan mengisi *ratio* yang sudah dilihat sebelumnya pada statistik training saat *set role* sebelumnya.



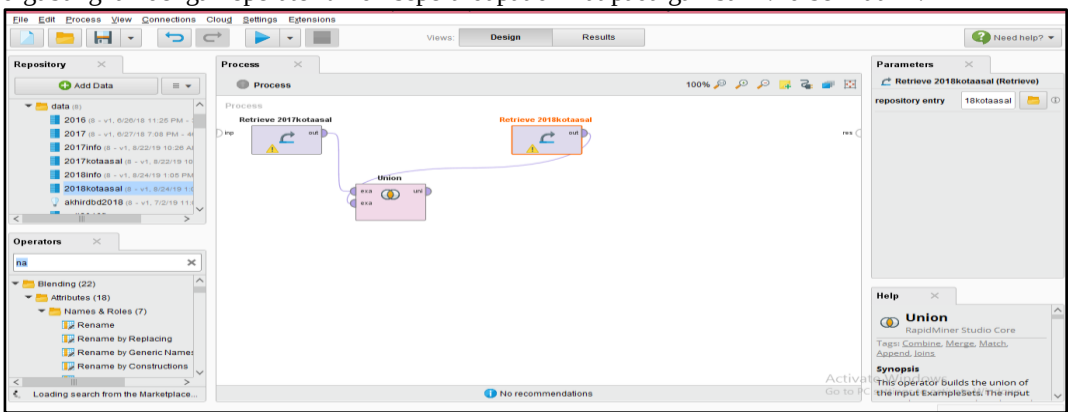

Gambar 4.4 *edit parameters*

Pada gambar diatas dalam *operator split data* pada *partitions* akan dilkukan *edit enumeration* yaitu memasukkan *fraction* kinerja yang telah dicatat sebelumnya.



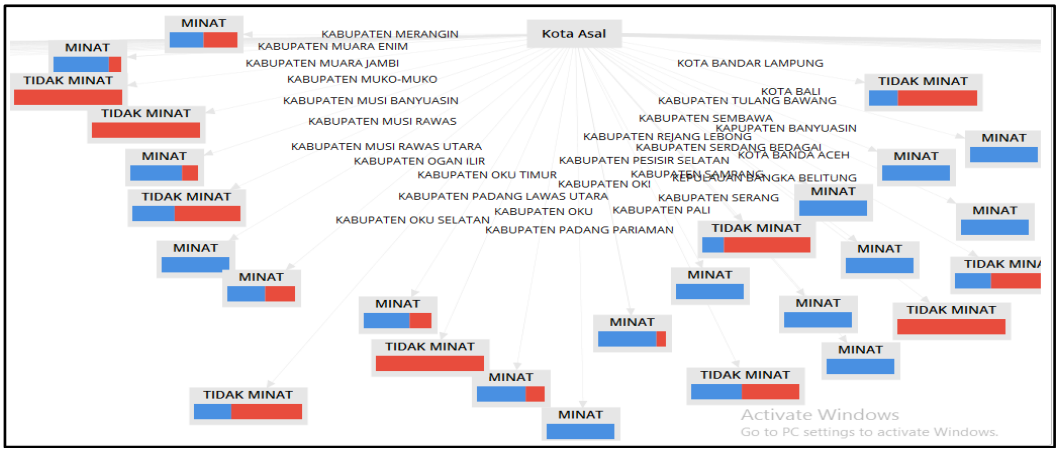

Gambar 4.5 *ID3* dan *apply model*

Pada gambar diatas *operator apply model* digunakan untuk menjalankan model *ID3* untuk mendapatkan *knowledge* yang diinginkan dan dimengerti oleh user. Terdapat sedikit perbedaan pada pengelolahan data prediksi pada tahap ini data tahun 2017, tahun 2018 dan tahun 2019 digabungkan dengan operator *union* seperti dapat dilihat pada gambar 4.10 berikut ini.




Gambar 4.6 *Operator Union* Kota Asal

Setelah dilakukan penggabungan maka *rapidminer* akan menghasilkan data prediksi kota asal tahun 2019.



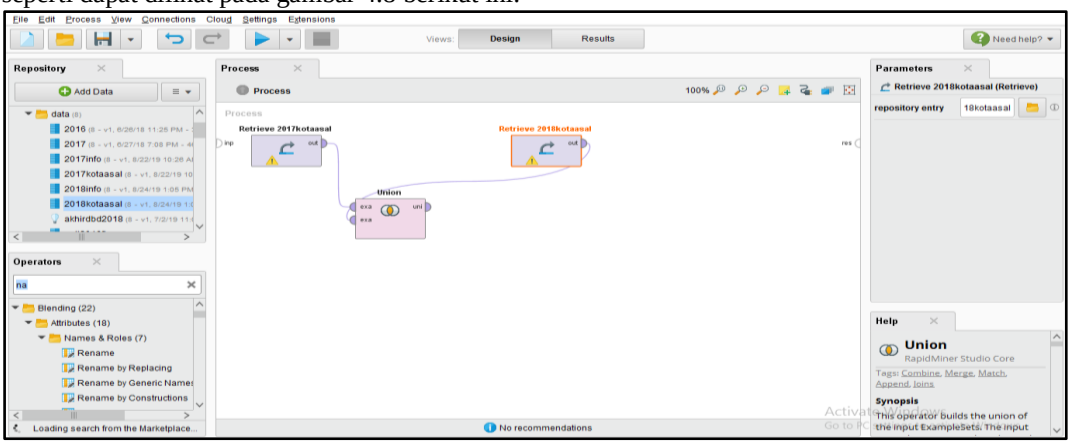

Gambar 4.7 Prediksi kota asal 2019

Pohon keputusan yang didapatkan untuk memprediksi tahun 2019 dengan 3 kota asal tertinggi dan 3 kota asal terendah, 3 kota asal tertinggi Palembang dengan 1353 peminat, Oki 185

peminat dan Muara Enim 135 peminat, sedangkan untuk 3 kota asal terendah adalah Musi Banyuasin 30 Tidak Minat, Prabumulih 23 Tidak Minat dan PagarAlam 22 Tidak Minat.

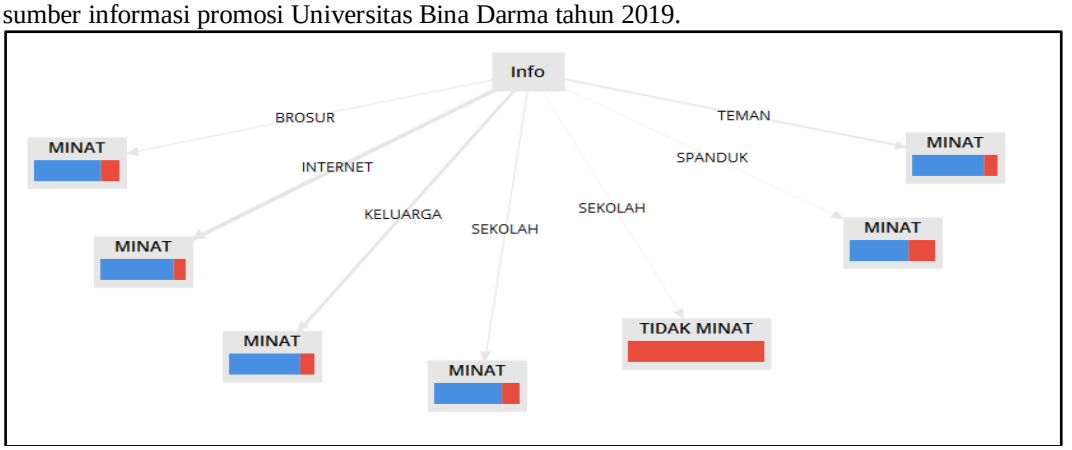
### 4.2.2 *RapidMiner* Sumber Informasi

Pada tahap ini akan memperlihatkan gambar dari tahapan implementasi algoritma *ID3* dengan menggunakan *tools rapidminer* untuk lokasi terbanyak Peminat dan paling sedikit Peminat terhadap Universitas Bina Darma Palembang. Terdapat sedikit perbedaan pada pengelolahan data prediksi pada tahap ini data tahun 2017 dan tahun 2018 digabungkan dengan operator *union* seperti dapat dilihat pada gambar 4.8 berikut ini.




Gambar 4.8 *Operator Union* Info

Setelah dilakukan penggabungan maka *rapidminer* akan menghasilkan data prediksi sumber informasi promosi Universitas Bina Darma tahun 2019.




Gambar 4.9 Prediksi Info 2019

Setelah dilakukan perhitungan sebelumnya maka didapatkan untuk memprediksi tahun 2019 dengan sumber info promosi tertinggi adalah internet dengan 876 peminat.

## 5. KESIMPULAN

Dari uraian bab-bab yang sudah dibahas sebelumnya dapat ditarik kesimpulan yaitu:

1. Algoritma *ID3* yang diproses dengan *tools rapidminer* dapat mengidentifikasi lokasi palinga banyak peminat dan paling sedikit peminat terhadap Universitas Bina Darma Palembang dilakukan dengan sangat benar serta dapat memprediksi lokasi berpotensi peminatan tersebut untuk tahun 2019 dengan sangat baik.
2. Pemilihan atribut yang akan digunakan dalam menentukan sebuah keputusan dan prediksi juga sangat mempengaruhi *knowledge* yang dihasilkan.

3. *Tools rapidminer* berhasil menentukan lokasi paling sedikit peminat pada tahun 2017 terdapat di Lahat, Banyuasin dan Pali dengan hasil akurasi kota asal tahun 2017 sebesar 87.39%, Tahun 2017 memperlihatkan lokasi paling sedikit peminatan tersebut berada pada Ogan Ilir, Lahat dan Musi Banyuasin dengan hasil akurasi sebesar 85.77%, Tahun 2018 memperlihatkan lokasi peminat paling sedikit berada di wilayah Prabumulih, Pagar Alam dan Musi Banyuasin dengan hasil akurasi sebesar 86.27%. Sedangkan untuk sumber informasi yang paling diminati oleh mahasiswa ataupun masyarakat luas pada umumnya adalah bersumber dari internet dengan rata-rata hasil akurasi sebesar 83.12%.

**DAFTAR PUSTAKA**

[1] Indrajani, S.Kom., M.M. *Database Systems Case Study All In One*. Jakarta: Media Komputindo, 2014.

[2] Nurcholifah, Ita. "STRATEGI MARKETING MIX DALAM PERSPEKTIF SYARIAH." *Journal Of Islamic Studies* 4 (2014): 14.

[3] Rahmat, Diding. "Implementasi Kebijakan Program Bantuan Hukum Bagi Masyarakat tidak Mampu di Kabupaten Kuningan," 1 Januari 2017.

[4] "Sejarah." *Portal Web UBD | Universitas Bina Darma, Palembang, Indonesia* (blog), 24 Maret 2016. https://www.binadarma.ac.id/sejarah/.

[5] Shiddiq, Ahmad, Ratih Kumalasari Niswatin, dan Intan Nur Farida. "Analisa Kepuasan Konsumen Menggunakan Klasifikasi Decision Tree Di Restoran Dapur Solo (Cabang Kediri)," 2018, 10.

[6] Vulandari, Retno Tri, S.Si.,M.Si. *Data Mining Teori dan Aplikasi Rapidminer*. Yogyakarta: GAVA MEDIA, 2017.