

Received : 21-05-2021

Revised : 30-06-2021

Published : 29-07-2021

Komparasi Algoritma Klasifikasi untuk Orientasi Minat Mahasiswa dalam Penuntasan Studi

Muhammad Rizki Fahdia¹, Dwiza Riana², Fachri Amsury³, Irwansyah Saputra⁴,
Nanang Ruhjana⁵

STMIK Nusa Mandiri

rizki.muz@nusamandiri.ac.id¹, dwiza@nusamandiri.ac.id², fachri.fcy@nusamandiri.ac.id³,
irwansyah.iys@nusamandiri.ac.id⁴, nanang.ngy@nusamandiri.ac.id⁵

Abstrak:

Politeknik Tri Mitra Karya Mandiri adalah salah satu perguruan tinggi vokasi yang berada di wilayah Cikampek Kabupaten Karawang yang pada tahun akademik 2017/2018 mempunyai jumlah mahasiswa mencapai 987 orang mahasiswa. Namun sayangnya dari total jumlah mahasiswa tidak seluruhnya mempunyai orientasi minat yang sesungguhnya untuk kuliah, banyak factor yang mempengaruhinya. Tingginya tingkat orientasi minat mahasiswa yang tidak memilih kuliah, inilah yang membuat diadakan penelitian tentang sebab-sebab mengapa mahasiswa berkuliah dikampus ini serta mencari solusi guna mengurangi jumlah mahasiswa yang menjadi non aktif ketika diketahui mempunyai orientasi minat yang bukan untuk kuliah. Dengan melakukan komparasi menggunakan 3 algoritma yang termasuk dalam metode klasifikasi data mining yaitu; *Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor* penelitian ini mencari nilai akurasi dan *Area Under Curve* (AUC) yang terbaik dari ketiga algoritma yang dikomparasi guna ditentukan model yang digunakan pada penentuan orientasi minat mahasiswa. Hasil dari komparasi yang dilakukan dalam penelitian ini adalah; algoritma *Decision Tree C4.5* mempunyai nilai akurasi sebesar 91,75% dan AUC sebesar 0,969, *Naïve Bayes* mempunyai nilai akurasi sebesar 86,77% dan AUC sebesar 0,930 sedangkan *K-Nearest Neighbor* mempunyai nilai akurasi sebesar 88,61% dan AUC sebesar 0,500. Melalui uji beda yang dilakukan menggunakan operator *T-test* pada *Rapid Miner* ditemukan hasil bahwa algoritma *Decision Tree C4.5* merupakan algoritma terbaik dari 3 buah algoritma yang digunakan, maka dalam penelitian ini digunakan rule *Decision Tree C4.5* untuk diterapkan pada *deployment* yang dilakukan.

Kata kunci: orientasi minat mahasiswa; komparasi; auc; decision tree c4.5; naïve bayes; k-nearest neighbor



COMPARISON OF CLASSIFICATION ALGORITHM FOR ORIENTATION OF INTEREST STUDENTS IN STUDY ASSESSMENT

Abstract:

Tri Mitra Karya Mandiri Polytechnic is one of the vocational colleges located in the Cikampek area of Karawang Regency, which in the 2017/2018 academic year has 987 students. But unfortunately from the total number of students not all have real interest orientation for college, many the factors that influence it. The high level of orientation of the interest of students who did not choose college, this is what made a study of the reasons why students study at the campus and find solutions to reduce the number of students who become non-active when they are known to have interest in orientation that is not for college. By comparing using 3 algorithms included in the data mining classification method, namely; This Decision Tree C4.5, Naïve Bayes and K-Nearest Neighbor research looked for the best value of accuracy and Area Under Curve (AUC) of the three comparable algorithms to determine the model used in determining the orientation of student interest. The results of the comparison carried out in this study are; Decision Tree C4.5 algorithm has an accuracy value of 91.75% and AUC of 0.969, Naïve Bayes has an accuracy value of 86.77% and AUC of 0.930 while K-Nearest Neighbor has an accuracy value of 88.61% and AUC of 0.500 . Through different tests conducted using the T-test operator on Rapid Miner, the results found that the Decision Tree C4.5 algorithm is the best algorithm of the 3 algorithms used, then in this study used the Decision Tree C4.5 rule to be applied to the deployment carried out.

Keywords: *student interest orientation; comparative; auc; decision tree c4.5; naive bayes; k-nearest neighbor*

PENDAHULUAN

Politeknik Tri Mitra Karya Mandiri adalah salah satu perguruan tinggi vokasi yang berada di wilayah Cikampek Kabupaten Karawang. Politeknik ini telah berdiri selama kurang lebih 17 tahun di mulai dari tahun 2001. Jumlah mahasiswa Politeknik Tri Mitra Karya Mandiri tiap tahun mengalami kenaikan walaupun tidak signifikan. Salah satu faktor yang mendorong keinginan dari calon mahasiswa menjadi mahasiswa di kampus ini adalah minat dari mahasiswa itu sendiri, terkadang minat yang dimiliki juga tidak menjadi dasar pertimbangan calon mahasiswa masuk ke kampus ini. Beberapa faktor menjadi penentu dari terealisasinya calon mahasiswa menjadi mahasiswa, dari mulai faktor finansial hingga faktor hobi dari calon mahasiswa tersebut. Pada penelitian ini digali keinginan sesungguhnya dari mahasiswa baru yang telah terdaftar pada kampus ini, sehingga diharapkan didapat pemetaan berdasarkan orientasi sesungguhnya mahasiswa tersebut dalam melanjutkan studinya dan faktor-faktor pendukungnya yang menyebabkan calon mahasiswa tersebut mendaftar di Politeknik Tri Mitra Karya Mandiri.

Minat adalah suatu rasa lebih suka dan rasa keterikatan pada suatu hal atau aktivitas, tanpa ada yang menyuruh. Minat pada dasarnya adalah penerimaan akan suatu hubungan antara diri sendiri dengan sesuatu di luar diri. Semakin kuat atau dekat hubungan tersebut, semakin besar minat. Suatu minat dapat diekspresikan melalui suatu pernyataan yang menunjukkan bahwa siswa lebih menyukai suatu hal daripada hal lainnya, dapat pula dimanifestasikan melalui partisipasi dalam suatu aktivitas. Siswa yang memiliki minat

terhadap subyek tertentu cenderung untuk memberikan perhatian yang lebih besar terhadap subyek tersebut. Minat menunjukkan adanya suatu ketertarikan terhadap sesuatu (Chusanawati, 2015). Ketika seorang siswa memutuskan untuk melanjutkan studinya ke jenjang lebih tinggi, maka banyak faktor yang mempengaruhi keputusan tersebut, dimana minat seseorang juga terkadang menjadi faktor utama dalam pengambilan keputusan yang dihasilkan untuk melanjutkan studi tersebut. Menyikapi hal tersebut maka minat seseorang tidak dapat diabaikan sebagai salah satu variabel pengambilan keputusan, pada Politeknik Tri Mitra Karya Mandiri terdaftar 987 mahasiswa dengan 865 orang mahasiswa tercatat sebagai mahasiswa aktif dan 122 orang sebagai mahasiswa non aktif dari berbagai latar belakang sosial dan berbagai penyebab sehingga mahasiswa tersebut melanjutkan studi di kampus ini. Untuk penelitian ini sendiri digunakan data dari kuisioner yang dibagikan kepada 219 orang mahasiswa baru tahun akademik 2017/2018 yang diharapkan dapat mengetahui korelasi dari orientasi minat sesungguhnya mahasiswa dengan faktor lain yang menyebabkan mereka melanjutkan studinya di kampus ini.

Analisa mengenai orientasi minat mahasiswa telah dilakukan menggunakan berbagai metode. Penelitian sebelumnya telah banyak dilakukan dengan menerapkan metode klasifikasi *data mining* oleh peneliti dalam melakukan Analisa klasifikasi orientasi data mahasiswa beserta akurasi, seperti penelitian yang dilakukan (Hastuti, 2012) dimana peneliti tersebut melakukan komparasi untuk mengetahui tingkat akurasi dari masing-masing algoritma. Algoritma yang digunakan adalah *logistic regression*, *decision tree*, *naïve bayes* dan *neural network*. Data yang digunakan sebanyak 3861 mahasiswa program studi Teknik Informatika, Sistem Informasi dan Desain Komunikasi Visual Universitas Dian Nuswantoro. Hasil dari proses klasifikasi dievaluasi dengan menggunakan *cross validation*, *confusion matrix*, *ROC Curve* dan *T-Test* untuk mengetahui algoritma klasifikasi *data mining* yang paling akurat untuk prediksi mahasiswa non aktif. Penelitian lainnya dilakukan oleh (Tahyudin et al., 2013) yang membandingkan beberapa algoritma klasifikasi data mining, khususnya *Decision Tree (DT)*, *Naive Bayes (NB)*, *Artificial Neural Network (ANN)*, *algoritma Support Vector Machine (SVM)* dan *Logistic Regression (LR)* dengan evaluasi validasi silang dan *T-Test* untuk memprediksi kelulusan siswa tepat waktu. Metode yang digunakan adalah metode perbandingan. Penelitian lainnya dari Daniel Swanjaya dan Abidatul Izzah, pada penelitian ini digunakan untuk memprediksi minat dan bakat kerja yang dimiliki oleh mahasiswa dengan menggunakan Decision Tree. Data yang digunakan dalam paper ini adalah data nilai 8 mata kuliah wajib dari 42 alumni jurusan S1 Matematika. Sedangkan lapangan kerja yang tersedia meliputi pengajar, pegawai, dan entrepreneur. Hasil evaluasi *classifier* ini menunjukkan nilai AUC diatas 0.9 untuk masing-masing profesi (Daniel Swanjaya, 2015). Penelitian lainnya dari Liliana Swastina metelusuri tentang pemilihan jurusan yang tidak sesuai dengan kemampuan, kepribadian, minat dan bakat yang dapat mempengaruhi mahasiswa dalam mengikuti perkuliahan. Parameter pemilihan jurusan adalah Indeks Prestasi Kumulatif Semester 1 dan 2, hasil eksperimen dan evaluasi menunjukan bahwa Algoritma Decision Tree C4.5 akurat diterapkan untuk penentuan kesesuaian jurusan mahasiswa dengan tingkat akurasi 93,31 % dan akurasi rekomendasi jurusan sebesar 82,64% (Swastina, 2013).

Untuk mendapatkan hasil yang diinginkan pada penelitian ini, digunakan algoritma *Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor* yang merupakan tiga buah algoritma yang dapat memberikan hasil analisa klasifikasi sesuai yang diharapkan pada penelitian ini, ke tiga metode tersebut merupakan sepuluh klasifikasi data mining paling

populer (Xindong Wu, n.d. 2009). Analisa akan dilakukan dengan membandingkan beberapa metode klasifikasi data mining, diantaranya yaitu Algoritma *Decision Tree C4.5*, *Naive Bayes*, dan *K-Nearest Neighbor*, ketiga metode tersebut digunakan untuk mengetahui model orientasi minat mahasiswa apakah bekerja atau kuliah dan faktor lainnya yang menyebabkan mereka memutuskan melanjutkan studinya di Politeknik Tri Mitra Karya Mandiri berdasarkan nilai akurasi dan *Area Under Curve*. Dengan mengambil beberapa tahapan penelitian dan pengambilan data dari banyaknya orientasi minat mahasiswa yang diteliti sehingga dapat diketahui model dan skema dari orientasi kerja dan orientasi kuliah, analisa juga akan membandingkan hasil akurasi dan *Area Under Curve* dari ketiga algoritma tersebut sehingga dapat diketahui algoritma yang lebih baik dari hasil penelitian dan pengujian data tersebut. Hasil penelitian akan diusulkan kepada pihak manajemen Politeknik Tri Mitra Karya Mandiri tentang kriteria penyebab mahasiswa memilih kampus ini sebagai tujuan studinya berdasarkan orientasi minat mahasiswa sesungguhnya dan faktor-faktor lain yang mendukung, sehingga diharapkan dimasa yang akan datang dapat diperoleh strategi promosi yang tepat, strategi mempertahankan mahasiswa yang aktif dari godaan non aktif ketika mendapatkan pekerjaan, strategi mempertahankan mahasiswa aktif berdasarkan minat dan bakatnya serta hal lainnya yang dirasa perlu berdasarkan kriteria klasifikasi yang dihasilkan.

Masalah Penelitian

Orientasi minat mahasiswa adalah salah satu keinginan yang mendasar dari mahasiswa ketika sebelumnya dia berada pada posisi Pendidikan dibawah perguruan tinggi. Minat yang berbeda, keinginan orang tua, faktor finansial dan juga ketersediaan akan terselurnya bakat yang dimilikinya ketika melanjutkan studi menjadi variable yang turut menentukan jadi tidaknya seorang calon mahasiswa melanjutkan studinya khususnya dikampus Politeknik Tri Mitra Karya Mandiri.

Permasalahan yang timbul adalah dari jumlah 219 mahasiswa baru yang terdaftar pada tahun akademik 2017/2018 ketika diadakan tanya jawab secara langsung ternyata tidak 100% mahasiswa baru tersebut berorientasi untuk kuliah, banyak juga diantara mereka yang berorientasi untuk bekerja. Hal ini tentunya menarik untuk diteliti karena bila pola orientasi mahasiswa ini dapat ditemukan, maka akan banyak hal yang bisa diperbaiki terutama untuk strategi mempertahankan mahasiswa tetap aktif akibat dari godaan dunia industri sebelum mereka selesai kuliah dan juga strategi promosi dan rekrutmen kepada calon mahasiswa baru di masa yang akan datang.

Pada penelitian ini difokuskan pada pencarian algoritma terbaik berdasarkan nilai akurasi dan *Area Under Curve*, dimana hasil dari penelitian tersebut akan ditemukan pola klasifikasi dari *orientasi minat mahasiswa* yang diteliti. Hasil dari pola tadi akan dikaji dan diajukan usulan ke pihak manajemen Politeknik Tri Mitra Karya Mandiri untuk dapat melakukan langkah antisipasi bila terjadi gangguan pada aktifitas perusahaan di masa yang akan datang berdasarkan usulan dari hasil penelitian ini.

Identifikasi Masalah

Analisa klasifikasi orientasi minat mahasiswa telah banyak dibahas pada penelitian-penelitian sebelumnya, tetapi untuk studi kasus yang diangkat pada penelitian ini belum banyak analisa yang dilakukan. Hal ini disebabkan karena klasifikasi yang dihasilkan mencakup dari variabel-variabel standar yang jarang dieksploitasi dan usulan yang akan diberikan dalam melalui suatu model atau pola yang ada. Penelitian ini dilakukan karena

adanya orientasi minat mahasiswa yang ternyata berbeda dan menyebabkan mereka kuliah di kampus ini, sehingga dikuatirkan akan timbul efek yang kurang baik dimasa yang akan datang. Analisa masalah dilakukan dengan menggunakan komparasi tiga buah algoritma yaitu *Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor*.

Untuk mendapatkan solusi dari permasalahan di atas, penelitian difokuskan untuk menjawab pertanyaan riset sebagai berikut :

“*Bagaimana Komparasi Dan Akurasi Dari Algoritma Klasifikasi Untuk Orientasi Minat Mahasiswa Dalam Penuntasan Studi ?* “

Ruang Lingkup Penelitian

Analisa permasalahan yang akan dikaji pada tesis ini dibatasi pada analisa orientasi minat mahasiswa yang ada di Politeknik Tri Mitra Karya Mandiri khususnya mahasiswa tahun akademik 2017/2018, dimana analisa akan dilakukan dengan melakukan komparasi algoritma *Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor* yang digunakan untuk mengetahui tingkat akurasi dan AUC yang terbaik dari ketiga algoritma tersebut berdasarkan orientasi minat mahasiswa yang terjadi, untuk data tes dari hasil development digunakan dengan data dari tahun 2018/2019.

Tujuan dan Manfaat Penelitian

Penelitian ini bertujuan untuk menganalisa seberapa besar pengaruh model yang dihasilkan dalam membuat klasifikasi orientasi minat mahasiswa mempengaruhi minat sesungguhnya dari mahasiswa tersebut. Dengan menggunakan algoritma *Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor* sebagai komparatornya untuk mendapat hasil akurasi dan *Area Under Curve* yang diinginkan. Manfaat dari penelitian ini adalah dengan ditemukannya tingkat akurasi dan AUC yang tepat dari ketiga algoritma yang digunakan diharapkan dapat digunakan untuk membantu dalam menentukan berbagai macam strategi kemahasiswaan di kampus ini pada masa yang akan datang.

Manfaat untuk Politeknik Tri Mitra Karya Mandiri dari penelitian ini adalah dengan didapatnya model atau *rule* dari orientasi minat mahasiswa maka akan lebih mudah bagi pihak manajemen kampus untuk melakukan pemetaan orientasi siswa, menghasilkan strategi yang tepat untuk rekrutmen mahasiswa, yang aktif dan non aktifan mahasiswa serta berbagai macam problema mahasiswa yang berhubungan dengan orientasi minatnya dimasa yang akan datang.

LANDASAN TEORI

Slameto (2010:180) mengemukakan bahwa minat adalah suatu rasa lebih suka dan rasa ketertarikan pada suatu hal atau aktivitas, tanpa ada yang menyuruh. Minat pada dasarnya adalah penerimaan akan suatu hubungan antara diri sendiri dengan sesuatu di luar diri. Semakin kuat atau dekat hubungan tersebut, semakin besar minat. Suatu minat dapat diekspresikan melalui suatu pernyataan yang menunjukkan bahwa siswa lebih menyukai suatu hal daripada hal lainnya, dapat pula dimanifestasikan melalui partisipasi dalam suatu aktivitas. Siswa yang memiliki minat terhadap subyek tertentu cenderung untuk memberikan perhatian yang lebih besar terhadap subyek tersebut. Minat menunjukkan adanya suatu ketertarikan terhadap sesuatu (Chusanawati, 2015). Menurut Crow and Row dalam Djaali mengatakan bahwa minat berhubungan dengan gaya gerak yang mendorong seseorang untuk menghadapi atau berurusan dengan orang lain, benda, kegiatan, pengalaman yang dirangsang oleh

kegiatan itu sendiri (Djaali, 2007:121). Sedang menurut Syah (2009:175) minat melanjutkan studi ke perguruan tinggi adalah ketertarikan siswa untuk melanjutkan studi ke perguruan tinggi yang tumbuh secara sadar dalam diri siswa tersebut (Chusanawati, 2015). Ketertarikan tersebut menyebabkan siswa memberikan perhatian yang lebih terhadap perguruan tinggi yang akan mereka masuki.

Besarnya kecilnya minat seseorang pada sesuatu dapat dipengaruhi oleh beberapa hal seperti siswa itu sendiri, lingkungan yang mendukung, teman, dan keluarga. Syah (2008: 136) di kutip dari Reber (1988), minat tidak termasuk istilah populer dalam psikologi karena ketergantungannya yang banyak pada faktor-faktor internal lainnya seperti: pemusatan perhatian, keingintahuan, motivasi, dan kebutuhan. Menurut Monks, Knoers, dan Haditono (2004) minat dipengaruhi oleh dua faktor (Chusanawati, 2015):

a. Faktor dari dalam (intrinsik) Suatu perbuatan yang memang diinginkan karena seseorang senang melakukannya. Disini minat datang dari dalam diri orang itu sendiri. Orang senang melakukan perbuatan itu demi perbuatan itu sendiri. Seperti: rasa senang, mempunyai perhatian lebih, semangat, motivasi, emosi.

b. Faktor dari luar (ekstrinsik) Suatu perbuatan dilakukan atas dorongan/pelaksanaan dari luar. Orang melakukan perbuatan itu karena ia didorong/dipaksa dari luar. Seperti: lingkungan, orang tua, guru, teman.

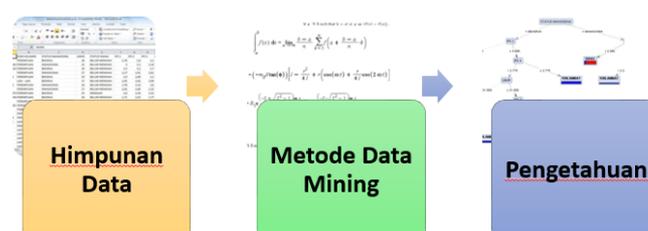
Data Mining

Data Mining adalah kajian yang meliputi kegiatan pengumpulan, pembersihan, pemrosesan, dan analisa sekumpulan data sehingga dengan kegiatan tersebut dapat diperoleh pemahaman yang mendalam akan data (Aggarwal, 2015).

Data mining telah banyak menarik perhatian di dunia sistem informasi dan di masyarakat secara keseluruhan dalam beberapa tahun ini, karena ketersediaan luas dalam jumlah besar data dan kebutuhan segera untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan. Informasi dan pengetahuan yang diperoleh dapat digunakan untuk aplikasi mulai dari pasar analis, deteksi penipuan, dan retensi pelanggan, untuk pengendalian produksi dan ilmu pengetahuan eksplorasi (Han, Kamber, & Pei, 2012).

Banyaknya data, ditambah dengan kebutuhan untuk alat analisis data yang kuat, telah digambarkan sebagai kaya data tapi miskin informasi. Jumlah data yang tumbuh secara cepat, dikumpulkan dan disimpan dalam repositori data yang besar dan banyak, telah jauh melampaui kemampuan manusia untuk memahami data-data tersebut tanpa mampu mengelola data tersebut. Akibatnya, data yang dikumpulkan dalam repositori data yang besar menjadi “kuburan data” (Han et al., 2012).

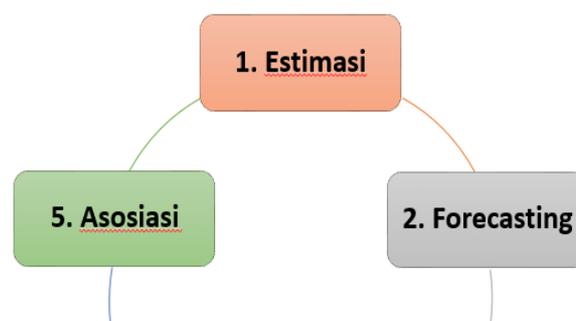
Data mining adalah disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar (Han et al., 2012).



Gambar 1. Konsep Data Mining

Metode Data mining yang banyak digunakan ada 5 metode, yaitu :

1. **Estimasi**
Estimasi adalah metode dalam data mining yang bertujuan untuk menerka sebuah nilai yang belum diketahui (Han et al., 2012). Syarat digunakannya metode ini bila datanya bersifat numerik dan memiliki label, misal menerka penghasilan seseorang ketika informasi mengenai orang tersebut diketahui. Algoritma yang digunakan antara lain *Linear Regression, Neural Network, Support Vector Machine*.
2. **Prediksi**
Prediksi digunakan untuk memperkirakan nilai masa mendatang (Han et al., 2012). Syarat digunakannya metode ini bila datanya numerik dan memiliki time series, misalnya memprediksi harga saham pada kurun waktu tertentu. Algoritma yang digunakan antara lain *Linear Regression, Neural Network, Support Vector Machine*.
3. **Klasifikasi**
Klasifikasi adalah sebuah proses analisa data yang menghasilkan model-model untuk menggambarkan kelas-kelas yang terkandung di dalam data (Han et al., 2012). Syarat digunakannya metode ini apabila atribut datanya numerik atau nominal dan label data nominal, misalnya klasifikasi kelulusan mahasiswa. Algoritma yang digunakan antar lain *Naïve Bayes, K-Nearest Neighbor, Decession Tree C4.5, ID3, CART, Linear Discriminat Analisis, Logistic Regression*.
4. **Klastering**
Clustering (pengelompokan), yaitu pengelompokan mengidentifikasi data yang memiliki karakteristik tertentu (Han et al., 2012). Syarat digunakannya metode ini apabila dataset tidak memiliki label, misalnya klasifikasi bunga iris. Algoritma yang digunakan antara lain *K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means*.
5. **Asosiasi**
Asosiasi dinamakan juga analisis keranjang pasar, dimana fungsi ini mengindetifikasikan item-item produk yang kemungkinan dibeli konsumen bersamaan dengan produk lain atau adakah hubungan antar item dalam suatu transaksi, misalnya ketika konsumen belanja susu pda sebuah mini market, maka seberapa besar peluang konsumen tersebut juga membeli sikat gigi. Algoritma yang digunakan antara lain *Apriori, Generalized Sequential Pattern (GSP), FP-Growth* dan *GRI Algorithm*



Gambar 2. Metode Data Mining

Tahapan Proses Data Mining

Data mining memiliki tiga tahapan proses yang satu sama lain saling terkait, yaitu:

1. Pengumpulan data
Pada tahapan pertama ini, data mentah dikumpulkan menggunakan berbagai perangkat yang sesuai dengan jenis datanya, misalnya program *Crawler* untuk data berbentuk situs web. Pengumpulan data yang baik akan sangat mempengaruhi proses data mining selanjutnya (Aggarwal, 2015).
2. *Pre-processing* data
Data yang telah dikumpulkan pada tahap pertama seringkali tidak dalam bentuk yang sesuai untuk proses analisa menggunakan algoritma *Data Mining*. Oleh karena itu, data tersebut harus di *pre-processing*. Hasil akhir *pre-processing* ini disebut *dataset*, yakni data yang sudah dalam bentuk yang sesuai dengan algoritma yang akan digunakan untuk analisa (Aggarwal, 2015).
3. Analisa data
Tahap ketiga ini merupakan proses yang sangat esensial dimana sebuah metode yang pintar diaplikasikan untuk mengekstrak pola-pola yang terkandung dalam data (Han et al., 2012).
4. Penyajian Pengetahuan
Hasil dari tahap ketiga adalah pola-pola menarik dan pengetahuan dari data yang selanjutnya disajikan kepada user menggunakan teknik-teknik penyajian dan visualisasi pengetahuan (Han et al., 2012). Bentuk penyajian ini bisa berbentuk pohon untuk algoritma *Decision Tree*,

Klasifikasi

Klasifikasi adalah sebuah proses analisa data yang menghasilkan model-model untuk menggambarkan kelas-kelas yang terkandung di dalam data (Han et al., 2012). Model - model tersebut disebut *Classifier*. Jadi, *Classifier* inilah yang akan digunakan untuk menyusun kelas-kelas yang terkandung di dalam data. Ada banyak jenis algoritma klasifikasi, tiga diantaranya adalah *Decision Tree C4.5*, *Naïve Bayes* dan *k-Nearest Neighbour (k-NN)*.

Klasifikasi terdiri memeriksa fitur dari objek yang baru disajikan dan menugaskannya ke salah satu rangkaian kelas yang telah ditentukan. Objek yang akan diklasifikasikan umumnya diwakili oleh catatan dalam tabel *database* atau *field*, dan tindakan klasifikasi dari dari menambahkan kolom baru dengan kode kelas dari beberapa jenis. Tugas klasifikasi dicirikan oleh definisi kelas yang didefinisikan dengan baik, dan seperangkat pelatihan yang

terdiri dari contoh-contoh yang telah diklasifikasikan. Tugasnya adalah membangun model dari beberapa jenis yang dapat diterapkan pada data yang tidak terklasifikasi untuk mengklasifikasikannya.

Klasifikasi *Data mining* adalah suatu metode pembelajaran, untuk memprediksi nilai dari sekelompok atribut dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah learning (fase training), dimana algoritma klasifikasi dibuat untuk menganalisa data training lalu direpresentasikan dalam bentuk rule klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari rule klasifikasi (Han et al., 2012). Dalam klasifikasi data mining, pemrosesan klasifikasi terbagi menjadi empat bagian yaitu (Gorunescu, n.d.):

- a. Kelas yang terdapat pada objek.
Contohnya: resiko penyakit diabetes, resiko kredit, customer loyalty, jenis gempa.
- b. Predictor Variabel independen yang direpresentasikan oleh karakteristik (atribut) data.
Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
- c. Training dataset
Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
- d. Testing dataset
Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat dan akurasi klasifikasi dievaluasi

Decision Tree

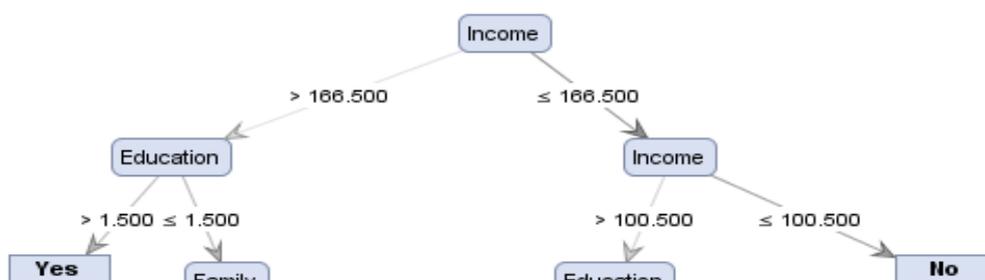
Decision Tree digunakan untuk mempelajari klasifikasi dan prediksi pola dari data dan menggambarkan relasi dari variabel atribut x dan variabel target y dalam bentuk pohon (Ye, 2015). *Decision Tree* adalah struktur menyerupai *flowchart* dimana setiap *internal node* (*node* yang bukan *leaf* atau bukan *node* terluar) merupakan pengujian terhadap variabel atribut, tiap cabangnya merupakan hasil dari pengujian tersebut, sedangkan *node* terluar yakni *leaf* menjadi labelnya (Han et al., 2012).

Algoritma C4.5 dan pohon keputusan merupakan dua model yang tak terpisahkan, karena untuk membangun sebuah pohon keputusan dibutuhkan algoritma C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 dan ID3 diciptakan oleh seorang peneliti di bidang kecerdasan buatan bernama J. Rose Quinlan pada akhir tahun 1970-an. Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, di mana atribut paling atas merupakan akar (*root*), dan yang paling bawah dinamakan daun (*leaf*).

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

1. Pilih atribut sebagai simpul akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Pemilihan atribut sebagai simpul, baik akar (*root*) atau simpul internal didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada.



Gambar 3. Contoh *Decision Tree*

Naïve Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema *Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Marlina, Putera, & Siahaan, 2016). Keuntungan penggunaan *Naive Bayes* adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. *Naive Bayes* sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan. Bentuk umum dari teori *Naive Bayes* seperti dibawah ini:

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)} \quad (2.1)$$

Dimana:

X :Data dengan class yang belum diketahui

H :Hipotesis data X merupakan suatu class spesifik

P(H|X):Probabilitas hipotesis H berdasar kondisi X (posteriori probability)

P(H) :Probabilitas hipotesis H (prior probability)

P(X|H):Probabilitas X berdasar kondisi pada hipotesis H

P(X) :Probabilitas dari X

***k*-Nearest Neighbors (*k*-NN)**

Algoritma k-Nearest Neighbors (k-NN) jenis metode yang tidak mendapatkan model dari *data training*. *k-NN* hanya menyimpan *dataset*, *k-NN* terjadi pada saat prediksi dengan adanya *test case* baru. Dengan adanya *test case* baru, ramalannya didapat mencari kasus serupa dalam *data training* yang tersimpan (Torgo, n.d.).

Dalam masalah klasifikasi, prediksi ini biasanya diperoleh dengan memilih dan dengan demikian angka ganjil untuk *K* diinginkan. Jenis model ini sangat bergantung pada pengertian kesamaan antar kasus. Gagasan ini biasanya didefinisikan dengan bantuan metrik di atas ruang input yang ditentukan oleh *variable predictor*. Metrik ini adalah fungsi jarak yang bisa menghitung angka yang mewakili “perbedaan” antara dua pengamatan. Ada banyak fungsi jarak, namun pemilihan yang agak sering adalah fungsi jarak *Euclidean* yang didefinisikan sebagai berikut:

$$d(X_i, X_j) = \sqrt{\sum_{k=1}^p (X_{i,k} - X_{j,k})^2} \quad (2.2)$$

Dimana *p* adalah jumlah *prediktor*, dan *x_i* dan *x_j* adalah dua pengamatan. Metode ini sangat sensitif terhadap metrik yang dipilih dan juga adanya variabel yang tidak relevan. Selain itu skala variabel harus seragam jika tidak, kita mungkin meremehkan beberapa perbedaan dalam variabel dengan nilai rata-rata lebih rendah. Pilihan jumlah (*k*) juga merupakan parameter penting dari metode ini, nilai yang sering termasuk angkadi himpunan {1, 3, 5, 7, 11} tapi jelas ini hanya *heuristik*. Namun, bisa dikatakan bahwa nilai (*k*) yang lebih besar harus dihindari karena ada resiko menggunakan kasus yang ada.

K-Fold Cross Validation

K-Fold Cross Validation adalah teknik untuk mengevaluasi model dengan membagi data asli menjadi *data training* dan *data testing* yang ditetapkan untuk mengevaluasinya. *K-Fold cross validation* data asli dibagi secara acak menjadi *subset* dengan ukuran yang sama, validasi yang membagi data ke dalam *k* bagian dan kemudian masing-masing bagian akan dilakukan proses klasifikasi. Metode *cross validation* biasa digunakan untuk menghindari *overlapping* pada *data testing*, pada *cross validation* *data testing* dan *data training* otomatis melakukan pemisahan.

Confusion Matrix

Confusion matrix merupakan ringkasan hasil prediksi pada masalah klasifikasi, dimana jumlah prediksi yang benar dan salah dirangkum dengan nilai hitungan dan dipecah oleh masing-masing kelas. *Confusion matrix* adalah teknik untuk meringkas kinerja algoritma klasifikasi. Keakuratan klasifikasi bisa menyesatkan jika memiliki jumlah pengamatan tidak sama di setiap kelas atau memiliki lebih dari 2 (dua) kelas dalam suatu dataset contohnya *Y* atau *X*. Keakuratan klasifikasi yang spesifik bisa dilihat pada tabel 2.1 model *confusion matrix*

Tabel 1. Model *Confusion Matrix*

| | | Actual Class | |
|-----------------|---|-----------------------------|------------------------------|
| | | Y | N |
| Predicted Class | Y | TP (<i>True Positive</i>) | FP (<i>False Positive</i>) |
| | X | | |

| | | |
|------------|------------------------|--------------------|
| (Expected) | FN (False Negative) | TN (True Negative) |
|------------|------------------------|--------------------|

Dari tabel 2.1 tentang prediksi *confusion matrix* akan dijelaskan pada keterangan sebagai berikut

1. Kelas yang diprediksi adalah Y, dan kelas sebenarnya adalah Y ini bernilai “ True Positive” atau TP
2. Kelas yang diprediksi adalah Y, dan kelas sebenarnya adalah N ini bernilai “ False Positive” atau FP
3. Kelas yang diprediksi adalah N, dan kelas sebenarnya adalah Y ini bernilai “ False Negative” atau FN
4. Kelas yang diprediksi adalah N, dan kelas sebenarnya adalah N ini bernilai “ True Negative” atau TN

Confusion matrix disusun secara matriks 2x2 seperti ditunjukkan pada tabel 2.1 Kelas yang diprediksi disusun secara horizontal dalam baris dan kelas sebenarnya disusun secara vertical dalam kolom, walaupun terkadang sebaliknya.

Keempat contoh kasus tersebut akan digunakan untuk memahami dan menjelaskan kinerja klasifikasi, keakuratan klasifikasi merupakan pengukuran yang menunjukkan seberapa baik pengklasifikasi mengidentifikasi dengan benar dari suatu objek. Menghitung *confusion matrix* dapat memberikan suatu gagasan yang lebih baik tentang model suatu algoritma klasifikasi semakin tepat dan jenis mengetahui jenis kesalahan apa yang dibuatnya. Adapun rumusan *confusion matrix* adalah:

- a. *Accuracy* didefinisikan sebagai kemampuan pengklasifikasi untuk memilih semua kasus yang dipilih dan menolak semua kasus yang ditolak. Untuk pengklasifikasi dengan akurasi 100%, berarti FN (False Negative) = FP (False Positive) = 0. *Accuracy* yang diberikan sebagai berikut: $(TP+TN) / (TP+FP+TN+FN)$.
- b. *Sensitivity* adalah kemampuan pengklasifikasi untuk memilih semua kasus yang perlu dipilih. Klasifikasi yang sempurna akan memilih semua produk Y yang sebenarnya dan tidak melewatkan produk Y yang sebenarnya. Dengan kata lain tidak ada False Negative. Pada kenyataannya, setiap klasifikasi akan kehilangan beberapa Y dan dengan demikian ada beberapa false negative. *Sensitivity* dinyatakan sebagai rasio (atau persentase) dihitung sebagai berikut: $TP / (TP + FN)$.
- c. *Specificity* adalah kemampuan pengklasifikasi untuk menolak semua kasus yang perlu ditolak. Klasifikasi yang sempurna akan menolak semua Y dan tidak akan memberikan hasil yang diterapkan. Dengan kata lain, tidak akan ada false positive. Pada kenyataan, setiap pengklasifikasi akan memilih beberapa kasus yang perlu ditolak dengan demikian memiliki beberapa kasus yang perlu ditolak dan dengan demikian ada beberapa false positive. *Specificity* dinyatakan sebagai rasio (atau persentase) dihitung sebagai berikut: $TN / (TN + FP)$.
- d. *Precision* adalah proporsi kasus ditemukan yang sebenarnya relevan. Contoh kasus dengan angka 70 dan dengan demikian presisi adalah 70/100 atau 70%. 70 dokumennya

- adalah TP (*True Positive*), sedangkan sisanya 20 adalah FP (*False Positive*). Oleh karena itu *Precision* dihitung sebagai berikut $TP / (TP + FP)$.
- e. *Recall* didefinisikan sebagai proporsi kasus relevan yang sebenarnya ditemukan diantara semua kasus yang relevan. Contoh kasus 70 dari total 110 (70 ditemukan + 40 terjawab) kasus yang benar-benar ditemukan, sehingga memberi *recall* $70/110 = 62.62\%$. Oleh karena itu *Recall* dihitung sebagai berikut $TP / (TP + FN)$.

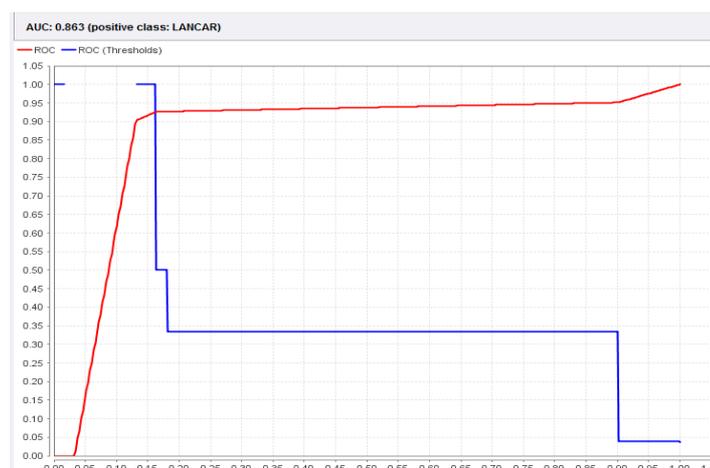
Tabel 2. *Evaluation Measure*

| Term | Calculation |
|--------------------|-------------------------|
| <i>Sensitivity</i> | $TP/(TP+FN)$ |
| <i>Specificity</i> | $TN/(TN+FP)$ |
| <i>Precision</i> | $TP/(TP+FP)$ |
| <i>Recall</i> | $TP/(TP+FN)$ |
| <i>Accuracy</i> | $(TP+TN)/(TP+TN+FP+FN)$ |

Tabel 2. merangkum semua rumus perhitungan dari *accuracy*, *sensitivity*, *specificity*, *precision*, dan *recall*.

Kurva ROC

Kurva ROC adalah grafik antara sensitifitas (*true positive*) pada sumbu Y dengan 1-spesifisitas pada sumbu X (*false positive*). Kurva ROC menggambarkan *trade-off relative* antara *true positive* dan *false positive*. Tetapi untuk mempresentasikan grafis yang menentukan klasifikasi yang lebih baik, digunakan metode yang menghitung luas daerah dibawah kurva ROC yang disebut AUC (*Area Under Curve*). AUC mengukur kinerja dengan memperkirakan *probabilitas output* dari sampel yang dipilih secara acak dari populasi positif atau negatif, semakin besar nilai AUC, semakin baik klasifikasi yang digunakan. Berikut contoh Gambar AUC.



Gambar 4. Contoh Kurva ROC – AUC (Sumber rapidminer)

Gambar 4 menggambarkan kurva ROC dimana sumbu X mewakili *false positive* dan sumbu Y mewakili *true positive*. Untuk mengklasifikasi keakuratan algoritma dengan menggunakan AUC bisa dilihat pada tabel klasifikasi AUC.

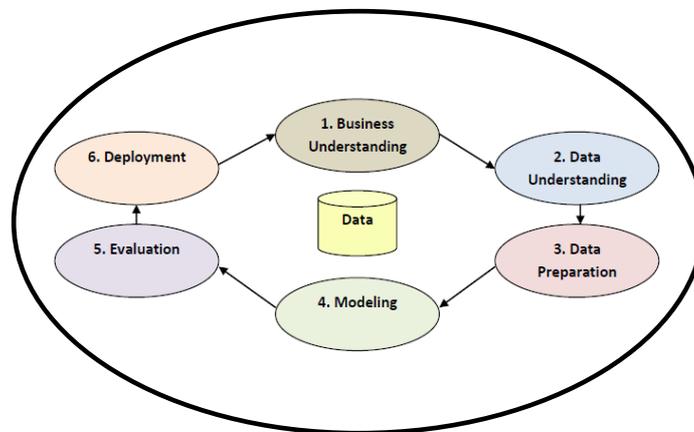
Tabel 3. Klasifikasi AUC

| Performance | Klasifikasi |
|-------------|---------------------------------|
| 0.90 – 1.00 | <i>excellent classification</i> |
| 0.80 – 0.90 | <i>good classification</i> |
| 0.70 – 0.80 | <i>fair classification</i> |
| 0.60 – 0.70 | <i>poor classification</i> |
| 0.50 – 0.60 | <i>failure</i> |

Pada tabel 3. klasifikasi AUC dimana performance yang memiliki nilai 0.50 – 0.60 termasuk dalam klasifikasi *failure*, untuk nilai 0.60 – 0.70 termasuk dalam *poor classification*, untuk nilai 0.70 – 0.80 termasuk dalam *fair classification*, untuk nilai 0.80 – 0.90 termasuk dalam *good classification* dan untuk nilai 0.90 – 1.00 termasuk dalam *excellent classification* (Gorunescu, n.d.).

CRIPS-DM

CRISP-DM atau *Cross Industry Standard Process for Data Mining* adalah sebuah cara yang ditempuh untuk secara sistematis menyelesaikan masalah penelitian. CRISP-DM merupakan hasil kolaborasi dari beberapa perusahaan, diantaranya Daimler-Benz, HRA, NCR Corp., dan SPSS Inc. yang mulai dirintis sejak tahun 1999 (Indra Purnama. Ragil Saputra, 2014).



Gambar 5. Proses CRIPS-DM

Adapun tahapan-tahapan dalam CRIPS-DM adalah sebagai berikut:

1. *Business Understanding*

Pada tahap ini dibutuhkan pemahaman tentang substansi dari kegiatan *data mining* yang akan dilakukan, kebutuhan dari perspektif bisnis. Kejadiannya antara lain: menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.

2. *Data Understanding*

Pada tahap ini dilakukan pengumpulan data awal, mempelajari data untuk bisa mengenal data yang akan dipakai. Fase ini mencoba mengidentifikasi masalah

yang berkaitan dengan kualitas data, mendeteksi subset yang menarik dari data untuk membuat hipotesa awal.

3. *Data Preparation*

Tahap ini sering disebut sebagai tahap yang padat karya. Aktivitas yang dilakukan antara lain memilih tabel dan *field* yang akan ditransformasikan ke dalam *database* baru untuk bahan *data mining* (set data mentah).

4. *Modelling*

Tahap ini menentukan tehnik *data mining* yang digunakan, menentukan *tools data mining*, teknik *data mining*, algoritma *data mining*, menentukan parameter dengan nilai yang optimal.

5. *Evaluation*

Tahap ini interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.

6. *Deployment*

Tahap dimana hasil dari seluruh tahapan sebelumnya digunakan secara nyata.

METODE PENELITIAN

Jenis Penelitian

Penelitian secara umum dapat diartikan sebagai sebuah usaha untuk mencari pengetahuan. Pencarian pengetahuan ini melalui metode pencarian solusi dari permasalahan secara objektif dan sistematis (Kothari, n.d.). Adapun dua jenis pendekatan penelitian yang utama adalah:

a. Metode penelitian Kualitatif

Metode penelitian kualitatif adalah metode penelitian yang berhubungan dengan penilaian subjektif dari sikap, pendapat, dan perilaku. Secara umum teknik yang digunakan adalah interview pada kelompok tertentu dan wawancara yang mendalam (Kothari, 2004).

b. Metode Penelitian Kuantitatif

Metode penelitian kuantitatif adalah metode penelitian yang melakukan penelitian pada sample tertentu, pengumpulan data menggunakan instrumen penelitian, analisis data bersifat kuantitatif/statistik dengan tujuan untuk menguji hipotesis yang telah ditentukan (Kothari, 2004).

Untuk itu dalam penelitian ini digunakan metode penelitian kualitatif, adapun desain riset yang digunakan dalam penelitian ini adalah eksperimen. Metode penelitian eksperimen dapat diartikan sebagai metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendalikan. Adapun data yang dicari didapatkan melalui cara penyebaran kuisioner, ini dilakukan untuk mencari hasil dari variable yang menyebabkan seseorang terpengaruh dalam mengambil suatu.

Jenis penelitian eksperimen dibagi dua, yaitu:

a. Eksprimen Absolut

Eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen, misalnya pengaruh honor dosen terhadap kinerja (Kothari, 2004).

b. Eksperimen Komparatif. Eksperimen

Eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing mana yang lebih baik (Kothari, 2004).

Pada penelitian ini, jenis penelitian yang diambil adalah Eksperimen Komparatif. Dalam penelitian ini mengkomparasi penerapan metode *Decision Tree C4.5*, *Naïve Bayes*, dan *K-Nearest Neighbor* yang dalam hal pengujian ketiga metode akan dipilih salah satu metode yang paling baik tingkat akurasi serta *Area Under Curve* dari masing-masing algoritma.

Metode Pemilihan Populasi dan Sampel

Populasi

Populasi adalah wilayah generalisasi yang terdiri atas: obyek/subyek yang mempunyai kualitas dan karakteristik tertentu yang ditetapkan oleh peneliti untuk dipelajari dan kemudian ditarik kesimpulannya (Kothari, 2004). Populasi yang ada dalam penelitian ini adalah populasi mahasiswa aktif yang tercatat pada Politeknik Tri Mitra Karya Mandiri Tahun Akademik 2017/2018, dimana data mahasiswa tersebut didapat dari data internal bagian akademik Politeknik Tri Mitra Karya Mandiri.

Sampel

Sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi tersebut. Penelitian dapat menggunakan sampel yang diambil dari populasi. Apa yang dipelajari dari sampel itu, kesimpulannya akan dapat diberlakukan untuk populasi. Untuk itu sampel yang diambil dari populasi harus betul-betul *representatif* (mewakili) (Kothari, 2004).

Sampel yang digunakan untuk penelitian ini adalah mahasiswa Politeknik Tri Mitra Karya Mandiri tahun akademik 2017/2018 semester 1, dimana mahasiswa tersebut merupakan mahasiswa baru pada perguruan tinggi ini sehingga lebih *representative* guna kepentingan pengambilan sampel data pada pembuatan tesis ini. Data ini adalah data kuisioner yang disebarkan pada mahasiswa baru pada saat matkulasi dilakukan di lingkungan kampus, dan yang digunakan untuk penelitian ini dan telah mendapatkan persetujuan dari pihak manajemen Politeknik Tri Mitra Karya Mandiri khususnya bagian akademik.

Adapun data tersebut dapat diketahui pada tabel 6. dibawah ini

Tabel 6. Data Mahasiswa Politeknik TMKN

| PROGRAM STUDI | JUMLAH MAHASISWA | AKTIF | TIDAK AKTIF |
|---------------|------------------|-----------------|-----------------|
| 4 | 987 | 865 (87,63%) | 122 (12,37%) |

Adapun data sampel yang diambil ditunjukkan pada tabel 7. dibawah ini

Tabel 7. Data Sampel Mahasiswa Non Aktif

| PROGRAM STUDI | JUMLAH MAHASISWA | AKTIF | TIDAK AKTIF |
|---------------|------------------|---------------|-------------|
| 4 | 219 | 219 (100%) | 0 (0%) |

Data sampel yang diambil masih mencapai 100% karena seluruh mahasiswa baru hadir pada saat penyebaran kuis dan mengembalikan jumlah kuisisioner yang diedarkan sesuai dengan jumlah yang dibagikan yaitu 219 kuisisioner, diharapkan data sampel tersebut mewakili dari data penelitian yang akan dilakukan

Metode Pengumpulan Data

Metode pengumpulan data dibagi menjadi dua sumber data yaitu data primer dan data sekunder. Data primer yaitu data yang dikumpulkan pertama kali, melalui observasi, interview, kuesioner, dan lain lain. Data sekunder adalah data yang telah dikumpulkan dan dianalisis oleh orang lain baik yang telah dipublikasikan maupun yang belum dipublikasikan, misalnya dari dokumentasi, literatur, buku, jurnal, dan informasi lainnya yang ada hubungannya dengan masalah yang diteliti.

Dalam penelitian ini metode pengumpulan data untuk mendapatkan sumber data yang digunakan adalah metode pengumpulan data primer dengan cara kuisisioner, sedangkan data pendukung lainnya didapat dari buku, jurnal dan publikasi lainnya.

Instrumen Penelitian

Pada tesis ini digunakan beberapa instrumen penelitian yang menunjang hasil dari tesis ini. Adapun instrument-instrumen yang digunakan adalah;

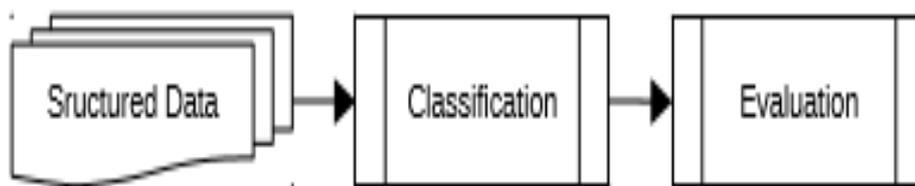
1. Penelitian ini menggunakan data primer berupa data orientasi minat mahasiswa yang dihasilkan dari kuisisioner yang diedarkan dan akan digunakan sebagai instrumen guna memperoleh data dalam proses orientasi minat mahasiswa
2. Data disajikan dalam bentuk Tabulasi model dan variabel masing- masing sebanyak 219 orientasi mahasiswa terdiri dari 63,50% orientasi mahasiswa kuliah dan 36,50% orientasi mahasiswa kerja.
3. Perangkat lunak yang digunakan untuk menganalisis adalah *Rapidminer*

Kerangka Kerja Penelitian

Metodologi penelitian adalah sebuah cara yang ditempuh untuk secara sistematis menyelesaikan masalah penelitian. Melalui metodologi penelitian ini kita dapat mempelajari langkah-langkah umum yang dilaksanakan oleh para peneliti dalam mempelajari permasalahan riset mereka (Kothari, 2004). Untuk penelitian *data mining*, telah ada metodologi standar yang disebut CRISP-DM atau *Cross Industry Standard Process for Data*

Mining. CRISP-DM merupakan hasil kolaborasi dari beberapa perusahaan, diantaranya Daimler-Benz, HRA, NCR Corp., dan SPSS Inc. yang mulai dirintis sejak tahun (Indra Purnama, Ragil Saputra, 2014).

Menggunakan klasifikasi berbasis ML menggunakan *dataset* yang bentuknya sudah terstruktur sebagai masukan untuk algoritma klasifikasi (Bramer, 2013). Kerangka kerja klasifikasi secara garis besar dapat digambarkan dalam diagram sebagai berikut (Sutoyo, 2012):



Gambar 6. Framework Penelitian Klasifikasi

1. Structured Data

Data masukan untuk algoritma klasifikasi yang tipe data atribut dan labelnya telah sesuai dengan algoritma klasifikasi yang akan digunakan. Untuk *Decision Tree* misalnya maka atributnya *numerik (integer, real)* sedangkan labelnya kategori (*binomial, polinomial*).

2. Classification

Berdasarkan data masukan yang tipe data atribut dan labelnya telah sesuai maka algoritma klasifikasi akan memproses data tersebut untuk menghasilkan pengetahuan. Untuk *Decision Tree* misalnya maka pengetahuan yang didapatkan direpresentasikan dalam bentuk Pohon Keputusan dengan aturan if-then.

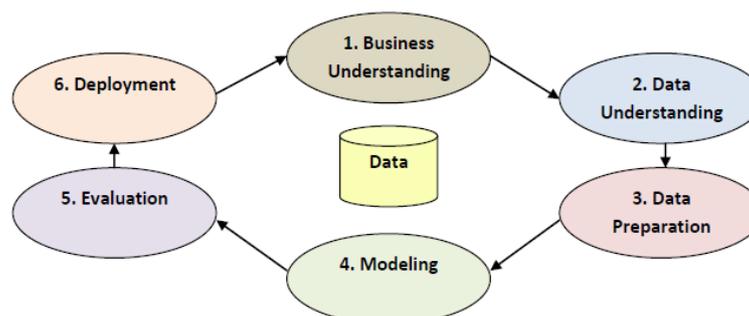
3. Evaluation

Pengetahuan yang telah didapatkan selanjutnya dievaluasi untuk diverifikasi kehandalannya. Contoh bentuk evaluasi ini adalah *k-fold cross validation* dengan k adalah jumlah pengulangan proses evaluasi.

Contohnya, 10-fold cross validation yang akan mengulang percobaan sebanyak 10 kali dengan hasil akhir adalah nilai rata-rata dari 10 kali percobaan tersebut.

Tahapan Analisa

Secara umum tahapan yang akan dilakukan dalam melakukan analisa efektifitas sesuai dengan metode CRIPS-DM adalah sebagai berikut:



Gambar 7. CRIPS-DM Yang Dimodifikasi

1. *Business Understanding*

Pada tahap ini mulai disusun kerangka penelitian yang akan dilakukan, tujuan instansi yang akan dilibatkan, data yang akan digunakan, waktu pelaksanaan serta hasil apa yang ingin dicapai. Pada tahap ini pula struktur pekerjaan dari sisi manajerial diolah dan dan ditetapkan.

Berdasarkan dari hasil data yang dihasilkan dari kuisisioner tentang orientasi minat mahasiswa dapat diketahui bahwa jumlah data yang di catat sebanyak 219 data sesuai jumlah mahasiswa baru, dimana tercatat 139 data mahasiswa berorientasi kuliah dan 80 mahasiswa berorientasi Hasil data akan diolah dan dibuatkan model untuk mengetahui pola type orientasi minat mahasiswa. Kemudian data akan diolah untuk mencari nilai akurasi dan AUC dari algoritma yang dipilih. Hal ini dilakukan dengan memanfaatkan teknik klasifikasi data mining model *Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor*.

2. *Data Understanding*

Pada tahap ini dilakukan seting data yang akan diambil dan juga melakukan pengumpulan data. Adapun Pengumpulan data dilakukan dengan mengambil data pada Politeknik Tri Mitra Karya Mandiri khususnya mahasiswa baru tahun akademi 2017/2018 sebagai objek penelitian.

3. *Data Preparation*

Pada tahap ini mulai dibagi data yang didapat diatas, data dibagi dalam label dan atribut yang telah ditentukan dan membuang data-data yang tidak dibutuhkan dalam keperluan Analisa, sehingga didapat hasil data yang siap dianalisa.

4. *Modelling*

Pada tahap ini mulai ditentukan algoritma yang digunakan dan melakukan analisis data berdasarkan algoritma yang telah ditentukan. Untuk penelitian digunakan *algoritma Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor* sebagai alat ukur perbandingan tingkat akurasi nan nilai AUC analisa penelitian.

5. *Evaluation*

Pada tahap ini diuji hasil dari Analisa yang dilakukan serta memperbaiki bila ditemukan kesalahan dalam penggunaan data ataupun algoritma. Pada tahap ini pula dipastikan bahwa Analisa tersebut dapat digunakan sebagaimana yang diinginkan dan mendapat hasil sesuai dengan apa yang direncanakan

6. *Deployment*

Pada tahap ini akan dilakukan uji kecocokan antara hasil yang keluar dari algoritma berdasarkan tool *Rapidminer* dengan hasi yang keluar melalui MS. Excel berdasarkan dari rule algoritma yang terpilih, kemudiah hasil penelitian akan diserahkan kepada pihak/instansi terkait dalam hal ini manajemen Politeknik Tri Mitra Karya Mandiri guna dapat ditindak lanjuti solusi kedepan yang diinginkan berdasarkan pola model yang didapat dari penelitian tersebut.

HASIL DAN PEMBAHASAN

Implementasi Metodologi

Sesuai dengan metodologi penelitian yang telah dipaparkan pada bab 3, berikut implementasi metodologi tersebut pada penelitian yang kami laksanakan:

Business Understanding

Tahapan I dari CRISP-DM adalah *Business Understanding*. Pada tahapan pertama ini kami membuat beberapa pertanyaan.

- a. Pertanyaan pertama, apa tujuan dari penelitian ini?
Jawabannya adalah tujuan dari penelitian ini adalah mengetahui klasifikasi dari minat sesungguhnya dari mahasiswa di Politeknik Tri Mitra Karya Mandiri, apakah keinginan dari mahasiswa tersebut adalah kuliah atau bekerja berdasarkan kategori-kategori yang telah ditentukan dari hasil penelitian.
- b. Mengapa minat mahasiswa bekerja atau kuliah sebagai objek penelitian?
Jawabannya adalah karena terus menurunnya jumlah mahasiswa per angkatan yang aktif mengikuti perkuliahan dari mulai pendaftaran hingga masuk ke semester akhir. Hal ini tentu saja berimbas pada menurunnya jumlah pendapatan yang dihasilkan dari mahasiswa dan juga mengakibatkan penilaian yang kurang baik dari LLDIKTI terhadap kinerja kampus dengan banyaknya mahasiswa yang menjadi non aktif.
- c. Data orientasi minat mahasiswa manakah yang akan menjadi pilihan untuk diteliti?
Jawabannya adalah data mahasiswa tahun akademik 2017/2018.
- d. Manakah algoritma yang terbaik untuk membuat model klasifikasi orientasi minat mahasiswa?

Untuk mendapatkan jawabannya maka diperlukan penelitian yang sifatnya perbandingan dengan membandingkan tiga jenis algoritma. Pada penelitian ini digunakan algoritma *Decision Tree C4.5*, *Naïve Bayes* dan *K- Nearest Neighbor*

Data Understanding

Tahapan II dari CRISP-DM adalah *Data Understanding*. Pada tahapan kedua ini kami meneliti minat mahasiswa yang dilakukan pada mahasiswa tahun akademik 2017/2018 dengan menggunakan metode kuis atau angket untuk menghasilkan data yang diinginkan, kemudian menyusunnya untuk digunakan pada tahap selanjutnya, yaitu tahap *Data Preparation*. Berikut langkah-langkah yang telah di laksanakan.

1. Melakukan kuis atau angket kepada mahasiswa dan menjadikan data hasil kuis tersebut sebagai basis data yang akan digunakan dalam penelitian ini.

Tabel 8. Data Hasil Angket/Kuis Minat Mahasiswa Tahun Akademik 2017/2018

| LAPORAN KUISIONER MINAT MAHASISWA | | | | | | |
|------------------------------------|-----|------|---------------------|-----------------------|---------------|-------------|
| TAHUN AKADEMIK 2017/2018 | | | | | | |
| POLITEKNIK TRI MITRA KARYA MANDIRI | | | | | | |
| NO. URUT | L/P | USIA | PEKERJAAN ORANG TUA | PENGHASILAN ORANG TUA | FAKTOR UTAMA | ORIENTATION |
| 0001 | L | 18 | PNS | 4,500,000 | ORANG TUA | KERJA |
| 0002 | L | 18 | KARYAWAN | 3,500,000 | ORANG TUA | KERJA |
| 0003 | L | 17 | PETANI | 4,500,000 | ORANG TUA | KERJA |
| 0004 | L | 18 | PNS | 4,500,000 | ORANG TUA | KERJA |
| 0005 | L | 17 | WIRASWATA | 4,500,000 | PROGRAM STUDI | KERJA |
| 0006 | L | 17 | WIRASWATA | 4,500,000 | PROGRAM STUDI | KERJA |
| 0007 | L | 18 | PNS | 5,000,000 | PROGRAM STUDI | KERJA |
| 0008 | L | 17 | WIRASWATA | 4,500,000 | PROGRAM STUDI | KERJA |
| 0009 | L | 18 | PNS | 5,000,000 | PROGRAM STUDI | KERJA |
| 0010 | L | 17 | PETANI | 4,500,000 | ORANG TUA | KERJA |
| 0011 | L | 18 | PNS | 5,000,000 | ORANG TUA | KERJA |
| 0012 | L | 18 | KARYAWAN | 3,500,000 | ORANG TUA | KERJA |
| 0013 | L | 18 | PNS | 5,000,000 | ORANG TUA | KERJA |
| 0014 | L | 17 | PETANI | 4,500,000 | ORANG TUA | KERJA |
| 0015 | L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| 0016 | L | 18 | KULIAH | 4,500,000 | OLAHRAGA | KERJA |
| 0017 | L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| 0018 | L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| 0019 | P | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| 0020 | L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| 0021 | L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| 0022 | L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| 0023 | L | 18 | KARYAWAN | 3,500,000 | OLAHRAGA | KERJA |
| 0024 | L | 17 | PNS | 3,500,000 | OLAHRAGA | KERJA |

- Melakukan verifikasi dan klarifikasi data kepada pihak Politeknik Tri Mitra Karya Mandiri khususnya pada bagian akademik tentang hasil kuisisioner terhadap atribut-atribut data yang dirasa *ambigu*, sehingga didapatkan data yang valid yang sesuai.
- Memberikan tambahan data yang sesuai berdasarkan hasil verifikasi dan klarifikasi sehingga data dapat diolah untuk penelitian guna mendapatkan hasil yang diharapkan.

Data Preparation

Tahapan III dari CRISP-DM adalah *Data Preparation*. Pada tahapan ketiga ini dilakukan cleaning dan pembersihan data sehingga didapatkan *dataset* yang akan siap diolah pada tahap berikutnya yaitu modeling. Tahapan ini dilakukan dengan menghilangkan attribute-atribut yang tidak mempengaruhi proses pengolahan data. Data hasil *cleaning* terdiri dari 5 atribut dan 1 buah label. Adapun pembagian atribut dan label dapat dilihat pada tabe 4.2 dibawah ini.

Tabel 9. Pembagian Atribut Dan Label

| Nama | Tipe | Jenis |
|-----------------------|------------|---------|
| L/P | Binominal | Atribut |
| Usia | Text | Atribut |
| Pekerjaan Orang Tua | Polynomial | Atribut |
| Penghasilan Orang Tua | Integer | Atribut |
| Faktor Utama | Polynomial | Atribut |
| Orientation | Binominal | Label |

Dari data yang ada dan di olah berdasarkan atribut dan label serta dibersihkannya data dari variable-variabel yang tidak diperlukan, maka *dataset* yang dihasilkan dari proses ini adalah sebagai berikut

Tabel 10. Dataset Minat Mahasiswa Politeknik Tri Mitra Karya Mandiri Tahun Akademik 2017/2018

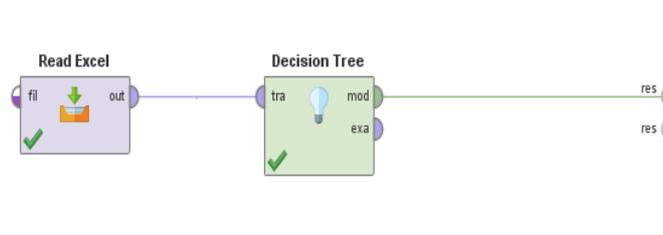
| L/P | USIA | PEKERJAAN ORANG TUA | PENGHASILAN ORANG TUA | FAKTOR UTAMA | ORIENTATION |
|-----|------|---------------------|-----------------------|---------------|-------------|
| L | 18 | PNS | 4,500,000 | ORANG TUA | KERJA |
| L | 18 | KARYAWAN | 3,500,000 | ORANG TUA | KERJA |
| L | 17 | PETANI | 4,500,000 | ORANG TUA | KERJA |
| L | 18 | PNS | 4,500,000 | ORANG TUA | KERJA |
| L | 17 | WIRASWATA | 4,500,000 | PROGRAM STUDI | KERJA |
| L | 17 | WIRASWATA | 4,500,000 | PROGRAM STUDI | KERJA |
| L | 18 | PNS | 5,000,000 | PROGRAM STUDI | KERJA |
| L | 17 | WIRASWATA | 4,500,000 | PROGRAM STUDI | KERJA |
| L | 18 | PNS | 5,000,000 | PROGRAM STUDI | KERJA |
| L | 17 | PETANI | 4,500,000 | ORANG TUA | KERJA |
| L | 18 | PNS | 5,000,000 | ORANG TUA | KERJA |
| L | 18 | KARYAWAN | 3,500,000 | ORANG TUA | KERJA |
| L | 18 | PNS | 5,000,000 | ORANG TUA | KERJA |
| L | 17 | PETANI | 4,500,000 | ORANG TUA | KERJA |
| L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| L | 18 | KULIAH | 4,500,000 | OLAHRAGA | KERJA |
| L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| P | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |
| L | 17 | PETANI | 3,500,000 | OLAHRAGA | KERJA |

Modelling

Tahapan IV dari CRISP-DM adalah *Modelling*. Pada tahapan ini, *dataset* yang telah dibuat pada tahap sebelumnya digunakan sebagai masukan untuk algoritma klasifikasi. Pada penelitian ini akan digunakan tiga jenis algoritma, yaitu, *Decision Tree*, *C4.5*, *Naïve Bayes* dan *k-Nearest Neighbor*. Adapun untuk Analisa dan pengujian dilakukan dengan menggunakan *Rapidminer* sebagai *software* bantu Analisa dan pengujiannya

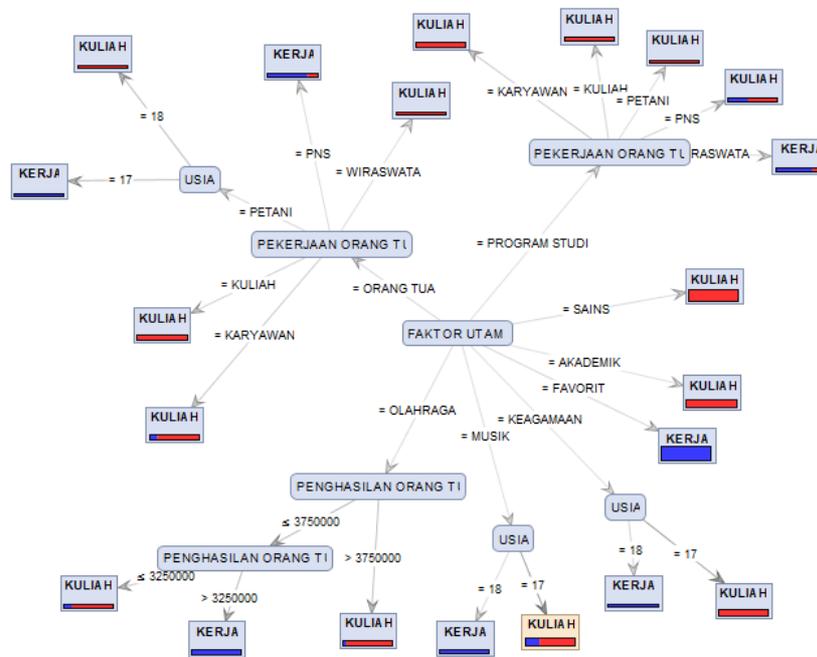
Decision Tree C4.5

Tujuan utama dari menganalisis data dengan menggunakan algoritma *Decision Tree C4.5* ini yaitu ingin mendapatkan *rule* (Ralf Kilkenberg Markus Hofmann, 2014), dimana dari *rule* yang dihasilkan tersebut akan dimanfaatkan untuk pengambilan keputusan pada data baru. *Rule* ini akan didapat berdasarkan dari model pada gambar 4.1.



Gambar 8. Proses *Decision Tree*

Dari hasil proses *Decision Tree* tersebut akan menghasilkan suatu model pohon keputusan yang dihasilkan oleh *Rapidminer* seperti terlihat pada gambar 4.2 dibawah ini.



Gambar 9. Model Decision Tree C4.5 (Pohon Keputusan)

Selain menghasilkan pohon keputusan dalam bentuk grafik, model ini juga menampilkan uraian dari gambar tersebut kedalam bentuk deskripsi rule seperti yang terlihat pada gambar 4.3 dibawah ini **Tree**

```

FAKTOR UTAMA = AKADEMIK: KULIAH {KERJA=0, KULIAH=17}
FAKTOR UTAMA = FAVORIT: KERJA {KERJA=43, KULIAH=0}
FAKTOR UTAMA = KEAGAMAAN
| USIA = 17: KULIAH {KERJA=0, KULIAH=12}
| USIA = 18: KERJA {KERJA=3, KULIAH=0}
FAKTOR UTAMA = MUSIK
| USIA = 17: KULIAH {KERJA=6, KULIAH=15}
| USIA = 18: KERJA {KERJA=3, KULIAH=0}
FAKTOR UTAMA = OLAHRAGA
| PENGHASILAN ORANG TUA > 3750000: KULIAH {KERJA=1, KULIAH=12}
| PENGHASILAN ORANG TUA ≤ 3750000
| | PENGHASILAN ORANG TUA > 3250000: KERJA {KERJA=9, KULIAH=0}
| | PENGHASILAN ORANG TUA ≤ 3250000: KULIAH {KERJA=1, KULIAH=5}
FAKTOR UTAMA = ORANG TUA
| PEKERJAAN ORANG TUA = KARYAWAN: KULIAH {KERJA=2, KULIAH=10}
| PEKERJAAN ORANG TUA = KULIAH: KULIAH {KERJA=0, KULIAH=10}
| PEKERJAAN ORANG TUA = PETANI
| | USIA = 17: KERJA {KERJA=3, KULIAH=0}
| | USIA = 18: KULIAH {KERJA=0, KULIAH=3}
| PEKERJAAN ORANG TUA = PNS: KERJA {KERJA=4, KULIAH=1}
| PEKERJAAN ORANG TUA = WIRASWATA: KULIAH {KERJA=0, KULIAH=4}
FAKTOR UTAMA = PROGRAM STUDI
| PEKERJAAN ORANG TUA = KARYAWAN: KULIAH {KERJA=0, KULIAH=7}
| PEKERJAAN ORANG TUA = KULIAH: KULIAH {KERJA=0, KULIAH=4}
| PEKERJAAN ORANG TUA = PETANI: KULIAH {KERJA=0, KULIAH=2}
| PEKERJAAN ORANG TUA = PNS: KULIAH {KERJA=2, KULIAH=3}
| PEKERJAAN ORANG TUA = WIRASWATA: KERJA {KERJA=3, KULIAH=1}
FAKTOR UTAMA = SAINS: KULIAH {KERJA=0, KULIAH=33}
    
```

Gambar 10. Deskripsi Decision Tree

Dari *rule* yang terlihat pada pohon keputusan dan *Deskripsi Decision Tree* terlihat pola/model minat mahasiswa. Menggunakan status orientation yang dijadikan label dapat diketahui model orientation mahasiswa Politeknik Tri Mitra Karya Mandiri Tahun Akademik 2017/2018. Adapun *rule* atau model yang didapat adalah:

1. Bila faktor utama berdasarkan akademik atau beasiswa maka mahasiswa tersebut akan memilih kuliah sebagai orientasinya.
2. Bila faktor utama berdasarkan perguruan tinggi favorit atau kampus unggulan maka mahasiswa tersebut akan memilih kerja sebagai orientasinya.
3. Bila faktor utama berdasarkan adanya kegiatan keagamaan yang baik di kampus, dan memiliki usia 17 tahun maka mahasiswa tersebut akan memilih kuliah sebagai orientasinya dan bila diatas 17 tahun maka akan memilih kerja sebagai orientasinya.
4. Bila faktor utama berdasarkan adanya kegiatan dibidang music atau kesenian di kampus, dan memiliki usia 17 tahun maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya dan bila diatas 17 tahun maka akan memilih kerja sebagai orientasinya.
5. Bila faktor utama berdasarkan adanya kegiatan dibidang olahraga di kampus, dan orangtuanya mempunyai penghasilan diatas 3.750.000 maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya dan bila orangtuanya mempunyai penghasilan dibawah 3.750.000 namun diatas 3.250.00 maka akan memilih kerja sebagai orientasinya, tapi bila mempunyai penghasilan dibawah 3.250.00 maka akan memilih kuliah sebagai orientasinya.
6. Bila faktor utama berdasarkan pengaruh dari orang tua dan pekerjaan orang tua dari mahasiswa tersebut adalah karyawan, maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya, bila pekerjaan orang tua dari mahasiswa tersebut adalah pekerja lapangan, maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya, bila pekerjaan orang tua dari mahasiswa tersebut adalah PNS, maka mahasiswa tersebut cenderung untuk memilih kerja sebagai orientasinya, bila pekerjaan orang tua dari mahasiswa tersebut adalah wiraswata, maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya, namun bila pekerjaan orang tua dari mahasiswa tersebut adalah petani dan berusia 17 tahun, maka mahasiswa tersebut cenderung untuk memilih bekerja sebagai orientasinya, bila diatas 17 tahun maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya
7. Bila faktor utama berdasarkan program studi yang ada di kampus dan pekerjaan orang tua dari mahasiswa tersebut adalah karyawan, maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya, bila pekerjaan orang tua dari mahasiswa tersebut adalah pekerja lapangan, maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya, bila pekerjaan orang tua dari mahasiswa tersebut adalah PNS, maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya, bila pekerjaan orang tua dari mahasiswa tersebut adalah wiraswata, maka mahasiswa tersebut cenderung untuk memilih kerja sebagai orientasinya, namun bila pekerjaan orang tua dari mahasiswa tersebut adalah, maka mahasiswa tersebut cenderung untuk memilih kuliah sebagai orientasinya.
8. Bila faktor utama berdasarkan prestasi dibidang sains dari kampus ini maka mahasiswa tersebut akan memilih kuliah sebagai orientasinya.

Bentuk *rule* atau model yang terbentuk dari pohon keputusan dan deskripsi *Decision Tree* adalah model yang terjadi pada type incident log pada server dari tahun 2016 hingga 2018

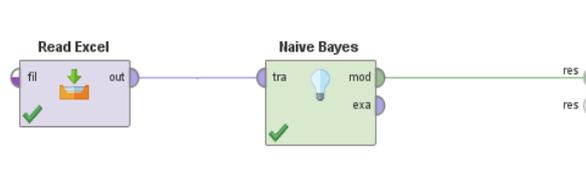
Naïve Bayes

Dalam membuat model *Naive Bayes* terlebih dahulu kita mencari probabilitas hipotesis untuk masing-masing Kelas P(H). Hipotesis yang ada yaitu mahasiswa beorientasi kuliah dan kerja. *Dataset* yang digunakan sama seperti pengujian algoritma *Decision Tree C4.5*, dengan total data yaitu 219 data orientasi mahasiswa dengan 139 mahasiswa beorientasi kuliah dan 80 mahasiswa berorientasi kerja perhitungan probabilitas yaitu seperti dibawah ini:

$$P(\text{Kuliah}) = 139 : 219 = 0.634703196347032$$

$$P(\text{Kerja}) = 80 : 219 = 0.365296803652968$$

Proses penggunaan *RapidMiner* untuk algoritma *Naive Bayes* dapat dilihat pada gambar 11 dibawah ini:



Gambar 11. Proses *Naive Bayes*

Dari proses ini didapat simple distribution yang digambarkan pada gambar dibawah ini.

SimpleDistribution

Distribution model for label attribute ORIENTATION

Class KERJA (0.365)
5 distributions

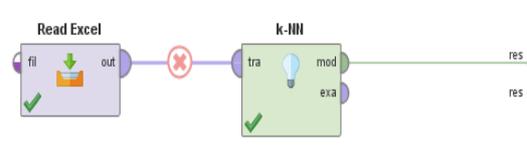
Class KULIAH (0.635)
5 distributions

Gambar 12. Simple Distribution *Naive Bayes*

Probabilitas 0,635 dari 6 distribusi dan orientasi mahasiswa kerja mendapat nilai probabilitas 0,365 dari 6 distribusi. Artinya perhitungan manual probabilities hipotesis untuk kelas orientasi mahasiswa kuliah dan kerja sesuai dengan hasil yang keluar dari proses aplikasi *Rapidminer*.

K- Nearest Neighbor

Adapun model yang terbentuk dari proses *K-Nearest Neighbor* adalah sebagai berikut:



Gambar 13. Proses *K-Nearest Neighbor*

KNNClassification

1-Nearest Neighbour model for classification.

The model contains 219 examples with 5 dimensions of the following classes:

KERJA
KULIAH

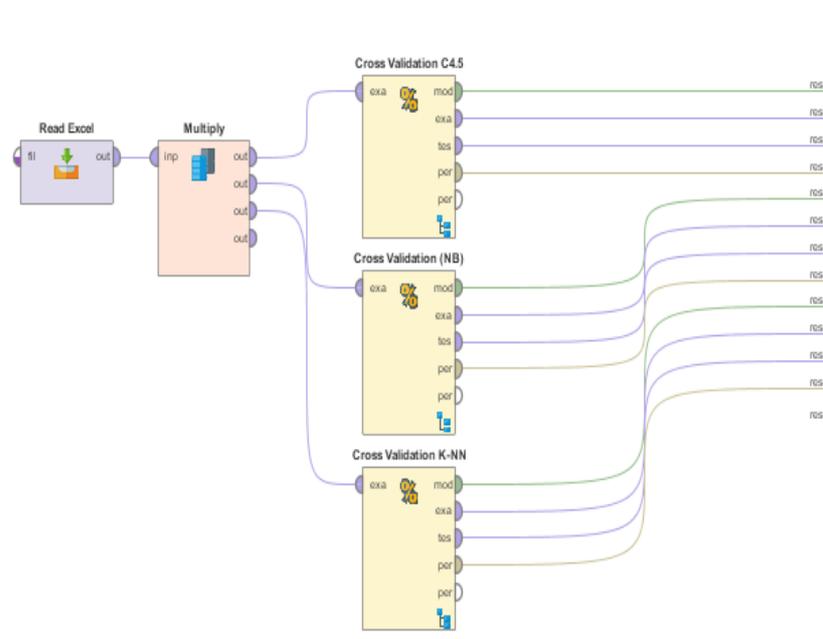
Gambar 14. Klasifikasi *K-Nearest Neighbor*

Dari *KNN Classification* dapat dijabarkan bahwa dari 219 data dengan 5 buah dimensi/atribut dan 1 buah label menghasilkan 2 buah kelas yaitu kelas Kerja dan Kuliah.

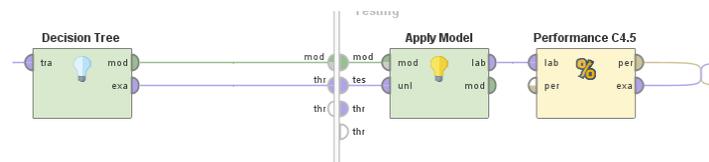
Evaluation

Tahap V dari CRISP-DM adalah *evaluation*. Evaluasi bertujuan untuk menentukan nilai kegunaan dari model yang telah berhasil kita buat pada langkah sebelumnya. Untuk evaluasi digunakan *10-fold cross validation* untuk dapat menghasilkan nilai akurasi dan nilai *Area Under Curve* sebagai bahan perbandingan dalam menentukan algoritma terbaik yang digunakan. *10-fold cross validation* bekerja dengan membagi *dataset* masukan menjadi 10 bagian yang sama rata. 9 bagian kemudian di-training sedangkan yang 1 bagian lainnya digunakan untuk testing. Proses ini diulang sebanyak 10 kali untuk setiap bagian sehingga setiap bagian dari kesepuluh bagian pernah menjadi data untuk testing. *Operator Cross Validation* melakukan proses *10-fold cross validation* ini untuk ketiga algoritma yang digunakan. Untuk setiap percobaan akan dihitung akurasi. Akurasi akhir adalah nilai rata-rata dari akurasi sepuluh percobaan tersebut. Hasilnya dapat disajikan dalam bentuk *confusion matrix*.

Adapun desain penggunaan *10-fold cross validation* dapat dilihat pada gambar 4.8 dibawah ini

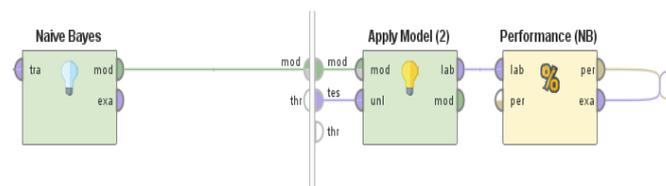


Gambar 15. Proses Perbandingan C4.5, Naive Bayes dan k-NN



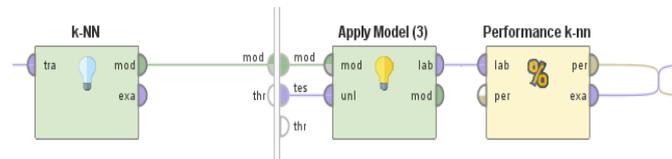
Gambar 16. Desain Proses 10 Fold Cross Validation untuk C4.5

Gambar 16 adalah desain proses yang berada di dalam operator *Cross Validation C4.5* yang terdapat pada gambar 15. Algoritma *Decision Tree C4.5* digunakan untuk melakukan training terhadap *dataset* masukan dimana kemudian hasilnya digunakan untuk melakukan *testing* menggunakan *dataset* yang sama.



Gambar 17. Desain Proses 10 Fold Cross Validation untuk Naive Bayes

Gambar 17 adalah desain proses yang berada di dalam operator *Cross Validation Naive Bayes* yang terdapat pada gambar 15. Algoritma k-NN digunakan untuk melakukan training terhadap *dataset* masukan dimana kemudian hasilnya digunakan untuk melakukan *testing* menggunakan *dataset* yang sama.



Gambar 18. Desain Proses 10 Fold Cross Validation untuk K-NN

Gambar 18. adalah desain proses yang berada di dalam operator *Cross Validation k-NN* yang terdapat pada gambar 15. Algoritma k-NN digunakan untuk melakukan training terhadap *dataset* masukan dimana kemudian hasilnya digunakan untuk melakukan *testing* menggunakan *dataset* yang sama.

Dari hasil proses yang ada pada gambar 15 dapat dihasilkan nilai akurasi dan AUC dari ketiga algoritma yang dibandingkan tersebut. Adapun hasil-hasilnya dapat dilihat pada gambar-gambar di bawah ini;

Akurasi dan AUC Algoritma *Decision Tree C4.5*

Hasil proses pada gambar 15 untuk algoritma *Decision Tree C4.5* pada bagian akurasi adalah sebesar 91,75 persen. Hal ini ditampilkan pada gambar 21 dibawah ini.

accuracy: 91.75% +/- 4.54% (mikro: 91.78%)

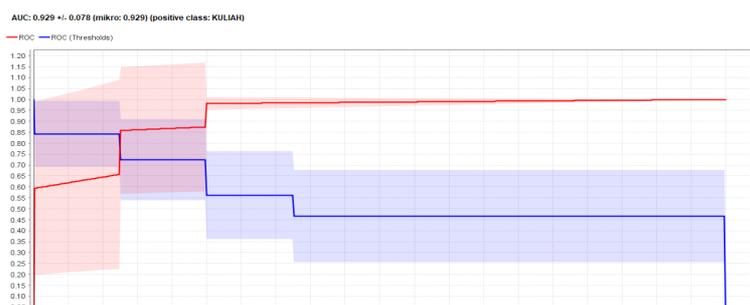
| | true KERJA | true KULIAH | class precision |
|--------------|------------|-------------|-----------------|
| pred. KERJA | 67 | 5 | 93.06% |
| pred. KULIAH | 13 | 134 | 91.16% |
| class recall | 83.75% | 96.40% | |

Gambar 19. Hasil Akurasi *Decision Tree C4.5*

Pada Gambar 19 *confusion matrix decision tree* menghasilkan nilai *accuracy* sebesar 91,75% dan dari Gambar 19 mendapatkan prediksi sebagai berikut:

- Prediksi Kerja – True Kerja (TP) = 67
- Prediksi Kerja – True Kuliah (TN) = 5
- Prediksi Kuliah – True Kerja (FP) = 13
- Prediksi Kuliah – True Kuliah (FN) = 134

Untuk nilai AUC *Decision Tree C4.5* berdasarkan hasil proses gambar 15 adalah sebesar 0.929. Hal ini ditampilkan pada gambar 20 dibawah ini.



Gambar 20. Hasil AUC *Decision Tree C4.5*

Akurasi dan AUC Algoritma *Naïve Bayes*

Hasil proses pada gambar 15 untuk algoritma *Naïve Bayes* pada bagian akurasi adalah sebesar 86,77 persen. Hal ini ditampilkan pada gambar 21 dibawah ini.

accuracy: 86.77% +/- 5.13% (mikro: 86.76%)

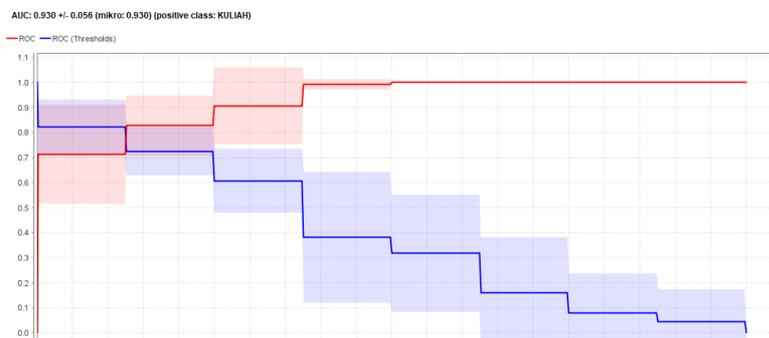
| | true KERJA | true KULIAH | class precision |
|--------------|------------|-------------|-----------------|
| pred. KERJA | 54 | 3 | 94.74% |
| pred. KULIAH | 26 | 136 | 83.95% |
| class recall | 67.50% | 97.84% | |

Gambar 21. Hasil Akurasi *Naïve Bayes*

Pada Gambar 21 *confusion matrix decision tree* menghasilkan nilai *accuracy* sebesar 86,77% dan dari Gambar 21 mendapatkan prediksi sebagai berikut:

- Prediksi Kerja – True Kerja (TP) = 54
- Prediksi Kerja – True Kuliah (TN) = 3
- Prediksi Kuliah – True Kerja (FP) = 26
- Prediksi Kuliah – True Kuliah (FN) = 136

Untuk nilai AUC *Naïve Bayes* berdasarkan hasil proses gambar 15 adalah sebesar 0.930. Hal ini ditampilkan pada gambar 22 dibawah ini.



Gambar 22. Hasil AUC *Naïve Bayes*

Akurasi dan AUC *K-Nearest Neighbor*

Hasil proses pada gambar 4.8 untuk algoritma *K-Nearest Neighbor* pada bagian akurasi adalah sebesar 88,61 persen. Hal ini ditampilkan pada gambar 4.16 dibawah ini.

accuracy: 88.61% +/- 6.49% (mikro: 88.58%)

| | true KERJA | true KULIAH | class precision |
|--------------|------------|-------------|-----------------|
| pred. KERJA | 77 | 22 | 77.78% |
| pred. KULIAH | 3 | 117 | 97.50% |
| class recall | 96.25% | 84.17% | |

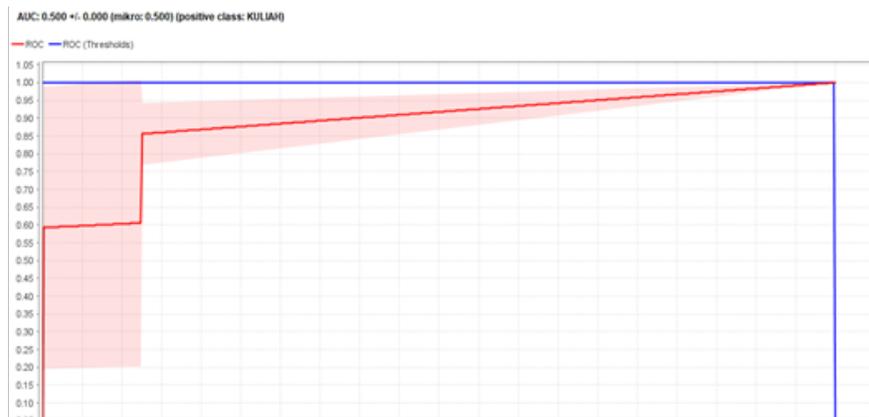
Gambar 23. Hasil Akurasi *K-Nearest Neighbor*

Pada Gambar 23 *confusion matrix decision tree* menghasilkan nilai *accuracy* sebesar 88,61% dan dari Gambar 23 mendapatkan prediksi sebagai berikut:

- Prediksi Kerja – True Kerja (TP) = 77
- Prediksi Kerja – True Kuliah (TN) = 22

- c) Prediksi Kuliah – True Kerja (FP) = 3
- d) Prediksi Kuliah – True Kuliah (FN) = 117

Untuk nilai AUC *K-Nearest Neighbor* berdasarkan hasil proses gambar 15 adalah sebesar 0.500. Hal ini ditampilkan pada gambar 24 dibawah ini.



Gambar 24. Hasil AUC *K-Nearest Neighbor*

HASIL

Berdasarkan pada gambar 19, gambar 21 dan gambar 23 tentang nilai *confusion matrix Decision Tree C4.5, Naïve Bayes* dan *K-Nearest Neighbord* maka didapatkan nilai perbandingan dari tiap seperti pada tabel 4.4 tentang perbandingan nilai *confusion matrix* 3 (tiga) algoritma

Tabel 11. Perbandingan *Confusion Matrix Decision Tree, Naïve Bayes* dan *k-Nearest Neighbor*

| Prediksi | <i>DT C4.5</i> | <i>NB</i> | <i>K-NN</i> |
|---------------------------------------|----------------|-----------|-------------|
| Pred Lancar - <i>true</i> Lancar (TP) | 67 | 54 | 77 |
| Pred Macet - <i>true</i> Macet (TN) | 5 | 3 | 22 |
| Pred Lancar - <i>true</i> Macet (FP) | 13 | 26 | 3 |
| Pred Macet - <i>true</i> Lancar (FN) | 134 | 136 | 117 |

Berdasarkan pada tabel 11 perbandingan *confusion matrix Decision Tree C4.5, Naïve Bayes* dan *K-Nearest Neighbord* maka maka didapatkan nilai untuk menghitung dari nilai *accuracy, sensitivity, specificity, precision, dan recall* yang bisa dilihat dari tabel 12 berikut

Tabel 12 Perbandingan Nilai *Accuracy, Sensitivity, Specificity, Precision, dan Recall confusion matrix Decision Tree C4.5, Naïve Bayes* dan *K-Nearest Neighbord*

| | <i>DT C4.5 (%)</i> | <i>NB (%)</i> | <i>K-NN (%)</i> |
|--------------------|--------------------|---------------|-----------------|
| <i>Accuracy</i> | 91,75 | 86,77 | 88,61 |
| <i>Sensitivity</i> | 33,33 | 28,42 | 39,69 |
| <i>Specificity</i> | 27,77 | 10,34 | 88,00 |
| <i>Precision</i> | 83,75 | 67,50 | 96,25 |
| <i>Recall</i> | 33,33 | 28,42 | 39,69 |

Berdasarkan pada tabel 4.5 tentang perbandingan nilai *accuracy*, *sensitivity*, *specificity*, *precision*, dan *Recall*, maka nilai *Accuracy Decision Tree C4.5* mendapatkan nilai akurasi sebesar 91,75%, *Naïve Bayes* mendapatkan nilai akurasi sebesar 86,77% dan *K-Nearest Neighbord* mendapatkan nilai akurasi 88,61%.

Setelah melalui proses *10-fold validation* seperti yang diterapkan pada desain gambar 15, maka dapat diperoleh hasil akurasi dan AUC dari masing-masing algoritma yang digunakan. Nilai Akurasi dan AUC digunakan sebagai langkah awal perbandingan pencarian algoritma terbaik yang akan dihasilkan. Pebandingan hasil perhitungan nilai akurasi dan AUC untuk metode *Decision Tree C4.5*, *Naïve Bayes*, dan *K-Nearest Neighbor* dapat dilihat pada Tabel 13.

Tabel 13. Komparasi Nilai Akurasi dan AUC

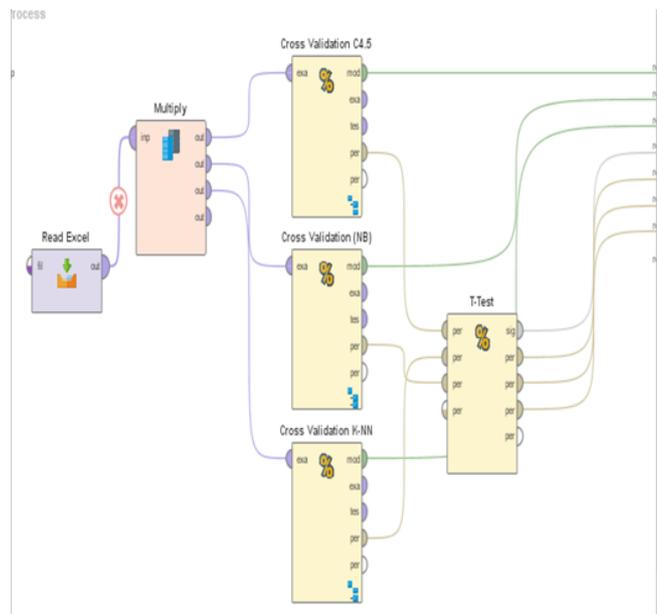
| | C4.5 | Naïve Bayes | k-NN |
|----------------|-------------|--------------------|-------------|
| AKURASI | 91,75 % | 86,77 % | 88,61 % |
| AUC | 0,929 | 0,930 | 0.500 |

Pada sisi akurasi Tabel 4.6 menggambarkan bahwa hasil perbandingan nilai akurasi, menunjukkan bahwa algoritma *Decision Tree C4.5* mendapatkan nilai tertinggi yaitu 91,75%, diikuti dengan *K-Nearest Neighbor* dengan nilai 88,61%, dan *Naïve Bayes* dengan nilai 86,77%. Dari ke tiga algoritma yang diuji berdasarkan nilai akurasi maka didapatkan bahwa algoritma *Decision Tree C4.5* memperoleh nilai terbaik dari sisi akurasinya.

Untuk AUC Tabel 4.6 menggambarkan bahwa hasil perbandingan nilai AUC, menunjukkan bahwa algoritma *Naïve Bayes* mendapatkan nilai tertinggi yaitu 0,930, diikuti dengan *Decision Tree C4.5* dengan nilai 0,929 dan *K-Nearest Neighbor* dengan nilai 0,500. Dari ke tiga metode yang dibandingkan *Decision Tree C4.5* dan *Naïve Bayes* mendapat nilai AUC klasifikasi *Excellent*, dikarenakan memiliki nilai diantara 0,9 – 1,0, sedangkan *K-Nearest Neighbord* Mendapat nilai AUC klasifikasi *Failure* karena memiliki nilai antar 0,5 – 0,6 (Gorunescu, n.d.).

Uji Beda

Walaupun secara hasil uji pada *Rapid Miner* menunjukkan bahwa algoritma *Naïve Bayes* menunjukkan hasil yang terbaik dari sisi AUC yaitu 0,930 dan dari sisi akurasi *Decision Tree C4.5* mempunyai nilai terbaik yaitu 91,75%, namun pada penelitian ini yang menggunakan perbandingan tiga buah algoritma tetap dilakukan uji beda. Uji beda ini dilakukan untuk mengetahui seberapa signifikan perbedaan dari masing-masing algoritma dan seberapa besar kemungkinan algoritma tersebut mempunyai pengaruh terhadap algoritma lainnya, oleh karena itu maka dilakukan uji beda guna mendapatkan algoritma terbaik dari ketiga algoritma yang diuji. Ketiga algoritma yang digunakan akan di uji dengan operator *T-test* dari *Rapidminer* sehingga dapat menghasilkan penilaian tentang algoritma terbaik yang akan digunakan. Hal ini dapat dilihat pada desain uji beda pada gambar 25 dibawah ini.



Gambar 25. Proses T-Test C4.5, Naive Bayes dan K-NN

Setelah proses pada *Rapidminer* dijalankan maka didapatkan hasil *T-test* seperti pada gambar 26 dibawah ini.

| A | B | C | D |
|-----------------|-----------------|-----------------|-----------------|
| | 0.918 +/- 0.045 | 0.886 +/- 0.065 | 0.868 +/- 0.051 |
| 0.918 +/- 0.045 | | 0.226 | 0.034 |
| 0.886 +/- 0.065 | | | 0.491 |
| 0.868 +/- 0.051 | | | |

Probabilities for random values with the same result.

Values with a colored background are smaller than $\alpha=0.050$ which indicates a probably significant difference between the actual mean values.

Gambar 26. Hasil Uji Beda

Dari hasil uji beda terlihat bahwa algoritma *Decision Tree* mempunyai nilai terbaik yaitu 0,226 dibandingkan dengan algoritma *Naive Bayes* dan nilai 0,034 dibandingkan *K-Nearest Neighbor*, sementara *K-Nearest Neighbor* mempunyai nilai 0,491 dibandingkan dengan *Naive Bayes*. Dengan demikian dari uji beda dapat ditarik kesimpulan bahwa algoritma *Decision Tree C4.5* adalah algoritma terbaik yang dapat digunakan pada penelitian ini. Hal tersebut ditunjukkan pula bahwa algoritma *Decision Tree C4.5* mempunyai nilai probabilitas terbaik diantara ketiga algoritma yang diuji karena memiliki nilai $\alpha \leq 0,05$.

Deployment

Pada tahap ini deployment dilakukan dengan menggunakan algoritma *Decision Tree C4.5*. Hal ini terjadi karena algoritma tersebut adalah algoritma terbaik setelah dilakukan uji beda dan mendapat nilai dibawah alpha sehingga algoritma *Decision Tree C4.5* akan digunakan sebagai rule dari *deployment* yang akan kita bangun. Proses ini sendiri dilakukan untuk membuktikan hasil dari rule/model algoritma terhadap data sesungguhnya serta mengetahui nilai akurasi yang dihasilkan oleh algoritma yang diuji. Pada tahap ini penelitian ini digunakan rule/model dari *Decision Tree C4.5* sebagai landasan uji pada uji kecocokan algoritma.

Deployment pada penelitian ini dilakukan dengan melakukan uji kecocokan algoritma. Uji kecocokan yang dilakukan pada penelitian ini bertujuan untuk menguji sejauh mana nilai akurasi yang dihasilkan oleh algoritma *Decision Tree C4.5* sebagai salah satu yang terpilih dari ketiga algoritma yang diuji jika dibandingkan dengan *dataset* yang ada dan juga *deployment* sederhana yang dilakukan pada *Microsoft Excell*. Pada tahap ini akan dilihat seberapa besar kemiripan atau kecocokan yang terjadi dari hasil perhitungan algoritma *Decision Tree C4.5* dengan *deployment* yang akan dilakukan di *Microsoft Excell* menggunakan *rule* atau *model* yang dihasilkan oleh algoritma *Decision Tree C4.5*.

Dataset yang di uji coba menggunakan menggunakan *tool RapidMiner* melalui operator *ExampleSet* sesuai dengan *rule* yang telah ditentukan dengan hasil 63,50% berorientasi kuliah dan 36,50% berorientasi kerja. Bila kita jabarkan maka terdapat 139 orientasi mahasiswa kuliah dan 80 orientasi mahasiswa kerja dari total 219 data orientasi mahasiswa yang diuji.

ExampleSet (219 examples, 1 special attribute, 5 regular attributes)

| Row No. | ORIENTATION | L/P | USIA | PEKERJAAN... | PENGHASIL... | MINAT MAH... |
|---------|-------------|-----|------|---------------|--------------|--------------|
| 1 | KERJA | L | 18 | PNS | 4500000 | ORANG TUA |
| 2 | KERJA | L | 18 | KARYAWAN | 3500000 | ORANG TUA |
| 3 | KERJA | L | 17 | PETANI | 4500000 | ORANG TUA |
| 4 | KERJA | L | 18 | PNS | 4500000 | ORANG TUA |
| 5 | KERJA | L | 17 | WIRASWATA | 4500000 | PROGRAM S... |
| 6 | KERJA | L | 17 | WIRASWATA | 4500000 | PROGRAM S... |
| 7 | KERJA | L | 18 | PNS | 5000000 | PROGRAM S... |
| 8 | KERJA | L | 17 | WIRASWATA | 4500000 | PROGRAM S... |
| 9 | KERJA | L | 18 | PNS | 5000000 | PROGRAM S... |
| 10 | KERJA | L | 17 | PETANI | 4500000 | ORANG TUA |
| 11 | KERJA | L | 18 | PNS | 5000000 | ORANG TUA |
| 12 | KERJA | L | 18 | KARYAWAN | 3500000 | ORANG TUA |
| 13 | KERJA | L | 18 | PNS | 5000000 | ORANG TUA |
| 14 | KERJA | L | 17 | PETANI | 4500000 | ORANG TUA |
| 15 | KERJA | L | 17 | PETANI | 3500000 | OLAHRAGA |
| 16 | KERJA | L | 18 | KERJA LAPA... | 4500000 | OLAHRAGA |

Gambar 27. Hasil Uji *Dataset* Menggunakan Algoritma *Decision Tree C4.5*

Adapun hasil dari uji *dataset* menggunakan algoritma *Decision Tree C4.5* dengan 219 data mahasiswa dengan 5 regular atribut dengan 1 buah label yang mempunyai dua buah nilai label, dapat dilihat pada gambar dibawah ini.

| Index | Nominal value | Absolute count | Fraction |
|-------|---------------|----------------|----------|
| 1 | KULIAH | 139 | 0.635 |
| 2 | KERJA | 80 | 0.365 |

Gambar 28. Hasil Uji Dataset Klasifikasi Orientasi Menggunakan Algoritma Decision Tree C4.5

Penerapan rule algoritma Decision Tree C4.5 menggunakan Microsoft Excell yang digunakan sebagai alat uji perbandingan kecocokannya menghasilkan nilai sebesar 68,50% mahasiswa berorientasi kuliah dan 31,50% mahasiswa berorientasi kerja. Bila kita jabarkan maka terdapat 150 orang mahasiswa berorientasi kuliah dan 59 orang mahasiswa berorientasi kerja dari total 219 orang mahasiswa yang diteliti, hal ini dapat dilihat pada gambar 29.

| L/P | USIA | PEKERJAAN ORANG TUA | PENGHASILAN ORANG TUA | FAKTOR UTAMA | ORIENTATION | DT C4.5 | MS. EXCELL |
|-----|------|---------------------|-----------------------|---------------|-------------|---------|------------|
| L | 18 | PNS | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 18 | KARYAWAN | 3.500.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 17 | PETANI | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 18 | PNS | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 17 | WIRASWATA | 4.500.000 | PROGRAM STUDI | KERJA | KERJA | KERJA |
| L | 17 | WIRASWATA | 4.500.000 | PROGRAM STUDI | KERJA | KERJA | KERJA |
| L | 18 | PNS | 5.000.000 | PROGRAM STUDI | KERJA | KERJA | KULIAH |
| L | 17 | WIRASWATA | 4.500.000 | PROGRAM STUDI | KERJA | KERJA | KERJA |
| L | 18 | PNS | 5.000.000 | PROGRAM STUDI | KERJA | KERJA | KULIAH |
| L | 17 | PETANI | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 18 | PNS | 5.000.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 18 | KARYAWAN | 3.500.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 18 | PNS | 5.000.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 17 | PETANI | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIAH |
| L | 17 | PETANI | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| L | 18 | KERJA LAPANGAN | 4.500.000 | OLAHRAGA | KERJA | KERJA | KULIAH |
| L | 17 | PETANI | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| L | 17 | PETANI | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| P | 17 | PETANI | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| L | 17 | PETANI | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| L | 17 | PETANI | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| L | 17 | PETANI | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| L | 18 | KARYAWAN | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |
| L | 17 | PNS | 3.500.000 | OLAHRAGA | KERJA | KERJA | KERJA |

Gambar 29. Hasil Uji Dataset Menggunakan Microsoft Excell

Dari hasil uji dataset diatas maka dapat dilihat dalam bentuk tabel perbandingan dari hasil uji algoritma Decision Tree dan juga Ms. Excel yang menggunakan rule Decision Tree seperti pada tabel 14 dibawah ini

Tabel 14. Hasil Perbandingan Uji Dataset

| | KULIAH | % | KERJA | % |
|------------|--------|--------|-------|--------|
| Dataset | 139 | 63,50% | 80 | 36,50% |
| C4.5 | 139 | 63,50% | 80 | 36,50% |
| Ms. Excell | 150 | 68,50% | 69 | 31,50% |

Adapun hasil perbandingan *dataset* dari awal hingga dilakukan uji kecocokan dapat dilihat pada gambar 30. Hasil kecocokan antara *dataset* yang diolah oleh algoritma *Decision Tree* dan *Microsoft Excell* sebanyak 174 data atau 79,50% data cocok dan tidak cocok sebesar 45 data atau 20,50% data. Hal ini menunjukkan bahwa ada perbedaan hasil pengolahan dari algoritma terpilih yaitu *Decision Tree C4.5* dengan *MS. Excel*, dimana dari tiap data yang diuji tidak semuanya mempunyai kesesuaian yang sama, sehingga didapat data ke tidak cocok yang mencapai 20,50% data mahasiswa yang diuji.

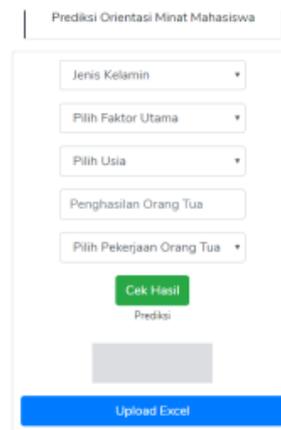
| L/P | USIA | PEKERJAAN ORANG TUA | PENGHASILAN ORANG TUA | FAKTOR UTAMA | ORIENTA TION | DT C4.5 | M.S. EXCEL | HA SIL |
|-----|------|---------------------|-----------------------|---------------|--------------|---------|------------|--------------|
| L | 18 | PNS | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | KARYAWAN | 3.500.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | PETANI | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | PNS | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | PNS | 5.000.000 | PROGRAM STUDI | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | PNS | 5.000.000 | PROGRAM STUDI | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | PETANI | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | PNS | 5.000.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | KARYAWAN | 3.500.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | PNS | 5.000.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | PETANI | 4.500.000 | ORANG TUA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 18 | KERJA LAPANGAN | 4.500.000 | OLAHRAGA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | PNS | 3.000.000 | OLAHRAGA | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | PNS | 3.500.000 | MUSIK | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | KARYAWAN | 4.500.000 | MUSIK | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | WIRASWATA | 4.500.000 | MUSIK | KERJA | KERJA | KULIA H | TIDA K COCOK |
| L | 17 | WIRASWATA | 4.500.000 | MUSIK | KERJA | KERJA | KULIA H | TIDA K COCOK |
| P | 17 | KARYAWAN | 4.000.000 | FAVORIT | KERJA | KULIAH | KERJA | TIDA K COCOK |
| L | 17 | KARYAWAN | 4.000.000 | FAVORIT | KERJA | KULIAH | KERJA | TIDA K COCOK |
| L | 17 | KARYAWAN | 4.000.000 | FAVORIT | KERJA | KULIAH | KERJA | TIDA K COCOK |
| L | 17 | WIRASWATA | 4.000.000 | FAVORIT | KERJA | KULIAH | KERJA | TIDA K COCOK |
| L | 18 | PNS | 4.500.000 | MUSIK | KERJA | KULIAH | KERJA | TIDA K COCOK |
| L | 17 | KARYAWAN | 3.500.000 | FAVORIT | KERJA | KULIAH | KERJA | TIDA K COCOK |
| L | 17 | KERJA LAPANGAN | 5.500.000 | FAVORIT | KERJA | KULIAH | KERJA | TIDA K COCOK |

Gambar 30. Hasil Uji Kecocokan Algoritma

Dari hasil perbandingan uji kecocokan diatas dapat dilihat bahwa tingkat perbedaan dari penghitungan menggunakan *Microsoft Excell* yang menggunakan mengakomodir rule *Decision Tree C4.5*, menunjukkan hasil berbeda yang cukup signifikan jika dibandingkan dengan hasil perhitungan yang keluar dari algoritma *Decision Tree C4.5* hasil dari proses pada pada *RapidMiner*. Hal ini disebabkan oleh sebaran data yang sangat dinamis pada rule *Decision Tree C4.5* yang dikeluarkan oleh *RapidMiner*, sehingga ketika di aplikasikan ke *Microsoft Excell* terjadi perbedaan yang cukup signifikan namun masih berada pada hasil diatas 75% tingkat kecocokan.

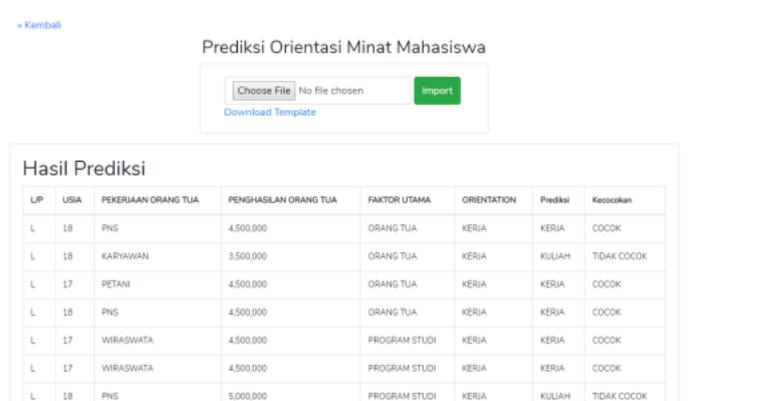
Rancangan GUI

Berdasarkan hasil percobaan yang dilaksanakan, dikarenakan algoritma C4.5 memiliki akurasi yang lebih tinggi, dalam deployment akan digunakan hasil dari algoritma C4.5. Implementasi graphical user interface (GUI) dapat dilakukan dengan menginputkan data uji kecocokan, jika disubmit akan muncul uji kecocokan yang diinputkan. Rancangan GUI dapat melakukan prediksi beberapa data sekaligus, dengan cara upload file sesuai dengan data.



Gambar 31.Tampilan Upload Data Orientasi

Pada halaman utama ini, GUI dirancang untuk dapat digunakan menguji kebenaran pola yang dihasilkan dalam bentuk tampilan web menggunakan bahasa pemrograman PHP.



| LIP | USIA | PEKERJAAN ORANG TUA | PENGHASILAN ORANG TUA | FAKTOR UTAMA | ORIENTATION | Prediksi | Kecocokan |
|-----|------|---------------------|-----------------------|---------------|-------------|----------|-------------|
| L | 18 | PNS | 4.500.000 | ORANG TUA | KERJA | KERJA | COCOK |
| L | 18 | KARYAWAN | 3.500.000 | ORANG TUA | KERJA | KULIAH | TIDAK COCOK |
| L | 17 | PETANI | 4.500.000 | ORANG TUA | KERJA | KERJA | COCOK |
| L | 18 | PNS | 4.500.000 | ORANG TUA | KERJA | KERJA | COCOK |
| L | 17 | WIRASWATA | 4.500.000 | PROGRAM STUDI | KERJA | KERJA | COCOK |
| L | 17 | WIRASWATA | 4.500.000 | PROGRAM STUDI | KERJA | KERJA | COCOK |
| L | 18 | PNS | 5.000.000 | PROGRAM STUDI | KERJA | KULIAH | TIDAK COCOK |

Gambar 32. Hasil Uji Kecocokan Algoritma By Alikasi

Implikasi Penelitian

Hasil dari penelitian yang telah dilakukan ini diharapkan dapat memberikan masukan bagi pihak Politeknik Tri Mitra Karya Mandiri untuk dapat memanfaatkan model yang terbentuk dari penelitian ini dalam membuat langkah strategis guna mencegah terjadinya penurunan atau pengurangan jumlah mahasiswa secara masif di semester-semester lanjutan yang disebabkan oleh kesempatan kerja dengan merekrut mahasiswa yang memang sebelumnya lebih berorientasi kerja daripada kuliah.

Adapun usulan yang dapat diberikan berdasarkan *rule model* yang terjadi pada algoritma *Decision Tree C4.5* antara lain:

- a. Lebih melakukan penyaringan kepada mahasiswa baru berdasarkan hasil penelitian ini, sehingga dapat dihasilkan calon mahasiswa yang memang ingin melanjutkan studi dan mempunyai kemampuan finansial yang baik, sehingga mengurangi adanya masalah kemahasiswaan di kemudian hari.

- b. Pihak Politeknik diharapkan dapat mencari strategi promosi yang jitu sehingga dapat memperoleh calon mahasiswa yang mempunyai kriteria yang diinginkan berdasarkan hasil penelitian ini.
- c. Lebih menggalakan dinamika kampus melalui berbagai kegiatan yang sesuai dengan minat mahasiswa, karena minat mahasiswa yang tersalurkan membuat mahasiswa tersebut cenderung untuk menekuni minatnya dan menyelesaikan kuliah tepat pada waktunya.
- d. Menyikapi kemampuan orangtua berdasarkan kemampuan finansial dan pekerjaannya, sehingga dapat ditemukan solusi yang baik yang tidak terlalu memberatkan orangtua dari sisi finansial dan tidak mengurangi ataupun menghilangkan penghasilan bagi universitas dari sisi finansial.
- e. Mengadakan even-even baik dari sisi akademis seperti even sains dan ketrampilan juga even-even keagamaan, olah raga dan seni, dimana selain untuk menyalurkan minat mahasiswa juga sebagai ajang promosi kepada calon mahasiswa dari berbagai SMU?SMK di wilayah Cikampek dan sekitarnya.

SIMPULAN

Berdasarkan pengujian komparasi Algoritma *Decision Tree C4.5*, *Naïve Bayes* dan *K-Nearest Neighbor*, dapat diketahui bahwa nilai akurasi orientasi minat mahasiswa di Politeknik Tri Mitra Karya Mandiri mendapatkan hasil bahwa Algoritma *Decision Tree C4.5* adalah Algoritma terbaik dari ketiga algoritma tersebut. Hal ini dapat dilihat pada hasil uji beda dari ketiga algoritma tersebut, dimana didapatkan hasil untuk algoritma *Decision Tree* dengan nilai sebesar 0,226 ketika diuji dengan algoritma *Naïve Bayes* dan nilai sebesar 0,034 ketika di uji dengan algoritma *K-Nearest Neighbor*. Sehingga setelah melalui uji beda didapat kesimpulan bahwa algoritma *Decision Tree C4.5* adalah algoritma yang terbaik Dibandingkan dengan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* karena algoritma *Decision Tree C4.5* menunjukkan nilai perbedaan yang cukup signifikan karena mempunyai nilai $\alpha < 0,05$. Karena hal tersebut maka dengan menggunakan rule Algoritma *Decision Tree C4.5* dapat digunakan untuk pengujian *dataset* orientasi minat mahasiswa, sehingga dengan menggunakan rule itu algoritma *Decision Tree C4.5* pula penulis melakukan deployment sederhana pada penelitian ini.

DAFTAR RUJUKAN

- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing.
- Baradwaj, B. K. (2011). Mining Educational Data to Analyze Students " Performance, 2(6), 63–69.
- Bramer, M. (2013). *Principles Of Data Mining*.
- Chusanawati, T. (2015). Pengaruh Promosi Terhadap Minat Kuliah Studi Kasus Di Politeknik Negeri Jakarta, 12(2), 105–110.
- Daniel Swanjaya, A. I. (2015). Educational Data Mining Untuk Mengetahu Pola Minat Kerja Mahasiswa, 978–979.
- Gorunescu, F. (2009). *Data Mining Concepts, Models And Techniques*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, itd: Morgan Kaufmann.

- Hastuti, K. (2012). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif, *2012(Semantik)*, 241–249.
- Indra Purnama, Ragil Saputra, A. W. (2014). Implementasi Data Mining Menggunakan CRIPS-DM Pada Sistem Informasi Eksekutif Dinas Kelautan Dan Perikanan Provinsi Jawa Tengah.
- Indra Purnama. Ragil Saputra, A. W. (2014). Implementasi Data Mining Menggunakan CRIPS-DM Pada Sistem Informasi Eksekutif Dinas Kelautan Dan Perikanan Provinsi Jawa Tengan, 384.
- Kothari, C. R. (2009). *Research Methodology*.
- Kothari, C. R. (2004). *Research Methodology, Methods And Technique (Second Revised Edition)*.
- Markus Hofmann, R. K. (2006). *Rapid Miner Data Mining Use Cases And Business Analytics Application*.
- Markus Hofmann, R. K. (2014). *Data Mining and Knowledge Discovery Series Edited by*.
- Marlina, L., Putera, A., & Siahaan, U. (2016). Data Mining Classification Comparison (Naïve Bayes and C4 . 5 Algorithms), *38(7)*, 380–383.
- Sutoyo. (2012). Perbandingan Algoritma Kkalsifikasi C4.5 Dengan K-NN Untuk Text Categoritation Menggunakan Clasiccal Arabic Corpus, 1–81.
- Swastina, L. (2013). Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa, *2(1)*.
- Tahyudin, I., Utami, E., Amborowati, A., Tahyudin, I., Utami, E., & Amborowati, A. (2013). Comparing Clasification Algorithm Of Data Mining to Predict the Graduation Students on Time, (December), 2–4.
- Torgo, L. (n.d.). *Data Mining with R Learning with Case Studies Data Mining and Knowledge Discovery Series*.
- Xindong Wu, V. K. (n.d.). *The Top Ten ALgorithms in Data Mining*.
- Ye, N. (2015). *Data Mining Theories, Algorithms and Examples. IEEE Potentials (Vol. 16)*.