

LAPORAN

PREDIKSI KANKER PAYUDARA MENGGUNAKAN METODE MACHINE LEARNING DENGAN ALGORITMA *BINARY CLASSIFICATION (NEURAL NETWORK)*

Dosen pengampu: Dr. Amalia S.T., M.T.



Oleh:

Yessica Angelita Situmorang	211401025
Johana Paskalina Sihotang	211401035
Rodiatul Husna Br Sitepu	211401038
A. Nurcahaya Tampubolon	211401071

PROGRAM STUDI ILMU KOMPUTER

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS SUMATERA UTARA

2023

ABSTRAK

Kanker payudara merupakan salah satu penyakit yang memiliki dampak signifikan terhadap kesehatan perempuan di seluruh dunia. Deteksi dini kanker payudara memainkan peran krusial dalam meningkatkan tingkat kesembuhan dan mengurangi angka kematian. Dalam upaya meningkatkan efektivitas deteksi dini, penelitian ini bertujuan untuk mengembangkan model prediksi kanker payudara menggunakan teknik *neural network*. Pendekatan ini memanfaatkan *Binary Classification* dalam proses membedakan antara kasus kanker payudara dan non-kanker. Pada penelitian ini, kami mengumpulkan dataset yang berkaitan dengan kanker payudara. Program breast cancer detection yang telah dibuat menggunakan *binary classification (neural networks)* mampu mencapai tingkat akurasi antara 94% hingga 96% dengan tingkat kehilangan (*model loss*) sebesar 9% hingga 18%. Hasil eksperimen menunjukkan bahwa model yang dihasilkan mampu memberikan prediksi yang akurat dan dapat diandalkan dalam klasifikasi kanker payudara. Dengan adanya model ini, diharapkan dapat meningkatkan efisiensi dan akurasi proses diagnosis kanker payudara, serta memberikan kontribusi positif dalam upaya penanggulangan penyakit yang memiliki dampak besar terhadap kesehatan masyarakat.

Kata Kunci: Kanker Payudara, *Binary Classification*, Klasifikasi, *Machine Learning*

ABSTRACT

Breast cancer is a disease that has a significant impact on women's health throughout the world. Early detection of breast cancer plays a crucial role in increasing the cure rate and reducing the death rate. In an effort to increase the effectiveness of early detection, this research aims to develop a breast cancer prediction model using *neural network* techniques. This approach utilizes *Binary Classification* in the process of distinguishing between breast cancer

and non-cancer cases. In this study, we collected a dataset related to breast cancer. The breast cancer detection program that has been created using *binary classification (neural networks)* is able to achieve an accuracy level of between 94% to 96% with a loss rate (model loss) of 9% to 18%. Experimental results show that the resulting model is able to provide accurate and reliable predictions in breast cancer classification. With this model, it is hoped that it can increase the efficiency and accuracy of the breast cancer diagnosis process, as well as make a positive contribution to efforts to control a disease that has a major impact on public health.

Keywords: Breast Cancer, *Binary Classification*, Classification, *Machine Learning*

DAFTAR ISI

LAPORAN.....	1
ABSTRAK.....	2
DAFTAR ISI.....	4
BAB I.....	5
PENDAHULUAN.....	5
1.1 Latar Belakang.....	5
1.2 Alasan Pemilihan Topik.....	6
1.3 Rumusan Masalah.....	7
1.4 Rumusan Masalah.....	8
BAB II.....	9
TINJAUAN PUSTAKA.....	9
BAB III.....	13
METODOLOGI PENELITIAN.....	13
4.1 Hasil.....	30
4.2 Pembahasan.....	30
BAB V.....	33
KESIMPULAN.....	33
DAFTAR PUSTAKA.....	34

DAFTAR TABLE

Tabel 3.1 Variabel Independen (X)	21
Tabel 3.2 Variabel Dependen (Y)	22

DAFTAR GAMBAR

Gambar 3.1 Metode Penelitian	17
Gambar 3.2.1 Cancer Breast Prediction Dataset	18
Gambar 3.2.2 Lanjutan Cancer Breast Prediction Dataset	19
Gambar 3.3 Pemisahan Kolom yang Diperlukan	27
Gambar 3.4 Hasil Proses Data Missing	28
Gambar 3.5 Hasil Pengecekan Data Duplikat	29
Gambar 3.6 Hasil Pengecekan Outliers menggunakan boxplot	29
Gambar 3.7 Hasil Outliers menggunakan IQR	30
Gambar 3.8 Hasil Penghapusan Outliers	31
Gambar 3.9 Replace variabel diagnosis	32
Gambar 3.10 Info Data setelah di Replace	32
Gambar 3.11 Pembuatan Variabel X dan Y	33
Gambar 3.12 Pelatihan Model	33
Gambar 3.13 Hasil Evaluasi Model	34
Gambar 4.1 Grafik Model Akurasi	37
Gambar 4.2 Grafik Model Loss	37

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kanker payudara merupakan jenis tumor ganas yang hingga kini masih menjadi pembunuh nomor satu dimana proses penyebaran atau metastasis memakan waktu yang lama, sehingga apabila diketahui dan dilakukan pencegahan sejak dini maka dapat menekan angka penderita kanker payudara (Sarina et al., 2020). Kasus kanker payudara pada wanita di negara maju terjadi lebih sedikit daripada negara berkembang yakni sebanyak 794.000 kasus, sedangkan pada negara berkembang kasus kanker payudara sebanyak 833.000 kasus. Kanker payudara sendiri umumnya menyerang perempuan dan merupakan salah satu kanker terbanyak yang terjadi di Indonesia (KEMENKES RI, 2015). Jumlah penderita kanker payudara menunjukkan bahwa terdapat peningkatan setiap tahunnya, dimana diperkirakan akan terus meningkat hingga sebesar 4 (empat) kali lipat jumlahnya pada tahun 2020 (American Cancer Society, 2016.).

Untuk mengurangi resiko peningkatan jumlah penderita kanker perlu dilakukan deteksi dini, beberapa metoda dapat digunakan untuk membantu proses pendeteksian diawal seperti cancer screening atau dengan metode komputasi seperti machine learning. Beberapa metode machine learning yang banyak dipilih untuk menyelesaikan kasus prediksi kanker payudara ini terutama algoritma klasifikasi diantaranya *Support Vector Machine (SVM)*, *Artificial Neural Network (ANN)*, *Neural Network*, dan *naive bayes*. Diantara berbagai teknik klasifikasi yang disarankan selama ini, *Neural Network* telah menjadi salah satu metode diagnosis yang paling populer dan konsisten dalam memecahkan permasalahan klasifikasi (Medicine and 2002).

Neural network (jaringan syaraf tiruan) adalah sistem yang belajar untuk menerima informasi dengan cara yang serupa dengan jaringan saraf manusia. Implementasi *neural network* menggunakan program komputer memungkinkan untuk melakukan berbagai proses perhitungan. Salah satu aplikasi penting dari *neural network* adalah dalam pengenalan pola, yang memungkinkan peniruan cara kerja manusia. *Neural network* adalah suatu sistem pengolahan informasi yang terdiri dari banyak unit sederhana yang bekerja secara paralel. Sistem ini mampu menyimpan pengetahuan yang diperoleh melalui eksperimen dan dapat digunakan untuk berbagai tujuan. *Neural Network* adalah salah satu metode kontrol yang menawarkan beberapa kelebihan yang dapat digunakan untuk membangun suatu model nonlinear antar input dan output (Zhang et al.) sehingga memiliki sifat dan karakteristik dengan proses yang ditinjau.

Binary classification merupakan suatu tugas dalam *machine learning* yang bertujuan untuk memisahkan data ke dalam dua kelas atau kategori yang berbeda. Dalam konteks ini, terdapat dua kelompok utama yang disebut sebagai kelas positif dan kelas negatif. Tujuan dari *binary classification* adalah untuk mengembangkan model yang dapat mempelajari pola atau fitur dari data pelatihan dan kemudian memprediksi kelas yang sesuai untuk data baru yang belum pernah dilihat sebelumnya.

Dari uraian diatas maka tujuan penelitian ini adalah untuk memprediksi atau mendiagnosis dini penyakit kanker payudara dengan menggunakan algoritma *binary classification neural network* sebagai proses training, learning dan testingnya. Dengan memprediksi kanker payudara diharapkan dapat membantu masyarakat dalam menentukan strategi perawatan serta pencegahan yang tepat.

1.2 Alasan Pemilihan Topik

Terdapat beberapa hal yang menjadi alasan mengapa penulis memilih prediksi kanker payudara menggunakan *Binary Classification* sebagai topik pada proyek akhir mata kuliah kecerdasan buatan dan sistem cerdas, diantaranya:

a. Implementasi Teknologi AI

Kecerdasan buatan atau *Artificial Intelligence* (AI) merupakan salah satu area penelitian yang terus berkembang pesat. Dengan menggunakan *Binary Classification* untuk memprediksi kanker payudara memungkinkan penulis dapat mengaplikasikan pengetahuan yang telah diperoleh selama satu semester untuk mendemonstrasikan teknik-teknik kecerdasan buatan dan sistem cerdas secara nyata.

b. Dampak Sosial dan Kemanusiaan

Kanker payudara memiliki dampak yang signifikan pada kualitas hidup individu dan masyarakat secara keseluruhan. Dengan membangun sebuah model prediksi kanker payudara menggunakan *Binary Classification* diharapkan dapat membantu dalam upaya pencegahan, pengelolaan, dan perawatan penyakit ini. Dengan adanya model prediksi kanker payudara ini penulis berharap jika nantinya diberi kesempatan untuk mengembangkan model prediksi kanker payudara menjadi lebih kompleks dapat memberikan banyak dampak positif pada kesehatan dan kesejahteraan masyarakat.

1.3 Rumusan Masalah

Adapun rumusan masalah dalam pembuatan prediksi kanker payudara menggunakan *Binary Classification(Neural Network)* adalah sebagai berikut:

- a. Bagaimana membangun model *Binary Classification(Neural Network)* untuk memprediksi risiko kanker payudara dengan akurasi tinggi?
- b. Bagaimana memilih dan menyesuaikan arsitektur *Binary Classification(Neural Network)* yang paling sesuai untuk prediksi kanker payudara?
- c. Bagaimana menafsirkan hasil prediksi yang dihasilkan oleh model *Binary Classification* untuk memberikan pemahaman yang lebih baik tentang faktor risiko kanker payudara?

1.4 Rumusan Masalah

Dalam pembuatan prediksi kanker payudara menggunakan *Binary Classification(Neural Network)* terdapat beberapa batasan masalah seperti berikut:

- a. Pembuatan model prediksi kanker payudara untuk saat ini hanya berfokus pada prediksi untuk penyakit kanker payudara umum saja yang akan membantu penulis dalam menyempitkan ruang lingkup penelitian dan analisis.
- b. Pembuatan model prediksi kanker payudara ini menggunakan dataset yang telah tersedia dan relevan untuk prediksi kanker payudara. Namun, penggunaan dataset yang lebih luas dan representatif dapat digunakan untuk penelitian lanjutan yang lebih kompleks.

- c. Pembuatan model prediksi kanker payudara ini menggunakan *Binary Classification(Neural Network)* sebagai metode utama untuk memprediksi kanker payudara.

BAB II

TINJAUAN PUSTAKA

Penelitian berjudul "Perbandingan Teknik Klasifikasi *Neural Network*, *Support Vector Machine*, dan *Naïve Bayes* dalam Mendeteksi Kanker Payudara," dilakukan oleh Derisma *et al.* (2020). Penelitian ini menggunakan pemrograman Orange Python dan set data Wisconsin Breast Cancer untuk pemodelan klasifikasi kanker payudara. Metode *data mining* yang diterapkan mencakup *Neural Network*, *Support Vector Machine*, dan *Naïve Bayes*. Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki nilai keakuratan klasifikasi (CA) yang lebih tinggi dibandingkan dengan algoritma lainnya. SVM mencapai nilai keakuratan tertinggi, yaitu sebesar 98.9%, sementara *Neural Network* (NN) mencapai 97.9%, dan *Naïve Bayes* (NB) hanya mencapai 96.1%. Dengan demikian, dapat disimpulkan bahwa SVM merupakan model terbaik dalam memprediksi kanker payudara dibandingkan dengan *Neural Network* dan *Naïve Bayes*.

Sebuah penelitian yang berjudul "Analisis Algoritma Klasifikasi *Neural Network* untuk Diagnosis Penyakit Kanker Payudara" dilakukan oleh Fitra *et al.* (2019). Penelitian ini bertujuan untuk mengklasifikasikan diagnosa penyakit kanker payudara menggunakan alat RapidMiner 9.0 pada dataset *breast cancer Coimbra*, yang diuji dengan menggunakan metode *Neural Network*. Hasil penelitian menunjukkan nilai akurasi yang baik, yaitu sebesar 71.83%.

Dalam penelitian yang dilakukan oleh Hardian *et al.* (2019) berjudul "Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma *Naive Bayes*," disimpulkan bahwa algoritma *Naive Bayes* mampu mengklasifikasi kanker payudara dengan akurasi yang baik, mencapai 96,9%. Meskipun demikian, peneliti menyarankan untuk melakukan pengembangan lebih lanjut dengan

mengintegrasikan *Naive Bayes* dengan metode seleksi fitur guna meningkatkan performa klasifikasi.

Penelitian berjudul "*Data Mining untuk Klasifikasi Penderita Kanker Payudara Berdasarkan Data dari University Medical Center Menggunakan Algoritma Naïve Bayes*," yang dilakukan oleh Ibnu *et al.* (2020), memanfaatkan alat RapidMiner 9.3 untuk menghitung nilai akurasi dan menerapkan algoritma *Naïve Bayes*. Hasil penelitian menunjukkan nilai akurasi yang cukup baik, yakni sebesar 71,43%.

Pada penelitian yang berjudul “Deteksi Dini Kanker Payudara Menggunakan Metode *Convolution Neural Network* (CNN)” yang dilakukan oleh Merinda *et al.* (2022), menerapkan *Deep Learning* yang saat ini memberikan hasil paling signifikan dalam pengenalan citra, yaitu *Convolutional Neural Network* (CNN). Penelitian ini menghasilkan nilai akurasi yang tinggi, mencapai 80%.

Dalam penelitian yang dilakukan oleh Afrizal *et al.* (2023), yang berjudul “Implementasi Metode *Convolutional Neural Network* untuk Klasifikasi Breast Cancer pada Citra Histopatologi” menerapkan algoritma *Convolutional Neural Network* (CNN) pada set data *The Breast Cancer Histopathological Image Classification* (BreakHis). Hasil dari penelitian menjelaskan bahwa CNN dapat melakukan klasifikasi kasus *breast cancer* dengan capaian nilai akurasi sebesar 85%.

Dalam penelitian yang dilakukan oleh Reynold *et al.* (2020), yang berjudul “Klasifikasi Kanker Payudara Menggunakan *Residual Neural Network*”, melakukan pengujian pada set data BreakHis menunjukkan bahwa model yang diusulkan dapat menghasilkan akurasi rata-rata

hingga 99,3% dan bersaing dengan *state-of-the-art* saat ini pada pengklasifikasian multi-class dengan tingkat akurasi rata-rata sebesar 94,6%.

Penelitian yang berjudul “Penerapan Algoritma *Decision Tree* Untuk Memprediksi Kanker Payudara menggunakan *Data Mining* dan *Machine Learning*”, dilakukan oleh Islmai *et al.* (2022). Peneliti menggunakan algoritma *Decision Tree* dengan model ID3, CART dan C4.5. Penerapan *pruning* dan *pre-pruning* dilakukan untuk mengevaluasi akurasi model yang dibangun. Hasilnya menunjukkan akurasi sebesar 98,25% pada algoritma ID3 dan CART, baik dengan maupun tanpa penggunaan *pruning* dan *pre-pruning*. Meskipun demikian, penelitian ini belum mencapai target peneliti, yakni mencapai nilai akurasi sebesar 99%. Peneliti menyatakan bahwa kendala mungkin disebabkan oleh beberapa parameter yang belum mendapatkan nilai yang sesuai.

Penelitian berjudul "Penerapan Metode *Neural Network* Berbasis Algoritma Genetika Untuk Prediksi Penyakit Kanker Payudara," dilakukan oleh Ela Nurelasari (2018). Peneliti menggunakan Algoritma Genetika sebagai metode optimasi untuk memperoleh nilai parameter yang lebih baik. Hal ini terbukti melalui peningkatan nilai akurasi pada model algoritma *Neural Network* sebesar 95,42%, sedangkan pada algoritma *Neural Network* berbasis Algoritma Genetika, akurasi meningkat menjadi 96,85%, dengan selisih akurasi sebesar 1,43%. Oleh karena itu, dapat disimpulkan bahwa penerapan teknik optimasi Algoritma Genetika mampu meningkatkan nilai akurasi pada algoritma *Neural Network*.

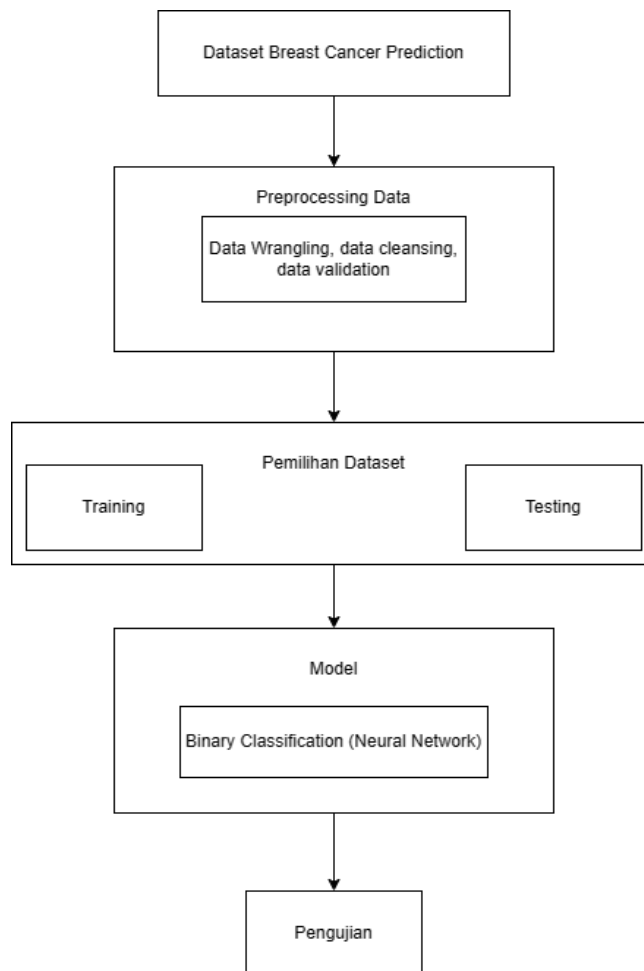
Penelitian yang berjudul "Optimalisasi Klasifikasi Kanker Payudara Menggunakan *Forward Selection* pada *Naïve Bayes*," dilakukan oleh Lastri *et al.* (2020). Peneliti menggunakan metode *Forward Selection* untuk meningkatkan efisiensi penggunaan algoritma *Naïve Bayes*. Hal

ini dilakukan karena algoritma *Naïve Bayes* memiliki kelemahan, yaitu prediksi hasil probabilitas yang tidak optimal dan kurangnya pemilihan fitur yang relevan terhadap klasifikasi sehingga akurasi rendah. Hasil penelitian menunjukkan peningkatan nilai akurasi setelah menerapkan metode *Forward Selection*, yang digunakan untuk memilih fitur atau mereduksi dimensi. Nilai akurasi menggunakan algoritma *Naïve Bayes* meningkat menjadi 93,57%. Sedangkan, nilai akurasi menggunakan algoritma *Naïve Bayes* dengan metode *Forward Selection* mencapai nilai akurasi sebesar 96,49% dengan selisih 2,92%.

BAB III

METODOLOGI PENELITIAN

Tujuan dari penelitian ini adalah membuat sebuah sistem yang dapat mendeteksi apakah seseorang berisiko mengalami kanker payudara atau tidak berdasarkan data yang diberikan. Dengan menggunakan model *Binary Classification (Neural Network)*, penelitian ini diharapkan dapat membantu dalam memprediksi kemungkinan seseorang mengidap kanker payudara berdasarkan faktor risiko tertentu.



Gambar 3.1 Metode Penelitian

Penelitian ini melibatkan beberapa tahap dalam pelaksanaannya, yaitu:

1. Menentukan Dataset

Dalam penelitian ini, penulis menggunakan dataset yang diperoleh dari Kaggle berupa data *cancer breast prediction* dataset dengan alamat *website*: [dataset breast cancer.csv](#). Adapun dataset yang digunakan terdiri dari 11 variabel, yaitu *diagnosis*, *radius mean*, *texture mean*, *perimeter mean*, *area mean*, *smoothness mean*, *compactness mean*, *concavity mean*, *concave points mean*, *symetry mean*, dan *fractal dimension mean*. Pada diagnosis 'M' digunakan untuk menunjukkan status malignant yang artinya kanker tersebut bersifat ganas sedangkan 'B' menunjukkan status benign yang artinya kanker tersebut bersifat jinak. Dataset yang digunakan berjumlah sebanyak 569 baris dan 11 kolom. Disini 'M' digunakan untuk menunjukkan status malignant yang artinya kanker tersebut bersifat ganas sedangkan 'B' menunjukkan status benign yang artinya kanker tersebut bersifat jinak.

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean
0	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419
1	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812
2	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069
3	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597
4	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809
...
564	M	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726
565	M	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752
566	M	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590
567	M	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397
568	B	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587

569 rows x 11 columns

Gambar 3.2.1 *Cancer Breast Prediction Dataset*

symmetry_mean	fractal_dimension_mean
0.2419	0.07871
0.1812	0.05667
0.2069	0.05999
0.2597	0.09744
0.1809	0.05883
...	...
0.1726	0.05623
0.1752	0.05533
0.1590	0.05648
0.2397	0.07016
0.1587	0.05884

Gambar 3.2.2 Lanjutan *Cancer Breast Prediction* Dataset

Dari gambar di atas dapat dilihat bahwa *cancer breast prediction* dataset memiliki 11 variabel yang terdiri dari 10 variabel independen dan 1 variabel dependen. Di bawah ini adalah daftar dari masing-masing variabel independen (X) dan variabel dependen (Y).

No.	Kolom Dataset	Tipe Data	Deskripsi
1	<i>Radius mean</i>	float64	Variabel ini merujuk rata-rata jarak dari pusat ke titik-titik di tepi tumor. Nilai yang lebih tinggi dapat

			menunjukkan tumor yang lebih besar atau tidak teratur.
2	<i>Texture mean</i>	float64	Variabel ini merujuk rata-rata nilai piksel dalam gambar grayscale dari tumor. Nilai yang lebih tinggi dapat menunjukkan variasi yang lebih besar dalam intensitas piksel, yang mungkin menunjukkan heterogenitas dalam struktur sel.
3	<i>Perimeter mean</i>	float64	Variabel ini merujuk rata-rata ukuran keliling tumor. Nilai yang lebih tinggi dapat menunjukkan tumor yang lebih besar.
4	<i>Area mean</i>	float64	Variabel ini merujuk rata-rata ukuran area tumor. Nilai yang lebih tinggi dapat menunjukkan tumor yang lebih besar.

5	<i>Smoothness mean</i>	float64	Variabel ini merujuk rata-rata variasi lokal dalam panjang radius. Nilai yang lebih tinggi dapat menunjukkan variasi yang lebih besar dalam panjang radius, yang mungkin menunjukkan tumor yang lebih tidak teratur.
6	<i>Compactness mean</i>	float64	Variabel ini merujuk rata-rata ukuran untuk $\text{perimeter}^2 / \text{area} - 1.0$. Nilai yang lebih tinggi dapat menunjukkan tumor yang lebih padat.
7	<i>Concavity mean</i>	float64	Variabel ini merujuk rata-rata ukuran indentasi atau cekungan dalam kontur tumor. Nilai yang lebih tinggi dapat menunjukkan adanya cekungan yang lebih banyak atau lebih dalam, yang mungkin menunjukkan invasi sel kanker ke

			jaringan sekitarnya.
8	<i>Concave points mean</i>	float64	Variabel ini merujuk rata-rata ukuran untuk jumlah cekungan kontur tumor. Nilai yang lebih tinggi dapat menunjukkan adanya cekungan yang lebih banyak, yang mungkin menunjukkan invasi sel kanker ke jaringan sekitarnya.
9	<i>Symetry mean</i>	float64	Variabel ini merujuk rata-rata ukuran simetri tumor. Nilai yang lebih tinggi dapat menunjukkan simetri yang lebih rendah, yang mungkin menunjukkan pertumbuhan sel kanker yang tidak teratur.
10	<i>Fractal dimension mean</i>	float64	Variabel ini merujuk rata-rata ukuran “coastline approximation” - 1. Nilai yang lebih tinggi dapat menunjukkan struktur yang lebih

			kompleks atau tidak teratur, yang mungkin menunjukkan pertumbuhan sel kanker yang tidak teratur.
--	--	--	--

Tabel 3.1 Variabel Independen (X)

No.	Kolom Dataset	Tipe Data	Deskripsi
1	<i>Diagnosis</i>	object	<p>Variabel ini menunjukkan apakah seseorang mengidap penyakit kanker yang bersifat ganas atau jinak.</p> <p>Variabel ini memiliki nilai object, yaitu 'M' digunakan untuk menunjukkan status malignant yang artinya kanker tersebut bersifat ganas sedangkan 'B' menunjukkan status benign yang artinya kanker tersebut bersifat jinak</p>

			Dikarenakan dalam <i>binary classification</i> membutuhkan <i>input</i> berupa nilai numerik, maka dalam penelitian ini variabel di <i>replace</i> untuk mengganti M menjadi 1 dan B menjadi 0.
--	--	--	---

Tabel 3.2 Variabel Dependen (Y)

2. Studi Literatur

Setelah memperoleh dataset, peneliti melakukan pengumpulan informasi melalui studi literatur yang relevan untuk menjadi tolak ukur penelitian yang dilakukan. Pencarian literatur dilakukan dengan tujuan untuk mendapatkan pemahaman yang lebih baik tentang prediksi kanker payudara menggunakan *Binary Classification Neural Network* (BCNN), serta melihat pendekatan, metode, dan temuan yang telah dilakukan dalam penelitian sebelumnya terkait kanker payudara.

Studi literatur ini membantu peneliti dalam mengidentifikasi kerangka kerja yang telah digunakan oleh penelitian sebelumnya dalam memprediksi kanker payudara dengan BCNN. Informasi dari literatur tersebut memberikan wawasan tentang variabel yang penting, metode preprocessing yang efektif, serta teknik pengoptimalan dan pengukuran kinerja model BCNN yang dapat diterapkan dalam penelitian ini.

Dengan merujuk pada literatur sebelumnya yang berkaitan dengan prediksi kanker payudara menggunakan BCNN, peneliti dapat memperoleh pandangan yang lebih dalam tentang langkah-langkah yang perlu diambil, tantangan yang mungkin dihadapi, dan bagaimana menginterpretasi hasil dari model BCNN untuk prediksi kanker payudara. Hal ini menjadi landasan penting dalam merencanakan dan melaksanakan penelitian prediksi kanker payudara menggunakan BCNN secara lebih efisien dan efektif.

3. Menentukan Metode Klasifikasi

Penelitian ini memilih BCNN untuk prediksi kanker payudara karena algoritma ini memiliki kemampuan dalam mengeksplorasi pola-pola kompleks dan hubungan non-linear di antara berbagai variabel yang digunakan dalam analisis data kanker payudara. BCNN dapat mempelajari pola-pola ini dan melakukan klasifikasi dengan tingkat akurasi yang dapat diandalkan.

Selain itu, BCNN juga cocok digunakan dalam kasus-kasus di mana data yang tersedia mungkin tidak lengkap atau tidak terstruktur. Algoritma ini mampu mengatasi data yang tidak lengkap dengan mengisi nilai yang hilang secara otomatis dan mengekstraksi fitur-fitur yang relevan dari data yang tidak terstruktur, yang dapat membantu dalam penentuan prediksi kanker payudara.

Oleh karena itu, penggunaan BCNN sebagai algoritma klasifikasi biner dalam prediksi kanker payudara dapat dijustifikasi karena kemampuannya dalam menangkap pola-pola yang kompleks, mengatasi data yang tidak lengkap, serta melakukan klasifikasi dengan tingkat akurasi yang baik berdasarkan data yang tersedia.

4. Pengolahan Data Awal (*Preprocessing*)

Dalam *preprocessing* untuk pembuatan prediksi kanker payudara menggunakan *Binary Classification (Neural Network)*, terdapat beberapa tahapan yang terkait dengan data *wrangling*, data *cleansing*, dan data *validation*.

a. Data *Wrangling*

Data *wrangling* adalah proses yang melibatkan transformasi dan pengaturan data untuk mempersiapkan data sebelum digunakan dalam model prediksi. Langkah yang dilakukan dalam tahap ini adalah pemisahan kolom dataset yang diperlukan. Pada langkah ini kita menampilkan kolom dataset, kemudian memilih kolom yang diperlukan untuk proses pengujian, selanjutnya akan ditampilkan kembali dataset yang sudah difilter kolomnya.

```
[ ] # menampilkan kolom dataset
df.columns

Index(['id', 'diagnosis', 'Radius_mean', 'Texture_mean', 'perimeter_mean',
      'area_mean', 'smoothness_mean', 'compactness_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', 'fractal_dimension_worst'],
      dtype='object')
```

```
[ ] # memilih kolom yang diperlukan untuk proses pengujian
df = df[['diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
        'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
        'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean']]
```

```
[ ] # menampilkan kembali dataset yang sudah difilter kolomnya
df
```

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	f
0	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	
1	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	
2	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	

[]	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	f
0	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	
1	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	
2	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	
3	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	
4	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	
...	
564	M	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	
565	M	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	
566	M	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	
567	M	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	
568	B	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	

569 rows x 11 columns

```
# menampilkan isi dari kolom diagnosis
df['diagnosis'].unique()
array(['M', 'B'], dtype=object)
```

Gambar 3.3 Pemisahan Kolom yang Diperlukan

Disini 'M' digunakan untuk menunjukkan status malignant yang artinya kanker tersebut bersifat ganas sedangkan 'B' menunjukkan status benign yang artinya kanker tersebut bersifat jinak

b. Data Cleansing

Data *cleansing* atau pembersihan data adalah proses menghapus *noise*, mengatasi data yang hilang atau tidak lengkap, dan menangani *outlier* atau nilai ekstrim dalam dataset. Dalam penelitian ini, terdapat langkah dalam melakukan data *cleaning*, diantaranya adalah melihat seluruh kolom beserta dengan *type* data nya, jumlah, *unique values* dan *missing value* nya diurutkan secara *descending* berdasarkan *missing values*. Dan hasilnya tidak terdapat data yang *missing*.

	variable	dtype	count	unique	missing
0	diagnosis	object	569	2	0
1	radius_mean	float64	569	456	0
2	texture_mean	float64	569	479	0
3	perimeter_mean	float64	569	522	0
4	area_mean	float64	569	539	0
5	smoothness_mean	float64	569	474	0
6	compactness_mean	float64	569	537	0
7	concavity_mean	float64	569	537	0
8	concave points_mean	float64	569	542	0
9	symmetry_mean	float64	569	432	0
10	fractal_dimension_mean	float64	569	499	0

Gambar 3.4 Hasil Proses Data Missing

1. Pengecekan Data Duplikat

Tahap ini memastikan apakah terdapat data duplikat atau tidak. Data duplikat dalam proses data validation dapat menyebabkan bias dan mengganggu hasil analisis yang akurat. Jika terdapat data duplikat, kemungkinan besar data tersebut akan diperhitungkan secara berlebihan dalam analisis. Hal ini dapat menghasilkan estimasi yang tidak akurat dan mengganggu interpretasi yang benar dari data. Dari hasil pengecekan data duplikat pada dataset yang digunakan, tidak terdapat data duplikat .

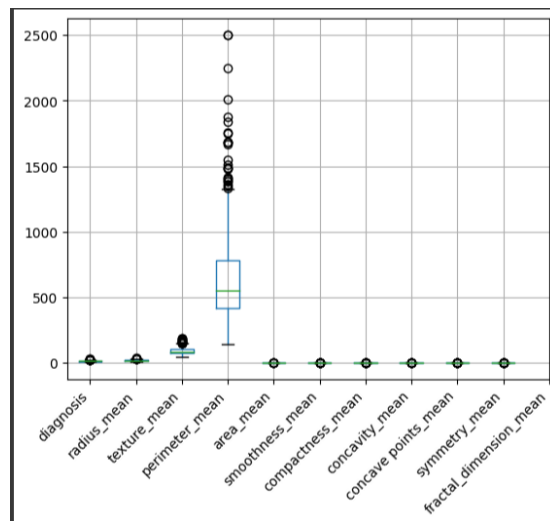
```
# cek apakah ada data yang duplikat
df.duplicated().sum()

0
```

Gambar 3.5 Hasil Pengecekan Data Duplikat

2. Pengecekan Outliers

Pengecekan outliers dalam prediksi kanker payudara melalui *Binary Classification Neural Network* (BCNN) penting untuk memastikan keakuratan dan keandalan model yang dibangun. Pada tahap ini, tujuan utama adalah mendeteksi dan menangani nilai ekstrim yang mungkin mempengaruhi hasil prediksi. Disini digunakan boxplot untuk melihat apakah ada outliers di tiap kolom dari dataset.



Gambar 3.6 Hasil Pengecekan Outliers menggunakan boxplot

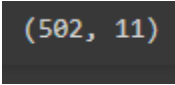
Outliers bisa di cek juga menggunakan IQR, kemudian membuat sebuah list dengan nama `index_list` yang akan menampung index dari outliers dari semua feature yang terdapat pada data. Didapati outliers pada list indeks.

```
[82, 503, 568, 567,
108, 521, 0, 0,
122, 23, 3, 68,
164, 82, 9, 78,
180, 108, 12, 82,
202, 122, 14, 108, 60,
212, 164, 78, 112, 78,
236, 180, 82, 122, 108,
339, 202, 108, 152, 122,
352, 212, 122, 180, 146,
369, 236, 181, 202, 150,
461, 250, 190, 212, 152,
503, 265, 258, 258, 258,
521, 272, 351, 351, 288,
219, 339, 352, 352, 323,
232, 352, 400, 400, 424,
239, 368, 567, 461, 561,
259, 369, 0, 563, 3,
265, 372, 68, 567, 9,
455, 373, 78, 78, 68,
562, 393, 82, 82, 71,
82, 449, 108, 108, 78,
108, 461, 112, 122, 151,
122, 503, 122, 180, 152,
164, 521, 152, 202, 176,
180, 563, 180, 212, 258,
202, 564, 202, 352, 318,
212, 3, 212, 393, 376,
236, 105, 258, 461, 379,
339, 122, 351, 3, 504,
352, 504, 352, 22, 505,
461, 520, 400, 25, 507]
```

Gambar 3.7 Hasil Outliers menggunakan IQR

3. Penghapusan Outliers

Jika terdapat outliers di dalam dataset, maka diperlukan penghapusan untuk memastikan bahwa validasi data dilakukan dengan objektif dan akurat. Jika outliers dibiarkan, maka akan memberikan bobot yang berlebihan pada evaluasi dan menghasilkan penilaian yang bias dan tidak akurat. Setelah dilakukan penghapusan outliers, selanjutnya data ditampilkan untuk mengecek jumlah baris dan kolom.



(502, 11)

Gambar 3.8 Hasil Penghapusan Outliers

c. Data Validation

Data *validation* melibatkan pengecekan integritas dan kualitas data untuk memastikan apakah data yang digunakan dalam pembuatan model adalah valid dan dapat digunakan. Terdapat beberapa langkah dalam melakukan data *validation*, diantaranya:

1. *Replace* variabel kategorikal ke dalam variabel numerik

Dalam kasus ini terdapat variabel kategorikal, yaitu variabel *diagnosis*. Dikarenakan dalam *binary classification* membutuhkan *input* berupa nilai numerik, maka dalam penelitian ini variabel *diagnosis* di *replace* untuk mengganti M menjadi 1 dan B menjadi 0. Kemudian akan ditampilkan info tipe data dan frekuensi tiap kolom.

```
# kita gunakan replace untuk mengganti M menjadi 1 dan B menjadi 0 agar
# nantinya dapat digunakan untuk proses binary classification
df['diagnosis'].replace(['M', 'B'], [1, 0], inplace=True)
```

Gambar 3.9 Replace variabel *diagnosis*

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 502 entries, 1 to 566
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   diagnosis                             502 non-null    int64
1   radius_mean                           502 non-null    float64
2   texture_mean                           502 non-null    float64
3   perimeter_mean                         502 non-null    float64
4   area_mean                             502 non-null    float64
5   smoothness_mean                       502 non-null    float64
6   compactness_mean                      502 non-null    float64
7   concavity_mean                        502 non-null    float64
8   concave points_mean                   502 non-null    float64
9   symmetry_mean                         502 non-null    float64
10  fractal_dimension_mean                 502 non-null    float64
dtypes: float64(10), int64(1)
memory usage: 47.1 KB
```

Gambar 3.10 Info Data setelah di *Replace*

2. Membuat Variabel X dan Y

Dalam tahap ini dibuat variabel X yang akan menyimpan variabel independen dan variabel Y yang akan menyimpan variabel dependen.

```
X = df[['radius_mean', 'texture_mean', 'perimeter_mean',
        'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
        'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean']]
y = df['diagnosis']
```

Gambar 3.11 Pembuatan Variabel X dan Y

5. Pelatihan Model *Binary Classification (Neural Network)*

Sebelumnya dataset dibagi menjadi training set dan test set dimana 70% data dipakai untuk training dan 30% untuk testing. Selanjutnya dilakukan pelatihan model BCNN dengan menggunakan data training. Model dilatih untuk mengenali pola dari fitur-fitur yang ada dalam dataset guna memprediksi kanker payudara. Proses ini melibatkan iterasi melalui beberapa epoch (iterasi) dan mengoptimalkan parameter model untuk meminimalkan loss function. Proses pelatihan model dan hasilnya dapat dilihat pada gambar di bawah ini.

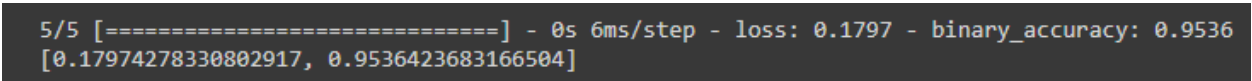
```
[ ] # Melatih model
h = model.fit(X_train, y_train, epochs=30)

Epoch 1/30
11/11 [=====] - 1s 21ms/step - loss: 0.5369 - binary_accuracy: 0.7721
Epoch 2/30
11/11 [=====] - 0s 24ms/step - loss: 0.2830 - binary_accuracy: 0.8775
Epoch 3/30
11/11 [=====] - 0s 23ms/step - loss: 0.1946 - binary_accuracy: 0.9231
Epoch 4/30
11/11 [=====] - 0s 22ms/step - loss: 0.1768 - binary_accuracy: 0.9288
Epoch 5/30
11/11 [=====] - 0s 22ms/step - loss: 0.1460 - binary_accuracy: 0.9345
Epoch 6/30
11/11 [=====] - 0s 24ms/step - loss: 0.1432 - binary_accuracy: 0.9259
Epoch 7/30
11/11 [=====] - 0s 25ms/step - loss: 0.1496 - binary_accuracy: 0.9487
Epoch 8/30
11/11 [=====] - 0s 24ms/step - loss: 0.1249 - binary_accuracy: 0.9430
Epoch 9/30
11/11 [=====] - 0s 22ms/step - loss: 0.1168 - binary_accuracy: 0.9459
Epoch 10/30
11/11 [=====] - 0s 20ms/step - loss: 0.1313 - binary_accuracy: 0.9402
Epoch 11/30
11/11 [=====] - 0s 22ms/step - loss: 0.1335 - binary_accuracy: 0.9430
Epoch 12/30
11/11 [=====] - 0s 24ms/step - loss: 0.1152 - binary_accuracy: 0.9573
Epoch 13/30
11/11 [=====] - 0s 22ms/step - loss: 0.1084 - binary_accuracy: 0.9573
```

Gambar 3.12 Pelatihan Model

6. Evaluasi Model

Jika pelatihan selesai, model BCNN selanjutnya akan dievaluasi menggunakan data pengujian yang telah dipisahkan sebelumnya. Hasil evaluasi seperti *binary accuracy* dan *loss* dapat digunakan untuk mengevaluasi performa model. Jika performa model dianggap memadai, model tersebut dapat digunakan untuk melakukan prediksi pada data baru.



```
5/5 [=====] - 0s 6ms/step - loss: 0.1797 - binary_accuracy: 0.9536  
[0.17974278330802917, 0.9536423683166504]
```

Gambar 3.13 Hasil Evaluasi Model

Dari hasil evaluasi, didapatkan nilai akurasi sebesar 95,36% persen yang menunjukkan bahwa model memiliki tingkat akurasi yang cukup tinggi dan nilai *loss* sebesar 0.1797 yang rendah yang menandakan bahwa model yang dibuat ini mempunyai hasil yang baik.

7. Prediksi

Setelah model BCNN untuk prediksi kanker payudara dievaluasi dan dinyatakan memiliki tingkat keakuratan yang baik, model tersebut dapat digunakan untuk melakukan prediksi pada data baru terkait kemungkinan seseorang mengalami kanker payudara.

Pada tahap prediksi, data baru yang berisi fitur-fitur yang relevan seperti diagnosis, radius mean, texture mean, perimeter mean, area mean, smoothness mean, compactness mean, concavity mean, concave points mean, symmetry mean, dan fractal dimension mean akan

dimasukkan ke dalam model yang sudah terlatih. Model BCNN akan melakukan prediksi berdasarkan informasi yang diberikan pada fitur-fitur ini.

Setelah proses prediksi selesai, model akan memberikan output berupa prediksi apakah kasus kanker payudara tersebut tergolong dalam kategori ganas atau tidak ganas (malignant atau benign), sesuai dengan variabel target yang telah ditentukan pada tahap pelatihan model.

Penting untuk diingat bahwa hasil prediksi dari model BCNN perlu dievaluasi dengan hati-hati. Interpretasi dan penggunaan hasil prediksi tersebut harus dilakukan secara hati-hati dengan mempertimbangkan keakuratan model dan konsultasi dengan profesional medis yang berwenang. Hal ini penting karena hasil prediksi dari model hanya berdasarkan pada data yang dijadikan acuan untuk pelatihan model, dan pengambilan keputusan medis yang sebenarnya memerlukan pertimbangan yang lebih luas dan mendalam.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Pertama-tama, dilakukan Pemuatan dan Pra-Pemrosesan Data. Pada tahap ini, dataset kanker payudara dimuat dan diproses untuk mempersiapkannya menjadi format yang dapat digunakan oleh model. Langkah-langkah melibatkan pemilihan fitur yang relevan, penghapusan outliers, dan konversi label diagnosis menjadi bentuk biner.

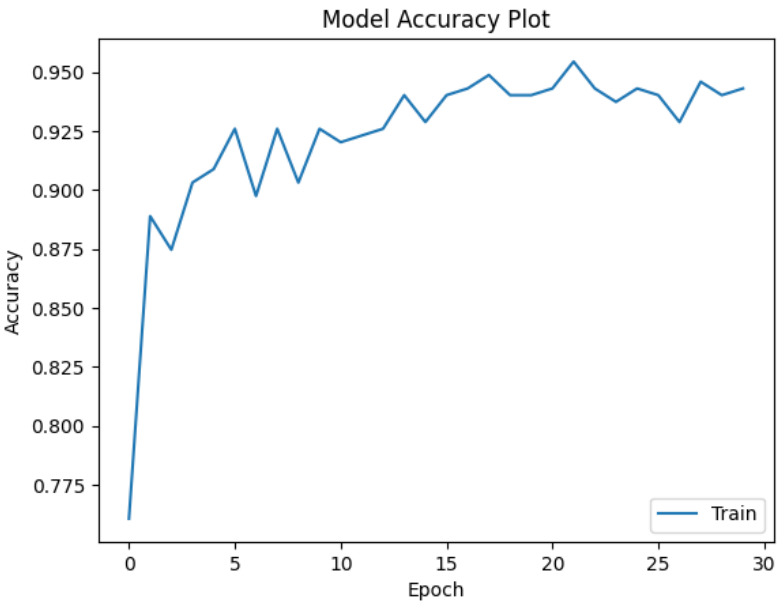
Setelah melalui proses pelatihan menggunakan metode optimasi Adam dan fungsi loss binary cross-entropy, model mencapai tingkat akurasi sebesar 95.36% pada data test. Nilai loss pada evaluasi model adalah sekitar 0.1797. Akurasi yang tinggi menandakan kemampuan model dalam membedakan antara tumor ganas dan jinak, sedangkan nilai loss yang rendah menunjukkan bahwa model memiliki tingkat kesalahan yang minim.

Grafik akurasi menunjukkan peningkatan yang signifikan selama proses pelatihan, mencapai puncaknya pada epoch ke-30. Peningkatan ini mencerminkan adaptasi model terhadap pola-pola dalam data training. Sementara itu, grafik loss menunjukkan penurunan yang stabil seiring dengan peningkatan akurasi. Hal ini mengindikasikan bahwa model berhasil mengurangi kesalahan prediksi secara konsisten selama pelatihan.

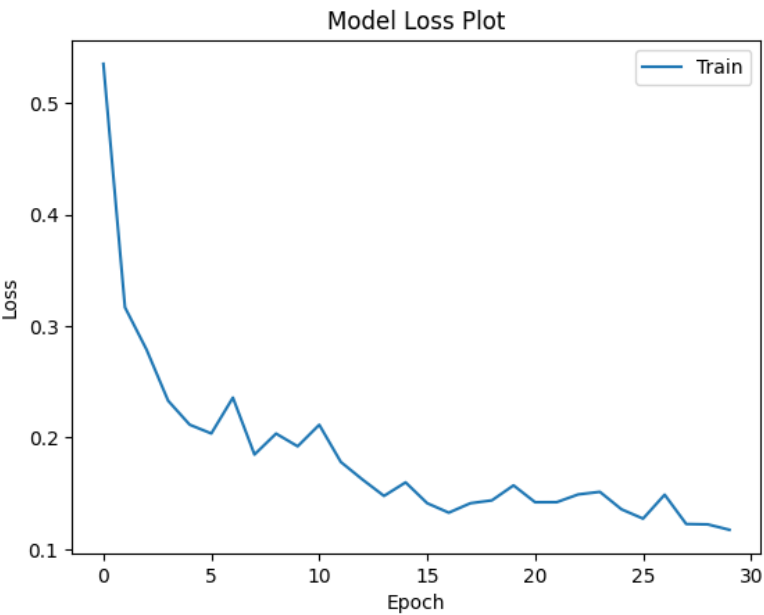
4.2 Pembahasan

Keberhasilan model dalam melakukan binary classification pada data kanker payudara dapat diukur dari tingkat akurasi yang tinggi. Hasil akurasi sebesar 95.36% menunjukkan

bahwa model mampu dengan baik mengklasifikasikan tumor sebagai ganas atau jinak berdasarkan fitur-fitur yang diberikan.



Gambar 4.1 Grafik Model Akurasi



Gambar 4.2 Grafik Model Loss

Model yang dikembangkan menunjukkan kemampuan yang baik dalam menanggapi perubahan data. Hal ini terlihat dari evaluasi pada data pengujian baru yang menghasilkan prediksi yang sesuai dengan harapan. Model tidak hanya menghafal data training tetapi mampu menggeneralisasi pola pada data yang tidak pernah dilihat sebelumnya.

Keberhasilan model dalam melakukan binary classification memberikan kontribusi signifikan dalam konteks deteksi dini kanker payudara. Dengan tingkat akurasi yang tinggi, model dapat digunakan sebagai alat bantu untuk memfilter dan mengidentifikasi kasus-kasus yang memerlukan perhatian lebih lanjut, memungkinkan deteksi dini dan intervensi yang lebih efektif.

BAB V

KESIMPULAN

Hasil evaluasi model prediksi kanker payudara yang menggunakan *Binary Classification Neural Network* (BCNN) menunjukkan bahwa model ini memiliki kemampuan yang cukup baik untuk memprediksi kemungkinan terkena kanker payudara yang jinak maupun ganas.

Analisis model BCNN menunjukkan bahwa meskipun terdapat kemungkinan *loss* yang lebih rendah, tingkat akurasi dan presisi yang tinggi menunjukkan bahwa model ini masih dapat diandalkan untuk memprediksi risiko kanker payudara, meskipun ada beberapa aspek yang perlu ditingkatkan. Analisis bobot variabel model memberikan pemahaman yang mendalam tentang komponen yang berkontribusi pada prediksi risiko kanker payudara.

Hasil evaluasi menunjukkan bahwa model yang menggunakan BCNN untuk memprediksi kanker payudara dapat digunakan sebagai alat prediksi. Disarankan untuk melakukan evaluasi lebih lanjut dengan memperluas penggunaan dataset dan melakukan pengujian yang lebih mendalam untuk memastikan keandalan model ini sebelum diterapkan secara luas.

Supaya dapat diaplikasikan secara luas dalam praktik medis untuk membantu memprediksi risiko kanker payudara, peningkatan pada metode pengujian, pengoptimalan parameter, dan eksplorasi arsitektur model yang lebih baik adalah penting untuk terus memperbaiki dan mengembangkan model ini.

DAFTAR PUSTAKA

Fitra, M., & Sri Rahayu. (2019). Analisis Algoritma Klasifikasi Neural Network untuk Diagnosis Penyakit Kanker Payudara. *DOI 10.33480/pilar.v15i2.601*. 15(2):149-156.

Hardian & Rahman. (2019). Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes. *Informatics Journal*. 4(3):117-120.

Ibnu & Kurniawati. (2020). Data Mining untuk Klasifikasi Penderita Kanker Payudara Berdasarkan Data dari University Medical Center Menggunakan Algoritma Naïve Bayes. *JURIKOM (Jurnal Riset Komputer)*. 7(1):21-27.

Merinda, Putri, & M. (2022). Deteksi Dini Kanker Payudara Menggunakan Metode Convolutional Neural Network (CNN). *Jurnal Teknologi Informasi dan Komunikasi*. 12(1):65-72.

M. & Merlinda. (2023). Implementasi Metode Convolutional Neural Network untuk Klasifikasi Breast Cancer pada Citra Histopatologi. *Jurnal Media Informatika Budidarma*. 7(1):41-49.

Reynold & Suyanto. (2020). Klasifikasi Kanker Payudara Menggunakan Residual Neural Network. *Ind. Journal on Computing*. 5(1):45-52.

Lastri, Imelda, Faradilla, & M. (2020). Optimalisasi Klasifikasi Kanker Payudara Menggunakan Forward Selection pada Naive Bayes. *JURNAL ILMIAH INFORMATIKA GLOBAL*. 11(2):63-67.

Ismail & Erna. (2022). Penerapan Algoritma Decision Tree Untuk Memprediksi Kanker Payudara menggunakan Data Mining dan Machine Learning . 11(2):103-112.

Ela. (2018). Penerapan Metode Metode Neural Network Berbasis Algoritma Genetika Untuk Prediksi Penyakit Kanker Payudara. *Journal Speed – Sentra Penelitian Engineering dan Edukasi*. 10(3):43-48.

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