

## MODEL PREDIKSI CANCER MENGGUNAKAN *AUTOENCODER*

Muhammad Ridho Fauzi, Mendy Tutu Aries Mia, Ari Kusuma Yahya

[ralhafizhi@gmail.com](mailto:ralhafizhi@gmail.com)

Mahasiswa Fakultas Teknik dan Sains Universitas Ibn Khaldun

### ABSTRAK

Algoritma Deep Learning merupakan bagian dari Machine Learning berbasis Jaringan Saraf Tiruan (JST). Metode yang akan dibahas pada penelitian ini yaitu metode Autoencoder, merupakan sebuah metode yang memiliki dua bagian utama yaitu Encoder dan Decoder. Autoencoder memiliki jumlah input dan output yang sama, dan selalu berbentuk simetris atau seperti jam pasir. Penderita kanker di Indonesia semakin meningkat setiap saat, menurut data terbaru dari Kementerian Kesehatan Indonesia penderita kanker lebih dari 347.000 penduduk. (Kemenkes, 2017). Penelitian mengenai kanker telah berkembang dengan pesat salah satunya dengan metode Deep Learning. Tujuan penelitian ini mengetahui nilai akurasi model prediksi yang dihasilkan. Adapun Manfaat yang dapat diambil dari penelitian ini diantaranya Mengetahui akurasi metode Deep Learning pada Autoencoder yang dapat dijadikan bahan acuan perbandingan.

***Kata Kunci:*** *Algoritma Deep Learning, Autoencoder, Penderita Kanker, Penelitian.*

### PENDAHULUAN

#### ***Latar Belakang***

Pada saat ini teknologi sangat berkembang dengan pesat, hal ini juga mempengaruhi pada perkembangan metode komputasi yang biasa diterapkan oleh seseorang atau peneliti pada pemecahan suatu masalah di berbagai bidang keilmuan. Saat ini metode komputasi yang banyak digunakan adalah Machine Learning atau biasa disebut dengan mesin pembelajaran. Seiring perkembangan metode komputasi, maka lahirlah sebuah metode baru yang lebih baik dari Machine Learning, yaitu metode Deep Learning. (Angermueller C, 2016; Sun T, 2017; Miotto R, 2017).

Algoritma Deep Learning merupakan bagian dari Machine Learning berbasis Jaringan Saraf Tiruan (JST). JST memiliki beberapa perbedaan yaitu jumlah hidden

layer yang banyak dapat mempengaruhi nilai output dan menghasilkan output yang akurat. Deep learning memiliki tiga bagian layer, yaitu input layer, hidden layer, dan output layer. Pada suatu jaringan Deep Learning ada yang sangat diperhatikan yaitu terletak pada hidden layer-nya, karena terdapat jumlah hidden layer yang berlapis-lapis pada satu rangkaian JST. Pada algoritma Deep Learning terdapat beberapa metode diantaranya, Autoencoder, Deep Convolutional Neural Networks (DCNN) untuk Image Classification, Deep Belief Network – Deep Neural Network (DBN-DNN) untuk Speech Recognition, Recurrent Neural Network (RNN) untuk penerjemahan bahasa, Query-Oriented Deep Extraction (QODE) yang berbasis Restricted Boltzmann Machine (RBM) untuk peringkasan. Metode yang akan

dibahas pada penelitian ini yaitu metode Autoencoder, merupakan sebuah metode yang memiliki dua bagian utama yaitu Encoder dan Decoder. Autoencoder memiliki jumlah input dan output yang sama, dan selalu berbentuk simetris atau seperti jam pasir. Selain itu juga memiliki layer code, dan jumlah neuron pada layer code merupakan jumlah dimensi yang dapat mengurangi dimensi yang terdapat pada data. (Fathul Kirom, 2016).

Penderita kanker di Indonesia semakin meningkat setiap saat, menurut data terbaru dari Kementerian Kesehatan Indonesia penderita kanker lebih dari 347.000 penduduk. (Kemenkes, 2017). Penelitian mengenai kanker telah berkembang dengan pesat salah satunya dengan metode Deep Learning.

### ***Rumusan Masalah***

Mengacu pada latar belakang yang telah dipaparkan, maka permasalahan yang diajukan adalah sebagai berikut :

1. Bagaimana penerapan algoritma Deep Learning dengan metode Autoencoder dalam mengidentifikasi penderita kanker payudara ?

## **TINJAUAN PUSTAKA**

### ***Deep Learning***

Machine Learning (ML) atau pembelajaran mesin ialah pendekatan dalam kecerdasan buatan (AI) yang banyak digunakan untuk menggantikan atau menirukan tingkah laku manusia untuk menyelesaikan suatu masalah secara otomatis. Ciri khas dari machine learning adalah proses latihan, pembelajaran, atau training. Oleh karena itu, Machine Learning membutuhkan data training untuk dipelajari. (Abu Ahmad Hania, 2017).

2. Bagaimana menghitung akurasi hasil menggunakan metode Autoencoder?

### ***Tujuan Penelitian***

Tujuan penelitian ini mengetahui nilai akurasi model prediksi yang dihasilkan dari metode Autoencoder pada permasalahan data penderita kanker payudara.

### ***Urgensi Penelitian***

Secara umum metode Deep Learning di bidang kesehatan sedang populer di dunia penelitian internasional dan masih sangat minim terutama pada penggunaan metode Autoencoder.

### ***Manfaat Penelitian***

Adapun Manfaat yang dapat diambil dari penelitian ini diantaranya sebagai berikut :

1. Sebagai bahan referensi untuk menjadi dasar penelitian selanjutnya
2. Mengetahui akurasi metode Deep Learning pada Autoencoder yang dapat dijadikan bahan acuan perbandingan metode untuk penelitian selanjutnya.

Seiring dengan kekurangan dari machine learning tersebut, dilakukanlah pengembangan-pengembangan metode pembelajaran mesin sehingga memunculkan metode baru yang dikenal dengan sebutan deep learning. (Ching T, 2018).

Saat ini, banyak penelitian yang mengarah kearah sistem kerja jaringan syaraf tiruan menggunakan konsep deep learning yang semakin berkembang. Deep learning sendiri menarik perhatian banyak aplikasi yang dapat digunakan untuk

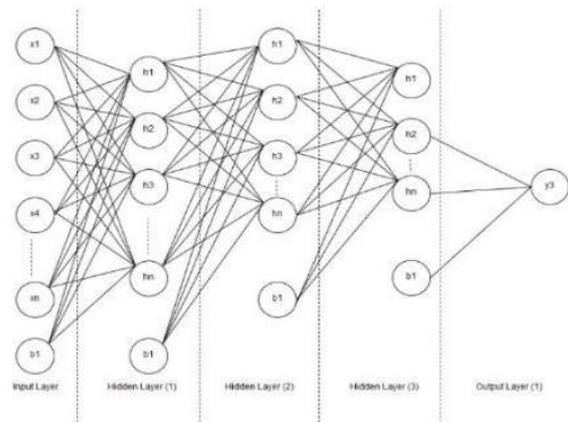
gambar, pengambilan keputusan, dan lain-lain. Faktanya, model komputasi yang digunakan deep learning terdiri dari jaringan syaraf tiruan multi layer. (Abu Ahmad Hania, 2017).

### ***Kanker***

Kanker merupakan penyakit tidak menular yang ditandai dengan pertumbuhan sel tidak normal dan tidak terkendali yang dapat merusak jaringan sekitarnya dan dapat menjalar ketempat yang jauh dari metastasis. Sel kanker bersifat ganas yang berasal dari jenis sel di tubuh manusia. (Kemenkes, 2017).

### ***Stacked Sparse Autoencoder***

SAE merupakan sebuah Jaringan Syaraf Tiruan yang terdiri dari hidden layer dari Sparse Autoencoder (SAE), output dari hidden layer terhubung ke lapisan input secara berturut turut. Pendekatan greedy layer wise dapat dijadikan sebagai pra-training dalam pelatihan setiap layer secara bergantian. Autoencoder dapat ditumpuk berlapis untuk melakukan pra-training dengan mentransformasikan setiap bobot pada jaringan. Autoencoder yang ditumpuk merupakan JST yang terdiri dari beberapa layer alat ukur otomatis tipis, dimana output setiap layer dihubungkan secara berturut turut pada input. (Aswindo Putra, 2018).



**Gambar 1.** Fitur tingkat tinggi menggunakan SSAE (Aswindo Putra, 2018).

Arsitektur autoencoder dibuat dengan tiga layer SAE yang menjadi pertimbangan, yang terdiri dari tiga hidden layer. SSAE memiliki lapisan SAE untuk fitur pembelajaran tingkat tinggi sesuai dengan Gambar 1 Jaringan syaraf feedforward melakukan pembelajaran secara efektif oleh satu layer pada satu waktu. Pembelajaran unsupervised learning dengan teknik greedy dapat meningkatkan kinerja dan inisialisasi bobot. Dalam penelitian terhadap kanker ini telah banyak dilakukan oleh peneliti, seperti penelitian terhadap kanker yang berjudul "Paralelisasi Klasifikasi Data Ekspresi Gen Kanker dengan Algoritma Deep Neural Network Menggunakan Stacked Sparse Autoencoder". Tujuan dari penelitian ini adalah untuk mengklasifikasikan penyakit berdasarkan gen yang ada. Penelitian di atas memiliki persamaan dengan penelitian yang akan kami teliti yaitu pada metodenya, sama-sama meneliti menggunakan metode autoencoder. Untuk perbedaannya terletak pada objek yang akan dijadikan input pada proses pre training. Jika penelitian di atas menggunakan gen sebagai input sedangkan penelitian kami berdasarkan hormon yang terdapat dalam tubuh yang dapat

menimbulkan kanker.(Aswindo Putra, 2018).

## METODE PENELITIAN

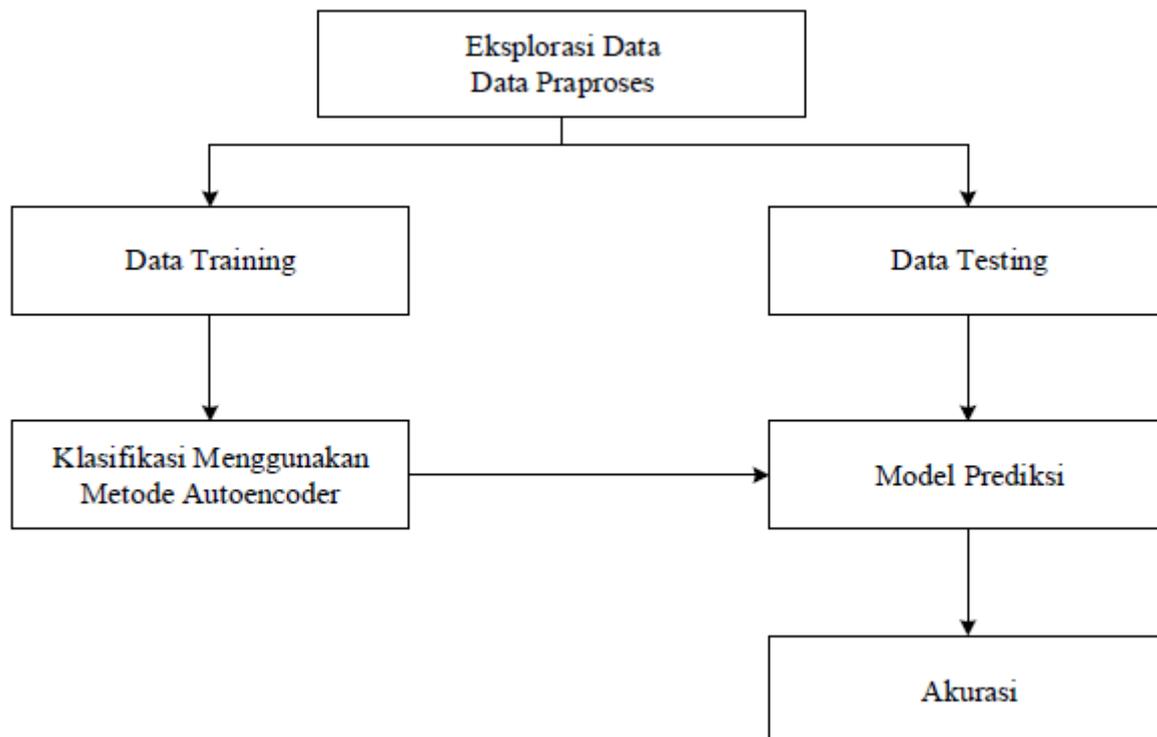
Penelitian ini dilaksanakan selama 7 bulan. Penelitian ini dilaksanakan di Laboratorium Sistem Informasi, Fakultas Teknik dan Sains, Universitas Ibn Khaldun Bogor.

### *Alat dan Bahan*

Bahan yang dibutuhkan pada penelitian ini adalah data kanker yang bersumber dari situs Uci Machine Learning Respository dengan data Breast Cancer

Coimbra Data Sets tahun 2018 yang datanya berjumlah 116 data. Alat-alat dan software (tool) yang digunakan diantaranya, seperangkat personal komputer (laptop) dengan spesifikasi Processor Intel Core i3-4030U @1,90 GHz (4 CPUs), RAM 6 GB, mouse, dan keyboard. Serta dengan menggunakan software Pyhton 3.7 dan Anaconda, Microsoft Office 2016 untuk pengetikan dan dokumentasi penelitian.

### *Prosedur Penelitian*

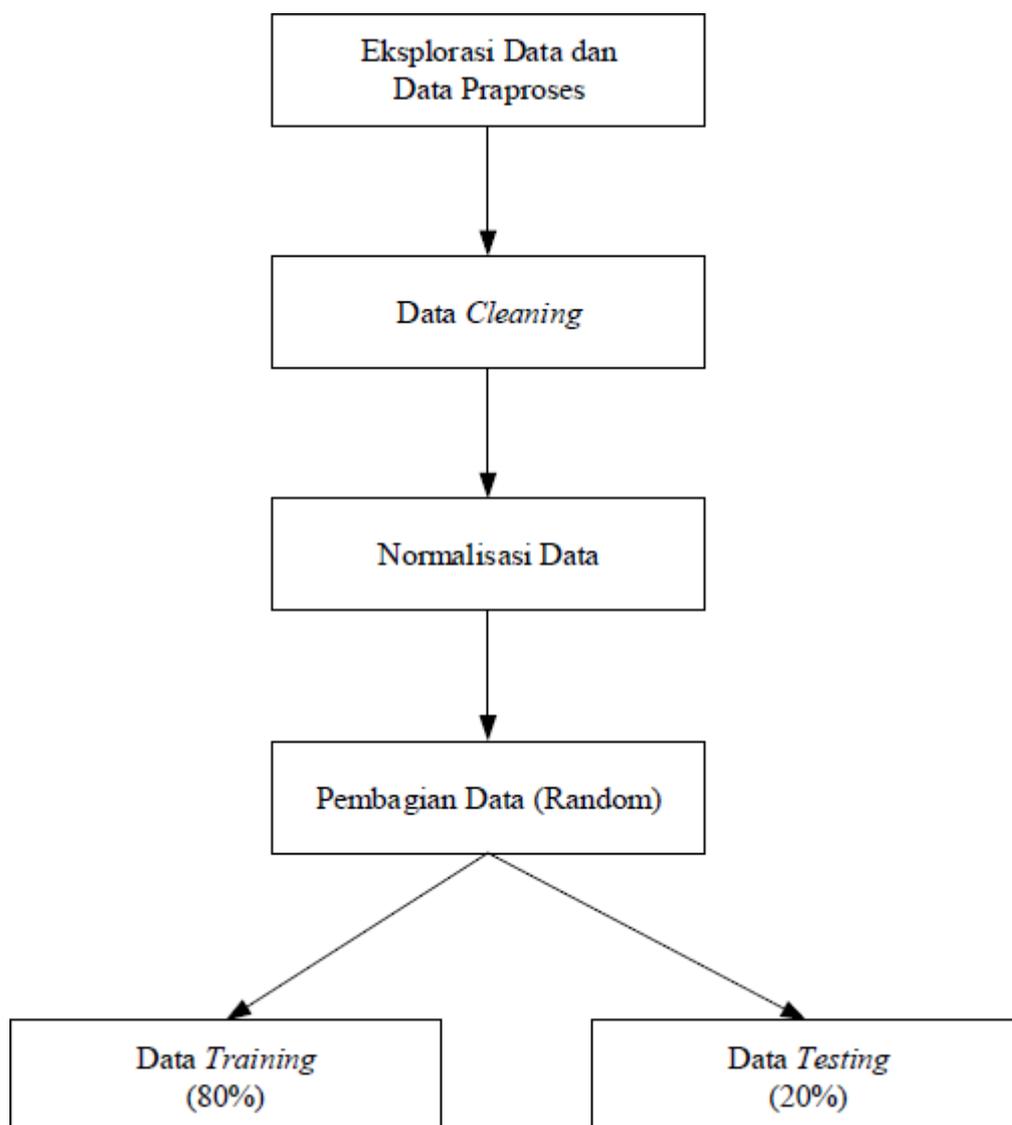


### *Eksplorasi Data*

Variabel dalam penelitian terdiri dari 10 variabel yang terdiri dari 9 atribut biasa yaitu, age (X1), BMI (X2), glucose (X3), insulin (X4), homa (X5), leptin (X6),

adiponectin (X7), resitin (X8), MCP.1 (X9) dan 1 atribut sebagai target atribut (label) yaitu classification.

No	Parameter (Variabel)	Satuan
1	Age	Years
2	BMI	kg/m <sup>2</sup>
3	Glucose	mg/dL
4	Insulin	μu/mL
5	Homa	-
6	Leptin	ng/mL
7	Adiponectin	μg/mL
8	Resitin	ng/mL
9	MCP.1	pg/dL
10	Classification (target)	-

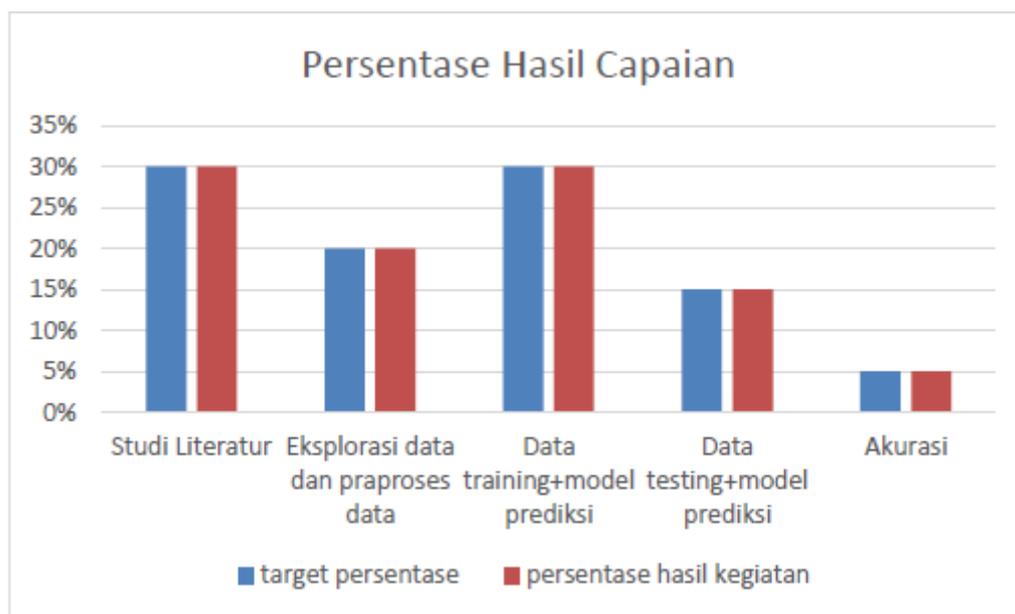


Gambar 3.2. Bagan Eksplorasi Data dan Pra Proses Data

## HASIL YANG DI CAPAI

### *Persentase Ketercapaian Hasil*

No	Kegiatan	Target Persentase	Persentase Hasil Kegiatan
1	Studi Literatur	30%	30%
2	Eksplorasi data dan praproses data	20%	20%
3	Data training+model prediksi	30%	30%
4	Data testing+model prediksi	15%	15%
5	Akurasi	5%	5%
<b>Total</b>		<b>100%</b>	<b>100%</b>



Gambar 4.1 Persentase Hasil Capaian

NO	EPOCHS	MSE (Mean Square Error)
1	500	0,0245278
2	1000	0,0250751
3	1500	0,0183424
4	2000	0,0180295
5	2500	0,0186244
6	3000	0,0172328
7	3500	0,0167198
8	4000	0,0181867

Hasil training di atas menghasilkan sebuah pendekatan model antara input dan output menggunakan metode Autoencoder untuk mendapatkan sebuah model atau bobot yang terbaik. Dari data di atas, dapat dilihat bahwa model yang terbaik adalah menggunakan sebanyak 3500 epochs dengan mempunyai Mean Square Error sebesar 0,0167198. Sehingga, model tersebut akan digunakan pada proses testing.

## KESIMPULAN

Hasil analisis dari data-data yang terkait dengan kanker payudara menunjukkan bahwa penyakit kanker payudara dapat disebabkan oleh Age, BMI, Glukosa, Insulin dan Leptin. Dari data-data tersebut kita dapat mendapatkan model prediksi dengan melakukan training dengan memasukan atribut-atribut tersebut kedalam lapisan encoder dan decoder sehingga menghasilkan sebuah model yang

## DAFTAR PUSTAKA

- Angermueller C, Pärnamaa T, Parts L, Stegle O. 2016. Deep learning for computational biology. "Molecular System Biology, Open Access". Vol. 12. Issue 7. DOI: 10.15252/msb.20156651.
- Ching T et al. 2018. Opportunities and Obstacles for Deep Learning in Biology and Medicine. J. R. Soc. Interface. 15: 20170387. DOI: 10.1098/rsif.2017.0387.
- Putra, Aswindo. 2018. Paralelisasi Klasifikasi Data Ekspresi Gen Kanker dengan Algoritma Deep

## Potensi Khusus

Penelitian ini akan menghasilkan artikel ilmiah yang bermanfaat untuk pengembangan atau bahan rujukan penelitian selanjutnya. Melalui hasil penelitian ini, diharapkan dapat dipublikasikan dalam jurnal ilmiah yang tersertifikasi.

meiliki MSE (Mean Square Error) yang terkecil yaitu sebesar 0,0167198 dengan menggunakan epochs sebanyak 3500. Model terbaik yang didapatkan dari proses training dimasukan ke dalam classifier Softmax untuk mengetahui tingkat akurasi model tersebut. Namun ,sampai saat ini kami belum mendapatkan nilai akurasi yang terbaik dari model yang didapatkan.

Neural Network Menggunakan Stacked Sparse Autoencoder. Universitas Telkom : Bandung

[http://www.academia.edu/31814136/fathul\\_kirom\\_metodologo\\_penelitian\\_review\\_peper.pdf](http://www.academia.edu/31814136/fathul_kirom_metodologo_penelitian_review_peper.pdf) (Diakses pada tanggal 5 Agustus 2019, pukul 19.55 WIB).

<http://www.depkes.go.id/article/print/17020200002/kementerian-kesehatan-ajak-masyarakat-cegah-dan-kendalikan-kanker.html> (Diakses pada tanggal 5 Agustus, pukul 22.55 WIB).