Perbandingan Algoritma Klasifikasi Untuk Penentuan Lokasi Promosi Penerimaan Mahasiswa Baru Pada IIB Darmajaya Lampung

Comparison of Classification Algorithm to Determine the Location of New Student Admission Promotion on IIB Darmajaya Lampung

Robby Toro[1] dan Sri Lestari[2]
[1,2]Teknik Informatika, Institut Informatika dan Bisnis Darmajaya
E-mail: robbytoro31@gmail.com, srilestari@darmajaya.ac.id

Abstrak

Tingkat pesaing perguruan tinggi setiap tahunnya mengalami peningkatan yang sangat pesat. Dalam proses mendapatkan mahasiswa baru setiap perguruan tinggi memiliki strategi promosi yang berbeda-beda, seiring perkembangan teknologi maka pihak perguruan tinggi dapat memanfaatkan data *mining* dalam upaya promosi. Data *mining* dapat menjadi sebuah dasar atau pedoman untuk menentukan kebijakan bisnis dalam upaya peningkatan pesaing bisnis *perusahaan*. Dalam upaya penghematan biaya promosi maka dilakukan pemetaan lokasi promosi. Penelitian ini membandingkan tiga alogoritma klasifikasi agar mendapatkan nilai akurasi yang tinggi sehingga informasi yang dihasilkan lebih akurat. Pemetaan lokasi promosi menggunakan algoritma klasifikasi yaitu, *decision tree, naïve bayes,* dan *k-nearest nighbour*, karena ketiga algoritma tersebut merupakan algoritma yang populer, serta menghasilkan akurasi yang tinggi. Data *set* yang digunakan berjumlah 1281 *record*, dan menggunakan variabel, nama, total, alumni, luar kota, dan asal sekolah, denganlabel yaitu, sangat potensi, potensi, dan kurang potensi. Hasil klasifikasi dapat digunakan pihak pemasaran untuk menentukan lokasi promosi pada tahun yang akan datang. Berdasarkan hasil evaluasi *crosss validation* dari ketiga alogoritma tersebut maka alogoritma yang memiliki nilai akurasi tertinggi yaitu *decision tree* sebesar 100%, selanjutnya adalah algoritma *k-nearest nighbour* sebesar 99,61%, dan yang terakhir algoritma *naïve bayes* hanya 84,78%.

Kata kunci: Data *Mining*, Akurasi, *Decision Tree, Naïve Bayes, K-Nearest Nighbour*

Abstract

The level of university competitors every year has increased very rapidly. In getting new students, each university has a different promotion strategy. Many of them use the location of promotion to calculate marketing costs. In line with technological developments, universities can use data mining in promotional efforts. This technique can be a basis or guideline for determining business policies to increase the company's business competitors. This study compares three classification algorithms to get a high accuracy value. The promotion location mapping uses a classification algorithm: decision tree, nave Bayes, and k-nearest neighbor, because the three algorithms are popular algorithms, and produce high accuracy. The data set used was 1281 records, and uses the variables, name, total, alumni, region/location, and school origin, with labels namely, very potential, potential, and less potential. The results of this classification can assist the marketing department to decide the location of the promotion in the coming year. Based on the cross-validation evaluation of the three algorithms, the highest accuracy value is the decision tree of 100%, the second is the k-nearest neighbor algorithm of 99.61%, and the lowest is the naive Bayes algorithm with only 84.78%.

Keywords: Data Mining, Accuracy, Decision Tree, Naïve Bayes, K-Nearest Neighbor

1. PENDAHULUAN

Tingkat persaingan perguruan tinggi setiap tahunnya mengalami peningkatan yang sangat pesat. Hingga saat ini sudah ada 3098 perguruan tinggi, baik negeri maupun swasta yang tersebar di seluruh Indonesia. Peningkatan ini harus diimbangi dengan seleksi yang maksimal untuk mendapatkan mahasiswa yang berkualitas [1]. Persaingan yang sangat ketat ini menuntut setiap perguruan tinggi khususnya swasta untuk meningkatkan strategi *marketing* [2].

Penerimaan mahasiswa baru pada Institut Informatika dan Bisnis (IIB) Darmajaya merupakan program yang wajib dilaksanakan setiap tahunnya untuk meregenerasi mahasiswa, sehingga mendapatkan mahasiswa baru sesuai dengan yang diharapkan, oleh sebab itu IIB Darmajaya melakukan promosi mulai dari sekolah-sekolah, hingga bekerja sama dengan berbagai media. Tahun 2020 dana yang dihabiskan untuk *road show* (kunjungan ke sekolah) mencapai Rp53.000.000 dengan tujuan 79 sekolah yang ada di provinsi Lampung. Berdasarkan hasil penerimaan mahasiswa baru pada tahun 2020 pihak pemasaran IIB Darmajaya hanya mampu memperoleh 55 sekolah dengan jumlah 758 mahasiswa baru. Penerapan data *mining* juga dapat digunakan sebagai strategi promosi yang lebih efektif dan efisien yang dapat diimplementasikan dalam berbagai bidang, baik ekonomi, kesehatan hingga pendidikan.

Data *mining* semakin popular dan berkembang dengan tren *big data* seiring dengan mudahnya mengakses informasi yang semakin luas dari waktu ke waktu. Data *mining* merupakan proses pencarian, pola, korelasi, dan tren yang menggabungkan bidang *machine learning,* teknik visualisasi, dan statistika melalui penyaringan dari sebuah data yang besar. Data *mining* juga menjadi berkembang dengan seiringnya tren baru dalam dunia teknologi informasi yang bertujuan untuk mendefinisikan data-data penting dan pengetahuan yang berada dalam sistem informasi. Teridentifikasinya data dapat menjadi sebuah informasi yang sangat penting untuk *menentukan* strategi dalam dunia bisnis. Data *mining* juga dapat menjadi sebuah dasar atau pedoman untuk menentukan kebijakan bisnis dalam upaya peningkatan pesaing bisnis perusahaan [3]. Peningkatkan strategi *marketing* di bagian pemasaran IIB Darmajaya dapat memanfaatkan data *mining* sebagai bahan pertimbangan untuk menentukan lokasi promosi yang lebih akurat dan dapat menghemat biaya promosi.

Beberapa penelitian telah dilakukan terkait data mining dan algoritmanya seperti yang dilakukan oleh [1] melakukan seleksi calon mahasiswa. Data yang digunakan dalam penelitian ini yaitu data mahasiswa yang masa belajarnya lebih dari 8 semester dan mengunakan data alumni. Atribut yang digunakan ada 6 yaitu status ketepatan waktu, nilai matematika, jenis kelamin, nilai ujian Bahasa Indonesia, jurusan, nilai ujian Bahasa Inggris. Metode data *mining* yang digunakan adalah *Rule Induction, K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, Linear Discriminant Analysis, Decision Stump, Neural Network, Decision Ttree, Random Forest,* dan *Linear Regression.* Hasil eksperimen menunjukkan algoritma terbaik yaitu *Support Vector Machine* dengan akurasi 65.00% namun nilai ini masih jauh dari nilai *excellent.*

Penelitian selanjutnya dilakukan oleh [4] yang melakukan diagnosa penyakit liver dan analisa algoritma. Algoritma yang digunakan yaitu algoritma *Neural Network, Naïve Bayes, Decision Tree, Nearest Neighbor.* Atribut yang gunakan dalam penelitian ini terdiri dari *Ratio, Age, A/G, Sgot, DB, Gender, ALB, Alkpho, TB, Dataset, Sgpt,* hingga *TP*. Berdasarkan hasil penelitian yang telah dilakukan algoritma yang paling tepat dalam pengklasifikasian pasien liver yaitu algoritma *Decision Tree,* dengan akurasi sebesar 72,89%.

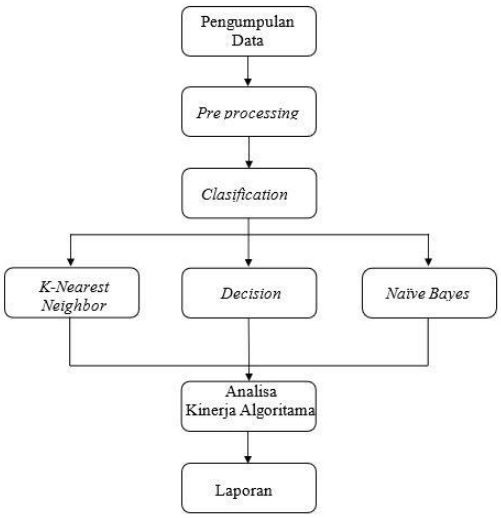
Penelitian oleh [5] memanfaatkan algoritma naïve bayes untuk penerima bantuan pangan non tunai. Penelitian ini bertujuan untuk mengklasifikasi penerima dan bukan perima bantuan pangan non tunai. Penelitian ini menggunkan dua agorima yaitu *naïve bayes* dan *decision tree*. Jumalah *field* yang digunakan terdiri dari 31 *field* dengan hasil nilai algoritma decision tree memiliki nilai akurasi sebesar 73,97% seangkan algorima naïve bayes memiliki nilai akurasi 58,29%.

Penelitian yang dilakukan juga menggunakan algoritma data mining untuk menentukan lokasi promosi yaitu algoritma klasifikasi diantaranya *Decision Tree, K-Nearest Neighbor,* dan *Naïve Bayes*. Mengacu pada penelitian terdahulu ketiga algorima tersebut cenderung sering

digunakan dalam perbandingan dikarenakan ketiga algorima berada pada posisi yang sama yaitu, klasifikasi, serta menghasilkan akurasi yang tinggi. Hasil evaluasi selanjutnya di komparasi untuk mengetahui algoritma yang lebih baik dalam merekomendasikan lokasi proposi. Aspek pembeda dalam penetian ini terletak pada indikator yang digunakan yaitu variabel nama, kota, pengasilan orang tua, jumlah alumni, asal sekolah dengan 3 label yang terdiri dari sangat potensi, potensi, kurang potensi. Pada variabel nama merupakan data pelengkap dari identitas calon mahasiswa.

2. METODE PENELITIAN

Tahapan penelitian ini diawali dengan pengumpulan data, kemudian dilakukan pre-processing, hingga dilakukannnya klasifikasi menggunakan *Decision Tree, K-Nearest Neighbor*, dan *Naïve Bayes* dan dilakukan evaluasi untuk mengetahui algoritma terbaik dengan nilai akurasi tertinggi. Secara runut tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Berikut penjelasan dalam tahapan penelitian yang dilakukan:

a. Pengumpulan Data

Dalam penelitian ini penulis mengunakan data mahasiswa baru pada tahun 2016-2020 pada IIB Darmajaya Lampung. Pengambilan data dilakukan dengan cara melakukan pengiriman surat permohonan izin dan data dikirim melalui surat eletronik. Data yang diberikan berjumlah 5685 *record* terdiri dari variabel nama, jurusan, jenis kelamin, kabupaten atau kota, kecamatan, sekolah dan pekerjaan orang tua. Data yang diterima tidak semua terekam dengan baik, sehingga akan dilakukan *pre-procesing.*

no	nama	jurusan	jk	ta	kabupaten	kecamatan	sekolah	pekerjaan ortu
1	Jenny Kharisma Putri	S1 - Sistem Informasi	Pr	20161	Kota Bandar Lampung	Kec. Teluk Betung Barat	LAINNYA	NULL
2	Liqya Fransisca	S1 - Akuntansi	Pr	20161	Kota Bandar Lampung	Kec. Rajabasa	LAINNYA	NULL
3	Sinta Aprilia	S1 - Akuntansi	Pr	20161	Kota Bandar Lampung	Kec. Rajabasa	LAINNYA	NULL
4	M Lutfi Nugraha Pratama	S1 - Teknik Informatika	Lk	20161	Kota Bandar Lampung	Kec. Rajabasa	LAINNYA	NULL
5	Alung Susanti	S1 - Akuntansi	Pr	20161	Kota Bandar Lampung	Kec. Rajabasa	LAINNYA	NULL
6	Made Ari Sucipto	S1 - Akuntansi	Lk	20161	Kota Bandar Lampung	Kec. Baradatu	LAINNYA	NULL
7	Rini Septiani	S1 - Teknik Informatika	Pr	20161	Kab. Lampung Timur	Kec. Margatiga	LAINNYA	TNI / POLRI
8	Kavin Pransiskus	S1 - Manajemen	Lk	20161	Kab. Tanggamus	Kec. Kota Agung	LAINNYA	NULL
9	Dani Salihin	S1 - Manajemen	Lk	20161	Kab. Lampung Timur	Kec. Batang Hari	LAINNYA	NULL
10	Dina Elisa Anggi Riatri Lubis	S1 - Manajemen	Pr	20161	Kota Bandar Lampung	Kec. Kemiling	LAINNYA	NULL
11	Evi Safitri	S1 - Akuntansi	Pr	20161	Kab. Lampung Selatan	Kec. Natar	LAINNYA	NULL
12	Rifki Kurniawan	S1 - Teknik Informatika	Lk	20161	Kab. Lampung Selatan	Kec. Natar	LAINNYA	NULL
13	Mita Hulzana	S1 - Manajemen	Pr	20161	Kota Bandar Lampung	Kec. Tanjung Karang Pusat	LAINNYA	NULL
14	Nindia Dwi Putri	S1 - Manajemen	Pr	20161	Kab. Tanggamus	Kec. Kota Agung	LAINNYA	PNS
15	A. Rivan Effendi	S1 - Manajemen	Lk	20161	Kota Bandar Lampung	Kec. Teluk Betung Utara	LAINNYA	NULL
16	Ivan Christian	S1 - Manajemen	Lk	20161	Kota Bandar Lampung	Kec. Teluk Betung Utara	LAINNYA	NULL
17	Novyansa Evan Supangka	S1 - Manajemen	Lk	20161	Kota Bandar Lampung	Kec. Tanjung Karang Pusat	LAINNYA	NULL
18	Bryan Arya Hutama	S1 - Teknik Informatika	Lk	20161	Kota Bandar Lampung	Kec. Kedaton	LAINNYA	NULL
19	Wakatanka	S1 - Manajemen	Lk	20161	Kota Bandar Lampung	Kec. Sukabumi	LAINNYA	NULL
20	Agus Suryanto Jauhari	S1 - Manajemen	Lk	20161	Kota Bandar Lampung	Kec. Kedaton	LAINNYA	NULL
21	Hamida Nurul Amelia	S1 - Manajemen	Pr	20161	Kota Bandar Lampung	Kec. Tanjung Karang Pusat	LAINNYA	NULL
22	Melinda Pratiwi	S1 - Manajemen	Pr	20161	Kota Bandar Lampung	Kec. Panjang	LAINNYA	NULL

Gambar 2. Data sebelum dilakukan *Pre-processing*

b. Pre-Processing

Proses pre-procesing yang benar akan mengasilkan informasi yang jelas dengan nilai akurasi yang baik. Proses pre-*processing* memiliki beberapa tahapan diantaranya, data *reduction* (proses mengurangi atribut yang tidak dibutuhkan), data *cleaning* (proses mengisi, memperbaiki, mengatur) data yang tidak sesuai dengan yang butuhkan, data *selection* hingga data balancing [6]. Penelitian ini melakukan reduksi dengan jumlah variable dari 9 atribut menjadi 5 yaitu nama, kabupaten atau kota, pekerjaan orang tua, jumlah alumni yang mendaftar di IIB Darmajaya, dan asal sekolah mahasiswa baru. Selein itu data yang digunakan adalah data 2018, 2019, dan 2020, sementara untuk data 2016 dan 2017 tidak digunakan karena dengan data tiga tahun sudah bisa mewakili. Proses cleaning peneliti menggunakan aplikasi *Microsoft Excel* dengan memfilter kolom satu perasatu dan mecari data yang kosong atau missing. Kolom sekolah terdapat banyak data yang tidak ter-record dengan baik. Selain itu juga di lakukan tranformasi dari variabel kabupaten bertujuan untuk mendapatkan pemetaan lokasi promosi pada data *training*, yang kemudian akan dilakukan transformasi data menjadi 2 jenis yaitu luar dan dalam kota dengan value: iya dan tidak. Pada penghasilan orang tua dikelompokkan menjadi tinggi, sedang, dan rendah berdasarkan jenis pekerjaan orang tua. Begitu pula pada asal sekolah menjadi Negeri dan Swasta. Pada total alumni menjadi dalam kota yaitu, ≥15, ≥9 dan ≤8, sementara untuk yang luar kota menjadi ≥12, ≥7 dan <7. Hasil pre-prosesing dapat dilihat pada Tabel 2.

Tabel 2. Setelah dilakukan *Pre-processing*

No.	Nama	Dalam Kota	Total Alumni	Pengasilan Orang Tua	Asal Sekolah	Label
1	Aldo Erlansyah	Tidak	≥7	Sedang	Swasta	Potensi
2	Aldo fernando	Iya	≥9	Sedang	Swasta	Potensi
3	Aldo Kusuma	Iya	≥15	Sedang	Swasta	Sangat Potensi
4	Aldy Bagus	Iya	≥15	Sedang	Swasta	Sangat Potensi
…	……	……	…….	……	……	……..
1281	Zuzlifatul Adnan	Tidak	<7	Rendah	Negeri	Potensi

Aturan pelabelan dalam penelitian ini berdasarkan hasil wawancara dengan kepala bagian pemasaaran IIB Darmajaya Lampung, dengan ketentuan sebagai berikut:

1) Sangat Potensi
   Sangat Potensi, jika lokasi sekolah berada di dalam kota dan jumlah alumni ≥15.
   Sangat Potensi, jika lokasi sekolah berada di luar kota dan jumlah alumni ≥12.
2) Potensi
   Pontensi, jika lokasi sekolah berada di dalam kota dan jumlah alumni ≥9, atau jumlah alumni ≤8 namun berasal dari sekolah negeri, atau jumlah alumni ≤8 tetapi penghasilan orang termasuk kedalam golongan sedang.
   Pontensi, jika lokasi sekolah berada di luar kota dan jumlah alumni ≥7, atau jumlah alumni <7 namun penghasilan orang tua dalam kategori sedang, atau jumlah alumni <7, penghasilan orang tua rendah, namun berasal dari sekolah negeri.
3) Kurang Potensi
   Kurang Potensi, jika lokasi sekolah berada di dalam kota, jumlah alumni ≤8 dan penghasilan orang tua tinggi, atau penghasilan orang tua rendah serta berasal dari sekolah swasta.
   Kurang Potensi, jika lokasi sekolah berada di luar kota, jumlah alumni <7 dan penghasilan orang tua tinggi, atau jumlah alumni <7 penghasilan orang tua rendah serta berasal dari sekolah swasta.

c. Clasification

Klasifikasi atau *Classification* merupakan upaya untuk mendapatkan model yang membedakan dan menjelaskan *class* data, atau proses pengklasifikasian satu atau beberapa *class* yang telah diidentifikasi sebelumnya. Berikut algoritma klasifikasi yang telah banyak digunakan yaitu, Naïve Bayes, *Decision Tree*, *K-Nearest Neighbor*, *Neural Rough Sets*, dan sebagainya [7]. Berikut algoritma klafisikasi yang digunakan dalam penelitian ini yaitu:

1) Algoritma *K-Nearest Neighbor*

Algoritma *K-nearest neighbor* salah satu algoritma dengan teknik klasifikasi data yang cukup baik. Algoritma ini mengggunakan data latih kemudian disimpan, sehingga jika ada data baru atau data *testing* maka dapat mengidentifikasi berdasarkan data yang disimpan sehingga mendapatkan nilai kemiripan. Data yang ada terlebih dahulu dilakukan transformasi data menjadi data numerik, sehingga bisa dihitung menggunakan *Euclidean distance* yang dapat dilihat pada Persamaan 1.

$$\sqrt{\sum_{i=1}^{n}(X_{i2} - X_{i1})^2} \quad (1)$$

X_{i2} : data uji
X_{i1} : data *training*
i : record (baris) ke- i dari tabel
n : jumlah data *training*

2) Algoritma *Naïve Bayes*

Naïve Bayes merupakan aloritma yang mudah diimplementasikan dalam data *mining* dibandingkan algoritma lainnya dalam lingkup klasifikasi. Algoritma ini juga dapat memproses teks dan angka [8]. Berikut adalah rumus dalam algoritma *naïve bayes*:

$$\mathrm{P}(H|X) = \frac{P(X|H)\,P(H)}{P(X)} \quad (2)$$

X : Data dengan *class* yang belum diketahui
H : Hipotesi data merupakan suatu *class* spesifik
$P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X *(posterior probabilitas)*
$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
$P(H)$: Probabilitas hipotesis H
$P(X)$: Probabilitas X

3) Algoritma *Decision Tree*

Konsep algoritma ini akan menjadikan struktur hirarki atau pohon yang kemudian dijadikan sebagai alur untuk menentukan pridiksi. Setiap pohon mempunyai cabang, untuk menuju cabang yang lain maka ada beberapa atribut yang harus di lengakapi dan berakhir ketika tidak ada cabang lagi atau biasa disebut daun [9]. *Entropy* terendah atau menghitung nilai *gain* yang tertinggi dari masing-masing atribut. Untuk menghitung nilai *indek entropy* menggunakan Persamaan 3, dan untuk menghitung *gain* menggunakan Persamaan 4.

$$Entropy\ (s) = \sum_{i=1}^{n} -P_i \operatorname{Log}_2 P_i \quad (3)$$

s : himpunan kasus
k : jumlah pasrtisi s
P_j : proporsi S*i* terhadap *S*

Menghitung nilai *gain* dengan rumus:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy\ (S_i) \quad (4)$$

S : himpunan kasus
A : Atribut
n : jumlah pada atribut A
$|S_i|$: jumlah kasus pada partisi ke-*i*
$|S|$: jumlah kasus dalam S

d. Analisa Kinerja Algoritma

Cross validation atau cross validasi merupakan sebuah metode statistik yang dapat digunakan dalam mengevaluasi kinerja algoritma atau model, dengan cara data akan dipisahkan menjadi 2 bagian data komoditas yaitu, 1 bagaian digunakan sebagai traing dan 1 bagian digunakan sebagai data testing [10]. Algoritma atau model dilatih oleh subset pembelajaran dan divalidasi oleh subset validasi dengan pemilihan jenis *Cross Validation* yang dapat didasarkan pada ukuran dataset. Biasanya *Cross Validation K-fold* digunakan karena dapat mengurangi waktu komputasi dengan tetap menjaga keakuratan estimasi [4]. Tabel 3 merupakan *confusion matrix* untuk mendapatkan performa *accuracy*, *precition*, dan *recall*:

Tabel 3. *Confution Matrix*

Acctual Class	**Predicted Class**	
	+	-
+	TP	FN
-	FP	TN

Keterangan:
TP : True Positif
TN : True Negatif
FP : False Positif
FN : False Negatif

Accuracy atau akurasi digunakan untuk mengukur dan mengevaluasi keakuratan dari hasil klasifikasi, semakin besar nilai akurasi maka semakin bagus tingkat pengklasifikasiannya, perhitungan akurasi menggunakan Persamaan 5.

$$Accuracy = \frac{TP+TN}{TP+FN+TP+TN} \quad (5)$$

Precision atau biasa disebut presisi atau biasa dikenal dengan nama *confidence* merupakan sebuah model perhitungan untuk mencari hasil dari proporsi jumlah kasus dengan hasil diagnosa positif. Perhitungan nilai presisi menggunakan Persamaan 6.

$$Precition = \frac{TP}{TP+FP} \quad (6)$$

Recall atau *sensitivity* merupakan sebuah model perhitungan untuk mencari hasil proporsi jumlah kasus positif yang diidentifikasikan dengan benar. Perhitungan nilai *recall* menggunakan Persamaan 7.

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

e. Laporan

Setelah melakukan beberapa tahapan pengumpulan data, pre-procesing, klasifikasi, dan evaluasi, maka langkah yang terakhir yaitu penyusunan laporan. Penyusunan laporan bertujuan untuk pendokumentasi penelitian dalam bentuk naskah dan sebagai bukti bahwa penelitian telah dilakukan sesuai dengan aturan-aturan, sesuai dengan teori-teori yang telah dikemukakan para ilmuwan.

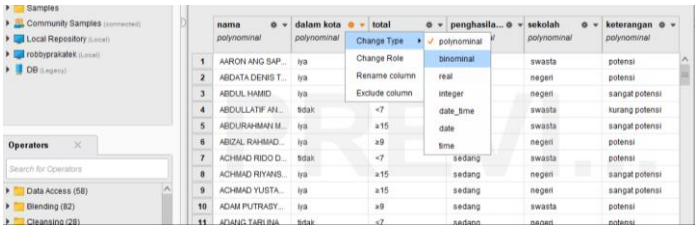
3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah klasifikasi dan sebuah informasi dalam penentuan lokasi promosi mahasiswa baru pada IIB Darmajaya Lampung. Dalam pengklasifikasian menggunakan tiga algoritma yang setara yaitu, *Decision Tree*, *Naïve Bayes*, dan *K-Nearest*

Neighbor dengan menggunakan ketiga algoritma tersebut diharapkan mendapatkan algoritma yang memiliki nilai akurasi yang baik, sehingga informasi yang dihasilkan dapat digunakan sebagai bahan pertimbangan bagian pemasaran untuk melakukan promosi pada tahun-tahun berikutnya.

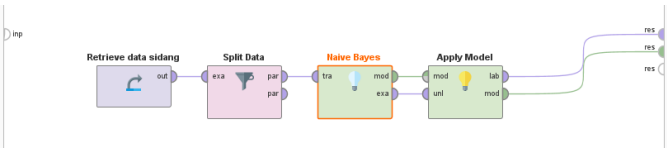
3.1. Penerapan Algoritma *K- Nearest Neighbor*

Penelitian ini menggunakan *software rapid miner*, untuk melakukan pre-prosesing data, dan pembangunan model klasifikasi, serta evaluasi.



Gambar 3 Proses seleksi atribut dari setiap variabel

Pada Gambar 3 dapat dilihat bahwa variabel dalam kota terisi atribut secara otomatis, namun atribut tersebut tidak sesuai dengan yang diharapkan, cara untuk merubahnya yaitu menekan simbol segitiga yang berada tepat di sabelah nama varibel tersebut, kemudian pilih *Change Type*, dan pilih *Binominal*, di karena *record* data hanya memiliki 2 pilihan yaitu, ya dan tidak. Varibel total untuk atribut datanya mengunakan *Polynominal* dikarenakan record yang ada pada varibel tersebut memiliki lebih dari 2 pilihan, seleksi atribut ini dilakukan pada semua varibel termasuk juga hingga pelabelan, sehingga atribut yang di pilih sesuai dengan *record* dari masing-masing varibel tersebut.



Gambar 4 *Design* penerapan algortima menggunakan *rapid miner*

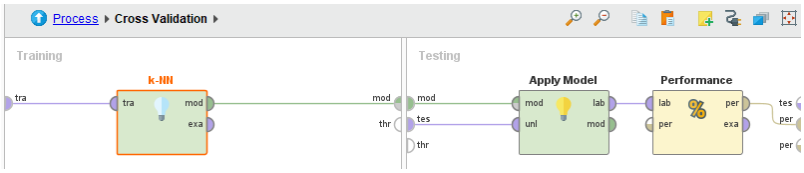
Gambar 4 merupakan langkah berikutnya yaitu penerapan alogritma *Naïve Bayes*, pada tahapan ini pilih *design*, kemudian pilih data yang telah disimpan pada *local repository*, selanjutnya dalam *operators* gunakan *split data*, yang berguna untuk menentukan data *testing* dan trening, dalam *split data* menggunakan perbadingan 70% dan 30%. Tahapan berikutnya cari algoritma *Naïve Bayes* pada bagian *operators*, dan masukan kedalam *design*, yang terakhir yaitu masukan *apply model* kedalam *design* untuk mendapatkan nilai *confidence* serta hubungkan dari data, *spit data*, algoritma *Naïve Bayes*, *apply model* dan *output*.

Row No.	keterangan	prediction(k...	confidence(...	confidence(...	confidence(...	nama	dalam kota	total	penghasilan...
1	potensi	potensi	1	0	0	AARON ANG ...	iya	≥9	sedang
2	potensi	potensi	1	0	0	ABDATA DEN...	iya	≥9	sedang
3	sangat potensi	sangat potensi	0	1	0	ABDUL HAMID	iya	≥15	sedang
4	kurang potensi	kurang potensi	0	0	1	ABDULLATIF ...	tidak	<7	rendah
5	sangat potensi	sangat potensi	0	1	0	ABDURAHMA...	iya	≥15	sedang
6	potensi	potensi	1	0	0	ABIZAL RAH...	iya	≥9	sedang
7	potensi	potensi	1	0	0	ACHMAD RID...	tidak	<7	sedang
8	sangat potensi	sangat potensi	0	1	0	ACHMAD RIY...	iya	≥15	sedang
9	sangat potensi	sangat potensi	0	1	0	ACHMAD YU...	iya	≥15	sedang
10	potensi	potensi	1	0	0	ADAM PUTRA...	iya	≥9	sedang
11	potensi	potensi	1	0	0	ADANG TAR...	tidak	<7	sedang
12	sangat potensi	sangat potensi	0	1	0	ADE DEL ALVI	iya	≥15	rendah
13	potensi	potensi	1	0	0	ADE KURNIA...	tidak	<7	sedang
14	potensi	potensi	1	0	0	ADE SAPUTRO	tidak	<7	sedang

Gambar 5 Hasil dari Algoritma *K- Nearest Neighbor*

Berdasarkan Gambar 5 dapat dijelaskan bahwa data *testing* atas nama Aaron Ang Saputra memiliki nilai *confidence* sangat potensi sebesar 0 dan nilai *confidence* potensi sebesar 1 serta nilai *confidence* kurang potensi 0 sehingga atas nama Aaron Ang Saputra dapat di kelompokan dalam lokasi promosi potensi. Hal serupa juga untuk nama-nama yang lainnya disesuikan dengan nilai dari masing-masing variable. Hasil prediksi tersebut merupakan proses kinerja untuk algoritma *K-Nearest Neighbor* berdasrkan Persamaan (1).

Dalam melatih dan menguji pemodelan *Cross Validation* menggunakan tiga operator yaitu, pada bagian *training* digunakan untuk algoritma *K-Nearest Neighbor*, dan untuk bagian *testing* digunakan fitur *Apply Model* yang berguna untuk untuk menampilkan *confusion table*, dan operator *performance* digunakan untuk menampilkan hasil dari *accuracy*, *precition*, dan *recall*, yang dapat dilihat pada Gambar 6.



Gambar 6. Susunan operator *cross validation* algoritma *K-Nearest Neighbor*

Table View Plot View

accuracy: 99.61% +/- 0.76% (micro average: 99.61%)

	true potensi	true sangat potensi	true kurang potensi	class precision
pred. potensi	374	3	0	99.20%
pred. sangat potensi	0	795	2	99.75%
pred. kurang potensi	0	0	107	100.00%
class recall	100.00%	99.62%	98.17%	

Gambar 7. Hasil *cross validation* algoritma *K-Nearest Neighbor*

Berdasarkan Gambar 7 di atas dapat di lihat bahwa algoritma *K-Nearest Neighbor* memiliki nilai akurasi sebesar 99,61% hal ini dapat dibuktikan dengan perhitungan manual dengan rumus Persamaan (5).

$$Accuracy = \frac{374+395+107}{374+3+795+2+107} = \frac{1276}{1281} \times 100\% = 99{,}61\%$$

Nilai presisi untuk prediksi label potensi sebesar 99,20% hal ini dapat dibuktikan dengan perhitungan manual dengan rumus Persamaan (6).

$$Presisi = \frac{374}{374+3} = \frac{374}{377} \times 100\% = 99{,}20\%$$

Nilai *recall* untuk prediksi label potensi sebesar 97,98% hasil tersebut dibuktikan dengan perhitungan manual dengan rumus Persamaan (7).

$$Recall = \frac{374}{374+0} = \frac{374}{374} \times 100\% = 100\%$$

Hasil tersebut berdasarkan ketentuan K=5 dan weighted dengan type mixed measures, dan mixed measure mixed euclidean disatance, serta *cross validation number of folds* sebesar 10 dan *sampling type automatic*.

3.2. Penerapan Algoritma Naïve Bayes

Dalam implementasi algorima *Naïve Bayes* tahap awal sama dengan Gambar 3 dan *design* untuk susunan operator sesuai dengan Gambar 4 yang membedakan hanya algorima yang digunakan.

Row No.	keterangan	prediction(keterangan)	confidence(potensi)	confidence(sangat potensi)	confidence(kurang potensi)	nama	dalam kota
1	kurang potensi	kurang potensi	0.000	0.000	1.000	ABDULLATIF ...	tidak
2	sangat potensi	sangat potensi	0.000	1.000	0.000	ACHMAD YU...	iya
3	potensi	potensi	1.000	0.000	0.000	ADE KURNIA...	tidak
4	potensi	potensi	1.000	0.000	0.000	ADELIA ALIC...	iya
5	sangat potensi	sangat potensi	0.000	1.000	0.000	ADHIT ARI SA...	iya
6	kurang potensi	kurang potensi	0.007	0.000	0.993	ADI PRASET...	tidak
7	potensi	potensi	1.000	0.000	0.000	ADI SYUMAN ...	iya
8	sangat potensi	sangat potensi	0.000	1.000	0.000	ADILLA SETY...	iya
9	potensi	potensi	1.000	0.000	0.000	ADITYA PAN...	tidak
10	sangat potensi	sangat potensi	0.000	1.000	0.000	AELY GUSNITA	iya
11	sangat potensi	sangat potensi	0.000	1.000	0.000	AGUNG PRIM...	iya
12	potensi	potensi	1.000	0.000	0.000	AHMAD ARD...	iya
13	potensi	potensi	1.000	0.000	0.000	AHMAD IRPA...	iya
14	potensi	potensi	1.000	0.000	0.000	AJENG BELL...	iya

Gambar 8 Hasil proses algoritma *naïve bayes*

Berdasarkan Gambar 8 dapat dijelaskan bahwa data *testing* atas nama Abdullatif Angga Saputra memiliki nilai *confidence* sangat potensi sebesar 0,000 dan nilai *confidence* potensi 0,000 serta nilai *confidence* kurang potensi sebesar 1,000 sehingga atas nama Abdullatif Angga Saputra dapat di kelompokan dalam lokasi promosi kurang potensi. Hal serupa juga untuk nama-nama yang lainnya disesuikan dengan nilai dari masing-masing variable. Hasil prediksi tersebut merupakan proses kinerja algorima Naïve Bayes sesuai dengan Persamaan (2).

Pengujian algorima *Naïve Bayes* menggunakan pemodelan *cross validation*, untuk operator yang digunakan sama dengan Gambar 6 yang membedakan adalah algortima yang digunakan.

accuracy: 84.78% +/- 4.09% (micro average: 84.78%)

	true potensi	true sangat potensi	true kurang potensi	class precision
pred. potensi	370	185	4	66.19%
pred. sangat potensi	4	613	2	99.03%
pred. kurang potensi	0	0	103	100.00%
class recall	98.93%	76.82%	94.50%	

Gambar 9. Hasil *cross validation* algorima *Naïve Bayes*

Berdasarkan Gambar 9 dapat di lihat bahwa algoritma *Naïve Bayes* memiliki nilai akurasi sebesar 84,78% hal ini dapat dibuktikan dengan perhitungan manual dengan menggunakan Persamaan (5).

$$Accuracy = \frac{370+613+103}{370+185+4+4+613+2+103} = \frac{1086}{1281} \times 100\% = 84{,}78\%$$

Nilai presisi untuk prediksi label potensi sebesar 99,20% hal ini dapat dibuktikan dengan perhitungan manual menggunakan Persamaan (6).

$$Presisi = \frac{370}{370+185+3} = \frac{370}{559} \times 100\% = 66{,}18\%$$

Nilai *recall* untuk prediksi label potensi sebesar 97,98% hasil tersebut dibuktikan dengan perhitungan manual menggunakan Persamaan (7).

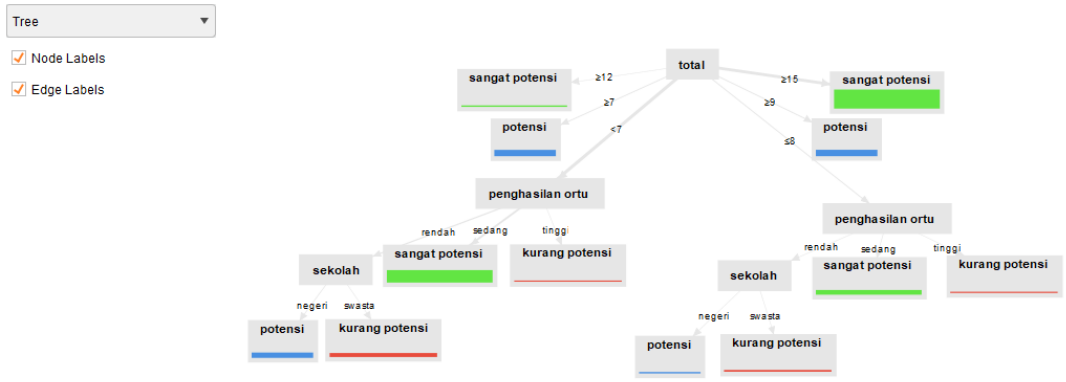
$$Recall = \frac{370}{370+4} = \frac{370}{374} \times 100\% = 98{,}93\%$$

Hasil tersebut dengan ketentuan *cross validation number of folds* sebesar 10 dan *sampling type automatic*.

Penerapan Algoritma *Decision Tree*

Pada implementasi algorima *Decision Tree* tahap awal yang dilakukan seperti pada Gambar 3 dan *design* untuk susunan operator sesuai dengan Gambar 4 yang membedakan hanya algorima yang digunakan.

Hasil dari proses algoritma *Decision Tree* atau pohon keputusan dapat digunakan untuk mengetahui indikator apa saja yang berpengaruh dalam data yang telah digunakan [5].



Gambar 10. Hasil proses algoritma *Decision Tree*

Berdasarkan Gambar 10 dapat dijelaskan bahwa node pertama merupakan variabel total alumni, jika alumni lebih dari 12 dan 15 maka dapat disimpulkan bahwa asal sekolah tersebut sangat potensi, namun jika alumni lebih dari 7 dan lebih dari 9 maka asal sekolah mahasiswa tersebut masuk dalam kelompok potensi, dan jika alumni kurang dari 7 dan 9 maka dilihat kembali pengasilan orang tua, jika berpengasilan tinggi maka masuk kedalam kelompok kurang potensi, dan jika penghasilan orang tua sedang masuk kedalam potensi dan jika pengasilan orang tua rendah maka dapat dilihat kembali asal sekolah mahasiswa tersebut, jika berasal dari sekolah negeri maka masuk kedalam kelompok potensi dan jika berasal dari sekolah swasta maka masuk kedalam kelompok kurang potensi, namun ada variabel yang tidak muncul yaitu variabel dalam kota, varibel tersebut tidak muncul dikarenakan nilai *gain* terlampau kecil, yang berarti varibael tersebut tidak terlalu berpengaruh dalam pohon keputusan. Proses kinerja algorima *Decision Tree* sesuai dengan Persamaan (3) dan Persamaan (4).

Proses pengujian algorima *Decision Tree* menggunakan pemodelan *cross validation*, untuk operator yang digunakan sama dengan Gambar 6 yang membedakan hanya algortima yang digunakan.

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true potensi	true sangat potensi	true kurang potensi	class precision
pred. potensi	374	0	0	100.00%
pred. sangat potensi	0	798	0	100.00%
pred. kurang potensi	0	0	109	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 11. Hasil *cross validation* algorima *Decision Tree*

Berdasarkan Gambar 11 dapat di lihat bahwa algoritma *Decision Tree* memiliki nilai akurasi sebesar 100% hal ini dapat dibuktikan dengan perhitungan manual dengan menggunakan Persamaan (5).

$$Accuracy = \frac{374+798+109}{374+798+109} = \frac{1281}{1281} \times 100\% = 100\%$$

Nilai presisi untuk prediksi label potensi sebesar 99,20% hal ini dapat dibuktikan dengan perhitungan manual menggunakan Persamaan (6).

$$Presisi = \frac{374}{374+0} = \frac{374}{374} \times 100\% = 100\%$$

Nilai *recall* untuk prediksi label potensi sebesar 97,98% hasil tersebut dibuktikan dengan perhitungan manual menggunakan Persamaan (7).

$$Recall = \frac{374}{374+0} = \frac{374}{374} \times 100\% = 100\%$$

Hasil tersebut menggunakan ketentuan *criterion gain ratio*, *apply pruning* dengan *confidence* 0,1 dan *apply prepruning* 0,01 dan *maximal depth* 10 serta *cross validation number of folds* sebesar 10 dan *sampling type automatic*.

3.3. Hasil Pengujian Algorima

Berdasarkan hasil uraian sebelumnya maka hasil evaluasi dari masing-masing algoritma dapat dilihat pada Tabel 3.

Tabel 3 Hasil pengukuran kinerja algorima menggunkan *cross validation*

Algoritma	Akurasi	Presisi	*Recall*
Decision Tree	100%	100%	100%
K-Nearest Neighbor	99,61%	99,20 %	100%
Naïve Bayes	84,78%	66,18%	98,93%

Berdasarkan Tabel 3 menunjukkan bahwa algoritma yang memiliki nilai akurasi, presisi, *recall*, yang baik yaitu algoritma *Decision Tree* dengan nilai masing-masing adalah 100%, namun untuk variabel luar kota tidak muncul pada *rapid miner*, setelah di analisa menunjukkan bahwa variabel luar kota memiliki nilai gain yang cukup rendah, sehingga variabel luar kota dianggap tidak berpengaruh. Berdasarkan eksperiman maka dapat disimpulkan algoritma *Decision Tree* lebih unggul di bading algritma *K-Nearest Neighbor* dan algoritma *Naïve Bayes*.

4. KESIMPULAN DAN SARAN

Penelitian ini melakukan komparasi hasil evaluasi dari algoritma *Decision Tree, Naïve Bayes, dan K-Nearest Neighbor* pada penentuan lokasi promosi. Hasil klasifikasi dapat digunakan sebagai rujukan oleh pihak pemasaran untuk penentuan lokasi promosi serta dapat memetakan lokasi promosi yang sangat potensi, potensi, dan kurang potensi sehingga dapat dijadikan bahan pertimbangan untuk melakukan promosi penerimaan mahasiswa baru ditahun yang akan datang.

Algoritma *Decision Tree* lebih unggul baik secara akurasi, presisi, maupun *recall* dibandingkan dengan algoritma *K-Nearest Neighbor* dan *Naïve Bayes*, dengan selisih nilai berturut-turut untuk akurasi yaitu 0,39% dan 15,22%. Selisih nilai presisi berturut-turut yaitu 0.8% dan 33,82%. Sementara untuk recall *Decision Tree* dan *K-Nearest Neighbor* memiliki nikai yang sama yaitu 100%, namun untuk *Naïve Bayes* dengan selisih nilai 1.07%. Atas dasar hal tersebut maka algoritma *Decision Tree* lebih direkomendasikan dalam pemilihan lokasi promosi untuk penerimaan calon mahasiswa baru di IIB Darmajaya.

Pekerjaan selanjutnya yang akan dilakukan adalah dengan menambahkan variable baru untuk mengetahui pengaruhnya terhadap peningkatan akurasi.

DAFTAR PUSTAKA

[1] A. Saifudin, "Metode Data Mining Untuk Seleksi Calon Mahasiswa," vol. 10, no. 1, pp. 25–36, 2018.

[2] T. T. Chasanah and Widiyono, "Penentuan Strategi Promosi Penerimaan Mahasiswa Baru Dengan Algoritma Clustering K-Means," *IC-Tech*, vol. 12, no. 1, pp. 39–44, 2017.

[3] H. Kurniawan and J. S. Informasi, "Aplikasi Datamining Untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Apriori Di Ibi Darmajaya Bandar Lampung," *J. Teknol. Inf. Magister Darmajaya*, vol. 2, no. 01, pp. 79–93, 2016.

[4] A. P. Ayudhitama and U. Pujianto, "Analisa 4 Algoritma Dalam Klasifikasi Penyakit Liver Menggunakan," *J. Inform. Polinema*, vol. 6, pp. 1–9, 2020.

[5] C. A. Sugianto, F. R. Maulana, and D. Mining, "Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)," vol. 18, no. 4, pp. 321–331, 2019.

[6] H. Said, N. H. Matondang, and H. N. Irmanda, "Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi," *Techno.Com*, vol. 21, no. 2, pp. 256–267, 2022, doi: 10.33633/tc.v21i2.5901.

[7] A. Y. Saputra and Y. Primadasa, "Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *Techno.Com*, vol. 17, no. 4, pp. 395–403, 2018, doi: 10.33633/tc.v17i4.1864.

[8] N. Sagala and H. Tampubolon, "Komparasi Kinerja Algoritma Data Mining pada Dataset Konsumsi Alkohol Siswa," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 2, p. 98, 2018, doi: 10.23917/khif.v4i2.7061.

[9] D. Sartika and D. Indra, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *J. Tek. Inform. Dan Sist. Inf.*, vol. 1, no. 2, pp. 151–161, 2017.

[10] E. Ismanto and M. Novalia, "Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas," *Techno.Com*, vol. 20, no. 3, pp. 400–410, 2021, doi: 10.33633/tc.v20i3.4576.