

Penilaian Apartemen Pada Perusahaan Konsultan Properti Menggunakan Metode *Naïve Bayes*

Lala Nilawati ^{1,*}, Martin ²

¹ Sistem Informasi; Universitas Bina Sarana Informatika; Jl. Kramat Raya No.98, Senen, Jakarta Pusat, 021-23231170; e-mail: lala.lni@bsi.ac.id

² Ilmu Komputer; Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri Jakarta; Jl. Jatiwaringin Raya No. 2, Jakarta Timur, 021-28534236; e-mail: martin.auf@gmail.com

* Korespondensi: e-mail: lala.lni@bsi.ac.id

Diterima : 17 Maret 2020; Review: 25 Maret 2020; Disetujui: 09 April 2020

Cara sitasi: Nilawati L, Martin. 2020. Penilaian Apartemen Pada Perusahaan Konsultan Properti Menggunakan Metode *Naïve Bayes*. Information System for Educators and Professionals. 4 (2): 114 – 123.

Abstrak: Penilaian terhadap aset berupa properti sangat diperlukan. Penilaian properti dapat dimanfaatkan dalam semua segi kehidupan. Perusahaan konsultan properti mempunyai peran memberikan masukan, ataupun bimbingan mengenai kondisi properti tertentu. Salah satu jenis properti yang banyak digemari adalah apartemen. Metode *Naïve Bayes* merupakan metode pengklasifikasian dengan probabilitas dalam bidang memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Pada penelitian menggunakan metode *Naïve Bayes* terhadap prediksi penilaian apartemen pada perusahaan konsultan properti. Faktor-faktor yang mempengaruhi penilaian apartemen akan dilihat dari *rent price* (harga), *city* (lokasi apartemen/kota), *size* (ukuran), *furnisihing* (kelengkapan), *bedroom* (jumlah kamar tidur), *bathroom* (jumlah kamar mandi) dan *maid badroom* (jumlah kamar pembantu) dengan hasil *output* penilaian nya adalah *condition* atau kondisi apartemen yaitu *excellent*, *best* dan *good*. Pengujian *data set* model algoritma *Naïve Bayes* menggunakan *software Rapidminer* menggunakan *cross validation (cv)* dengan nilai parameter 10. Pengujian model yang terbentuk akan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi. Berdasarkan hasil pengujian terlihat hasil akurasi penggunaan metode *Naïve Bayes* cukup tinggi, untuk kasus prediksi penilaian apartemen yaitu sebesar 92.24%.

Kata kunci: *data mining*, *naïve bayes*, penilaian apartemen

Abstract: Valuation of assets in the form of property is now very necessary. Property valuation can be utilized in all aspects of life. Property consulting companies have the role of providing input, or guidance regarding certain property conditions. One type of property that is currently in high demand is the apartment. Meode *Naïve Bayes* is a classification method with probabilities in the field of predicting future opportunities based on past experience. In this research, *Naïve Bayes* method will only be used to predict apartment valuations at property consulting companies. Factors that influence apartment valuation will be seen from the rent price, city (location of the apartment / city), size (size), furnishing (completeness), bedroom (number of bedrooms), bathroom (number of bathrooms) and maid badroom (the number of servants' rooms) with the result of the assessment output is the condition or condition of the apartment which is excellent, best and good. Testing the data set of the *Naïve Bayes* algorithm model using *Rapidminer* software. In this model using *cross validation (CV)* with parameter value 10. Testing the formed model will use *Confusion Matrix* to determine the level of accuracy. Based on the test results, the accuracy of using the *Naïve Bayes* method is quite high, for an apartment prediction assessment case that is equal to 92.24%.

Keywords: *apartment valuation*, *data mining*, *naïve bayes*

1. Pendahuluan

Penilaian terhadap aset berupa properti sangat diperlukan baik di kalangan swasta maupun pemerintah, dimana tujuannya untuk meningkatkan akuntabilitas dan terciptanya tertib administrasi dalam mendukung pengelolaan (optimalisasi) aset ke arah yang lebih baik dan *modern*. Penilaian properti dapat dimanfaatkan dalam semua segi kehidupan, misalnya: Penilaian Properti untuk tujuan jual-beli maupun kelayakan sewa atau pengkajian sewa berjalan serta untuk menilai obyek pajak untuk menentukan besaran pajaknya, menilai aset daerah untuk kepentingan penyusunan neraca atau laporan keuangan daerah dan untuk menghitung komposisi hutang dan ekuitas dalam struktur permodalan suatu usaha. Perusahaan konsultan properti mempunyai peran memberikan masukan ataupun bimbingan mengenai kondisi properti tertentu. Salah satu jenis properti yang banyak digemari adalah apartemen. Pada proses memberikan masukan salah satunya perusahaan konsultan properti akan melakukan penilaian terhadap properti yang dibahas adalah apartemen. Bahan pertimbangan dalam penilaian sebuah apartemen biasanya konsultan properti memberikan studi pembangunan, analisa kelayakan, membuat laporan hingga analisa peluang investasi bagi para klien. Analisa yang diberikan konsultan properti biasanya menjadi pertimbangan bagi klien untuk membeli, menjual atau mengelola properti mereka.

Data mining adalah salah satu cabang ilmu yang dikategorikan masih baru tetapi sudah menghasilkan keuntungan yang cukup besar bagi kemajuan pengetahuan saat ini. Banyaknya manfaat dari penggunaan *data mining* sudah tidak diragukan lagi terlihat dari banyaknya kasus penelitian yang telah banyak diselesaikan [1]. Teknik yang digunakan dalam *data mining* adalah teknik klasifikasi untuk menemukan model yang menggambarkan dan membedakan konsep atau kelas-kelas data, agar model yang sudah dibentuk dapat digunakan untuk memprediksi. Pengelompokan teknik klasifikasi dikategorikan kedalam beberapa algoritma diantaranya Pohon Keputusan, Jaringan Saraf Tiruan (JST), *Naïve Bayes* dan lain-lain. Selain itu *data mining* merupakan suatu proses otomatis yang digunakan untuk mendapatkan pengetahuan berdasarkan nilai-nilai yang terdapat pada basis data berukuran besar, dan dari banyaknya teknik *data mining* yang dapat digunakan salah satunya dapat digunakan untuk memprediksi nilai yang akan keluar dari suatu *domain* yang diteliti [2]. Sebagian pihak ada yang beranggapan bahwa *data mining* difungsikan hanya sebagai *machine learning* atau analisa statistik yang berjalan di atas *database*, tetapi hal ini tidak disetujui karena disini lain ada juga yang berpendapat bahwa *database* berperan penting di *data mining* terutama dalam optimisasi *query*, karena *data mining* mengakses data yang ukurannya besar (bisa sampai *terabyte*). Salah satu kemampuan *data mining* yang dapat dirasakan pada penggunaannya yaitu terutama pada proses mengekstraksi pengetahuan yang diinginkan dari sebuah basis data, menganalisis, memahami, atau bahkan menggambarkan, melalui pencarian pengetahuan dalam basis data dalam proses identifikasi pola-pola yang tepat, berpotensi manfaat, sehingga dapat dipahami secara mudah bagi siapapun yang menggunakannya.

Tugas yang dimiliki oleh *data mining* dapat dikategorikan menjadi dua kategori utama, yaitu: a). Tugas prediksi, tugas ini difokuskan kepada prediksi nilai dari atribut tertentu berdasarkan nilai dari atribut lainnya. Adapun jenis atribut yang digunakan untuk membuat prediksi disebut atribut penjas atau *independent variable*, sedangkan atribut yang diprediksi dikenal dengan atribut target atau *dependent variable*. Tujuan dari tugas prediksi ini akan membangun pemodelan *prediktif* yang akan membangun sebuah model untuk *dependent variable*. Tugas dari *independent variable* adalah pemodelan prediksi yang didalamnya terdapat dua tipe yaitu klasifikasi dan regresi. Klasifikasi digunakan untuk proses *dependent variable* yang *diskrit*. Sedangkan regresi digunakan untuk proses *dependent variable* yang kontinyu. b) Tugas deskriptif, yaitu tugas yang difokuskan pada pembuatan pola yang dapat digunakan untuk menyimpulkan hubungan di dalam data. Dalam tugas deskriptif ini terdiri dari beberapa model diantaranya adalah analisa asosiasi yang difungsikan untuk menentukan pola yang mendeskripsikan fitur-fitur data yang saling berhubungan, yang berikutnya adalah analisa *cluster* yaitu merupakan proses untuk mencari kelompok-kelompok data, sehingga data yang berada dalam satu kelompok memiliki kemiripan dibandingkan data yang terletak pada kelompok lain, dan langkah terakhir adalah proses analisa deteksi *anomaly* yaitu merupakan proses identifikasi data yang memiliki perbedaan karakteristik yang signifikan dengan data yang lain.

Naïve Bayes merupakan teknik pengklasifikasian dengan metode probabilitas dan statistik (memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya)

[3]. Konsep dasarnya adalah *Teorema Bayes* yaitu menghitung probabilitas untuk melakukan pengklasifikasian. Perhitungan probabilitas menggunakan persamaan dari teorema Bayes [4] yaitu:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots\dots\dots (1)$$

Sumber: (Kusrini and Luthfi, 2009)

Keterangan:

- X : Data dengan *class* yang belum diketahui
- H : Hipotesis data merupakan suatu *class* yang spesifik
- P(H|X) : Probabilistik hipotesis H berdasar kondisi X (*posteriori* probabilistik)
- P(H) : Probabilistik hipotesis H (*prior* probabilitas)
- P(X|H) : Probabilistik hipotesis X berdasar kondisi pada hipotesis H
- P(X) : Probabilitas X

Penggunaan metode *Naïve Bayes* telah banyak digunakan oleh para peneliti sebelumnya. *Naïve Bayes* digunakan dalam klasifikasi *Tweet* berdasarkan keterkaitannya terhadap topik tertentu pada media sosial *Twitter* didapat hasil bahwa nilai akurasi yang didapatkan dari semua topik termasuk ke dalam kategori kurang baik dengan tingkat kesalahan diatas 40% dan bahkan ada yang mempunyai akurasi dibawah 50% [5]. Metode *Naïve Bayes* bisa dikategorikan salah satu metode klasifikasi berjenis teks pada penelitian tentang prediksi rating film. *Naïve Bayes* merupakan metode sederhana dan efisien, serta sangat familiar jika dipakai untuk pengklasifikasian teks dan mempunyai performa yang bagus pada banyak domain [6].

Tingkat akurasi dalam menganalisis Tingkat Mutu Pelayanan (TMP) untuk pelanggan di PT. PLN Batam menggunakan metode Algoritma *Naïve Bayes* didapat sebesar 76.46% [7]. Sistem melalui perhitungan dari *Naïve Bayes*, mampu memberikan rekomendasi penilaian pembelajaran kepada guru-guru berdasarkan hasil evaluasi angket yang dilakukan kepada siswa-siswi SMA *Nation Star Academy* [8]. Hasil penelitian lain menunjukkan bahwa pengujian model algoritma *Naïve Bayes*, diketahui memiliki nilai *accuracy* dan *AUC* cukup tinggi yaitu sebesar 96,67% dalam menentukan calon penerima beasiswa pada SMK Pasim Plus Sukabumi secara lebih efektif dan efisien [9]. Pada penelitian lain memperlihatkan bagaimana pengembangan keoptimalan *Naïve Bayes Classifier* pada studi kasus penilaian status gizi dengan mempertimbangkan independensi parameter, dengan pemodelan data menggunakan *Backward Feature Selection*. Kolerasi parameter mempengaruhi hasil klasifikasi [10]. Penggunaan metode *Naïve Bayes* pada beberapa penelitian sering diuji komparasi dari hasil akurasi yang dihasilkan, seperti pada penelitian menganalisa kelancaran pembiayaan pada KSPPS/BMT AL-Fadhila membandingkan metode *Naïve Bayes Classifier* dan metode *Decision Tree* (C4.5), dan di dapat hasil *Naïve Bayes Classifier* lebih mudah untuk digunakan karena hanya memiliki alur perhitungan yang tidak panjang sedangkan pada Metode Algoritma *Decision Tree* (C4.5) jika data diubah atau ditambah maka perhitungan akan memerlukan waktu yang lebih lama lagi [11]. Komparasi metode *Naïve Bayes* dan metode C4.5 dilihat juga dari sisi algoritmanya yang terdapat pada penelitian tentang prediksi penyakit liver. Penelitian membandingkan algoritma yang paling baik dengan menggunakan metode pengujian *Cross Validation* dan *Split Percentace* [12]. Hasil komparasi metode *Naïve Bayes* dengan metode *Support Vector Machine* berbasis algoritma PSO dalam penelitian tentang prediksi kesuburan menunjukkan bahwa algoritma *Naïve Bayes* menghasilkan tingkat akurasi sebesar 85% [13]. Penggunaan metode *Naïve Bayes* sering dikombinasikan juga untuk memperoleh hasil optimasi terbaik dengan metode lainnya, pada penelitian mengoptimalkan metode *Naïve Bayes* menggunakan *Particle Swarm Optimization* (PSO) serta untuk meningkatkan akurasi dalam memprediksi kredit macet di koperasi. Hasil optimasi PSO terhadap metode *Naïve Bayes* menunjukkan untuk memprediksi kredit macet meningkatkan akurasi sebesar 21,03% [14].

Berdasarkan beberapa pemaparan penggunaan metode *Naïve Bayes* oleh para peneliti sebelumnya, pada penelitian akan menggunakan metode *Naïve Bayes* terhadap prediksi penilaian apartemen pada perusahaan konsultan properti. Faktor-faktor yang mempengaruhi penilaian apartemen akan dilihat dari *rent price* (harga), *city* (lokasi apartemen/kota), *size*

(ukuran), *furnisihing* (kelengkapan), *bedroom* (jumlah kamar tidur), *bathroom* (jumlah kamar mandi) dan *maid bedroom* (jumlah kamar pembantu) dengan hasil *output* penilaiannya adalah *condition* atau kondisi apartemen yaitu *excellent*, *best* dan *good*. Penilaian apartemen akan dilakukan menggunakan metode *Naïve Bayes*, pengujian *data set* model algoritma *Naïve Bayes* akan menggunakan *software Rapidminer*, dan hasil akhirnya akan terlihat bagaimana akurasi penilaian yang didapat dalam penggunaan metode.

2. Metode Penelitian

Penelitian dilakukan dalam beberapa tahapan mulai dari 1) Pengumpulan data, untuk jenis data yang digunakan adalah data primer dan data sekunder. Data primer yaitu pengumpulan data dengan menjalankan penelitian langsung pada perusahaan konsultan properti yang menjadi objek penelitian. Sedangkan untuk data sekunder yaitu pengumpulan data dari beberapa literatur dan karya ilmiah yang berhubungan dengan pengolahan dan penyusunan hasil penelitian; 2) Pengolahan data awal yaitu proses pemilihan data yang berasal dari sumber data yaitu data perusahaan konsultan properti, yang berupa pembentukan sumber data acak (*set the random seed*) dan pembentukan variabel pemilihan (*partition variabel*). Proses pengambilan data sampel menggunakan teknik *Systematic Random Sampling* yaitu teknik yang merupakan modifikasi dari teknik *random sampling*, langkahnya adalah dengan cara memilih subjek dari daftar populasi secara sistematis dan bukan secara acak. Data yang didapat dari perusahaan konsultan properti masih berupa data yang terdiri dari berbagai parameter, sehingga harus diolah terlebih dahulu, kemudian dilakukan penyeleksian data, data dimodifikasi sesuai kebutuhan kemudian data ditransformasikan ke dalam bentuk data yang diinginkan, sehingga dapat dilakukan persiapan dalam pembuatan model.

Beberapa tahap yang dilakukan dalam proses pengolahan data adalah: a). *Data Validation*, yaitu proses yang digunakan untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*) dengan melakukan analisa terlebih dahulu; b). *Data Integration and Transformation*, yaitu proses yang dilakukan untuk meningkatkan akurasi dan efisiensi algoritma, kemudian data ditransformasikan ke dalam *software Rapidminer*; c). *Data Size Reduction and Discretization*, yaitu tahapan proses yang dilakukan untuk memperoleh data set dari sekumpulan data dengan memperhitungkan jumlah atribut dan *record* yang jumlahnya lebih sedikit tetapi bersifat informatif; 3) Penerapan metode, yaitu menerapkan metode *Naïve Bayes* untuk mendapatkan hasil optimal yang dilakukan dengan menggunakan *software rapidminer*. Data yang akan digunakan dibagi menjadi dua bagian, yaitu data *training* dan data *testing*; 4) Pengujian. Proses pada tahap ini adalah data yang telah diolah akan digunakan didalam pengujian algoritma metode *Naïve Bayes* dengan melihat tingkat akurasinya, semakin besar tingkat akurasinya maka akan semakin baik; 5) Evaluasi hasil akurasi, penilaian terhadap *dataset* untuk mengetahui tingkat akurasi metode *Naïve Bayes*.

3. Hasil dan Pembahasan

Dalam penelitian data yang digunakan berjumlah 1624 *record*, dibagi menjadi dua yaitu untuk data *training* (80%) dan data *testing* (20%), sebanyak 1299 *record* akan digunakan untuk data *training* (80%) dan 325 *record* digunakan sebagai data *testing* (20%). Selain itu dalam penelitian menggunakan *K-Fold Cross Validation*, yaitu teknik validasi dengan membagi data secara acak kedalam *k* bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak *k*. Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum pengujian nilai *k* dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian nilai *k* yang digunakan berjumlah 10 atau *10-fold Cross Validation*. Berikut ini adalah contoh tampilan data sebelum dilakukan *preprocessing*.

Tabel 1. Data Sebelum *Processing*

Apartments	Currency	Price	Place	Size	Furnishing	Bedroom	Bathroom	Maidbed	Condition
Shangri-La Residences	USD	15000	South Jakarta	773 m2	Furnished	4	3	1	Excellent
Senopati Suites	USD	8500	South Jakarta	540 m2	Furnished	4	3	1	Excellent
Pacific Place	USD	8000	South Jakarta	500 m2	Furnished	4	4	1	Excellent

Apartments	Currency	Price	Place	Size	Furnishing	Bedroom	Bathroom	Maidbed	Condition
Pacific Place	USD	8000	South Jakarta	500 m2	Unfurnished	4	4	1	Excellent
Four Seasons	USD	8000	South Jakarta	350 m2	Furnished	4	3	1	Excellent
Airlangga / Ritz Carlton	USD	7500	South Jakarta	440 m2	Furnished	4	3	2	Excellent
The Capital Residence	USD	6600	South Jakarta	320 m2	Furnished	4	3	1	Excellent
Kempinski Residence	USD	6500	Central Jakarta	261 m2	Furnished	3	2	1	Excellent
Kempinski Residence	USD	6500	Central Jakarta	261 m2	Furnished	3	2	1	Excellent
Pacific Place	USD	6500	South Jakarta	500 m2	Unfurnished	4	3	2	Excellent
The Pakubuwono House	USD	6500	South Jakarta	450 m2	Unfurnished	4	4	1	Excellent
Shangri-La Residences	USD	6200	South Jakarta	320 m2	Furnished	3	2	1	Excellent
Batavia	USD	6000	Central Jakarta	42 m2	Furnished	1	1	0	Good
Dharmawangsa Residences	USD	6000	South Jakarta	365 m2	Unfurnished	4	3	1	Excellent
The Pakubuwono residence	USD	6000	South Jakarta	303 m2	Furnished	3	3	1	Excellent
The Pakubuwono residence	USD	6000	South Jakarta	303 m2	Furnished	3	2	1	Excellent
The Pakubuwono residence	USD	6000	South Jakarta	303 m2	Furnished	3	2	1	Excellent
The Pakubuwono residence	USD	6000	South Jakarta	303 m2	Unfurnished	3	3	1	Excellent
The Capital Residence	USD	6000	South Jakarta	160 m2	Furnished	3	3	1	Excellent

Sumber: Hasil Penelitian (2020)

Pada Tabel 1. menunjukkan data awal yang di dapat dari perusahaan konsultan properti. Data ini kemudian akan mulai diproses sehingga akan menghasilkan data seperti pada tampilan Tabel 2.

Tabel 2. Data Setelah *Processing*

Kota	Harga	Ukuran	Kelengkapan	Kamar Tidur	Kamar Mandi	Kamar Pembantu	Kondisi
South Jakarta	Mahal	Luas	Furnished	Sedang	Banyak	Banyak	Excellent
South Jakarta	Mahal	Luas	Furnished	Banyak	Banyak	Sedang	Excellent
South Jakarta	Mahal	Sedang	Furnished	Sedang	Banyak	Sedikit	Best
South Jakarta	Mahal	Luas	Furnished	Sedang	Banyak	Banyak	Excellent
South Jakarta	Mahal	Luas	Furnished	Banyak	Banyak	Sedang	Excellent
South Jakarta	Murah	Sempit	Furnished	Sedikit	Sedang	Sedang	Best
South Jakarta	Murah	Sempit	Furnished	Sedikit	Sedang	Sedang	Best
South Jakarta	Murah	Sempit	Furnished	Sedikit	Sedang	Sedikit	Good
South Jakarta	Murah	Sempit	Furnished	Sedikit	Sedang	Sedikit	Good
South Jakarta	Murah	Sempit	Furnished	Sedikit	Sedang	Sedikit	Good
South Jakarta	Mahal	Sedang	Semi Furnished	Sedang	Banyak	Banyak	Excellent
Central Jakarta	Mahal	Sedang	Furnished	Sedikit	Sedang	Sedang	Best
Central Jakarta	Mahal	Sedang	Furnished	Sedikit	Sedang	Sedang	Best
Central Jakarta	Mahal	Sempit	Furnished	Sedikit	Sedikit	Sedikit	Good
Central Jakarta	Mahal	Sedang	Semi Furnished	Sedang	Sedang	Sedang	Best
Central Jakarta	Mahal	Sedang	Semi Furnished	Sedang	Sedang	Sedang	Best
Central Jakarta	Sedang	Sempit	Furnished	Sedikit	Sedikit	Sedikit	Good
Central Jakarta	Sedang	Sempit	Furnished	Sedikit	Sedikit	Sedikit	Good
Central Jakarta	Sedang	Sedang	Semi Furnished	Sedang	Sedang	Sedang	Best

Kota	Harga	Ukuran	Kelengkapan	Kamar Tidur	Kamar Mandi	Kamar Pembantu	Kondisi
Central Jakarta	Sedang	Sempit	Semi Furnished	Sedikit	Sedang	Sedikit	Best
Central Jakarta	Sedang	Sedang	Unfurnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Banyak	Banyak	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedang	Sedang	Sedang	Best
West Jakarta	Sedang	Sedang	Furnished	Sedikit	Sedang	Sedang	Best

Sumber: Hasil Penelitian (2020)

Pada Tabel 2. Menunjukkan tampilan contoh data setelah dilakukan *processing*, maka didapat hasil bahwa digunakan tujuh variabel *input* yaitu, kota, harga, ukuran, kelengkapan, kamar tidur, kamar mandi, kamar pembantu, dan satu *output* yaitu kondisi. Pada tahap *processing* isian setiap data variabel diubah nilainya dengan ketentuan yang terlihat pada tabel 3.

Tabel 3. Tabel Penilaian

Nama Variabel	Nilai	Kategori
<i>Place / Kota</i>	<i>South Jakarta</i>	<i>South Jakarta</i>
	<i>Central Jakarta</i>	<i>Central Jakarta</i>
	<i>West Jakarta</i>	<i>West Jakarta</i>
<i>Price / Harga</i> (dalam satuan mata uang US\$)	Mahal	3000-15000
	Sedang	1001-2999
	Murah	0-1000
<i>Size / Ukuran</i> (dalam m ²)	Luas	400-800
	Sedang	101-399
	Sempit	0-100
<i>Furnishing / Kelengkapan</i>	<i>Furnished</i>	<i>Furnished</i>
	<i>Semi Furnished</i>	<i>Semi Furnished</i>
	<i>Unfurnished</i>	<i>Unfurnished</i>
<i>Bedroom / Kamar Tidur</i>	Banyak	5-6
	Sedang	3-4
	Sedikit	0-2
<i>Bathroom / Kamar Mandi</i>	Banyak	3-4
	Sedang	2
	Sedikit	0-1
<i>Maid Bedroom / Kamar Pembantu</i>	Banyak	2
	Sedang	1
	Sedikit	0

Sumber: Hasil Penelitian (2020)

Pada Tabel 3. Menunjukkan informasi pemberian nilai untuk setiap isian dari variabel sesuai kategori. Sedangkan *output* yang dihasilkan akan memunculkan nilai *excellent*, *best* dan *good* untuk hasil penilaian dari apartemen.

Implementasi *Modeling Naïve Bayes*

Sebelum *data set* tersebut diimplementasi kedalam sebuah software *Rapidminer*, akan terlebih dahulu dihitung probabilitas hipotesis dari masing masing kelas data tersebut menggunakan persamaan dari teorama Bayes. Ada tiga kelas pada output kondisi yaitu *excellent*, *best* dan *good* dari sebuah penilaian apartemen. Tabel 4 adalah hasil perhitungan probabilitas.

Tabel 4. Hasil Perhitungan Probabilitas

Atribut	Jumlah Kasus	Excellent	Best	Good	P(X Ci)		
					Excellent	Best	Good
Total	1624	51	1371	202			
Kota							
South Jakarta	1533	51	1287	195	1	0.938730853	0.965346535
Central Jakarta	82	0	75	7	0	0.054704595	0.034653465
West Jakarta	9	0	9	0	0	0.006564551	0
Harga							
Mahal	337	51	285	1	1	0.207877462	0.004950495
Murah	83	0	34	49	0	0.024799416	0.242574257
Sedang	1204	0	1052	152	0	0.767323122	0.752475248
Ukuran							
Luas	16	7	9	0	0.137254902	0.006564551	0
Sedang	1144	44	1098	2	0.862745098	0.800875274	0.00990099
Sempit	464	0	264	200	0	0.192560175	0.99009901
Kelengkapan							
Furnished	1483	50	1237	196	0.980392157	0.902261123	0.97029703
Semi Furnished	37	1	35	1	0.019607843	0.025528811	0.004950495
Unfurnished	104	0	99	5	0	0.072210066	0.024752475
Kamar Tidur							
Banyak	4	2	2	0	0.039215686	0.001458789	0
Sedang	859	49	810	0	0.960784314	0.590809628	0
Sedikit	761	0	559	202	0	0.407731583	1
Kamar Mandi							
Banyak	122	51	71	0	1	0.051787017	0
Sedang	1236	0	1221	15	0	0.89059081	0.074257426
Sedikit	266	0	79	187	0	0.057622174	0.925742574
Kamar Pembantu							
Banyak	10	3	7	0	0.058823529	0.005105762	0
Sedang	1204	48	1146	10	0.941176471	0.835886214	0.04950495
Sedikit	410	0	218	192	0	0.159008023	0.95049505

Sumber: Hasil Penelitian (2020)

Pada tabel 4. bisa dilihat peluang atau probabilitas untuk tiga output pada prediksi penilaian apartemen, yaitu:

Probabilitas (*Excellent*): $51/1624=0.031403941$;

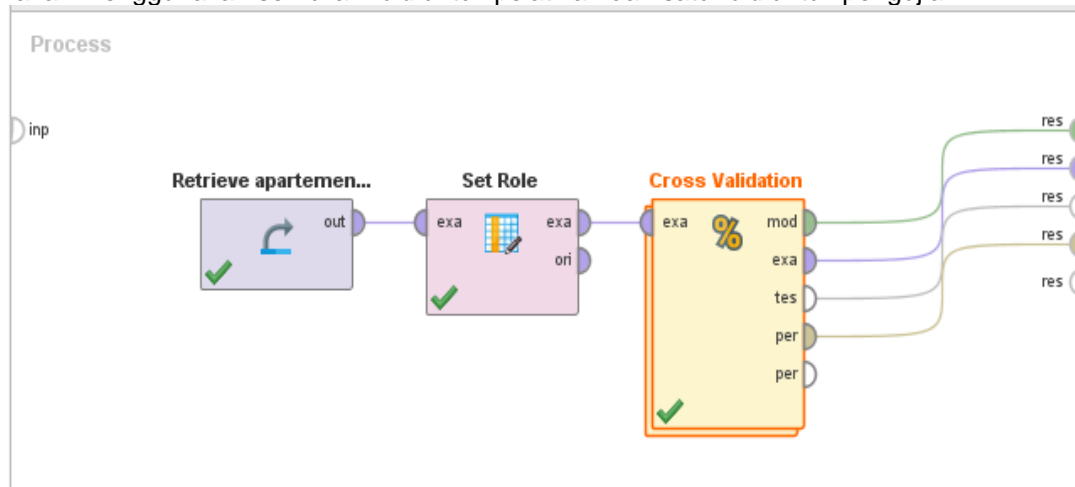
Probabilitas (*Best*): $1371/1624=0.844211823$;

Probabilitas (*Good*): $202/1624=0.124384236$.

Tahap Pengujian

Pada tahap ini *data set* kemudian diuji dengan membuat model algoritma *Naïve Bayes*. *Software* yang digunakan adalah *Rapidminer*. Pada model ini menggunakan *cross validation (cv)* dengan nilai parameter 10. 10 *fold cv* adalah salah satu *K fold cv* yang direkomendasikan untuk pemilihan model terbaik karena cenderung memberikan estimasi akurasi yang kurang

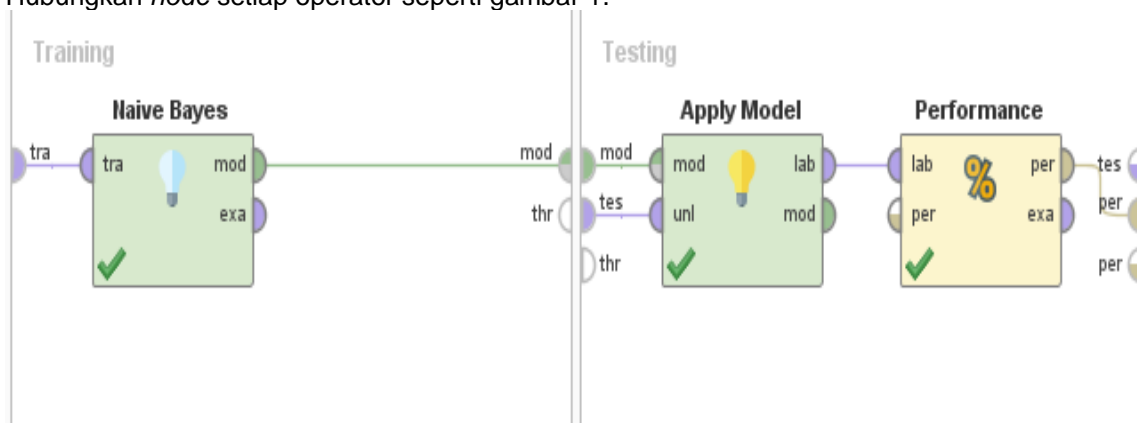
bias dibandingkan dengan *cv* biasa, *leave-one-out cv* dan *bootstrap*. Dalam 10 fold *cv*, data dibagi menjadi 10 *fold* berukuran kira-kira sama, sehingga kita memiliki 10 *subset* data untuk mengevaluasi kinerja model atau algoritma. Untuk masing-masing dari 10 subset data tersebut, *cv* akan menggunakan sembilan *fold* untuk pelatihan dan satu *fold* untuk pengujian.



Sumber: Hasil Penelitian (2020)

Gambar 1. Pengujian Model *Naive Bayes* Bagian Proses

Pada gambar 1. menunjukkan model dibentuk dengan data *input* yang di *import* ke dalam *rapidminer*, kemudian diberikan operator *set role* dan operator *cross validation*. Hubungkan *node* setiap operator seperti gambar 1.



Sumber: Hasil Penelitian (2020)

Gambar 2. Pengujian Model *Naive Bayes* Bagian *Cross Validation*

Gambar 2 memperlihatkan isi dari operator *cross validation*. Pada bagian *training* di tambahkan operator algoritma *Naive Bayes*, sedangkan dibagian *testing* ditambahkan operator *Apply Mode* dan *Performance*

Hasil Pengujian dan Evaluasi

Setelah data diolah maka model yang terbentuk akan diuji dengan menggunakan *Confusion Matrix* untuk mengetahui tingkat akurasi. *Confusion Matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar, prediksi positif yang salah, prediksi negatif yang benar, dan prediksi negatif yang salah. Akurasi akan dihitung dari seluruh prediksi yang benar (baik prediksi positif dan negatif). Semakin tinggi nilai akurasi, semakin baik pula model yang dihasilkan. Hasil pengujian dapat dilihat pada tampilan gambar 3.

accuracy: 92.24% +/- 2.91% (micro average: 92.24%)

	true Excellent	true Best	true Good	class precision
pred. Excellent	49	5	0	90.74%
pred. Best	2	1251	4	99.52%
pred. Good	0	115	198	63.26%
class recall	96.08%	91.25%	98.02%	

Sumber: Hasil Penelitian (2020)

Gambar 3. Tampilan Hasil Pengujian Akurasi Metode *Naïve Bayes*

Pada Gambar 3. Terlihat hasil akurasi penggunaan Metode *Naïve Bayes* cukup tinggi untuk kasus prediksi penilaian apartemen yaitu sebesar 92.24%.

4. Kesimpulan

Berdasarkan hasil pengujian menggunakan data lengkap dengan metode algoritma *Naïve Bayes*, dapat dilihat bahwa *Naïve Bayes* termasuk algoritma yang sangat baik untuk digunakan pada data prediksi penilaian apartemen, karena nilai akurasi yang dihasilkan termasuk sangat baik dengan nilai sebesar 92.24% dan dapat disimpulkan algoritma ini termasuk dalam kategori bagus. Pada penelitian selanjutnya akan dibuat komparasi hasil akurasi dengan metode lainnya, dan akan diuji hasil optimasi terbaiknya.

Referensi

- [1] P. P. Widodo, R. T. Handayanto, and Herlawati, *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains, 2013.
- [2] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Boston: Pearson Addison Wesley, 2006.
- [3] A. Mutoi Siregar and A. Puspabhuana, "Data Mining." CV Kekata Group, Surakarta, p. 200, 2017.
- [4] Kusriani and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [5] M. Baydhowi, W. Apriliah, and I. Kurniawan, "Klasifikasi Tweet Berdasarkan Keterkaitan Tweet Terhadap Topik Tertentu Pada Twitter Menggunakan *Naïve Bayes*," *Inf. Syst. Educ. Prof.*, vol. 4, no. 1, pp. 95–103, 2019.
- [6] R. W. Pratiwi and Y. S. Nugroho, "Prediksi Rating Film Menggunakan Metode *Naive Bayes*," *J. Tek. Elektro*, vol. 8, no. 2, pp. 60–63, 2016.
- [7] R. Harman, "Penerapan Metode Data Mining (*Naïve Bayes*) Untuk Menganalisis Tingkat Mutu Pelayanan (TMP) Pada Pelanggan Study Kasus PT. PLN Batam," *CBIS J.*, vol. 1, pp. 45–56, 2013.
- [8] H. Yulius, K. Brian, and I. B. Trisno, "Perancangan Sistem Informasi Kepuasan Layanan Pendidikan Dengan Pendekatan *Naïve Bayes* Studi Kasus Di Nation Star Academy," in *Prosiding SNST ke-10 Tahun 2019 Fakultas Teknik Universitas Wahid Hasyim Semarang*, 2019, pp. 95–100.
- [9] R. A. Saputra and S. Ayuningtias, "Penerapan Algoritma *Naive Bayes* Untuk Penentuan Calon Penerima Beasiswa Pada SMK Pasim Plus Sukabumi," *Swabumi*, vol. IV, no. 2, pp. 114–120, 2016.
- [10] A. Fitri Cahyanti, R. Saptono, and S. Widya Sihwi, "Penentuan Model Terbaik pada Metode *Naive Bayes Classifier* dalam Menentukan Status Gizi Balita dengan Mempertimbangkan Independensi Parameter," *J. Teknol. Inf. ITSmart*, vol. 4, no. 1, p. 28, 2016.
- [11] T. Rosandy, "Perbandingan Metode *Naive Bayes Classifier* Dengan Metode *Decision Tree (C4.5)* Untuk Menganalisa Kelancaran Pembiayaan (Study Kasus : KSPPS/BMT Al-Fadhila)," *J. Teknol. Inf. Magister Darmajaya*, vol. 2, no. 01, pp. 52–62, 2016.

- [12] E. Rahmawati, "Analisa Komparasi Algoritma Naive Bayes Dan C4.5 Untuk Prediksi Penyakit Liver," *Techno Nusa Mandiri*, vol. XII, no. 2, pp. 21–26, 2015.
- [13] E. Nurelasari, "Komparasi Algoritma Naive Bayes Dengan Support Vector Machine Berbasis Particle Swarm Optimization untuk Prediksi Kesuburan," *Bina Insa. ICT J.*, vol. 5, no. 1, pp. 61–70, 2018.
- [14] T. Mardiana, "Optimasi Naïve Bayes Dengan Particle Swarm Optimization Dan Stratified Untuk Prediksi Kredit Macet Pada Koperasi," *J. Ris. Inform.*, vol. 1, no. 1, pp. 43–50, 2018.