

**IMPLEMENTASI *ARTIFICIAL NEURAL NETWORK* DENGAN
ALGORITMA *BACKPROPAGATION* UNTUK KLASIFIKASI PENILAIAN
RATING PADA PENJUALAN BLACKMORES DI TOKOPEDIA**

(Skripsi)

Oleh

**DALFA HABIBAH NURUL AINI
1857031001**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

ABSTRACT

IMPLEMENTATION OF ARTIFICIAL NEURAL NETWORK WITH BACKPROPAGATION ALGORITHM FOR RATING CLASSIFICATION ON SALES OF BLACKMORES IN TOKOPEDIA

By

DALFA HABIBAH NURUL AINI

Classification is a process to find a model that can distinguish class data, with the aim that the model can be used to predict an unknown class of an object of observation. Random oversampling is a data resampling method that is able to overcome the problem of dataset imbalance in classification. Neural network is one method that has a good level of accuracy by working like a human brain neural network that has interconnected neurons to record data. In the neural network there is a backpropagation algorithm that is able to minimize loss by going backwards. In this research, an assessment of the sale of blackmores at tokopedia is carried out using the neural network method with the backpropagation algorithm and applying random oversampling to overcome the unbalanced amount of data. The results obtained indicate that the neural network method with the backpropagation algorithm is able to classify the available data with an accuracy rate of 84%. The application of resampling data using random oversampling and determining the amount of training data distribution, data testing, the number of epochs and the number of the right batch size affect the results obtained.

Keywords: *Artificial Neural Network, Backpropagation, Classification.*

ABSTRAK

IMPLEMENTASI *ARTIFICIAL NEURAL NETWORK* DENGAN ALGORITMA *BACKPROPAGATION* UNTUK KLASIFIKASI PENILAIAN RATING PADA PENJUALAN BLACKMORES DI TOKOPEDIA

Oleh

DALFA HABIBAH NURUL AINI

Klasifikasi merupakan proses untuk menemukan model yang dapat membedakan kelas data, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas yang belum diketahui dari suatu objek pengamatan. *Random oversampling* adalah salah satu metode *resampling* data yang mampu mengatasi permasalahan *imbalanced dataset* pada klasifikasi. *Neural network* merupakan salah satu metode yang memiliki hasil tingkat akurasi yang baik dengan bekerja seperti jaringan syaraf otak manusia yang memiliki *neuron-neuron* yang saling berhubungan untuk memproses sejumlah data. Di dalam *neural network* terdapat algoritma *backpropagation* yang mampu meminimalisir *loss* dengan proses secara mundur. Dalam penelitian ini dilakukan klasifikasi penilaian rating pada penjualan blackmores di tokopedia menggunakan metode *neural network* dengan algoritma *backpropagation* dan menerapkan *random oversampling* untuk mengatasi jumlah data yang tidak seimbang. Hasil yang didapatkan menunjukkan bahwa metode *neural network* dengan algoritma *backpropagation* mampu melakukan klasifikasi data yang tersedia dengan tingkat *accuracy* sebesar 84%. Penerapan *resampling* data menggunakan *random oversampling* serta penentuan jumlah pembagian data *training*, data *testing*, jumlah *epoch* dan jumlah *batch size* yang tepat mempengaruhi hasil yang diperoleh.

Kata kunci: *Artificial Neural Network*, *Backpropagation*, Klasifikasi.

**IMPLEMENTASI *ARTIFICIAL NEURAL NETWORK* DENGAN
ALGORITMA *BACKPROPAGATION* UNTUK KLASIFIKASI PENILAIAN
RATING PADA PENJUALAN BLACKMORES DI TOKOPEDIA**

Oleh

DALFA HABIBAH NURUL AINI

Skripsi

**Sebagai Salah Satu Syarat untuk Mencapai Gelar
SARJANA MATEMATIKA**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2022**

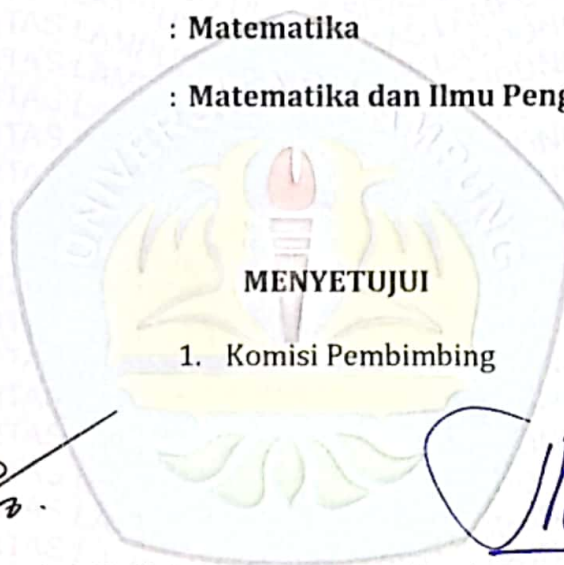
Judul Skripsi : **IMPLEMENTASI ARTIFICIAL NEURAL NETWORK
DENGAN ALGORITMA BACKPROPAGATION
UNTUK KLASIFIKASI PENILAIAN RATING PADA
PENJUALAN BLACKMORES DI TOKOPEDIA**

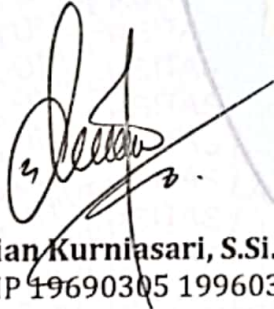
Nama Mahasiswa : **Dalfa Habibah Nurul Aini**


Nomor Pokok Mahasiswa : **1857031001**

Jurusan : **Matematika**

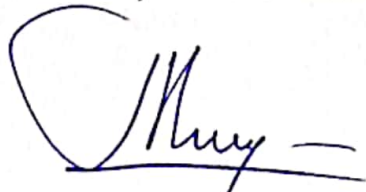
Fakultas : **Matematika dan Ilmu Pengetahuan Alam**




Dian Kurniasari, S.Si., M.Sc.
NIP 19690305 199603 2 001


Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

2. Ketua Jurusan Matematika


Dr. Aang Nuryaman, S.Si., M.Si.
NIP 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : Dian Kurniasari, S.Si., M.Sc.

Sekretaris : Dr. Aang Nuryaman, S.Si., M.Si.

**Penguji
Bukan Pembimbing : Prof. Drs. Mustofa Usman, M.A., Ph.D.**



2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Dr. Eng. Suripto Dwi Yuwono, S.Si., M.T.
NIP 19740705 200003 1 001

Tanggal Lulus Ujian Skripsi : 05 Juli 2022

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama Mahasiswa : Dalfa Habibah Nurul Aini

Nomor Pokok Mahasiswa : 1857031001

Jurusan : Matematika

Judul Skripsi : **IMPLEMENTASI ARTIFICIAL NEURAL NETWORK DENGAN ALGORITMA BACKPROPAGATION UNTUK KLASIFIKASI PENILAIAN RATING PADA PENJUALAN BLACKMORES DI TOKOPEDIA**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 28 Juli 2022

Penulis



Dalfa Habibah Nurul Aini

RIWAYAT HIDUP

Penulis bernama Dalfa Habibah Nurul Aini, dilahirkan di Kabupaten Bandung Barat, Provinsi Bandung pada 28 Agustus 2000. Penulis merupakan anak pertama dari tiga bersaudara, dengan pasangan Bapak Adnan Qohar dan Ibu Rohayati.

Penulis mengawali pendidikan di Taman Kanak-Kanak (TK) Nurul Azhar pada tahun 2004-2006. Kemudian menempuh pendidikan Sekolah Dasar (SD) di SDN 1 Kragilan pada tahun 2006-2012. Selanjutnya, penulis melanjutkan ke Sekolah Menengah Pertama di Pondok Pesantren Modern Kulliyatunnasiin Al-Islamiyah pada tahun 2012-2015 dan melanjutkan Sekolah Menengah Atas di MAN 1 Kragilan pada tahun 2015-2018.

Pada tahun 2018 penulis terdaftar sebagai mahasiswa S1 Jurusan Matematika FMIPA UNILA melalui jalur Seleksi Mandiri Masuk Perguruan Tinggi Negeri (SMMPTN) Barat. Selama menjadi mahasiswi penulis juga aktif dalam organisasi Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA UNILA. Pada tahun 2021 penulis melakukan Kuliah Praktik (KP) di Badan Pusat Statistik (BPS) Kabupaten Serang dan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Sawarna, Kecamatan Bayah, Kabupaten Lebak.

KATA INSPIRASI

“Dan Allah mengeluarkan kamu dari perut ibumu dalam keadaan tidak mengetahui sesuatu apapun, dan Dia memberimu pendengaran, penglihatan, dan hati agar kamu bersyukur”

(Q.S. An-Nahl: 78)

“Maka sesungguhnya bersama kesulitan itu ada kemudahan”

(Q.S. Al-Insyirah: 5)

“Dan barangsiapa yang bertakwa kepada Allah, niscaya Allah menjadikan baginya kemudahan dalam urusannya”

(Q.S. At-Talaq: 4)

“Bila kau cemas dan gelisah akan sesuatu masuklah ke dalamnya, sebab ketakutan menghadapinya lebih mengganggu dari pada sesuatu yang kau takuti sendiri”

(Ali bin Abi Thalib)

PERSEMBAHAN

Dengan mengucapkan rasa syukur atas segala puji dan kehadiran Allah SWT. yang telah melimpahkan nikmat serta hidayah-Nya sehingga skripsi ini dapat diselesaikan. Serta tak lupa juga sholawat serta salam selalu tercurahkan kepada junjungan kita Nabi Muhammad SAW. Dengan penuh ketulusan, penulis mempersembahkan rasa terimakasih ini kepada:

Bapak Adnan Qohar dan Ibu Rohayati

Seorang bapak dan ibu yang selalu memberikan dukungan dalam setiap keputusan dan keadaan, yang menerima segala kekurangan serta selalu memberikan doa pada setiap langkahku.

Dosen Pembimbing dan Pembahas

Terima kasih kepada dosen pembimbing dan pembahas yang sangat berjasa, selalu membantu, memberikan arahan, masukan dan ilmu yang sangat bermanfaat.

Sahabat-sahabatku

Terimakasih kepada semua sahabat-sahabatku yang telah banyak memberikan dukungan, semangat, motivasi, serta doa-doanya selama proses penyusunan skripsi ini.

Almamater Tercinta Universitas Lampung

SANWACANA

Puji syukur kehadirat Allah SWT. atas rahmat dan hidayah-Nya, shalawat serta salam selalu tercurahkan kepada baginda besar Nabi Muhammad SAW. Sehingga penulis dapat menyelesaikan skripsi dengan judul **“Implementasi *Artificial Neural Network* dengan Algoritma *Backpropagation* Untuk Klasifikasi Penilaian Rating Pada Penjualan Blackmores di Tokopedia”**. Penulis menyadari bahwa skripsi ini tidak akan terwujud tanpa adanya bantuan, bimbingan serta saran dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati penulis ingin menyampaikan banyak terima kasih kepada:

1. Ibu Dian Kurniasari, S.Si., M.Sc. selaku dosen pembimbing I yang selalu memberikan arahan, bantuan, bimbingan, motivasi dan saran yang mendukung sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku dosen pembimbing II atas saran dan masukan kepada penulis sehingga dapat menyelesaikan skripsi ini.
3. Bapak Prof. Drs. Mustofa Usman, M.A., Ph.D. selaku dosen penguji yang telah memberikan kritik dan saran yang membangun selama proses penyusunan skripsi.
4. Bapak Agus Sutrisno, S.Si., M.Si. selaku dosen pembimbing akademik yang telah memberikan bimbingan dan arahan selama masa perkuliahan.
5. Bapak Dr. Aang Nuryaman, S.Si., M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Bapak Dr. Eng. Suropto Dwi Yuwono, S.Si., M.T. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Seluruh dosen, staff, karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
8. Bapak, ibu, dan adik-adik yang selalu memberikan doa, dukungan, kasih sayang dan motivasi kepada penulis.

9. Vinny, Alifiah, Dila, Bunga, Nanda, Mega, dan Zeta yang telah menemani penulis dalam berbagai keadaan, terima kasih atas semua cerita dan kenangan selama masa perkuliahan.
10. Teman-teman seperbimbingan (Alifiah, Virda, Maydi, Sulis, Putri, Shavira, Luthfia, Febi, Oktin, Rekti, Farrel, Zaenal, Ferzy, Jhosua) yang telah memberikan dukungan dan motivasi selama ini.
11. Seluruh pihak terkait yang telah banyak membantu dan tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan guna penyempurnaan skripsi ini.

Bandar Lampung, Juli 2022
Penulis,

Dalfa Habibah Nurul Aini

DAFTAR ISI

	Halaman
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
I. PENDAHULUAN	1
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian	4
II. TINJAUAN PUSTAKA	5
2.1 <i>Machine Learning</i>	5
2.2 <i>Knowledge Discovery in Database</i>	7
2.3 <i>Resampling Data</i>	10
2.3.1 <i>Random Oversampling</i>	11
2.3.2 <i>Random Undersampling</i>	11
2.4 <i>Artificial Neural Network</i>	11
2.4.1 <i>Arsitektur Jaringan</i>	12
2.4.2 <i>Fungsi Aktivasi</i>	14
2.4.3 <i>Struktur Artificial Neural Network</i>	17
2.5 <i>Algoritma Backpropagation</i>	18
2.5.1 <i>Learning Rate, Epoch dan Batch Size</i>	20
2.6 <i>Confusion Matrix</i>	21
III. METODOLOGI PENELITIAN	24
3.1 Tempat dan Waktu Penelitian	24
3.2 Data Penelitian.....	24
3.3 Metode Penelitian	25
IV. HASIL DAN PEMBAHASAN	28
4.1 <i>Proses Knowledge Discovery in Database (KDD)</i>	28
4.1.1 <i>Data Selection</i>	28
4.1.2 <i>Pre-processing Data</i>	29
4.1.3 <i>Transformasi Data</i>	30
4.2 <i>Resampling Data</i>	31
4.3 <i>Pembagian Data Training dan Data Testing</i>	32
4.4 <i>Membangun Model Artificial Neural Network</i>	33

4.5	Menentukan Nilai Parameter	34
4.6	Menguji Model.....	36
4.7	Evaluasi Model.....	39
V.	KESIMPULAN	42
	DAFTAR PUSTAKA	43
	LAMPIRAN	

DAFTAR TABEL

Tabel	Halaman
1. Contoh <i>Label Encoder</i>	8
2. Contoh Data Sebelum <i>One Hot Encoder</i>	8
3. Contoh Data Setelah <i>One Hot Encoder</i>	8
4. Data Awal.....	28
5. <i>Categorical Encoding</i>	28
6. Transformasi Data.....	30
7. <i>Resampling Data</i>	32
8. Pembagian Data <i>Training</i> dan Data <i>Testing</i>	32
9. Perbandingan Nilai <i>Accuracy</i> dengan <i>learning rate</i> 0,001	34
10. Perbandingan Nilai <i>Accuracy</i> dengan <i>learning rate</i> 0,01	35
11. Perbandingan Nilai <i>Accuracy</i> dengan <i>learning rate</i> 0,1	35
12. Nilai <i>Precision</i> , Nilai <i>Recall</i> dan <i>F1-Score</i>	40

DAFTAR GAMBAR

Gambar	Halaman
1. <i>Single-Layer Perceptron</i>	13
2. <i>Multi-Layer Perceptron</i>	14
3. Fungsi Aktivasi ReLU.....	15
4. Fungsi Aktivasi Sigmoid.....	16
5. Struktur <i>Artificial Neural Network</i>	17
6. Arsitektur <i>Backpropagation</i>	19
7. <i>Confusion Matrix</i> Dasar	21
8. <i>Confusion Matrix Multiclass</i>	23
9. <i>Flowchart</i> Algoritma ROS	26
10. <i>Flowchart</i> Proses Metode <i>Artificial Neural Network</i>	27
11. Plot Data Penilaian Rating	31
12. Arsitektur <i>Backpropagation</i>	33
13. Grafik Model <i>Accuracy</i> Tanpa ROS	36
14. Grafik Model <i>Loss</i> Tanpa ROS	37
15. Grafik Model <i>Accuracy</i> ROS	37
16. Grafik Model <i>Loss</i> ROS	38
17. <i>Confusion Matrix</i>	39

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Keberadaan pandemi di Indonesia dari awal tahun 2020 mendorong peningkatan transaksi jual beli produk kesehatan seperti *hand sanitizer*, masker, *face shield*, dan suplemen yang mengandung vitamin. Namun kegiatan tersebut terhambat karena diberlakukannya kebijakan *social distancing* guna meminimalisir penyebaran *covid*, sehingga saat ini kegiatan manusia banyak dilakukan secara *digital* seperti transaksi jual beli yang ditawarkan melalui *marketplace* atau *e-commerce*. Salah satu *e-commerce* atau *marketplace* yang banyak dikunjungi yaitu tokopedia.

Tokopedia merupakan salah satu perusahaan yang menawarkan produknya melalui internet secara *online*. Tokopedia menyediakan beberapa fitur untuk para konsumen dalam memberikan ulasan mengenai barang yang telah dibeli, dimana ulasan tersebut dapat memberikan informasi untuk calon konsumen lainnya. Salah satu fitur yang dapat digunakan sebagai referensi para konsumen untuk membeli suatu produk yaitu rating.

Rating merupakan *feedback* dari konsumen yang diberikan dalam bentuk bintang yang bertujuan untuk menilai suatu produk. Rating biasanya menjadi alasan para konsumen dalam meninjau kembali bagus atau tidak produk yang akan dibeli. Semakin tinggi rating pada suatu produk, maka semakin tinggi minat beli para konsumen. Maka dari itu perlu dilakukan penelitian yang dapat membantu pelaku usaha untuk meningkatkan rating penjualan pada produk sehingga nantinya

mampu dalam meningkatkan minat beli para konsumen. Salah satu cara yang dapat digunakan untuk mengatasi permasalahan tersebut yaitu dengan cara mengklasifikasi rating penjualan suatu produk.

Klasifikasi adalah proses untuk menemukan model atau fungsi yang dapat menggambarkan dan membedakan kelas data atau konsep, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas yang belum diketahui dari suatu objek pengamatan (Sihombing dan Arsani, 2021). Jumlah data pada proses klasifikasi seringkali mengalami perbedaan pada setiap kelasnya atau biasa disebut *imbalanced dataset* (data yang tidak seimbang). Permasalahan tersebut dapat berpengaruh terhadap hasil dari klasifikasi yang didapatkan. *Imbalanced dataset* dapat diatasi dengan menerapkan *random oversampling* yang akan meningkatkan jumlah pada kelas minoritas sampai berjumlah sama dengan kelas mayoritas, sehingga hasil akurasi yang didapatkan pada proses klasifikasi akan menjadi lebih baik.

Metode yang sering digunakan untuk klasifikasi yaitu *Naïve Bayes* namun metode tersebut memiliki kelemahan dalam pengklasifikasian data numerik, *Naïve Bayes* cocok digunakan untuk pengklasifikasian *text* atau dokumen, sehingga diperlukan metode lain yang mampu digunakan untuk klasifikasi model data numerik. Salah satu metode lain yang memiliki algoritma klasifikasi prediksi yang akurasinya cukup tinggi adalah *artificial neural network*. Metode tersebut diimplementasikan dengan menggunakan program komputer sehingga mampu menyelesaikan sejumlah proses perhitungan (Purnama dkk, 2020), salah satunya digunakan untuk klasifikasi data numerik. *Artificial neural network* bekerja seperti jaringan saraf otak manusia dimana metode tersebut memiliki *neuron* yang saling berhubungan untuk memproses sebuah data. Dalam *artificial neural network* terdapat algoritma *backpropagation* yang termasuk ke dalam *multi layer perceptron* dimana algoritma tersebut memiliki *hidden layer* yang berada diantara *input layer* dan *output layer*. Algoritma *backpropagation* mempunyai tiga tahapan pada prosesnya yaitu tahap maju, tahap mundur, dan modifikasi bobot dimana ketiga

tahap tersebut dilakukan secara berulang-ulang hingga didapatkan *loss* terkecil dengan tingkat akurasi yang baik.

Beberapa penelitian mengenai *artificial neural network* telah dilakukan sebelumnya, seperti penelitian yang dilakukan oleh Purnama dkk (2020), mengenai klasifikasi penyakit diabetes menggunakan *artificial neural network* yang mendapatkan nilai akurasi 80% yang menunjukkan bahwa hasil klasifikasinya cukup, sehingga dari *dataset* tersebut dapat diprediksi mana saja yang mengidap penyakit diabetes dan mana yang bukan, sehingga dapat diprediksi dan menjadi tolak ukur diagnosis sehingga dapat dideteksi lebih dini. Kemudian Hadianto dkk (2019), juga melakukan penelitian untuk klasifikasi peminjaman nasabah bank menggunakan metode *artificial neural network* dan menghasilkan nilai akurasi 98.24%, yang menunjukkan bahwa klasifikasi yang dihasilkan sangat baik, sehingga nasabah dengan parameter yang ada dapat diprediksi menggunakan pola ini untuk menentukan nasabah yang layak diberikan pinjaman dari pihak bank.

Oleh karena itu, penulis akan mencoba menerapkan metode *artificial neural network* dengan algoritma *backpropagation* untuk klasifikasi penilaian rating pada penjualan *blackmores* di tokopedia untuk mengetahui aspek apa saja yang mempengaruhi hasil dari klasifikasi rating penjualan produk agar mendapatkan tingkat akurasi yang baik.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu:

1. Melakukan klasifikasi penilaian rating pada penjualan *blackmores* di tokopedia menggunakan metode *artificial neural network* dengan algoritma *backpropagation* untuk mendapatkan klasifikasi rating penjualan *blackmores* dengan tingkat akurasi yang baik.
2. Menerapkan *random oversampling* pada klasifikasi penilaian rating pada penjualan *blackmores* di tokopedia untuk mendapatkan hasil *accuracy*, *precision*, *recall*, dan *f1-score* yang baik dengan jumlah data yang seimbang.
3. Mengidentifikasi aspek apa saja yang mempengaruhi hasil klasifikasi penilaian rating pada penjualan produk di tokopedia menggunakan metode *artificial neural network* dengan algoritma *backpropagation*.

1.3 Manfaat Penelitian

Adapun manfaat dari penelitian ini yaitu:

1. Sebagai rujukan penelitian dalam klasifikasi penilaian rating penjualan produk di *e-commerce* serta menjadi bahan pertimbangan dan informasi tambahan bagi peneliti selanjutnya.
2. Sebagai bahan pertimbangan untuk para penjual dalam meningkatkan penilaian rating penjualan produk pada *e-commerce* sehingga nantinya mampu meningkatkan minat beli konsumen di *e-commerce*.

II. TINJAUAN PUSTAKA

2.1 *Machine learning*

Teknologi *Machine learning* (ML) merupakan pembelajaran terbaru yang dikembangkan untuk menyalurkan pengetahuan dan penalaran manusia dalam membentuk suatu model yang dapat membangun mesin dan rekayasa sistem secara otomatis (bisa belajar dengan sendirinya tanpa arahan dari penggunanya) (Deisenroth dkk, 2021). Hal ini menjadikan mesin tidak hanya mampu berperilaku mengambil keputusan, namun juga dapat beradaptasi dengan perubahan yang terjadi.

Menurut Deisenroth dkk (2021), pembelajaran mendasar pada *machine learning* membentuk empat pilar, yaitu:

1. Regresi

Regresi merupakan analisis yang bertujuan untuk menemukan fungsi yang memetakan input x ke nilai fungsi yang diamati, yaitu *output* y . Regresi merupakan permasalahan yang mendasar dalam *machine learning* dan masalah pada regresi muncul dalam beragam penelitian dan aplikasi, termasuk analisis *time series*, *control and robotics*, *reinforcement learning*, *optimization*, dan penerapan *deep learning*. Regresi memiliki input x dan *output* y dimana *outputnya* merupakan bilangan riil (Deisenroth dkk, 2021).

2. *Dimensionality Reduction* (Reduksi Dimensi)

Reduksi dimensi memiliki tujuan utama, yaitu untuk menemukan *dataset* yang baik dengan mengurangi dimensi *dataset* dari dimensi *dataset* x yang tinggi yang seringkali lebih mudah dianalisis dari pada data aslinya. Tidak seperti regresi,

reduksi dimensi hanya berkaitan dengan pemodelan data dan tidak mempunyai *output* yang berkaitan dengan data x (Densenroth dkk, 2021).

3. *Density estimation*

Tujuan dari *density estimation* adalah untuk menemukan distribusi probabilitas yang menggambarkan kumpulan data yang diberikan. Sama seperti reduksi dimensi, *density estimation* tidak mempunyai *output* yang berkaitan dengan data x namun, pada *density estimation* dilakukan bukan untuk mencari dimensi data yang rendah pada dataset sebaliknya, *density estimation* dilakukan untuk menggambarkan data (Densenroth dkk, 2021).

4. *Classification*

Classification (klasifikasi) merupakan tugas menggeneralisasi struktur yang sudah diketahui untuk diterapkan pada data baru. Mirip dengan regresi, klasifikasi juga memiliki input x dan *output* y namun, tidak seperti regresi yang *output*nya merupakan bilangan *real*, *output* pada klasifikasi merupakan bilangan bulat (Deisenroth dkk, 2021). Klasifikasi dibagi menjadi dua kategori yaitu *binary classification* dan *multiclass classification*. Perbedaan dari kedua kategori tersebut terletak pada jumlah kelas yang dihasilkan. *Binary classification* mengklasifikasikan elemen-elemen dari himpunan yang diberikan menjadi dua kelompok sedangkan *multiclass classification* mengklasifikasikan elemen-elemen ke dalam salah satu dari tiga atau lebih pada kelas yang dibentuk (Chaitra dan Kumar, 2018). *Loss function* yang biasa digunakan untuk *binary classification*, yaitu *binary cross entropy*, sedangkan *loss function* yang biasa digunakan untuk *multiclass classification*, yaitu *multiclass cross entropy* atau *categorical cross entropy*. *Multiclass cross entropy* digunakan untuk menghitung nilai *loss* secara terpisah untuk setiap *class* label pada setiap observasi dan menjumlahkan nilai *loss* tersebut. Berikut adalah persamaan dari *multiclass cross entropy*:

$$CE = -\sum_{k=1}^m P_k \ln(S_k) \quad (2.1)$$

dimana:

P_k = nilai probabilitas aktual *output layer*

S_k = nilai probabilitas prediksi *output layer*

2.2 *Knowledge Discovery in Database*

Istilah *Knowledge Discovery in Database* (KDD) dan data mining seringkali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain, dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining (Mardi, 2016).

Menurut Bramer *and* Max (2007), proses KDD secara garis besar dapat dijelaskan sebagai berikut:

1. *Data Selection*

Merupakan proses pemilihan data dengan tujuan proses pengolahan akan menjadi lebih baik sesuai dengan target penelitian yang akan dicapai.

2. *Pre-processing Data*

Pre-processing data mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak dan nilai yang hilang.

a. *Missing value*

Missing value atau nilai yang hilang menjadi salah satu masalah utama pada *pre-processing* data. *Missing value* dapat ditemukan karena kesalahan manusia, kesalahan mesin, atau karena kurangnya pembaruan pada data. Metode yang paling populer dalam mengatasi *missing value* yaitu dengan imputasi seperti mengganti menjadi *mean* dan modus (Sharma *et al*, 2020).

b. *Categorical Encoding*

Fitur dalam *dataset* dapat berisi satu atau lebih label dalam format numerik atau kata. Hal tersebut akan sulit dimengerti oleh komputer, oleh karena itu digunakan pengkodean (*encoding*) untuk memudahkan komputer dalam memahami label tersebut. Menurut Sharma *et al* (2020), terdapat dua *encoding* yang sering digunakan yaitu *Label Encoder* dan *One Hot Encoder*. *Label encoder* tepat digunakan ketika labelnya memiliki tingkatan yang berbeda. Berikut disajikan contoh penggunaan *label encoder* dalam Tabel 1.

Tabel 1. Contoh *Label Encoder*

Ukuran	<i>Label Encoder</i>
Kecil	1
Sedang	2
Besar	3

Sebaliknya, *one hot encoder* tepat digunakan ketika labelnya tidak memiliki tingkatan. *One hot encoder* merupakan metode untuk mengubah data kategorikal menjadi data sparse, yang memiliki segmen hanya antara satu dan nol, atau dapat disebut membuat matriks desain (Vijayalakshmi *and* Venkatachalapathy, 2019). Berikut disajikan contoh *one hot encoder* dalam Tabel 2 dan Tabel 3.

Tabel 2. Contoh Data Sebelum *One Hot Encoder*

Sample	Category	Numerical
1	Human	1
2	Human	1
3	Penguin	2
4	Octopus	3
5	Alien	4

Tabel 3. Contoh Data Sesudah *One Hot Encoder*

Sample	Human	Penguin	Octopus	Alien
1	1	0	0	0
2	1	0	0	0
3	0	1	0	0
4	0	0	1	0
5	0	0	0	1

3. Transformasi Data

Transformasi data merupakan proses yang digunakan untuk merubah skala pengukuran dari bentuk asli kedalam bentuk lain sehingga data dapat digunakan untuk analisis dan asumsi-asumsi tertentu. Transformasi bertujuan untuk menyesuaikan data yang diolah berdasarkan algoritma yang digunakan di dalam pengolahan data. Beberapa cara yang digunakan pada transformasi data yaitu:

- a. *Min-max normalization* merupakan proses perubahan linear pada data asli dengan rentang yang diberikan (Saranya and Manikandan, 2013). Persamaan pada *min-max normalization* yaitu:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2.2)$$

dimana:

x^* = nilai x baru

X_{min} = nilai x minimum

X_{max} = nilai x maksimum

- b. *Z-score normalization* merupakan metode normalisasi berdasarkan *mean* (nilai rata-rata) dan standar deviasi dari data (Nasution dkk, 2019). Persamaan pada *z-score normalization* yaitu:

$$Z = \frac{x - \mu}{\sigma} \quad (2.3)$$

dimana:

x = nilai yang diamati

μ = rata-rata

σ = standar deviasi

- c. *Decimal scaling* merupakan metode normalisasi dengan menggerakkan nilai desimal dari data ke arah yang diinginkan (Nasution dkk, 2019). Persamaan pada *decimal scaling* yaitu:

$$v' = \frac{v}{10^j} \quad (2.4)$$

dimana:

v' = nilai baru setelah *decimal scaling*

v = nilai masing-masing atribut

j merupakan bilangan bulat yang mendefinisikan pergerakan titik desimal dimana itu sama dengan jumlah digit yang ada dalam nilai maksimum dalam tabel data.

4. Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu.

5. *Interpretation / Evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

2.3 *Resampling Data*

Dalam *machine learning* dibutuhkan data yang akan digunakan untuk dipelajari oleh mesin agar mesin dapat memiliki kecerdasan untuk memprediksi data baru. Namun terkadang beberapa data memiliki jumlah yang tidak seimbang (*imbalanced*), hal tersebut terjadi karena beberapa kelas data memiliki berat sebelah atau terdapat kelas yang memiliki banyak data dan terdapat kelas lain yang memiliki sedikit data. Permasalahan pada *imbalanced dataset* dapat diatasi menggunakan *resampling* data. Menurut Mahmood (2015), teknik *resampling* terbagi menjadi 3, yaitu *oversampling*, *undersampling*, dan *hybrid* dari kedua teknik tersebut. *Oversampling* digunakan untuk menambahkan jumlah data pada kelas minoritas hingga jumlah data pada kelas minoritas dan kelas mayoritas seimbang, contohnya yaitu *random oversampling*. *Undersampling* digunakan untuk mengeliminasi beberapa data pada kelas mayoritas yang dianggap kurang relevan, contohnya yaitu *random undersampling*. *Hybrid* yaitu gabungan dari kedua teknik *sampling* tersebut sesuai dengan kebutuhan dan karakteristik data.

2.3.1 *Random Oversampling*

Random oversampling merupakan salah satu cara untuk mengatasi *imbalanced dataset* dengan menyeimbangkan jumlah data pada kelas mayoritas dan kelas minoritas. Berbeda dengan *random undersampling* yang secara acak mengeliminasi data pada kelas mayoritas, pada proses *random oversampling* akan dilakukan replikasi secara acak pada data di kelas minoritas hingga kelas minoritas mendapatkan jumlah data yang sama dengan kelas mayoritas. Penelitian menunjukkan penggunaan *random oversampling* dapat mengurangi terjadinya *overfitting* (Zheng dkk, 2015).

2.3.2 *Random Undersampling*

Random undersampling merupakan salah satu cara untuk mengatasi *imbalanced dataset* dengan mengeliminasi data. *Random undersampling* akan melakukan eliminasi secara acak pada data di kelas mayoritas hingga kelas mayoritas mendapatkan jumlah data yang sama dengan kelas minoritas. Adapun contoh proses *random oversampling* sebagai berikut:

2.4 *Artificial Neural Network*

Salah satu metode yang paling umum digunakan untuk klasifikasi yaitu *artificial neural network*. Terinspirasi dari otak manusia, ANN juga terdiri dari beberapa *neuron* dan terdapat hubungan antara *neuron-neuron* tersebut (Putri, 2021). ANN mempunyai kemampuan untuk memecahkan masalah yang berkaitan dengan pola atau klasifikasi dengan menyimpan pengetahuan yang didapatkan dari hasil proses pelatihan (Gerry *et al*, 2021).

Menurut Deisenroth dkk (2021), perancangan algoritma ANN pada *machine learning* memiliki dua perbedaan data yakni:

1. Pelatihan (*train/ training*)

Proses ini dimulai dengan memasukkan pola-pola pelatihan (data latih) ke dalam jaringan. Pola pelatihan tersebut akan mengubah bobot yang menjadi penghubung antar *neuron*. Semakin banyak data latih yang digunakan, akan semakin tinggi nilai akurasi yang didapatkan sesuai yang didapatkan pada pengujian partisi data (Khatami dkk, 2020).

2. Pengujian / prediksi (*test/ testing*).

Tahap pengujian (*testing*) merupakan pengujian suatu pola masukan yang belum pernah dilatihkan sebelumnya (data uji) menggunakan bobot-bobot yang telah dihasilkan pada tahap pelatihan (Prabowo dkk, 2020).

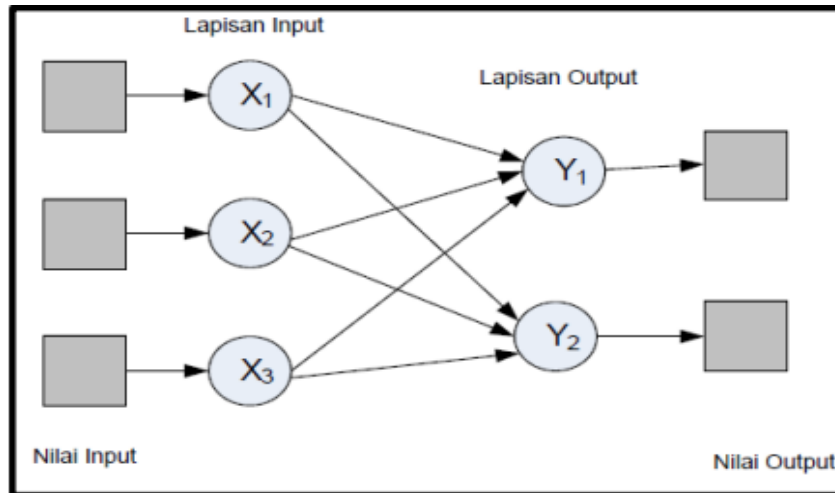
Menurut Susanti dan Maiyana (2021), *artificial neural network* ditentukan oleh 3 hal:

1. Pola hubungan antar *neuron* (arsitektur jaringan)
2. Metode untuk menentukan bobot penghubung (metode *training*)
3. Fungsi aktivasi

2.4.1 Arsitektur Jaringan

1. *Single Layer Perceptron*

Single layer perceptron merupakan salah satu arsitektur jaringan *artificial neural network* yang paling sederhana, yang terdiri dari lapisan *input* dan lapisan *output*.

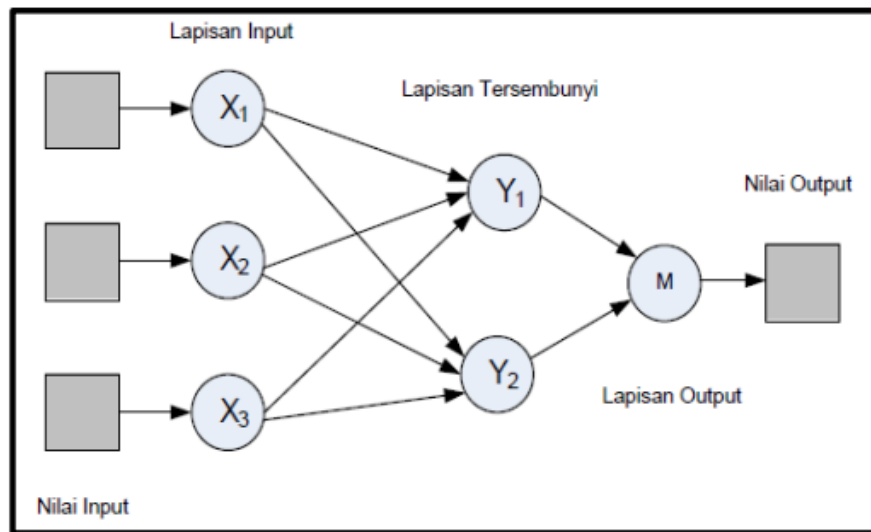


Gambar 1. *Single-Layer Perceptron*.(Sumber: Aprizal dkk, 2019)

Input layer yaitu lapisan yang menerima masukan (*input*) dari jaringan luar dan mendistribusikan sinyal ke semua *neuron* pada *output layer*. *Output* merupakan hasil dari semua perhitungan *input* dan bobot-bobot yang diberikan pada *neuron* yang saling berhubungan dengannya. Pada dasarnya *output* merupakan hasil akhir pemrosesan (Prabowo dkk, 2020).

2. *Multi Layer Perceptron*

Multi layer perceptron merupakan arsitektur jaringan *artificial neural network* yang terdiri atas tiga bagian utama yakni lapisan *input*, lapisan tersembunyi, dan lapisan *output*.



Gambar 2. Multi-Layer Perceptron. (Sumber: Aprizal dkk, 2019)

Lapisan tersembunyi atau *hidden layer* yaitu lapisan yang berfungsi untuk menyalurkan informasi dari *input layer* ke *output layer* dimana sebelumnya informasi tersebut sudah dilakukan pemrosesan untuk memberikan *output* yang optimal (Prabowo dkk, 2020).

2.4.2 Fungsi Aktivasi

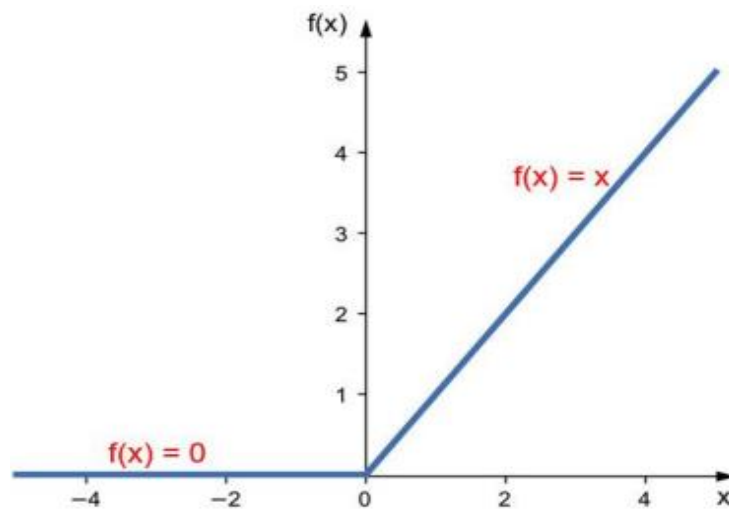
Menurut Putri (2021), fungsi aktivasi digunakan untuk mengaktifkan atau tidak mengaktifkannya *neuron*. Fungsi aktivasi yang umum digunakan pada *artificial neural network* yaitu fungsi aktivasi *Rectified Linear Unit* dan *sigmoid*. Fungsi aktivasi yang dapat digunakan untuk *multiclass* yaitu fungsi aktivasi *softmax*.

a. Fungsi Aktivasi *Rectified Linear Unit* (ReLU)

ReLU merupakan fungsi aktivasi yang digunakan untuk menormalisasikan nilai yang dihasilkan *layer*. Persamaan pada fungsi aktivasi ini yaitu:

$$f(x) = \max(0, x) \quad (2.5)$$

Setiap *input* negatif yang diberikan akan selalu dipetakan menjadi 0 dan setiap *input* positif akan dipertahankan nilainya, sehingga tidak terdapat hasil yang bernilai negatif. Adapun grafik fungsi aktivasi ReLU sebagai berikut:



Gambar 3. Fungsi Aktivasi ReLU. (Sumber: Maysanjaya, 2020)

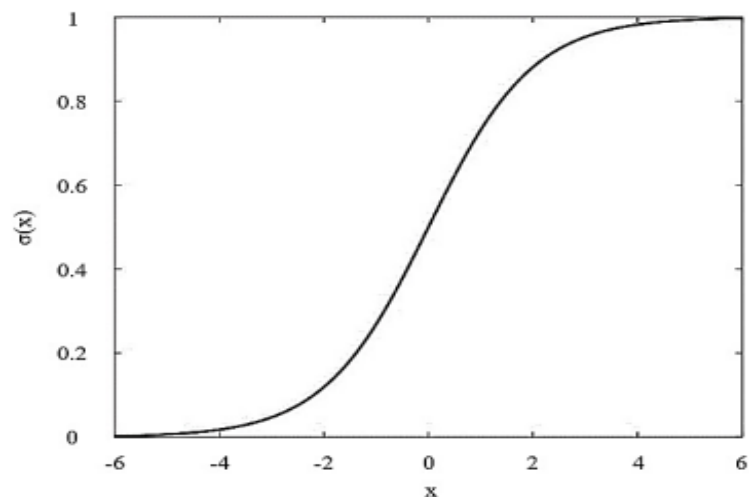
b. Fungsi Aktivasi *Sigmoid*

Fungsi sigmoid merupakan fungsi aktivasi yang sering digunakan pada *artificial neural network* dengan algoritma *backpropagation*. Fungsi ini mengambil kisaran antara 0 sampai 1. Bentuk dari fungsi *sigmoid* yaitu:

$$f(x) = \frac{1}{1+e^{-x}} \quad (2.6)$$

dengan turunan

$$\begin{aligned}
 f'(x) &= \frac{u(x)}{v(x)} = \frac{u'(x)v(x) - u(x)v'(x)}{[v(x)]^2} \\
 f'(x) &= \frac{((0)(1+e^{-x}) - (1)(-e^{-x}))}{(1+e^{-x})^2} \\
 &= \frac{e^{-x}}{(1+e^{-x})^2} \\
 &= \frac{1-1+e^{-x}}{(1+e^{-x})^2} \\
 &= \frac{1+e^{-x}}{(1+e^{-x})^2} - \frac{1}{(1+e^{-x})^2} \\
 &= \frac{1}{(1+e^{-x})} - \frac{1}{(1+e^{-x})^2} \\
 &= \frac{1}{(1+e^{-x})} \left(1 - \frac{1}{(1+e^{-x})}\right) \\
 f'(x) &= f(x)(1 - f(x)) \tag{2.7}
 \end{aligned}$$



Gambar 4. Fungsi aktivasi *Sigmoid*. (Sumber: Maysanjaya, 2020)

c. Fungsi Aktivasi *Softmax*

Softmax merupakan fungsi aktivasi yang digunakan untuk klasifikasi *multiclass*. *Output* pada fungsi aktivasi ini berupa probabilitas pada masing-masing *class* dengan rentang 0 hingga 1, yang jika dijumlahkan hasilnya sama dengan 1.

Berikut merupakan persamaan fungsi *softmax*.

$$S_k = \frac{e^{y_k}}{\sum_{k=1}^m e^{y_k}} \tag{2.8}$$

dimana:

S_k = nilai probabilitas prediksi *output layer*

y_k = sinyal *output* pada *output layer*

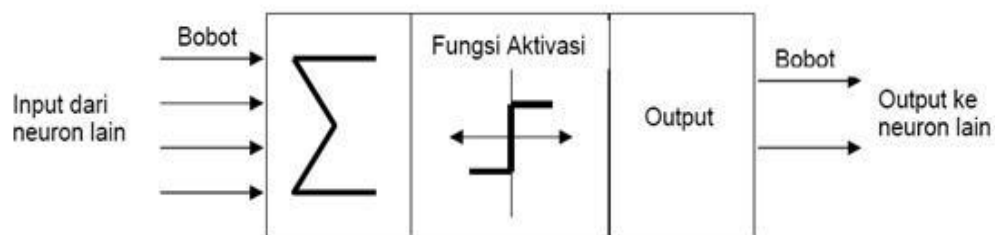
e = eksponensial, dimana memiliki ketetapan yaitu 2,7183

m = jumlah kelas pada *output layer*

$\sum_{k=1}^m e^{y_k}$ = jumlah eksponensial sinyal *output* pada *output layer*

2.4.3 Struktur *Artificial Neural Network*

Artificial neural network atau jaringan syaraf tiruan terdiri dari beberapa *neuron* dan terdapat penghubung antara *neuron-neuron* tersebut seperti halnya otak manusia yang dikenal dengan bobot (Kusumadewi, 2003). Disajikan struktur *artificial neural network* pada gambar di bawah ini:



Gambar 5. Struktur *Artificial Neural Network*. (Sumber: Kusumadewi, 2003)

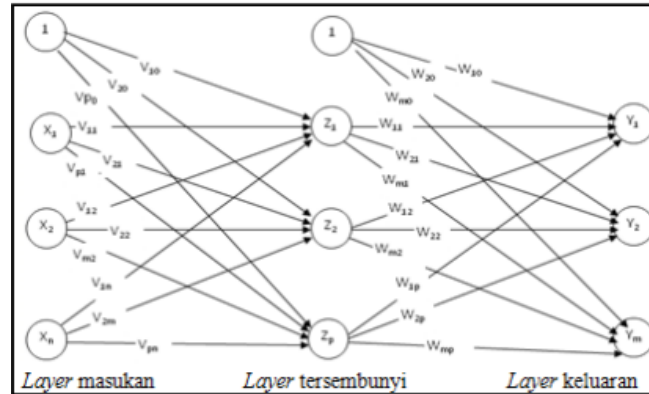
Sebagai suatu generalisasi model matematis dan pemahaman manusia atau suatu syarat biologis manusia, maka *artificial neural network* dibentuk atas dasar asumsi sebagai berikut :

1. Pemrosesan informasi terjadi pada elemen-elemen sederhana yang disebut *neuron*.
2. Sinyal dilewatkan antara *neuron* melalui jalur yang terhubung.
3. Setiap jalur yang terhubung akan memiliki bobot yang disesuaikan kemudian akan dikalikan dengan sinyal yang melewatinya.

4. Setiap *neuron* memiliki fungsi aktivasi dengan cara menjumlahkan bobot-bobot yang masuk untuk menentukan *outputnya*.

2.5 Algoritma *Backpropagation*

Menurut Prabowo dkk (2020), metode *backpropagation* merupakan salah satu metode yang sangat baik dalam menangani masalah pengenalan pola-pola kompleks. Algoritma *backpropagation* merupakan bagian dari *multi layer perceptron*, sehingga dalam struktur jaringannya terdapat *hidden layer* yang harus diperhatikan. *Backpropagation* merupakan teknik *supervised learning* yang memiliki target sebagai *output* dengan mengubah dan menyesuaikan nilai pada bobot-bobot yang terhubung dengan *neuron-neuron* yang ada pada *hidden layer*. *Backpropagation* memiliki dua cara utama dalam pembelajaran kesalahan-koreksi yaitu pergerakan maju dan mundur. Pada proses pergerakan maju, vektor input diterapkan pada layer input disetiap jaringan yang akan mempengaruhi semua jaringan berdasarkan lapis demi lapis kemudian pada tahap ini juga *neuron-neuron* diaktifkan dengan menggunakan fungsi aktivasi, kemudian didapatkan *error* atau selisih antara nilai setelah diaktivasi atau *output* dengan target. Kemudian *error* tersebut akan diminimalisir dengan cara mengubah bobot dan bias yang dihasilkan dengan pergerakan mundur dari *output* yang dihasilkan sebelumnya. Bobot disetiap lapisan pada pergerakan maju bernilai tetap sedangkan pada pergerakan mundur, pembobotan di setiap lapisan berubah berdasarkan aturan koreksi kesalahan (Hadianto dkk, 2019).



Gambar 6. Arsitektur *Backpropagation*. (Sumber: Sukarno dkk, 2014)

Berikut adalah model jaringan *backpropagation* dengan satu *hidden layer*:

$$Y_k = f_2 [W_m + \sum_{j=1}^p W_{jk} \cdot f_1 [V_p + \sum_{i=1}^n V_{ij} X_i]] \quad (2.9)$$

dimana:

V_p = bobot bias pada *input layer* ke *hidden layer*

W_m = bobot bias pada *hidden layer* ke *output layer*

V_{ij} = bobot pada lapisan *input* ke *hidden layer*

W_{jk} = bobot pada *hidden layer* ke *output layer*

f_1 = fungsi aktivasi yang digunakan pada *hidden layer*

f_2 = fungsi aktivasi yang digunakan pada *output layer*

X_i = data awal pada *input layer*

Proses pelatihan pada algoritma *backpropagation* meliputi tahapan berikut:

1. Tahap maju

Input layer akan menerima data awal $X_i, i = 1, 2, \dots, n$ yang akan diteruskan ke *hidden layer*, kemudian tiap *node* pada *hidden layer* $Z_j, j = 1, 2, \dots, p$ akan menjumlahkan sinyal-sinyal terbobot,

$$z_j = V_p + \sum_{i=1}^n V_{ij} X_i \quad (2.10)$$

digunakan fungsi aktivasi untuk menghitung sinyal *output* pada *hidden layer*.

$$Z_j = f_1 \cdot z_j \quad (2.11)$$

Kemudian sinyal tersebut akan diteruskan menuju *output layer* dan tiap *node* pada *output layer* $Y_k, k = 1, 2, \dots, m$ akan menjumlahkan sinyal-sinyal terbobot,

$$y_k = \left[W_m + \sum_{j=1}^p W_{jk} \right] \cdot Z_j \quad (2.12)$$

digunakan fungsi aktivasi untuk menghitung sinyal *output* pada *output layer*.

$$Y_k = f_2 \cdot y_k \quad (2.13)$$

2. Tahap mundur

Tiap *node* pada *output layer* $Y_k, k = 1, 2, \dots, m$ akan menerima target pola yang berhubungan dengan pola input dan tiap *hidden layer* $Z_j, j = 1, 2, \dots, p$ akan menghitung δ_j kemudian digunakan untuk menghitung bobot terkoreksi dan bias antara *input layer* dan *hidden layer*.

$$\frac{\partial E}{\partial W} = -(t_j - Y_k) * Y_k(1 - Y_k) * X_i \quad (2.14)$$

3. Modifikasi bobot

Masing-masing *node* pada *output layer* $Y_k, k = 1, 2, \dots, m$ akan memperbaharui nilai bobot dan nilai bias $j = 1, 2, \dots, p$ dan setiap *node* pada *hidden layer* memperbaharui nilai bobot dan nilai bias $i = 1, 2, \dots, n$ sehingga didapatkan nilai bobot dan nilai bias yang baru.

$$Bobot_{baru} = Bobot_{lama} - \eta \frac{\partial E}{\partial W} \quad (2.15)$$

$$Bias_{baru} = Bias_{lama} - \eta \frac{\partial E}{\partial W} \quad (2.16)$$

2.5.1 Learning Rate, Epoch dan Batch Size

Learning rate adalah salah satu parameter yang digunakan untuk tahap mundur pada algoritma *backpropagation* yang mempengaruhi sistem untuk menurunkan nilai *loss* pada proses *training* (Khatami dkk, 2020).

Menurut Vijayalakshmi and Venkatachalapathy (2019), *epoch* merupakan jumlah iterasi selama proses pelatihan yang memberikan input dari jaringan dan juga memperbarui bobot jaringan. Proses pelatihan tersebut berlangsung pada *artificial neural network* sampai ke awal lagi.

Batching adalah salah satu pendekatan umum untuk mempercepat komputasi *artificial neural network*. Proses ini melibatkan perhitungan *gradient* yang berisi beberapa contoh *training* dalam satu kali *forward pass* / *backward pass*. Untuk mengetahui efisiensi komputasi *batching*, misalkan terdapat ukuran *batch* sebesar 500, dengan 1000 contoh *training*. Maka hanya diperlukan 2 iterasi untuk menyelesaikan 1 *epoch* (Lewis, 2019).

2.6 Confusion Matrix

Confusion matrix yaitu matriks yang digunakan untuk menampilkan prediksi klasifikasi serta klasifikasi yang aktual. Pada dasarnya *confusion matrix* berdimensi 2x2 yang terdiri dari baris dan kolom dimana baris merupakan kelas prediksi dan kolom merupakan kelas sebenarnya. Kelas yang digunakan dalam matriks dasar terbagi menjadi dua yaitu kelas positif dan negatif (Arifiyanti dkk, 2018).

Confusion matriks dasar dapat dilihat pada gambar berikut:

		Predicted Values	
		POSITIVE	NEGATIVE
Actual Values	POSITIVE	TP (True positive)	FN (False negative)
	NEGATIVE	FP (False positive)	TN (True negative)

Gambar 7. *Confusion Matrix* Dasar. (Sumber: www.analyticsvidhya.com)

Dari *confusion matrix* tersebut akan didapatkan *accuracy*, *precision*, *recall* dan *f1-score* pada model klasifikasi. *Accuracy* merupakan kemampuan

pengklasifikasi dalam memprediksi kelas yang benar (Arifiyanti dkk, 2018), *precision* yaitu tingkat ketepatan antara informasi yang diminta oleh peneliti dengan *output* yang diberikan oleh sistem, *recall* yaitu merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi, dan *f1-score* yaitu perbandingan rata-rata nilai *precision* dengan nilai *recall* yang dapat digunakan sebagai acuan dalam melihat ketepatan hasil prediksi dan nilai sebenarnya ketika jumlah data FN dan FP memiliki perbedaan yang jauh.

$$Accuracy = \frac{TP+TN}{Total} \quad (2.17)$$

$$Precision = \frac{TP}{TP+FP} \quad (2.18)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.19)$$

dimana:

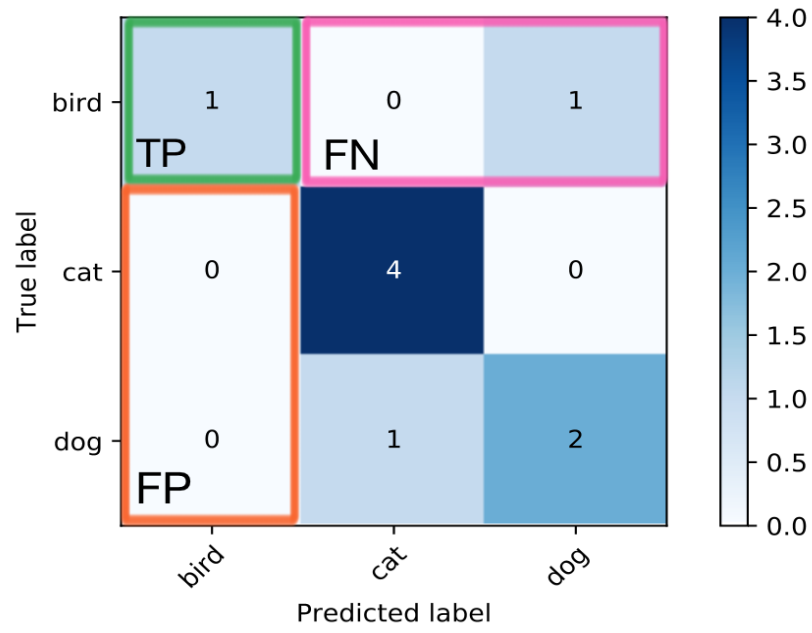
TP = jumlah kelas prediksi bernilai benar dan kelas sebenarnya juga bernilai Benar

TN = jumlah kelas prediksi bernilai salah dan kelas sebenarnya juga bernilai salah

FP = jumlah kelas prediksi bernilai benar dan kelas sebenarnya bernilai salah

FN = jumlah kelas prediksi bernilai salah dan kelas sebenarnya bernilai benar

Sedangkan *confusion matrix* untuk klasifikasi *multiclass* berbeda dengan *confusion matrix* dasar yang memiliki kelas positif dan negatif. Gambar *confusion matrix multiclass* sebagai berikut:



Gambar 8. *Confusion Matrix Multiclass*. (Sumber: www.mariakhalusova.com)

III. METODOLOGI PENELITIAN

3.1 Tempat dan Waktu Penelitian

Penelitian ini dilakukan pada semester genap tahun akademik 2021/2022, bertempat di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

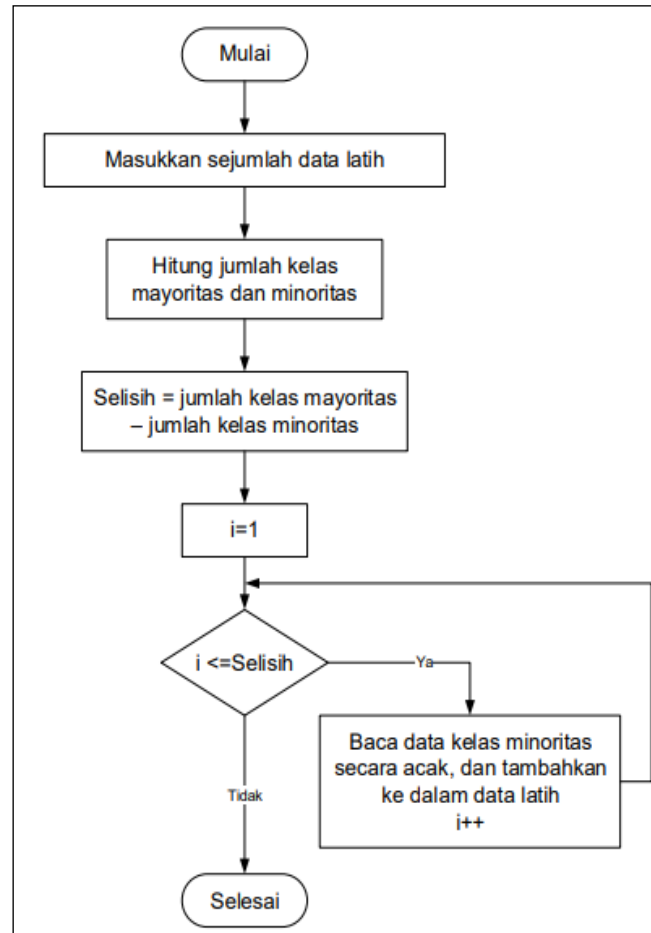
Data yang digunakan pada penelitian ini adalah data penilaian rating pada produk blackmores yang diperoleh dari <https://www.kaggle.com/>. Terdapat 5 variabel yang diperoleh dari *website* tersebut, yaitu harga produk, stok produk, *official store* (toko yang menjual produk asli) dengan bentuk data kategorik (*true* dan *false*), *power badge* (toko yang berlangganan *power merchant* pada tokopedia) dengan bentuk data kategorik (*true* dan *false*), dan rating yang terbagi menjadi 6 penilaian yaitu penilaian rating 0 (produk belum memiliki penilaian) dengan jumlah 11768 data, rating 1 dengan jumlah 2 data, rating 2 dengan jumlah 5 data, rating 3 dengan jumlah 7 data, rating 4 dengan jumlah 47 data, dan rating 5 dengan jumlah 844 data. Data awal pada penelitian ini berjumlah sebanyak 12673 data dan setelah dilakukan *resampling* data menggunakan *random oversampling* data akan berjumlah 70608 data dengan masing-masing rating berjumlah 11768 data.

3.3 Metode Penelitian

Penelitian ini menggunakan metode *artificial neural network* dengan algoritma *backpropagation* untuk klasifikasi rating penjualan produk *blackmores* di tokopedia dengan menggunakan *software python*. Langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut:

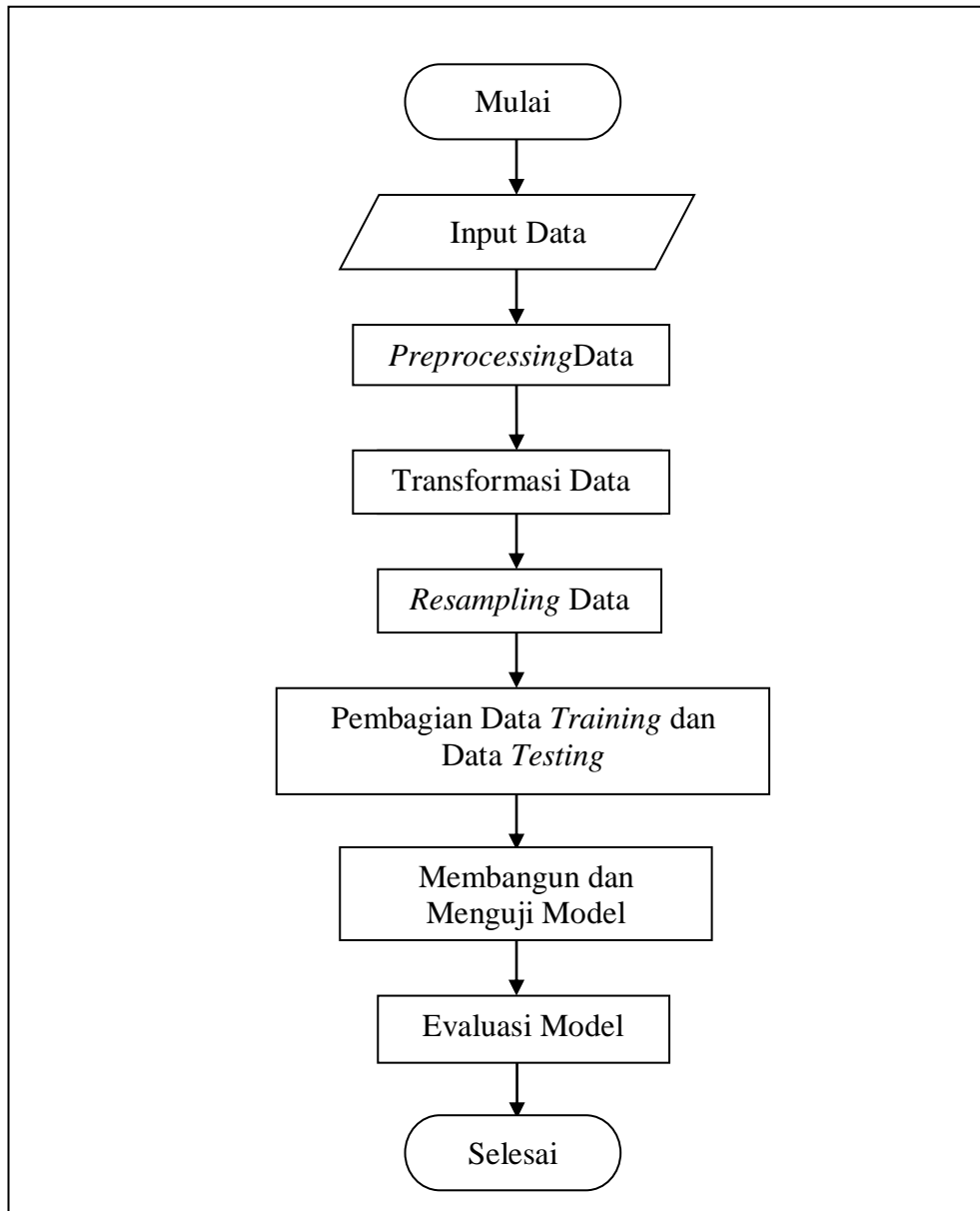
- a. Melakukan input data penilaian rating pada penjualan produk *blackmores* di tokopedia yang diperoleh dari *website kaggle*.
- b. Melakukan *Pre-processing* data dengan *label encoder* pada variabel yang berbentuk kategorik (*true* dan *false*) yaitu *official store* dan *power badge* menjadi data numerik yang bernilai 1 dan 0.
- c. Melakukan normalisasi data dengan *z-score normalization* untuk variabel harga produk dan stok produk.
- d. Melakukan *resampling* data menggunakan *random oversampling* pada penilaian rating yang akan menjadi target penelitian atau *output*.
- e. Membagi data *training* dan data *testing*.
- f. Membangun model *artificial neural network* dengan algoritma *backpropagation*.
- g. Menentukan nilai parameter yang akan digunakan seperti *learning rate*, *batch size* dan *epoch*.
- h. Menguji model dan membandingkan grafik *loss* model dan grafik *accuracy* model yang telah diperoleh pada data yang menerapkan *random oversampling* dengan data yang tidak menerapkan *random oversampling*.
- i. Melakukan evaluasi model menggunakan *confussion matrix*.
- j. Melihat tingkat akurasi model.

Adapun *flowchart* pada algoritma *random oversampling* dapat digambarkan sebagai berikut:



Gambar 9. *Flowchart* Algoritma ROS.

Berikut merupakan *flowchart* dari model secara singkat:



Gambar 10. *Flowchart* Proses Metode *Artificial Neural Network*

V. KESIMPULAN

Telah dikaji mengenai metode *artificial neural network* dengan algoritma *backpropagation* untuk klasifikasi penilaian rating pada penjualan *blackmores* di tokopedia. Data yang digunakan terdiri dari 5 variabel, yaitu harga produk, stok produk, *official store*, *power badge*, dan rating yang terdiri dari 6 penilaian. Berdasarkan hasil yang diperoleh maka dapat disimpulkan bahwa penerapan *resampling* data untuk variabel rating menggunakan *random oversampling* berpengaruh pada hasil penelitian. Karena data yang tidak menerapkan *random oversampling* mengalami *overfitting* sedangkan data yang menerapkan *random oversampling* tidak mengalami *overfitting* pada grafik model *accuracy* dan grafik model *loss* yang diperoleh pada penelitian. Kemudian beberapa aspek yang digunakan pada penelitian ini seperti penentuan jumlah pembagian data *training*, pembagian data *testing*, penentuan *learning rate*, jumlah *epoch* dan jumlah *batch size* dapat mempengaruhi hasil penelitian. Pembagian data 90% untuk data *training* dan 10% data *testing* dapat menghasilkan nilai *accuracy* terbaik dibandingkan dengan persentase pembagian data lainnya. Penggunaan *hypertuning* dalam menentukan *learning rate*, jumlah *epoch* dan *batch size*, dapat memperoleh *learning rate*, *epoch* dan *batch size* terbaik secara berurutan yaitu 0,001, 50 dan 32. Maka dapat disimpulkan bahwa metode *artificial neural network* dengan algoritma *backpropagation* mampu melakukan klasifikasi penilaian rating pada penjualan *blackmores* di tokopedia dengan baik dengan mendapatkan tingkat *accuracy* sebesar 84%. Berdasarkan data yang tersedia, maka model dapat mengklasifikasi penilaian rating pada produk dengan baik.

DAFTAR PUSTAKA

- Aprizal, Y., Zainal, R.I., dan Afriyudi. 2019. Perbandingan Metode *Backpropagation* dan *Learning Vector Quantization* (LVQ) Dalam Menggali Potensi Mahasiswa Baru Di STMIK Palcomtech. *Jurnal Matrik*. **18**(2):294-301.
- Arifiyanti, A.A., Pradana, R.M., dan Novian, I.F. 2018. Klasifikasi Produk Retour dengan Algoritma Pohon Keputusan C4.5. *Jurnal IPTEK*. **22**(1):79-86.
- Bharathi. 2021. *Confusion Matrix for Multi-class Classification*. <https://www.analyticsvidhya.com/blog/2021/06/confusion-matrix-for-multi-class-classification/>. Diakses pada 19 Juni 2022.
- Bramer dan Max. 2007. *Principles of Data Mining*. London: Springer.
- Chaitra, P.C and Kumar, S. 2018. *A Review Of Multi-Class Classification Algorithms*. *International Journal of Pure and Applied Mathematics*. **118**(14):17-26.
- Deisenroth, M.P., Faisal, A.A., dan Ong, C.S. 2021. *Mathematics for Machine Learning*. Cambridge University Press, England.
- Gerry, Santoso, B.S., dan Tanjung, J.P. 2021. *Classification of Wheat Seeds Using Artificial Neural Network Backpropagation Algorithm*. *Journal of informatics and telecommunication engineering*. **4**(2):335-342.
- Hadianto, N., Novitasari, H.B., dan Rahmawati, A. 2019. Klasifikasi Peminjaman Nasabah Bank Menggunakan Metode *Artificial Neural Network*. *Jurnal PILAR Nusa Mandiri*. **5**(2):163-170.
- Khalusova, M. 2021. *Machine Learning Model Evaluation Metrics Part 2: Multi-class Classification*. <https://www.mariakhalusova.com/posts/2019-04-17-ml-model-evaluation-metrics-p2/>. Diakses pada 19 Juni 2022.
- Khatami, F.A., Irawan, B., dan Setianingsih, C. 2020. Analisis Sentimen Terhadap *Review Aplikasi Layanan E-Commerce* Menggunakan Metode *Convolutional Neural Network*. *E-proceeding of Engineering*. **7**(2):4559-4566.

- Kusumadewi, S. 2003. *Artificial Intelligence Teknik dan Aplikasi*. Yogyakarta: Graha Ilmu.
- Lewis, N.D. 2017. *Artificial Neural Network for Time Series Forecasting with R*. CrateSpace Independent Publishing Platform, US.
- Mahmood, A.M. 2015. *ClassImbalance Learning in Data Mining – A Survey*. *International Journal of Communication Technology for Social Networking Services*. 3(2) : 17-38.
- Mardi, Y. 2016. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*. 2(2):213-219.
- Maysanjaya, D.M.I. 2020. Klasifikasi Pneumonia Pada Citra X-rays Paru-paru dengan *Convolutional Neural Network*. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*. 9(2):190-195.
- Nasution, D.A., Khotimah, H.H., dan Chamidah, N. 2019. Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Journal of Computer Engineering System and Science*. 4(1):78-82.
- Prabowo, J.R., Santoso, R., dan Yasin, H. 2020. Implementasi Jaringan Syaraf Tiruan *Backpropagation* Dengan Algoritma *Conjugate Gradient* Untuk Klasifikasi Kondisi Rumah (Studi Kasus Di Kabupaten Cilacap Tahun 2018). *Jurnal Gaussian*. 9(1):41-49.
- Purnama, J.J., Rahayu, S., Nurdiani, S., Haryanti, T., dan Mayangky, N.A. 2020. Analisis Algoritma Klasifikasi *Neural Network* Untuk Diagnosis Penyakit Diabetes. *Indonesian Journal on Computer and Information Technology*. 5(1):1-7.
- Putri, A.W. 2021. Implementasi *Artificial Neural Network (ANN) Backpropagation* Untuk Klasifikasi Jenis Penyakit Pada Daun Tanaman Tomat. *Jurnal Ilmiah Matematika*. 9(2):344-350.
- Saranya, C dan Manikandan, G. 2013. *A Study On Normalization Techniques For Privacy Preserving Data Mining*. *International Journal of Engineering and Technology*. 5(3):2701-2704.
- Sharma, N., Bhandari, H.V., Yadav, N.S., dan Shroff, H.V.J. 2020. *Optimization of IDS using Filter-Based Feature Selection and Machine learning Algorithms*. *International Journal of Innovative Technology and Exploring Engineering*. 10(2):96-102.
- Sihombing, P.R dan Arsani, A.M. 2021. Perbandingan Metode *Machine learning* Dalam Klasifikasi Kemiskinan Di Indonesia Tahun 2018. *Jurnal Teknik Informatika*. 2(1):51-56.

- Sukarno, N.M, Wirawan, P.W, dan Adhy, S. 2014. Perancangan dan Implementasi Jaringan Syaraf Tiruan Backpropagation Untuk Mendiagnosa Penyakit Kulit. *Jurnal Masyarakat Informatika*. **5**(10):9-18.
- Susantia, E.Y dan Maiyana, E. 2021. Pemanfaatan ANN untuk Prediksi Penjualan *Online* Industri Rumahan selama Pandemi Covid-19. *Jurnal Sains dan Informatika*. **7**(1):1-7.
- Vijayalakshmi, V dan Venkatachalapathy, K. 2019. *Deep Neural Network for Multi-Class Prediction of Student Performance in Educational Data*. *International Journal of Recent Technology and Engineering (IJRTE)*. **8**(2):5073-5081.
- Zheng, Z., Cai, Y. dan Li, Y. 2015. *Oversampling method for imbalanced classification*. *Computing and Informatics*. **34**(5):1017–1037.