

KLASIFIKASI PENENTUAN LOKASI STRATEGIS OUTLET BANK SYARIAH INDONESIA DENGAN METODE NAÏVE BAYES CLASSIFIER

Navioer Rizal^{1*}, Mohamat Fatekurohman², Dian Anggraeni³

^{1,2,3}Jurusan Matematika, Fakultas MIPA, Universitas Jember

*e-mail : navioerr@gmail.com

DOI: 10.14710/J.GAUSS.12.4.477-486

Article Info:

Received: 2022-03-23

Accepted: 2024-01-03

Available Online: 2024-04-01

Keywords:

Classification; Locations; Naïve Bayes Classifier; Outlet; Strategic

Abstract: Today, the development of the banking sector occurs in the conventional banking sector and the Islamic banking sector, one of which is developing Bank Syariah Indonesia. Bank Syariah Indonesia strives to develop a strategic new office network or branch outlet location that has not been optimal. This research aims to know and analyze the model and determination of variable importance and its effect on the strategic location classification of Bank Syariah Indonesia outlets using the Naïve Bayes Classifier method. The classification model of strategic location determination of offices or outlets obtained from the analysis results in calculating prior probability values and conditional probabilities. The results of the model evaluation test indicator for the Naïve Bayes Classifier method showed an accuracy value of 94,12% and an AUC score of 0,9808. The model was able to classify 16 of the 17 data. The model produces the results of variables importance 6 recommendations variables of the 7 variables used in the study it is location in office area, location in industrial area, populations density of the area, moslem populations of the area, distance from the security office, and distance from the market. The variable importance can be a consideration of Bank Syariah Indonesia optimizing indicators of the office location selection.

1. PENDAHULUAN

Bank Syariah Indonesia adalah contoh utama dari pertumbuhan industri perbankan syariah, yang berkembang seiring dengan perkembangan lembaga perbankan di seluruh dunia. Lokasi jaringan kantor baru atau outlet cabang yang strategis menjadi prioritas Bank Syariah Indonesia. Permasalahan yang dialami oleh Bank Syariah Indonesia sebagai hasil merger tersebut adalah jaringan kantor atau *outlet* Bank Syariah Indonesia saat ini masih belum optimal sebarannya di beberapa wilayah dibandingkan dengan kantor atau *outlet* bank konvensional. Menurut Kasmir (2010), karena suatu lokasi merupakan investasi jangka panjang bagi bank itu sendiri, banyak karakteristik yang harus dipertimbangkan ketika menentukan lokasi strategis suatu bank, termasuk tingkat efisiensi lokasi operasional. Menurut Faturahman (2018), permasalahan mengenai penentuan lokasi strategis untuk pendirian suatu kantor atau *outlet* bank cukup kompleks dengan pertimbangan beberapa variabel keputusan seperti jarak kantor dengan pusat perdagangan, pemerintahan, kantor keamanan, lokasi kantor berada di wilayah industri, dan lokasi kantor berada di wilayah perkantoran. Secara alami, faktor-faktor ini bisa dikategorikan untuk memungkinkan analisis dalam pengambilan keputusan.

Model klasifikasi yang dikembangkan dan dipakai untuk analisis penentuan lokasi, perlu dilaksanakan analisis yang lebih lanjut secara sistematis faktor mana yang paling penting. Teknik *Naive Bayes Classifier* merupakan salah satu cara untuk mengatasi permasalahan tersebut. Hasil studi literatur penelitian terdahulu membuktikan keefektifan dan keefisienan metode *Naive Bayes Classifier* dalam melakukan proses klasifikasi variabel atau parameter sebagai dasar analisis pengambilan keputusan. Penelitian yang dilaksanakan

oleh Imandasari dkk (2019) menandakan akurasi sebanyak 78,95% dalam proses klasifikasi penentuan lokasi pembangunan sumber air di daerah PDAM Tirta Lihou. Wulandari dan Iqbal (2021) juga melihat kelayakan situs promosi dalam Penerimaan Peserta Didik Baru (PMB) dan menemukan bahwa metode *Naive Bayes Classifier* memiliki akurasi 91,43 persen. Investigasi pendekatan *Naive Bayes Classifier* untuk mengidentifikasi kesesuaian rumah sebelumnya telah dilaksanakan oleh Fithri (2016). Kajian ini memberikan saran mengenai tiga faktor utama, antara lain keamanan, bebas banjir, dan harga hidup, dari tujuh karakteristik tambahan yang bisa dimanfaatkan masyarakat dalam memilih lokasi tinggal. Pendekatan *Naive Bayes Classifier* yang dikembangkan berdasarkan penelitian sebelumnya mampu memecahkan permasalahan kategorisasi data, termasuk kesulitan dalam mengidentifikasi lokasi optimal untuk konstruksi bangunan.

Tujuan utama dari riset ini adalah memakai *Naive Bayes Classifier* untuk mengembangkan model klasifikasi untuk mengidentifikasi penempatan cabang Bank Syariah Indonesia yang optimal. Model yang diperoleh akan dipakai sebagai proses prediksi kestrategisan lokasi kantor atau *outlet* Bank Syariah Indonesia. Riset ini juga bertujuan untuk mengetahui dan menganalisis hasil penentuan *variable importance* serta pengaruhnya terhadap klasifikasi lokasi strategis *outlet* Bank Syariah Indonesia memakai metode *Naive Bayes Classifier* sehingga hasil dari model tersebut akan lebih *interpretable* dan bisa menjadi rekomendasi untuk pemilihan indikator yang perlu dioptimalkan dalam pemilihan suatu kantor atau *outlet* Bank Syariah Indonesia.

2. TINJAUAN PUSTAKA

Teknik mengantisipasi data yang tidak mempunyai pelabelan jelas kelompok awal atau kelas salah satunya dengan klasifikasi, klasifikasi adalah proses pengembangan model dan fungsi yang dipakai untuk mengkarakterisasi atau membedakan data dalam berbagai kelompok. Klasifikasi pada suatu data secara umum terdiri dari dua proses. Proses pertama akan dilaksanakan pembangunan atau pembentukan model klasifikasi berdasarkan data sebelumnya. Proses kedua dalam klasifikasi adalah menentukan akurasi dari model klasifikasi yang dilaksanakan apakah layak untuk diterapkan pada data tersebut ataupun untuk data baru. Berikut adalah beberapa contoh algoritma metode proses klasifikasi yang sering dilaksanakan seperti *Naive Bayes Classifiers / Bayesian Classifiers, Decision/Classification Trees, Neural Networks, Algoritma Genetika, Rough Sets, K-Nearest Neighbor, Metode Rule Based, Memory Based Reasoning, dan Support Vector Machines (SVM)* (Han, 2011).

Teknik klasifikasi berdasarkan data mining salah satunya termasuk *Naive Bayes Classifier*. Teknik klasifikasi *Naive Bayes Classifier* ini memakai konsep penggabungan dari Teorema Bayes dengan konsep *Naif*. Konsep metode ini bisa disebut sebagai *class-conditional independent* untuk penyederhanaan perhitungan yang nantinya akan terlibat dalam proses klasifikasi. Konsep Teorema Bayes sebagai dasar dalam metode ini bisa dirumuskan sebagai berikut.

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

keterangan:

- X : Data dengan Kelas yang belum diketahui
- H : Hipotesis data X adalah suatu kelas spesifik
- $P(H|X)$: Probabilitas hipotesis data H berdasarkan kondisi X (*Posteriori Probability*)
- $P(H)$: Probabilitas hipotesis H (*Prior Probability*)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
 $P(X)$: Probabilitas X

Persamaan (1) merupakan bentuk umum Teorema Bayes untuk menentukan probabilitas munculnya suatu kejadian bersyarat yang dipengaruhi oleh kejadian lainnya yang saling memengaruhi satu sama lain. Teorema Bayes tersebut selanjutnya akan disesuaikan dengan konsep klasifikasi memakai algoritma *Naïve Bayes Classifier* dengan diberikan sejumlah petunjuk untuk menentukan kelas yang cocok bagi suatu data. Berikut adalah bentuk proses penyesuaian rumus metode *Naïve Bayes Classifier* dengan Teorema Bayes.

$$P(C|F_1 \dots F_n) = \frac{P(C) \cdot P(F_1 \dots F_n|C)}{P(F_1 \dots F_n)} \quad (2)$$

Menurut Saleh (2015), bahwa jika C merupakan suatu kelas spesifik dalam suatu data sedangkan $F_1 \dots F_n$ merupakan parameter atau karakteristik petunjuk klasifikasi suatu data ke dalam kelas spesifik tertentu tersebut, maka selanjutnya bisa dijelaskan bahwa $P(C|F_1 \dots F_n)$ sebagai probabilitas *posterior* atau peluang munculnya suatu karakteristik atau parameter tertentu dalam kelas C . Metode *Naïve Bayes Classifier* memakai suatu asumsi yang mengakibatkan masing- masing syarat atau parameter $F_1 \dots F_n$ saling bebas satu sama lain yaitu asumsi *naif*. Berdasarkan penerapan asumsi *Naif* ini maka Persamaan (2) akan lebih disederhanakan lagi menjadi :

$$P(C|F_1 \dots F_n) = P(C) \cdot \prod_{i=1}^n P(F_i|C) \quad (3)$$

Prosedur klasifikasi teknik ini didasarkan pada model teorema Naive Bayes yang diwakili oleh Persamaan (3). Data X diprediksi termasuk dalam kelas C_i oleh Naive Bayes Classifier jika dan hanya jika memenuhi kriteria yang ditetapkan oleh Han (2011).

$$P(X|C_i) \cdot P(C_i) > P(X|C_j) \cdot P(C_j) \text{ untuk } 1 \leq j \leq m, j \neq i \quad (4)$$

dengan kata lain prediksi X akan termasuk ke dalam kelas C_i jika nilai $P(X|C_i) \cdot P(C_i)$ bernilai maksimum.

Proses klasifikasi setelah dilaksanakan penentuan model klasifikasi dilanjutkan dengan proses evaluasi performa model. Evaluasi performa model bisa dilaksanakan dengan berbagai metode salah satunya adalah *confusion matrix*. *Confusion matrix* atau yang disebut juga sebagai matriks kesalahan adalah suatu bentuk tata letak tabel yang menyajikan visualisasi dari kinerja suatu algoritma. Tata letak setiap tabel berupa kolom dalam matriks ini akan mewakili kondisi kelas yang diprediksi, sedangkan baris mewakili kondisi sebaliknya yaitu kondisi kelas sebenarnya dari suatu data. Tabel *confusion matrix* sebagai metode evaluasi program bisa disajikan seperti pada Tabel 1 yang dituliskan untuk data yang terdiri dari dua kelas yaitu kelas positif dan negatif (Bramer, 2007)

Tabel 1. *Confusion matrix*

True Condition	Predicted Condition	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Tabel *confusion matrix* memakai beberapa *record* istilah seperti *True Positive* yang merupakan data jumlah *record* positif yang diklasifikasi sebagai positif. *False Positive* ialah data jumlah *record* negatif yang diklasifikasikan sebagai positif, *False Negative* ialah data jumlah *record* positif yang diklasifikasikan sebagai negatif, dan *True Negative* data jumlah *record* negatif yang diklasifikasikan negatif. Evaluasi performa model akan memakai beberapa indikator. Indikator yang akan dipakai untuk evaluasi performa model ini adalah akurasi dan *precision*. Akurasi akan menandakan persentase hasil klasifikasi data secara benar dan tepat memakai algoritma tersebut atau kedekatan kondisi aktual dengan kondisi hasil prediksi. *Precision* menandakan persentase data prediksi benar positif dengan keseluruhan data prediksi positif atau menandakan derajat reliabilitas model dalam memberikan prediksi positif. Menurut Abdurrohman (2018), *precision* akan baik dipakai jika model bisa membedakan antara kondisi aktual negatif dengan aktual positif serta menghindari terjadinya kondisi *false positive* yang besar. Perhitungan nilai akurasi dan *precision* bisa dilaksanakan dengan memakai Persamaan (5) dan Persamaan (6) sebagai berikut.

$$\text{Akurasi} = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

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

Ada dua macam faktor yang berperan dalam penelitian ini: variabel independen dan variabel dependen. Setiap variabel yang dipertimbangkan dalam riset ini dijelaskan secara rinci di bawah ini.

a. Variabel bebas atau *independen*

Suatu variabel yang menyebabkan atau memberikan kontribusi terhadap perubahan variabel lain, dapat diartikan bahwa variabel tersebut adalah variabel *bebas* atau independen. Beberapa variabel *independen* yang dipakai dalam penelitian ini, antara lain:

1. Populasi penduduk (Ribuan)

Istilah "populasi" mengacu pada jumlah total orang yang tinggal di suatu tempat tertentu pada saat tertentu. Variabel ini akan menjadi indikator wilayah tempat pendirian lokasi menjadi strategis untuk mengembangkan dan menawarkan produk-produk perbankan dalam menjangkau nasabah maupun calon nasabah yang lebih banyak di suatu wilayah. Data dalam variabel ini akan melihat jumlah populasi penduduk daerah tersebut berdasarkan data publikasi Badan Pusat Statistik wilayah atau daerah terkait.

2. Jumlah penduduk beragama islam (Ribuan)

Jumlah penduduk beragama islam di suatu wilayah akan menjadi salah satu parameter atau indikator pengambilan keputusan kali ini. Masyarakat beragama islam menjadi salah satu segmen penting dalam pengembangan segmentasi Bank Syariah Indonesia dalam memasarkan produk-produk perbankannya. Data dalam variabel ini akan melihat jumlah penduduk beragama islam berdasarkan data publikasi Badan Pusat Statistik wilayah atau daerah terkait.

3. Jarak *outlet* dengan pusat perdagangan atau pasar (*km*)

Variabel lokasi *outlet* bank saat ini banyak disesuaikan dengan jarak yang tidak begitu jauh dari pusat perdagangan seperti pasar. Data dalam variabel ini akan diukur jarak terdekat antara *outlet* dengan pusat perdagangan atau pasar di wilayah tersebut.

4. Jarak *outlet* dengan pusat pemerintahan (*km*)

Data dalam variabel ini akan diukur jarak terdekat antara outlet dengan pusat pemerintahan di wilayah tersebut seperti kantor kecamatan, kantor kabupaten, dan lain sebagainya. variabel ini didasarkan oleh salah satu segmen nasabah yang akan dijangkau oleh Bank Syariah Indonesia adalah Pegawai Pemerintahan berstatus ASN atau Aparatur Sipil Negara.

5. Lokasi kawasan industri

Data dalam variabel ini akan dilaksanakan pengecekan total jumlah industri mulai dari mikro, kecil, menengah, dan besar yang terbisa di cakupan wilayah lokasi *outlet* Bank Syariah Indonesia tersebut.

6. Lokasi kawasan perkantoran

Data dalam variabel ini akan dilaksanakan pengecekan total jumlah kantor atau perusahaan milik perorangan, swasta, maupun negara yang terbisa di cakupan wilayah lokasi *outlet* Bank Syariah Indonesia tersebut.

7. Jarak *outlet* dengan kantor keamanan (*km*)

Data dalam variabel ini akan diukur jarak terdekat antara *outlet* dengan kantor keamanan di wilayah tersebut. Variabel letak lokasi yang berdekatan dengan kantor keamanan seperti kantor polisi ataupun lainnya akan memberikan kenyamanan dalam hal keamanan kepada para nasabah ketika melakukan transaksi melalui kantor atau outlet bank tersebut.

b. Variabel terikat atau *dependen*

Variabel terikat atau variabel *dependen* adalah variabel yang dipengaruhi atau diprediksi oleh suatu variabel bebas atau variabel *independen*. Variabel *dependen* riset ini adalah proses klasifikasi kantor cabang Bank Syariah Indonesia. Dalam riset ini simbol variabel terikat yang mewakili pilihan lokasi outlet cabang Bank Syariah Indonesia adalah (Y).

3. METODE PENELITIAN

Riset ini memanfaatkan informasi dari 58 cabang atau kantor Bank yang berbeda di kota Jember, Bondowoso, Situbondo, dan Banyuwangi, serta kota Lumajang, Malang, Pasuruan, dan Probolinggo, yang semuanya bisa ditemukan di situs resmi Bank. Hukum Syariah di Indonesia. Informasi sekunder dan primer akan dipakai untuk mengisi kesenjangan dalam penelitian ini. Kabupaten atau Kota dalam Angka Badan Pusat Statistik Tahun 2021 menjadi sumber data sekunder variabel yang dipakai dalam penelitian ini, antara lain populasi penduduk di wilayah tersebut (X_1) dan jumlah penduduk beragama Islam (X_2). Jarak ke pasar atau pusat perdagangan (X_3), posisi pusat pemerintahan (X_4), lokasi kawasan industri (X_5), lokasi kawasan perkantoran (X_6), dan jarak ke kantor keamanan (X_7) merupakan variabel data utama. yang diperoleh melalui pengukuran jarak terdekat dalam satuan kilometer antara lokasi kantor atau *outlet* dengan lokasi tempat berdasarkan variabel yang dibutuhkan melalui citra peta digital *Google Maps*. Data dari masing-masing variabel penelitian akan dikategorikan untuk proses analisis data dan proses pelabelan kelas awal dalam klasifikasi. Indikator dan bobot dalam proses pengkategorian riset ini disajikan dalam Tabel 2 sebagai berikut.

Tabel 2. Pengkategorian data penelitian

Variabel	Kategori	Indikator	Bobot Kategori
Populasi Penduduk (X_1)	Tinggi	> 100.000 orang	1
	Rendah	≤ 100.000 orang	0
Jumlah penduduk beragama Islam (X_2)	Tinggi	> 100.000 orang	1
	Rendah	≤ 100.000 orang	0
Jarak <i>outlet</i> dengan pusat perdagangan atau pasar (X_3)	Dekat	≤ 2 Km	1
	Jauh	> 2 Km	0
Jarak <i>outlet</i> dengan pusat pemerintahan (X_4)	Dekat	≤ 2 Km	1
	Jauh	> 2 Km	0
Lokasi kawasan Industri (X_5)	Lokasi Industri	≥ 10 Tempat Industri	1
	Bukan Lokasi Industri	< 10 Tempat Industri	0
Lokasi kawasan perkantoran (X_6)	Lokasi di wilayah perkantoran	≥ 10 Kantor	1
	Bukan Lokasi wilayah perkantoran	< 10 Kantor	0
Jarak <i>outlet</i> dengan kantor keamanan (X_7)	Dekat	≤ 2 Km	1
	Jauh	> 2 Km	0
Lokasi (Y)	Strategis	Total bobot ≥ 4	1
	Tidak Strategis	Total bobot < 4	0

Pengklasifikasi dengan *Naive Bayes Classifier* akan dipakai untuk mengkategorikan informasi atau data yang dikumpulkan untuk riset ini. Paket perangkat lunak statistik R-studio akan dipakai untuk analisis data. Data pelatihan bisa mencakup sebanyak 70% dari keseluruhan data, sedangkan data pengujian bisa mencakup sebanyak 30% dari total data. Prosedur analisis data dalam riset ini adalah sebagai berikut:

- 1) Proses pengumpulan data dari sumber data ke dalam *Microsoft Excel*.
- 2) Persiapan data yang akan dilaksanakan pembobotan kategori pada masing-masing parameter sesuai dengan skala pengkategorian data.
- 3) Data akan dibagi menjadi dua kelompok data yaitu data *training* sebagai data latih untuk membentuk model pada data dengan jumlah sebanyak 70% dari total data. Kelompok data kedua merupakan data *testing* sebagai data uji dari model yang telah dibentuk sebelumnya sebanyak 30% dari total jumlah data.
- 4) Proses akan dilanjutkan dengan pembentukan model klasifikasi pada data *training* dengan memakai metode *Naïve Bayes Classifiers*. Proses pembentukan model akan dilaksanakan dengan memakai bantuan program aplikasi R-studio.
- 5) Model yang telah diperoleh pada tahapan pengujian dengan data *training* akan dilanjutkan penentuan nilai *k-fold* terbaik untuk diterapkan pada *k-fold cross validation*. *K-fold cross validation* dipakai untuk meminimalkan kesalahan proses klasifikasi. Proses akan menguji dua nilai *k-fold* yaitu *5-fold* dan *10-fold* untuk selanjutnya dipilih nilai dengan rata-rata akurasi terbesar.
- 6) Proses evaluasi performa model pada data *testing* dilanjutkan dengan perhitungan tingkat akurasi data memakai *confusion matrix*. Proses dilaksanakan dengan memakai program aplikasi R-studio. Evaluasi model juga akan dilaksanakan dengan melihat kurva performa *Receiver Operating Characteristic Curve* (ROC) dengan indikator nilai *Area Under the ROC Curve* (AUC).
- 7) Hasil proses klasifikasi penentuan lokasi kantor *outlet* Bank Syariah Indonesia selanjutnya ditentukan variabel yang memengaruhi terhadap proses klasifikasi

berdasarkan model yang diperoleh dengan memakai fungsi *varImp* yang telah tersedia dalam program aplikasi R-studio.

4. HASIL DAN PEMBAHASAN

Riset ini akan melakukan uji coba memakai 4 percobaan pengujian beberapa persentase pembagian data. Pengujian dilaksanakan untuk memilih persentase pembagian data yang menghasilkan evaluasi model berupa nilai akurasi dan *precision* terbaik. Indikator evaluasi model akan ditentukan dari hasil *confusion matrix* pengujian pada data *training* dan data *testing*. Hasil pengujian beberapa persentase pembagian data disajikan dalam Tabel 3 berikut.

Tabel 3. Hasil pengujian persentase data *training* dan data *testing*

Data Training	Data Testing	Akurasi Data Training	Akurasi Data Testing	Precision Data Training	Precision Data Testing
70%	30%	87,80%	94,12%	81,57%	92,86%
75%	25%	90,91%	85,71%	89,19%	84,62%
85%	15%	90%	87,5%	88,37%	85,71%
90%	10%	88,68%	80%	88,64%	80%

Hasil pemilihan persentase pembagian 70% data *training* dan 30% data *testing* menghasilkan nilai indikator akurasi dan *precision* yang cukup baik dan konsisten untuk meminimalkan kondisi *false positive* yang dihindari dalam penelitian ini. Kondisi *false positive* akan dihindari dalam hasil penentuan lokasi strategis yaitu sebagai suatu kesalahan proses prediksi dari kelas data aktual lokasi tidak strategis namun menjadi kelas strategis. Hal tersebut tentunya akan menjadi sebuah bentuk kesalahan model yang nantinya bisa mengganggu kegiatan operasional bank dalam penerapannya untuk menguji kestrategisan suatu lokasi. Persentase pembagian data 70% dan 30% menghasilkan nilai akurasi data *training* sebanyak 87,80% dan untuk data *testing* sebanyak 94,12%. Persentase pembagian data 70% dan 30% juga menghasilkan nilai *precision* data *training* sebanyak 81,57% dan untuk data *testing* sebanyak 92,86%. Kedua indikator tersebut menghasilkan nilai yang terbaik dibandingkan dengan persentase pembagian data lain dimana terdapat penurunan nilai akurasi dan *precision* saat pengujian pada data *testing*, hasil tersebut mengindikasikan terdapat beberapa kesalahan proses prediksi kelas yang terjadi dimana salah satunya adalah terjadinya kondisi *false positive* yang lebih banyak.

Model klasifikasi yang diperoleh dari analisis data training 70% pada riset ini adalah perhitungan nilai probabilitas *prior* dan *conditional probabilities*. Nilai probabilitas *prior* dan *conditional probabilities* ini yang akan menjadi dasar perhitungan nilai probabilitas *posterior* dalam memprediksi kelas suatu data berdasarkan Persamaan (3). Hasil pengambilan keputusan pengklasifikasian selanjutnya didasarkan pada konsep pemaksimalan nilai probabilitas *posterior*.

Model klasifikasi yang telah diperoleh selanjutnya akan dilaksanakan proses kedua dalam klasifikasi yaitu uji evaluasi performa model. Pengujian model yang diperoleh pada metode *Naïve Bayes Classifier* pada riset ini akan dilaksanakan dengan memakai data *testing* dari pembagian proporsi data sebanyak 30% dari total data. Data *testing* yang dipakai dalam pengujian ini berjumlah 17 data dengan pembagian 13 data masuk dalam kelas lokasi strategis dan 4 data untuk kelas lokasi tidak strategis. Pengujian uji evaluasi model pertama

akan dilaksanakan dengan memakai visualisasi kinerja algoritma *Naïve Bayes Classifier* berdasarkan model klasifikasi yang diperoleh sebelumnya untuk data *testing* dengan memakai *confusion matrix* dengan pemilihan nilai $k = 10$ untuk metode *k-fold cross validation* untuk meminimalkan kesalahan klasifikasi. Hasil pengujian evaluasi model pada *confusion matrix* ditampilkan melalui Tabel 4 berikut.

Tabel 4. *Confusion matrix* evaluasi model klasifikasi

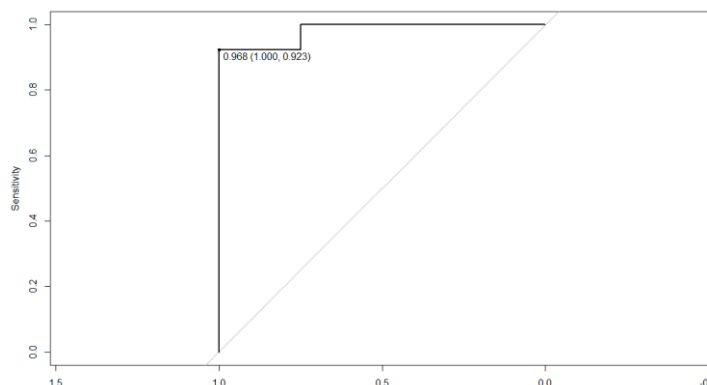
Kelas Data Aktual	Kelas Data Prediksi	
	Strategis	Tidak Strategis
Strategis	13	0
Tidak Strategis	1	3

Hasil *confusion matrix* pada Tabel 4 menandakan *record* data prediksi kelas lokasi kantor atau *outlet* Bank Syariah Indonesia memakai metode *Naïve Bayes Classifier*. *Record* data hasil prediksi kelas menandakan kondisi *true positive* sebanyak 13 data, *false positive* sebanyak 1 data, *true negative* sebanyak 3 data, dan tidak terdapat *record* data. Dari hasil *record* data pengujian evaluasi model selanjutnya bisa ditentukan indikator evaluasi model yaitu akurasi, *precision*, *sensitivity*, dan *specificity*. Hasil perhitungan indikator evaluasi model klasifikasi pada riset ini ditampilkan melalui Tabel 5 berikut.

Tabel 5. Indikator uji evaluasi model klasifikasi

Indikator Uji Evaluasi Model	Nilai
Akurasi	94,12%
<i>Precision</i>	92,86%

Hasil indikator uji evaluasi model pada data *testing* untuk metode *Naïve Bayes Classifier* menandakan nilai akurasi proses klasifikasi sebanyak 94,12%, hasil menandakan bahwa sebanyak 16 data dari 17 data *testing* yang dipakai diklasifikasikan dengan benar. Hal ini menandakan metode *Naïve Bayes Classifier* ini baik dalam melakukan klasifikasi penentuan lokasi strategis kantor atau *outlet* Bank Syariah Indonesia. Hasil uji evaluasi proses klasifikasi juga menghasilkan nilai *precision* sebanyak 92,86%, yang menandakan terdapat 1 kesalahan kelas prediksi kelas lokasi strategis.



Gambar 1. Kurva ROC

Uji evaluasi metode berikutnya dilaksanakan dengan melihat performa *Naïve Bayes Classifier* untuk melakukan klasifikasi dengan memakai kurva *Receiver Operating Characteristic Curve* (ROC) dengan indikator nilai *Area Under the ROC Curve* (AUC). Kurva ROC akan merepresentasikan grafik performa proses klasifikasi yang dilaksanakan oleh metode *Naïve Bayes Classifier* pada penelitian ini. Menurut Gorunescu (2011), Nilai AUC yang semakin besar mendekati 1,00 akan menandakan performa yang baik dan kuat dari proses klasifikasi yang dilaksanakan. Hasil pengujian performa model dengan kurva ROC pada Gambar 1 menandakan hasil nilai AUC sebanyak 0,9808. Hasil nilai AUC tersebut Menurut Gorunescu (2011) menandakan performa model klasifikasi dengan metode *Naïve Bayes Classifier* tersebut tergolong ke dalam klasifikasi berperforma klasifikasi yang sangat baik.

Hasil pengujian evaluasi performa model klasifikasi pada riset ini menandakan performa yang baik dalam melakukan klasifikasi. Proses akan dilanjutkan dengan penentuan variabel prediktor yang paling memengaruhi dalam proses dengan memakai konsep *variable importance*. Konsep *variable importance* ini akan melihat variabel-variabel mana yang memengaruhi berdasarkan model dan pola-pola klasifikasi yang dilaksanakan oleh metode *Naïve Bayes Classifier*. Penentuan *variable importance* ini dilaksanakan untuk membuat model hasil dari metode *Naïve Bayes Classifier* ini lebih *interpretable* untuk dijelaskan kepada pembaca. Penentuan *variable importance* akan menandakan hierarki dari ketujuh variabel bebas yang dijadikan parameter proses klasifikasi pada penelitian ini. Hasil penentuan *variable importance* ditampilkan pada Tabel 6 berikut.

Tabel 6. *Variable importance* proses klasifikasi

No.	Variabel	Importance
1.	Lokasi Di Kawasan Perkantoran	100,00
2.	Lokasi Di Kawasan Industri	87,80
3.	Populasi Penduduk	82,93
4.	Jumlah Penduduk Beragama Islam	46,34
5.	Jarak <i>Outlet</i> Dengan Kantor Keamanan Terdekat	37,80
6.	Jarak <i>Outlet</i> Dengan Pasar	12,80
7.	Jarak <i>Outlet</i> dengan Pusat Pemerintahan	0,00

Hasil pada Tabel 6 menandakan bahwa enam variabel dari tujuh variabel bebas yang dipakai sebagai parameter klasifikasi mempunyai pengaruh dalam menentukan dan membedakan kelas suatu data berdasarkan model prediksi yang didapatkan sebelumnya. Hasil hierarki variabel paling memengaruhi dalam penentuan lokasi strategis kantor atau *outlet* bisa diurutkan sebagai lokasi di kawasan perkantoran, lokasi di kawasan industri, populasi penduduk, jumlah penduduk beragama islam, jarak *outlet* dengan kantor keamanan terdekat, dan jarak *outlet* dengan pasar.

5. KESIMPULAN

Model klasifikasi penentuan lokasi strategis kantor atau *outlet* Bank Syariah Indonesia memakai metode *Naïve Bayes Classifier* diperoleh model nilai probabilitas *prior* dan *conditional probabilities*, sebagai dasar proses prediksi kelas kestrategisan suatu lokasi kantor atau *outlet* Bank Syariah Indonesia. Model menunjukkan tingkat akurasi sebesar 94,12% dengan nilai AUC uji performa model sebanyak 0,9808 yang menandakan proses

klasifikasi yang dilaksanakan tergolong kedalam kategori sangat baik. Model yang diperoleh mampu mengklasifikasikan data penentuan lokasi strategis sebanyak 16 data dengan benar dari total 17 data yang tersedia. Model menghasilkan hasil *variable importance* 6 rekomendasi hierarki variabel paling memengaruhi dari 7 variabel yang dipakai dalam penelitian. Hasil hierarki variabel paling memengaruhi dalam penentuan lokasi strategis kantor atau *outlet* bisa diurutkan sebagai lokasi di kawasan perkantoran, lokasi di kawasan industri, populasi penduduk, jumlah penduduk beragama islam, jarak *outlet* dengan kantor keamanan terdekat, dan jarak *outlet* dengan pasar. Hasil yang diperoleh pada penelitian menandakan metode *Naïve Bayes Classifier* bisa diterapkan untuk mengklasifikasi penentuan lokasi strategis kantor atau *outlet* Bank Syariah Indonesia.

DAFTAR PUSTAKA

- Abdurrohman, H., R. Dini., dan A. P. Muharram. 2018. Evaluasi Performa Metode *Deep Learning* untuk Klasifikasi Citra Lesi Kulit *The HAM10000*. *Prosiding Seminar Nasional Instrumentasi, Kontrol dan Otomasi (SNIKO)*. Desember 2019: 63-68.
- Bramer, M. 2007. *Principles of Data Mining*. London : Springer.
- Faturahman, M. D dan R. Z. Maliki. 2018. Evaluasi Kesesuaian Lokasi Bank BRI Di Wilayah Kantor Cabang Setiabudi Kota Bandung. *Majalah Geografi Indonesia*. 32(2) : 184 -197
- Fithri, D. L. 2016. Model *Data Mining* Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Memakai Metode *Naive Bayes*. *Jurnal Simetris*. 7(2) : 725-730
- Gorunescu, F. 2011. *Data Mining Concepts, Models and Techniques*. Berlin Heidelberg: Springer.
- Han, J, M. Kamber, dan J. Pei. 2011. *Data Mining Concept and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Imandasari, T, dkk. 2019. Algoritma *Naive Bayes* Dalam Klasifikasi Lokasi Pembangunan Sumber Air. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*. September 2019: 750-761.
- Kasmir. 2010. *Pemasaran Bank*. Jakarta : Kencana.
- Saleh, A. 2015. Implementasi Metode Klasifikasi *Naïve Bayes* Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*. 2(3): 207-217.
- Wulandari, S dan M. Iqbal. 2021. Analisis Kelayakan Lokasi Promosi Dalam Penerimaan Mahasiswa Baru (PMB) Dengan Algoritma *Naïve Bayes* dan *Decision Tree C4.5*. *KILAT*. 10(1) : 169-178.