

# Perbandingan Teknik Klasifikasi *Neural Network*, *Support Vector Machine*, dan *Naive Bayes* dalam Mendeteksi Kanker Payudara

Derisma <sup>1,\*</sup>, Fajri Febrian <sup>1</sup>

<sup>1</sup> Sistem Komputer Universitas Andalas, Limau Manis Kec. Pauh, Kota Padang, Sumatera Barat 25163; e-mail: [derisma@fti.unand.ac.id](mailto:derisma@fti.unand.ac.id), [fajrifebrian1312@gmail.com](mailto:fajrifebrian1312@gmail.com)

\* Korespondensi: e-mail: [derisma@fti.unand.ac.id](mailto:derisma@fti.unand.ac.id)

Diterima: 8 April 2020; Review: 12 April 2020; Disetujui: 19 April 2020

Cara sitasi: Derisma, Febrian Fajri. 2020. Perbandingan Teknik Klasifikasi *Neural Network*, *Support Vector Machine*, dan *Naive Bayes* dalam Mendeteksi Kanker Payudara. BINA INSANI ICT JOURNAL. Vol. 7 (1): 53-62.

---

**Abstrak:** Kanker payudara merupakan jenis kanker yang sering ditemukan oleh kebanyakan wanita. Di Indonesia Kanker payudara menempati urutan pertama pada pasien rawat inap di seluruh rumah sakit. Tujuan dari penelitian ini adalah melakukan diagnosis penyakit kanker payudara berbasis komputasi yang dapat menghasilkan bagaimana kondisi kanker seseorang berdasarkan akurasi algoritma. Penelitian ini menggunakan pemrograman orange python dan dataset Wisconsin Breast Cancer untuk pemodelan klasifikasi kanker payudara. Metode data mining yang diterapkan yaitu *Neural Network*, *Support Vector Machine*, dan *Naive Bayes*. Dalam penelitian ini didapat algoritma klasifikasi terbaik yaitu algoritma *Kernel SVM* dengan tingkat akurasi sebesar 98.9 % dan algoritma terendah yaitu *Naive Bayes* senilai 96.1 %.

**Kata kunci:** kanker payudara, neural network, support vector machine, naive bayes

**Abstract:** *Breast cancer is a type of cancer that mostly found in many women. In Indonesia, breast cancer ranks first in hospitalized patients at every hospital. This study aimed to conduct a computation-based diagnose of breast cancer disease that could produce the state of cancer of an individual based on the accuracy of algorithm. This study used python orange programming and Wisconsin Breast Cancer dataset for a modeling and application of breast cancer classification. The data mining methods that were applied in this study were Neural Network, Support Vector Machine, dan Naive Bayes. In this study, Kernel SVM's algorithm was the best classification algorithm of breast cancer disease with 98.9 % accuracy rate and Naive Bayes was the lowest with 96.1 % of accuracy rate.*

**Keywords:** *breast cancer, neural network, support vector machine, naive bayes*

## 1. Pendahuluan

Kanker merupakan penyebab nomor satu kematian. Pada tahun 2012 sekitar 8,2 juta kematian disebabkan oleh kanker. Kanker paru-paru, kolorektal, hati dan kanker payudara merupakan beberapa jenis kanker yang banyak menimbulkan kematian. Pada tahun 2018, setidaknya terdapat 58.256 kasus kanker payudara yang diderita wanita di Indonesia dari total 188.231 kasus kanker [7]. Selanjutnya disusul kanker serviks sebanyak 32.469 kasus dan kanker ovarium sebanyak 13.310 kasus.

Terdapat dua jenis kanker, yaitu : ganas (*malignant*) dan jinak (*benign*). Kanker ganas merupakan kanker yang menyerang secara agresif dan merusak sel di sekitarnya. Sedangkan kanker jinak merupakan kanker yang tidak menyebar atau menyerang sel di sekitarnya. Tidak ada gejala pasti yang dapat didiagnosis secara akurat untuk mendeteksi kanker payudara.

Tetapi, kanker payudara dapat diprediksi dengan melihat ciri-ciri yang terdapat pada payudara [6].

Tantangan penting dalam bidang data mining dan machine learning adalah membangun pengklasifikasi yang tepat dan efisien secara komputasi untuk aplikasi Medis [1]. Penelitian tentang kanker payudara menggunakan data mining dan machine learning telah dilakukan oleh Higa[6] menggunakan Decision Tree dan Artificial Neural Network, kedua algoritma tersebut berhasil mengklasifikasikan dengan benar lebih dari 92% kasus dalam 10 percobaan. Namun, algoritma Neural Network memiliki rata-rata tingkat akurasi prediktif yang lebih baik (tingkat klasifikasi yang benar adalah mencapai 95,9%). Peneliti [2][7][10] memprediksi diagnosis dini penyakit kanker payudara pada pengguna atau pasien menjadi dua kategori ganas atau jinak. Penelitian ini menggunakan algoritma Naive Bayes yang melakukan teknik klasifikasi menggunakan metode probabilitas dan statistik. Dari 140 data pelatihan yang diuji mendapatkan nilai prediksi yang benar sebanyak 137 dari semua data yaitu 47 data dari kelas Malignant dan 90 data dari kelas Benigna. Dari hasil perhitungan, dapat disimpulkan bahwa sistem memiliki akurasi 98% dengan tingkat kesalahan 2%. Derisma [5] juga menggunakan Neural Network dengan optimasi menggunakan Genetic Algorithm didapatkan akurasi sebesar 97.24 %

Penelitian serupa sebelumnya juga pernah dilakukan oleh Gharibdousti [3] pada tahun 2019 dengan judul Penelitian “Breast Cancer Diagnosis Using Feature Extraction Techniques With Supervised and Unsupervised Classification Algorithms”. Penelitian ini menggunakan teknik multivariat seperti analisis komponen utama, analisis diskriminan, dan regresi logistik untuk pengurangan fitur yang dikombinasikan dengan tool machine learning untuk mengklasifikasikan dan memprediksi jenis kanker. Metode klasifikasi yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM), regresi logistik, Neural Network (NN), Decision Trees (DT) dan Naive Bayes (NB).

Penelitian lainnya juga pernah dilakukan oleh Ibeni [7] pada tahun 2019 dengan judul “Comparative Analysis On Bayesian Classification For Breast Cancer Problem”. Makalah ini menyajikan pendekatan Bayesian penuh untuk menilai distribusi prediksi semua kelas menggunakan tiga pengklasifikasi; naïve bayes (NB), bayesian networks (BN), dan augmented naïve bayes (TAN) dengan tiga dataset; Kanker payudara, kanker payudara, dan jaringan payudara. Selanjutnya, akurasi prediksi pendekatan bayesian juga dibandingkan dengan tiga algoritma pembelajaran mesin standar dari literatur; K-Nearest Neighbors(K-NN), Support Vector Machine (SVM), dan Decision Tree (DT).

Penelitian [3][7] dijadikan rujukan untuk membuat penelitian ini. Tetapi sedikit berbeda karena yang penulis buat berfokus pada 3 algoritma saja, yaitu : *Naive Bayes*, *Neural Network*, dan *Support Vektor Machine*. Pemilihan algoritma ini berdasarkan hasil penelitian yang menunjukkan bahwa ketiga algoritma ini dapat memprediksi kanker payudara dengan akurasi yang tinggi.

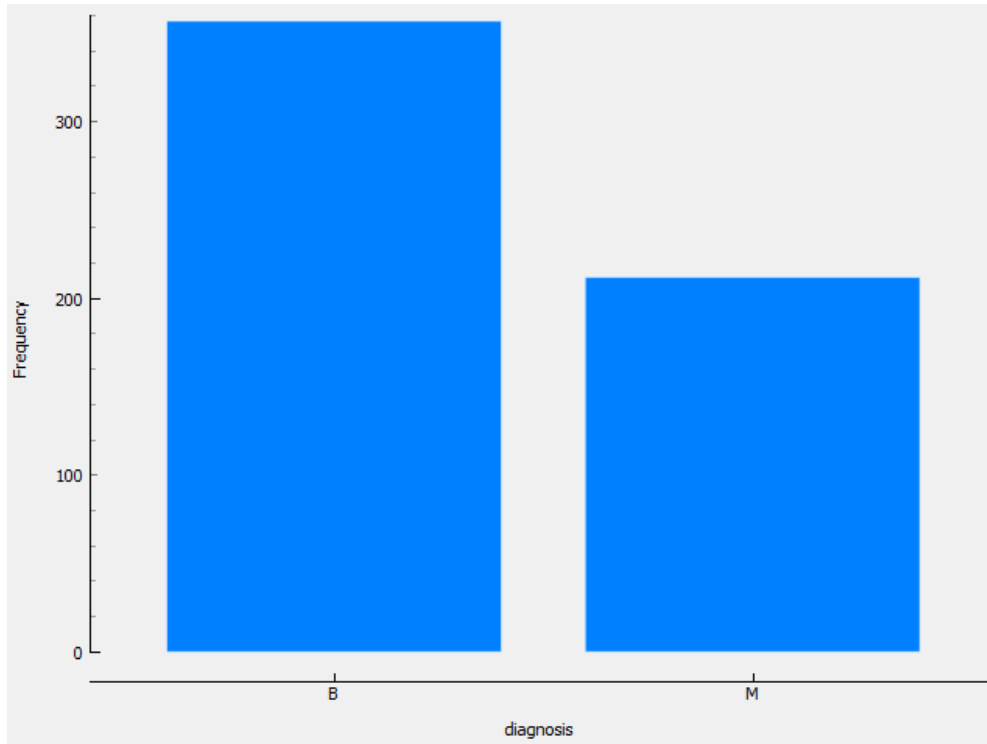
## 2. Metode Penelitian

Metodologi penelitian ini menggunakan pendekatan data mining untuk mengklasifikasikan data. Dalam hal ini, terdapat fase penting yang harus dilakukan untuk mendapatkan hasil penelitian terbaik. Sedangkan untuk model metodologi, penelitian ini menggunakan model KDD yaitu *Knowledge Discovery in Database*. Ada lima fase dalam KDD yaitu Seleksi, *Preprocessing*, Transformasi, Klasifikasi dan Evaluasi.

### **Dataset Selection**

Penelitian menggunakan dataset Wisconsin Breast Cancer yang jadi acuan kebanyakan penelitian. Dataset Wisconsin Breast Cancer ini berisi 569 instance. Dimana, terdapat 212 instance untuk jenis kanker ganas dan 357 instance untuk jenis kanker jinak. Pada dataset ini juga terdapat 30 atribut yang menunjukkan ciri-ciri bahwa seseorang menderita kanker payudara. Atribut tersebut antara lain : “*radius mean, texture mean, perimeter mean, area mean, smoothness mean, compactness mean, concavity mean, concavity mean, concave points mean, symmetry mean, fractal dimension mean, radius se, texture se, perimeter se, area se, smoothness se, compactness se, concavity se, concave points se, symmetry se, fractal dimension se, radius worst, texture worst, perimeter worst, area worst, smoothness worst, compactness worst, concavity worst, concave points worst, symmetry worst dan fractal*

*dimension worst*”. Langkah awal adalah memasukkan dataset ke dalam software orange python.

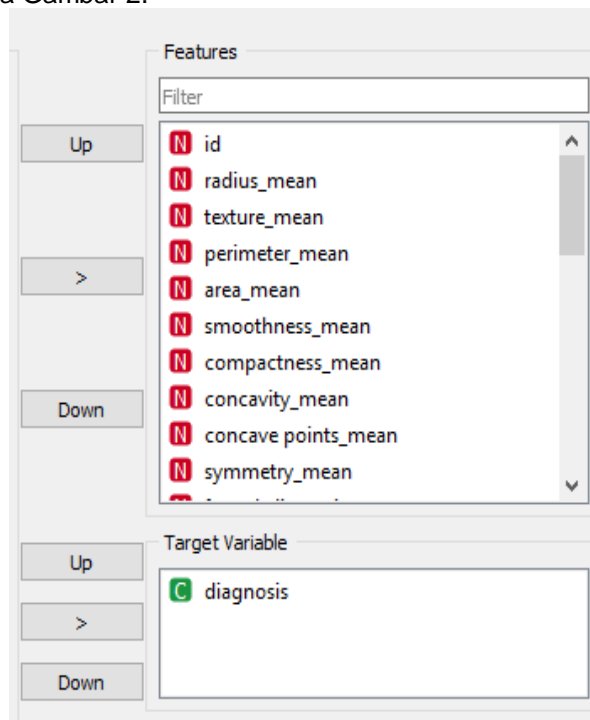


Sumber : Hasil Penelitian 2020

Gambar 1. Distribusi Data

### Preprocessing

Pada tahap ini, dilakukan pengelompokan antara kolom target dan kolom features. Kolom diagnosis dijadikan sebagai target dan kolom yang lainnya dijadikan features seperti yang dapat dilihat pada Gambar 2.

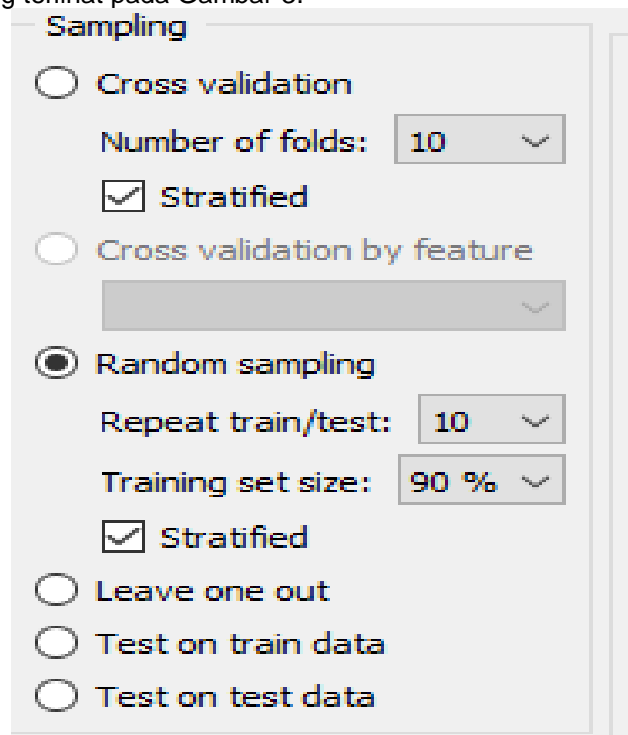


Sumber : Hasil Penelitian 2020

Gambar 2. Dataset Seleksi

### Transformation

Tahap selanjutnya adalah memisahkan antara data training dan data testing. Pada aplikasi orange sudah terdapat fitur untuk membagi data training dan data testing secara random. Seperti yang terlihat pada Gambar 3.



Sumber : Hasil Penelitian 2020

Gambar 3. Transformasi dataset

Dapat dilihat pada Gambar di atas, penulis memilih metode random sampling dengan ketentuan banyak sampling yang dijadikan kan data training adalah 90% dari jumlah data . Sehingga jumlah data training yang digunakan berjumlah 512 datadan kemudian dari data training yang ada di ambil 57 data sebagai data testing . Lalu, kedua data tersebut di-training dan di-testing sebanyak 10 kali. Sehingga total data training dan data testing juga meningkat 10 kali.

### Classification Algoritma

Dari hasil transformasi tersebut, selanjutnya dilakukan klasifikasi data menggunakan 3 algoritma klasifikasi yang memiliki akurasi paling tinggi berdasarkan penelitian terkait. Algoritma tersebut adalah *Naive Bayes*, *Neural Network* dan *Support Vektor Machine*.

#### Naive Bayes

Pengklasifikasi Naive Bayesian didasarkan pada teorema Bayes dengan asumsi independensi di antara para prediktor. Model Naive Bayesian mudah dibangun, tanpa estimasi parameter iteratif yang rumit yang membuatnya sangat berguna untuk dataset yang sangat besar. Terlepas dari kesederhanaannya, klasifikasi Naive Bayesian sering kali berhasil dengan sangat baik dan banyak digunakan karena sering mengungguli metode klasifikasi yang lebih canggih.

Teorema Bayes menyediakan cara menghitung probabilitas posterior,  $P(c | x)$ , dari  $P(c)$ ,  $P(x)$ , dan  $P(x | c)$ . Klasifikasi Naif Bayes mengasumsikan bahwa pengaruh nilai prediktor ( $x$ ) pada kelas tertentu ( $c$ ) tidak tergantung pada nilai-nilai predictor lain. Asumsi ini disebut independensi kondisional kelas. Untuk algoritma pada Naive Bayes dapat dilihat pada pada persamaan 1:

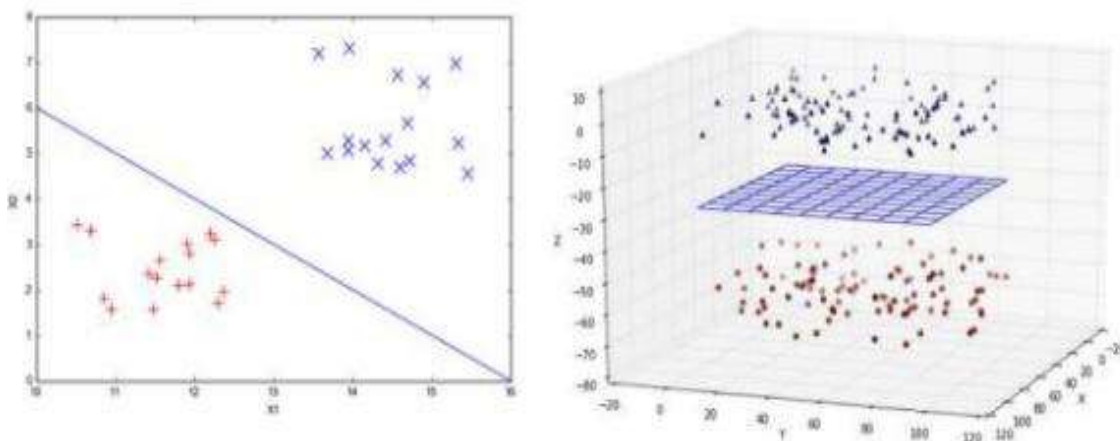
$$P(C|X) = \frac{P(x|c)P(c)}{P(x)} \dots\dots\dots (1)$$

Keterangan :

- x** : Data kelas yang belum diketahui
- c** : Hipotesis data
- P(c|x)** : Probabilitas hipotesis berdasar kondisi (posteriori probability)
- P(c)** : Probabilitas hipotesis (prior probability)
- P(x|c)** : Probabilitas berdasarkan kondisi pada hipotesis
- P(x)** : Probabilitas c

**Support Vector Machine (SVM)**

Support Vector Machine (SVM) merupakan salah satu teknik prediksi, baik dalam kasus klasifikasi maupun regresi [10]. Support Vector Machine (SVM) memiliki prinsip dasar linier classifier yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun saat sekarang ini Support Vector Machine (SVM) juga mampu memecahkan masalah non-linier dengan menambahkan konsep kernel pada workspace berdimensi tinggi. Pada ruang ini, akan dilakukan pencarian terhadap hyperplane (hyperplane) yang dapat memaksimalkan jarak antar kelas data.

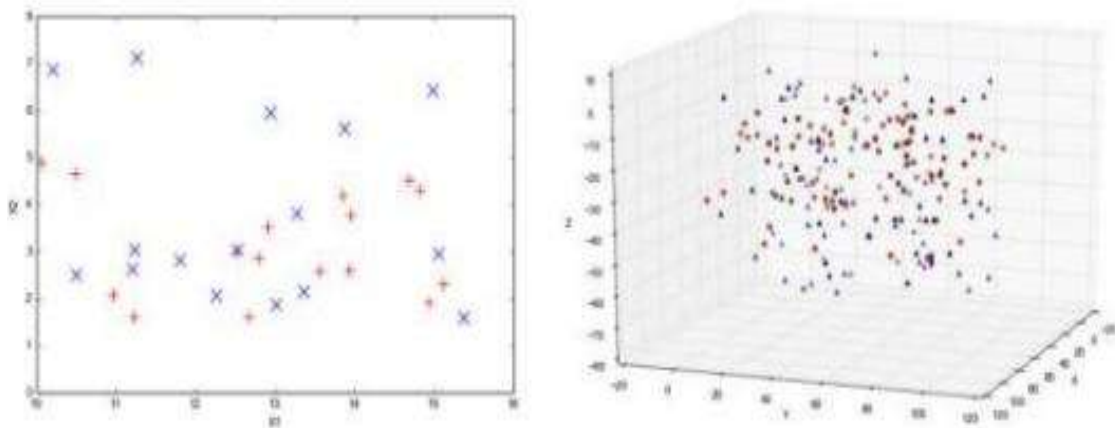


Sumber : Hasil Penelitian 2020

Gambar 4. Data Linier 2 Dimensi(kiri) dan 3 Dimensi (Kanan).

Gambar 4 adalah contoh lain dari data yang dipisahkan secara linier. Pada data linier bidang pemisah cenderung mudah untuk ditemukan pada ruang 2 dimensi ataupun 3 dimensi. Berbeda jika data tersebut adalah data yang tidak linier.

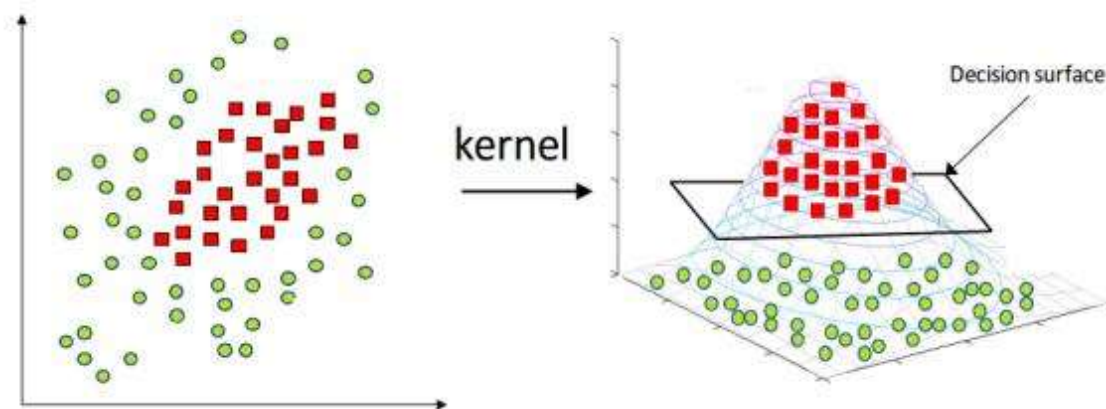
Ide dasar dari pembelajaran SVM adalah untuk menemukan pemisah antara kelas satu dengan kelas lainnya. pada gambar diatas sebelah kiri margin belum mendapatkan nilai yang maksimal sehingga hyperplane tidak optimal. Sedangkan pada gambar sebelah kanan margin telah mendapatkan nilai maksimal sehingga hyperplane optimal. gambar adalah contoh kasus dengan data yang dapat dipisahkan secara linear. Kenyataannya di dalam dunia nyata data yang didapat selalu tidak linier dan sulit untuk ditemukan bidang pemisahannya. Pada gambar adalah contoh kasus dengan data yang dapat dipisahkan secara linear. Kenyataannya di dalam dunia nyata data yang didapat selalu tidak linier dan sulit untuk ditemukan bidang pemisahannya.



Sumber : Hasil Penelitian 2020

Gambar 5. Data Tidak Linier 2 Dimensi(kiri) dan 3 Dimensi (Kanan).

Menurut Eko Prasetyo [9], SVM sebenarnya hanya dapat bekerja pada data yang dipisahkan secara linear saja. Ketika data tersebut non linier maka cara yang dapat digunakan adalah menggunakan kernel pada data. Kernel dapat diartikan sebagai suatu fungsi yang memetakan ulang data pada ruang dimensi asal kedalam ruang dimensi yang lebih Untuk sementara.



Sumber : Hasil Penelitian 2020

Gambar 6. Data Pada Ruang Dimensi Awal(kiri) dan Data Pada Ruang Dimensi Baru(Kanan).

Setelah dilakukan pemetaan pada fitur data lama kedalam fitur data baru, selanjutnya data tersebut dapat dilakukan proses pelatihan yang sama menggunakan algoritma SVM linear.

### **Neural Network (NN)**

*Neural Network (NN)* atau jaringan syaraf tiruan merupakan algoritma klasifikasi yang mampu menukar strukturnya agar mampu memecahkan suatu masalah berdasarkan hasil informasi yang berasal dari eksternal maupun internal yang melewati jaringan tersebut. Neural Network dapat memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola dari data [8].

### **Evaluation Metrics**

Hasil penelitian ini, disajikan ke dalam bentuk *Confusion Matrix*. Dimana, matrix ini menampilkan hasil prediksi dan nilai sebenarnya untuk mengetahui keakuratan dari masing-masing algoritma.

### 3. Hasil dan Pembahasan

Pertama algoritma *Naive Bayes*. Pada Tabel 1 dapat dilihat, terdapat 274 prediksi yang benar dan 11 prediksi yang salah dari total 285 data *testing*. Sehingga, tingkat akurasi algoritma ini dapat dihitung  $274/285 \times 100\% = 96\%$ .

Tabel 1. Hasil prediksi algoritma *Naive Bayes*

Naive Bayes	Prediksi		
		B	M
Nilai sebenarnya	B	174	6
	M	5	100

Sumber : Hasil Penelitian 2020

Selanjutnya algoritma Neural Network. Pada Tabel 2 dapat dilihat, terdapat 279 prediksi yang benar dan 6 prediksi yang salah dari total 285 data *testing*. Sehingga, tingkat akurasi algoritma ini dapat dihitung  $279/285 \times 100\% = 97\%$ .

Tabel 2. Hasil prediksi algoritma *Neural Network*

Neural Network	Prediksi		
		B	M
Nilai sebenarnya	B	177	3
	M	3	102

Sumber : Hasil Penelitian 2020

Terakhir algoritma Support Vektor Machine. Pada Tabel 3 dapat dilihat, terdapat 282 prediksi yang benar dan 3 prediksi yang salah dari total 285 data *testing*. Sehingga, tingkat akurasi algoritma ini dapat dihitung  $282/285 \times 100\% = 98\%$ .

Tabel 3. Hasil prediksi algoritma *Neural Network*

Support Vektor Machine	Prediksi		
		B	M
Nilai sebenarnya	B	180	0
	M	3	102

Sumber : Hasil Penelitian 2020

Hasil kinerja dari ketiga algoritma yaitu Neural Network, Naive Bayes, dan Support Vector Machine menggunakan dataset Wisconsin Breast Cancer menunjukkan hasil yang berbeda. Hasil kinerja dapat melacak kanker payudara jinak atau ganas diklasifikasikan dengan benar atau salah hasil prediksi tersebut. Seperti pada Tabel 4, yang menggambarkan hasil evaluasi dari masing-masing model.

Tabel 4. Perbandingan hasil prediksi algoritma

Model	AUC	CA	F1	Precision	Recall
SVM	0.999	0.989	0.989	0.990	0.989
NN	0.995	0.979	0.979	0.979	0.979
NB	0.988	0.961	0.961	0.962	0.961

Sumber : Hasil Penelitian 2020

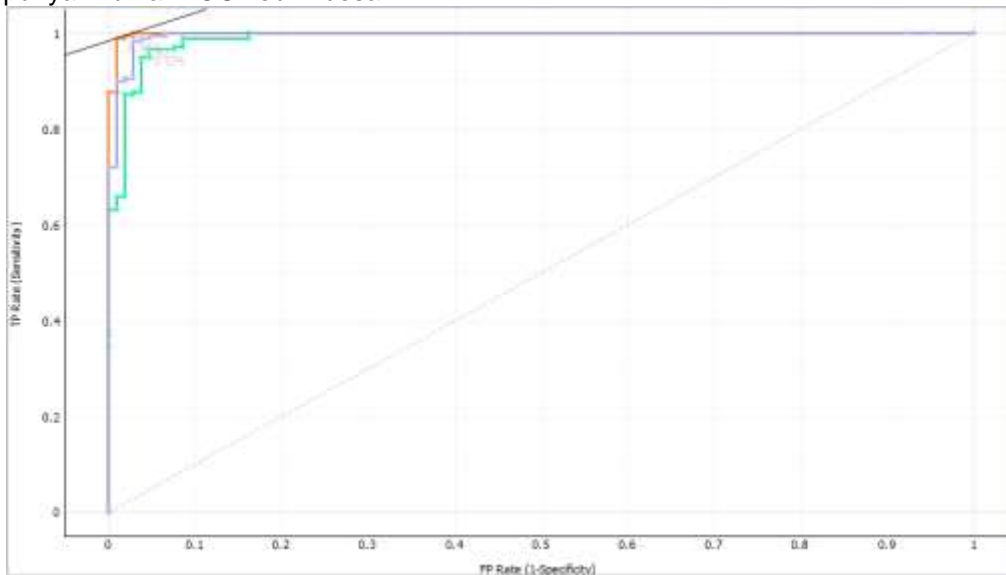
Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine (SVM) mempunyai nilai keakuratan klasifikasi (CA) yang lebih tinggi dibanding algoritma yang lainnya. Support Vector Machine (SVM) mempunyai nilai keakuratan yang paling tinggi, sebesar 98.9% sementara Neural Network (NN) mencapai 97,9% dan Naive Bayes (NB) hanya mencapai 96.1%. Perbandingan hasil dari makalah terkait menggunakan algoritma lain yaitu KNN hanya mencapai 94,9% dan algoritma DT 94,9%. Ini membuktikan algoritma Support Vector Machine (SVM) adalah klasifikasi terbaik menggunakan dataset Wisconsin Breast Cancer.

Presisi (Precision) adalah nilai prediktif untuk label kelas apakah positif atau negatif tergantung pada kelas itu dihitung. Ini pada dasarnya adalah kekuatan prediksi algoritma klasifikasi. Pada Tabel 4 menunjukkan presisi Support Vector Machine (SVM) adalah 99,0%, lebih tinggi dibanding Neural Network (NN) 97.9% dan Naive Bayes (NB) 96.2%.

Selanjutnya, Recall adalah fraksi dari instance yang relevan yang diambil. Nilai penarikan kembali yang tinggi menunjukkan bahwa algoritma klasifikasi dapat mengembalikan sebagian besar hasil yang relevan. Pada Tabel 4 menunjukkan Support Vector Machine (SVM) memiliki hasil recall yang lebih baik, sebesar 98,9% dibandingkan dengan algoritma lainnya Neural Network (NN) dengan 97,9% dan Naive Bayes (NB) dengan 96,1%.

Selanjutnya, F1 adalah ukuran akurasi yang mempertimbangkan ketepatan dan tingkat penarikan tes dan menghitung skor komposit. Skor yang baik mendukung algoritma dengan sensitivitas yang lebih tinggi dan menantang algoritma dengan spesifisitas yang lebih tinggi. Berdasarkan pertimbangan tersebut, dapat disimpulkan bahwa *Support Vector Machine* (SVM) lebih disukai daripada *Neural Network* (NN) dan *Naive Bayes* (NB). Pada Tabel 4 dapat dilihat, ukuran-F untuk *Support Vector Machine* (SVM) lebih tinggi dengan nilai 98,9% sementara *Neural Network* (NN) 97,9% dan *Naive Bayes* (NB) 96,1%.

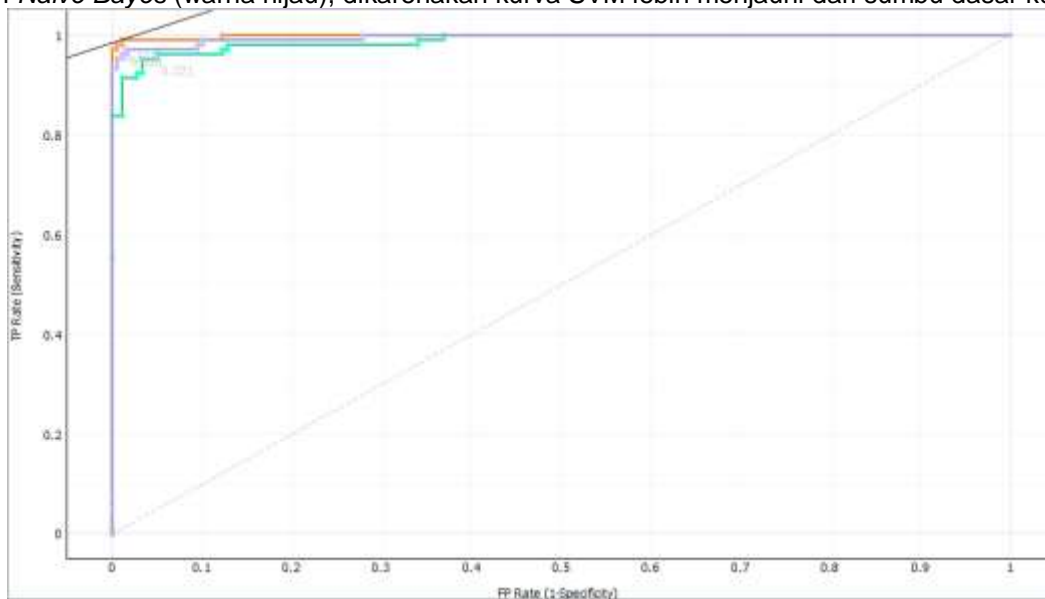
Kurva ROC merupakan teknik untuk memvisualisasi dan menguji kinerja pengklasifikasian berdasarkan performanya. Model klasifikasi yang lebih baik adalah yang mempunyai kurva ROC lebih besar.



Sumber : Hasil Penelitian 2020

Gambar 7. ROC Analysis Target class B

Dari gambar 7 class *benign* maupun class *malign*, hasil SVM (warna orange) diperoleh hasil kurva lebih baik dari *Neural Network* (warna ungu), dan Nilai *Neural Network* lebih baik dari *Naive Bayes* (warna hijau), dikarenakan kurva SVM lebih menjauhi dari sumbu dasar kurva.



Sumber : Hasil Penelitian 2020

Gambar 8. ROC Analysis Target class M



Dari gambar 8 baik class benign maupun class malign, hasil SVM (warna orange) diperoleh hasil kurva lebih baik dari *Neural Network* (warna ungu), dan Nilai *Neural Network* lebih baik dari *Naive Bayes* (warna hijau), dikarenakan kurva SVM lebih menjauhi dari sumbu dasar kurva.

#### 4. Kesimpulan

Klasifikasi merupakan teknik penting dari data mining yang dapat diaplikasikan di berbagai bidang. Makalah ini menyajikan percobaan perbandingan teknik klasifikasi *Neural Network*, *Support Vector Machine*, dan *Naive Bayes* dalam mendeteksi kanker payudara menggunakan dataset *Wisconsin Breast Cancer*. Dari ketiga teknik klasifikasi tersebut dicari teknik terbaik. Dalam mencari yang terbaik, langkah penelitian yang dilakukan menggunakan model KDD yaitu *Knowledge Discovery in Database*. Ada lima fase dalam KDD yaitu Seleksi, *Preprocessing*, Transformasi, Klasifikasi dan Evaluasi. Setelah data dilakukan klasifikasi maka akan didapatkan berbagai hasil evaluasi yaitu ROC analysis, confusion matrix dan calibration plot. Pada confusion matrix didapatkan data AUC, CA, F1, Precision dan Recall. AUC adalah luas daerah dibawah kurva ROC. Didapatkan hasil nilai AUC pada , Model *Support Vector Machine* = 0.999 , Model *Neural Network* = 0.995, Model *Naive Bayes* = 0.988. CA merupakan akurasi algoritma dalam diprediksi untuk sampel. Model *Support Vector Machine* = 0.989 , Model *Neural Network* = 0.979, Model *Naive Bayes* = 0.961. F1 merupakan ukuran akurasi yang mempertimbangkan ketepatan dan tingkat penarikan tes dan menghitung skor komposit. Skor yang baik mendukung algoritma dengan sensitivitas yang lebih tinggi dan menantang algoritma dengan spesifisitas yang lebih tinggi. Model *Support Vector Machine* = 0.989, Model *Neural Network* = 0.979, Model *Naive Bayes* = 0.961. Precision adalah nilai prediktif untuk label kelas apakah positif atau negatif tergantung pada kelas itu dihitung. Ini pada dasarnya adalah kekuatan prediksi algoritma klasifikasi. Model *Support Vector Machine* = 0.990 , Model *Neural Network* = 0.979, Model *Naive Bayes* = 0.962. Recall adalah tingkat keberhasilan sistem dalam menemukan sebuah informasi. Model *Support Vector Machine* = 0.989 , Model *Neural Network* = 0.979, Model *Naive Bayes* = 0.961. Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) mempunyai nilai keakuratan klasifikasi (CA) yang lebih tinggi dibanding algoritma yang lainnya. *Support Vector Machine* (SVM) mempunyai nilai keakuratan yang paling tinggi, sebesar 98.9% sementara *Neural Network* (NN) mencapai 97,9% dan *Naive Bayes* (NB) hanya mencapai 96.1%. Jadi SVM adalah model terbaik dalam memprediksi kanker payudara dibanding *Neural Network* dan *Naive Bayes*. Pada penelitian selanjutnya dapat menggunakan model selain SVM, *Neural Network* dan *Naive Bayes* sehingga didapatkan model yang lebih baik lagi untuk memprediksi kanker payudara.

#### References

- [1] Derisma, Silvana, M., & Imelda. (2018). Optimization of Neural Network with Genetic Algorithm for Breast Cancer Classification. 2018 International Conference on Information Technology Systems and Innovation, ICITSI 2018 - Proceedings, 398–403. <https://doi.org/10.1109/ICITSI.2018.8696014>
- [2] Frisetyo, T., & Kuswara, H. (2019). Diagnosa Penderita Penyakit Kanker Payudara Menggunakan Metode Naive Bayes. *Information Management for Educators And Professionals*, 4(1), 51–62.
- [3] Gharibdousti, M. S., Haider, S. M., Ouedraogo, D., & Lu, S. (2019). Breast cancer diagnosis using feature extraction techniques with supervised and unsupervised classification algorithms. *Applied Medical Informatics Original Research*, 41(1), 40–52.
- [4] Higa, A. (2018). Diagnosis of Breast Cancer using Decision Tree and Artificial Neural Network Algorithms. *International Journal of Computer Applications Technology and Research*, 7(1), 23–27. <https://doi.org/10.7753/ijcatr0701.1004>
- [5] International Agency for Research on Cancer, “Indonesia - Global Cancer Observatory” [Online]. Available: <https://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-fact-sheets.pdf>. [Accessed: 17-Maret-2020].
- [6] Ibeni, W. N. L. W. H., Salikon, M. Z. M., Mustapha, A., Daud, S. A., & Salleh, M. N. M. (2019). Comparative analysis on bayesian classification for breast cancer problem. *Bulletin of Electrical Engineering and Informatics*, 8(4), 1303–1311. <https://doi.org/10.11591/eei.v8i4.1628>

- [7] Liao. Recent Advances in Data Mining of Enterprise Data: Algorithms and Application . Singapore: World Scientific Publishing. 2007.
- [8] Lorena, S., Ginting, B. R., & Permana, A. A. (2016). Penerapan Data Mining Untuk Klasifikasi Kelayakan Nasabah Dalam Pengajuan Kredit Menggunakan. 1–10.
- [9] Prasetyo, Eko. (2012) “Data Mining Konsep dan Aplikasi menggunakan MATLAB”, Andi, Yogyakarta.
- [10] Alamsyah dkk (2017). Implementation Of Naive Bayes Method In Classification Of Breast Cancer Disease. 2(1), 191–194..