

**KLASIFIKASI TEKS ULASAN DENGAN METODE HYBRID
CONVOLUTIONAL NEURAL NETWORK (CNN) DAN
GATED RECURRENT UNIT (GRU)**

SKRIPSI



FIQIH FATHOR RACHIM

**PROGRAM STUDI S1 MATEMATIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**KLASIFIKASI TEKS ULASAN DENGAN METODE HYBRID
CONVOLUTIONAL NEURAL NETWORK (CNN) DAN
GATED RECURRENT UNIT (GRU)**

SKRIPSI



FIQIH FATHOR RACHIM

**PROGRAM STUDI S1 MATEMATIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA**

2022

**KLASIFIKASI TEKS ULASAN DENGAN METODE HYBRID
CONVOLUTIONAL NEURAL NETWORK (CNN) DAN
GATED RECURRENT UNIT (GRU)**

SKRIPSI

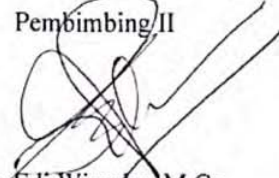
Sebagai Salah Satu Syarat Untuk Memperoleh
Gelar Sarjana Sains Bidang Matematika
pada Fakultas Sains dan Teknologi
Universitas Airlangga

Fiqih Fathor Rachim
NIM 081711233003

Disetujui Oleh :

Pembimbing I


Auli Damayanti, S.Si, M.Si
NIP. 197511072003122004

Pembimbing II


Drs. Edi Winarko, M.Cs.
NIP. 196705141993031001

LEMBAR PENGESAHAN NASKAH SKRIPSI

Judul : Klasifikasi Teks Ulasan dengan Metode *Hybrid Convolutional Neural Networks* (CNN) dan *Gated Recurrent Unit* (GRU)

Penyusun : Fiqih Fathor Rachim

NIM : 081711233003

Pembimbing I : Auli Damayanti, S.Si, M.Si.

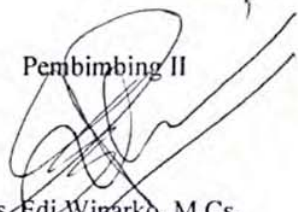
Pembimbing II : Drs. Edi Winarko, M.Cs.

Tanggal Seminar : 21 Juni 2022

Disetujui Oleh :

Pembimbing I

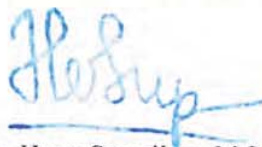

Auli Damayanti, S.Si, M.Si
NIP. 197511072003122004

Pembimbing II


Drs. Edi Winarko, M.Cs.
NIP. 196705141993031001

Mengetahui,

Ketua Departemen Matematika
Fakultas Sains dan Teknologi
Universitas Airlangga



Dr. Herry Suprajitno, M.Si.
NIP. 196804041994031020

PEDOMAN PENGGUNAAN SKRIPSI

Skripsi ini tidak dipublikasikan, namun tersedia di perpustakaan dalam lingkungan Universitas Airlangga, diperkenankan untuk dipakai sebagai referensi kepustakaan, tetapi pengutipan harus seizin penyusun dan harus menyebut sumbernya sesuai kebiasaan ilmiah. **Dokumen skripsi ini merupakan hak milik Universitas Airlangga.**

SURAT PERNYATAAN TENTANG ORISINALITAS

Yang bertanda tangan dibawah ini, saya :

Nama : Fiqih Fathor Rachim

NIM : 081711233003

Program Studi : S-1 Matematika

Fakultas : Sains dan Teknologi

Jenjang : Sarjana (S1)

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul :

“Klasifikasi Teks Ulasan dengan Metode Hybrid Convolutional Neural Networks (CNN) dan Gated Recurrent Unit (GRU)”

Apabila suatu saat terbukti melakukan tindakan plagiat, maka saya akan menerima sanksi yang telah ditetapkan.

Demikian surat ini saya buat dengan sebenar-benarnya.



Fiqih Fathor Rachim
NIM. 081711233003

KATA PENGANTAR

Puji syukur kepada Allah SWT atas berkah, rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan proposal skripsi yang berjudul “Klasifikasi Teks Ulasan dengan Metode *Hybrid Convolutional Neural Network* (CNN) dan *Gated Recurrent Unit* (GRU)” pada Program Studi S-1 Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga.

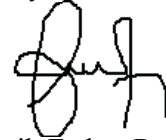
Selama penyusunan skripsi ini tidak sedikit penulis mendapat bantuan doa, dukungan, bimbingan, dan motivasi sehingga skripsi ini dapat terselesaikan dengan baik. Oleh karena itu penulis mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Prof. Dr. Mohammad Nasih SE., MT., AK., selaku rektor Universitas Airlangga
2. Prof. Dr. Moh Yasin, M.Si., selaku dekan Fakultas Sains dan Teknologi, Universitas Airlangga.
3. Dr. Herry Suprajitno, M.Si., selaku ketua departemen matematika, Fakultas Sains dan Teknologi, Universitas Airlangga sekaligus sebagai koordinator program studi S-1 matematika, Fakultas Sains dan Teknologi, Universitas Airlangga.
4. Dr. Liliek susilowati S.Si, M.Si., selaku dosen wali atas bimbingan, dan motivasi kepada penulis.
5. Auli damayanti, S.Si, M.Si., selaku dosen pembimbing I atas bimbingan dan saran yang diberikan kepada penulis.
6. Drs. Edi winarko, M.Cs., selaku dosen pembimbing II atas bimbingan dan saran yang diberikan kepada penulis.
7. Segenap dosen program studi S-1 matematika Universitas Airlangga atas segala ilmu, saran, serta motivasi kepada penulis.

8. Orang tua atas doa, bimbingan, kasih sayang serta motivasinya demi teraihnya kesuksesan penulis
9. Seluruh civitas akademika program studi S-1 matematika yang telah memberikan dukungan, serta saran yang telah diberikan
10. Serta seluruh pihak yang telah turut membantu penyusunan skripsi ini.

Penyusun menyadari bahwa skripsi ini masih banyak kekurangan dan kesalahan. Oleh karena itu, penulis mengharap segala kritik dan saran untuk perbaikan dan pengembangan kedepannya.

Surabaya, 12 Juni 2022



Fiqih Fathor Rachim

Fiqih Fathor Rachim, 2022, **Klasifikasi Teks Ulasan Dengan Metode *Hybrid Convolutional Neural Network (CNN)* Dan *Gated Recurrent Unit (GRU)***, Skripsi ini dibawah bimbingan Auli Damayanti, S.Si, M.Si dan Drs. Edi winarko, M.Cs. Departemen Matematika, Fakultas Sains dan Teknologi, Universitas Airlangga, Surabaya.

ABSTRAK

Ulasan konsumen merupakan pendapat dari pembeli kepada penjual berdasarkan kepuasan pelayanan atau kualitas barang. Semakin banyak ulasan konsumen menyebabkan proses menganalisa secara manual akan menyulitkan. Oleh karena itu, dibutuhkan sistem analisa sentimen secara otomatis. Setiap ulasan akan dikelompokkan kedalam kelas sentimen yang dibagi menjadi kelas positif dan negatif. Penelitian ini bertujuan untuk mengklasifikasi teks ulasan dengan metode *Convolutional Neural Network (CNN)* dan *Gated Recurrent Unit (GRU)*. Tahapan penelitian pada penelitian ini antara lain pengumpulan data teks ulasan Tokopedia, mengekstraksi informasi tersembunyi teks ulasan menggunakan CNN, melakukan pembelajaran pada teks ulasan menggunakan GRU. Sebanyak 1000 teks ulasan dibagi menjadi data latih sebesar 80% dan data uji sebesar 20%. Teks ulasan diubah menjadi matriks menggunakan algoritma *One Hot Encoding* kemudian diekstraksi menggunakan CNN. Proses CNN meliputi perhitungan *convolution*, perhitungan fungsi aktivasi *Rectified Linear Unit (ReLU)*, dan tahap *pooling*. Hasil ekstraksi pada proses CNN dilanjutkan pada proses GRU. Proses GRU meliputi menginisialisasi parameter, umpan maju GRU, perhitungan *Cross-Entropy Error*, umpan mundur GRU, dan memperbarui bobot dan bias. Bobot optimal diperoleh ketika nilai error pada pelatihan kurang dari minimal error yang diharapkan atau iterasi pelatihan telah mencapai maksimum iterasi yang ditentukan. Bobot optimal digunakan untuk uji validasi terhadap data uji. Implementasi klasifikasi teks ulasan dengan metode *hybrid Convolutional Neural Network (CNN)* dan *Gated Recurrent Unit (GRU)* dibuat program dengan bahasa pemrograman python. Akurasi yang diperoleh pada uji validasi sebesar 88.5%.

Kata Kunci: Klasifikasi Sentimen, Convolutional Neural network, Gated Recurrent Unit, Jaringan Syaraf tiruan

Fiqih Fathor Rachim, 2022, **Classification Of Review Text Using Hybrid Convolutional Neural Network and Gated Recurrent Unit Methods**, this undergraduate thesis is supervised by Auli Damayanti, S.Si, M.Si and Drs. Edi winarko, M.Cs. Mathematics Department, Faculty of Science and Technology, Airlangga University, Surabaya.

ABSTRACT

Consumer reviews are opinions from buyers to sellers based on service satisfaction and quality of purchased goods. The large number of consumer reviews make the manual analysis process will be difficult. Therefore, it needs a sentiment analysis system from consumer reviews automatically. Each review will be grouped into a sentiment class which is divided into positive and negative classes. This study aims to classify review texts using the Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) methods. The stages of this research are collecting data of review texts on Tokopedia application, extracting hidden information on review texts using CNN, learning of review texts using GRU. A total of 1000 review texts were divided into 80% for training data and 20% for test data. The review text is converted into matrix using One Hot Encoding algorithm and extracted using CNN. The CNN process includes the convolution calculation, calculation of the activation function of the Rectified Linear Unit (ReLU), and the pooling stage. The CNN extraction result were trained using the GRU. The GRU training phase includes initializing parameters, GRU feed forward, calculating Cross-Entropy Error, GRU feedback, and updating weights and biases. The optimal weight is obtained when the error value in the training is less than the expected minimum error or the training iteration has reached the specified maximum iteration. The optimal weight in the training process is used to validate the test data. The implementation of the review text classification using the hybrid Convolutional Neural Network (CNN) and Gated Recurrent Unit (GRU) method was made using the python programming language. The obtained accuracy of the validation test is 88.5%.

Keywords: Sentiment Classification, Convolutional Neural Network, Gated Recurrent Unit, Artificial Neural Network

DAFTAR ISI

LEMBAR JUDUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PEDOMAN PENGGUNAAN SKRIPSI.....	iii
LEMBAR PERNYATAAN	iv
KATA PENGANTAR	v
ABSTRAK	vii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan.....	3
1.4 Manfaat.....	3
1.5 Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Analisa Sentimen.....	5
2.2 Jaringan Syaraf Tiruan	6
2.3 <i>One Hot Encoding</i>	13
2.4 <i>Convolutional Neural Network</i>	13
2.5 <i>Gated Recurrent Unit</i>	16
2.6 <i>Cross-entropy Error</i>	17
2.7 <i>Backpropagation</i>	18
2.8 <i>Confusion Matrix</i>	19

2.9. Metode Inisialisasi Bobot.....	20
2.10. Bahasa Pemrograman Python	21
BAB III METODE PENELITIAN.....	23
BAB IV PEMBAHASAN.....	31
4.1 Prosedur Pelatihan <i>Hybrid Convolutional Neural Network</i> dan <i>Gated Recurrent Unit</i>	31
4.2 Data	41
4.3 Contoh Kasus Klasifikasi Sentimen Menggunakan <i>Hybrid Convolutional Neural Network</i> dan <i>Gated Recurrent Unit</i> Secara Manual.....	41
4.4. Implementasi Program Pada Kasus Klasifikasi Sentimen dengan <i>Hybrid Convolutional Neural Network</i> dan <i>Gated Recurrent Unit</i>	66
BAB V KESIMPULAN DAN SARAN.....	69
5.1 Kesimpulan.....	69
5.2 Saran.....	70
DAFTAR PUSTAKA	71
LAMPIRAN	74

DAFTAR TABEL

Nomor	Judul Tabel	Halaman
2. 1	<i>Confusion Matrix</i>	20
3. 1	Karakter yang digunakan pada pelatihan	24
4. 1	Contoh data pelatihan	42
4. 2	Indeks Karakter	43
4. 3	Representasi Matriks untuk Teks Ulasan 1	43
4. 4	Representasi Matriks Teks Ulasan	44
4. 5	Inisialisasi Kernel	45
4. 6	Hasil <i>Convolution</i>	48
4. 7	Hasil ReLU untuk Data ke-1	49
4. 8	Hasil ReLU untuk Setiap Data	49
4. 9	Hasil <i>Pooling</i> untuk Data ke-1	50
4. 10	Hasil <i>Pooling</i>	50
4. 11	Inisialisasi parameter <i>Gated Recurrent Unit</i> (GRU)	51
4. 12	Hasil Pencarian Nilai Terbesar dan Terkecil	52
4. 13	Hasil Normalisasi Data	52
4. 14	Hasil Pelatihan Iterasi ke-1 <i>Gated Recurrent Unit</i> (GRU)	63
4. 15	Hasil Pelatihan CNN-GRU	67
4. 16	Hasil Uji Validasi Terhadap Data Uji	68

DAFTAR GAMBAR

Nomor	Judul Gambar	Halaman
2.1	Struktur Jaringan Syaraf Tiruan	6
2.2	Prinsip dasar Jaringan Syaraf Tiruan	7
2.3	Fungsi aktivasi ReLU	8
2.4	Fungsi aktivasi <i>Sigmoid</i>	9
2.5	Fungsi aktivasi Tanh	9
2.6	Arsitektur jaringan syaraf tiruan dengan lapisan tunggal	11
2.7	Arsitektur jaringan syaraf tiruan dengan lapisan ganda	11
2.8	Arsitektur jaringan syaraf tiruan dengan lapisan kompetitif	12
2.9	Arsitektur Jaringan syaraf tiruan dengan lapisan berulang	12
2.10	Arsitektur <i>Convolutional Neural Network</i>	13
2.11	Skema pada lapisan <i>convolution</i>	15
2.12	Contoh pengambilan vektor <i>max pooling</i>	15
2.13	Arsitektur <i>Gated Recurrent Unit</i>	17
3.1	Flowchart pembentukan matriks representasi	27
3.2	Flowchart proses pelatihan data	28
3.3	Flowchart Prosedur <i>Convolution Neural Network</i>	29
3.4	Flowchart Prosedur <i>Gated Recurrent Unit</i>	30
4. 1	Prosedur Pelatihan CNN dan GRU	32
4. 2	Prosedur Penyusunan Masukan dan Keluaran Pelatihan	33
4. 3	Prosedur Inisialisasi Parameter CNN	34
4. 4	Prosedur Umpan Maju CNN	35
4. 5	Prosedur Tahap <i>Convolution</i>	36
4. 6	Prosedur Tahap ReLU	36
4. 7	Prosedur Tahap <i>Pooling</i>	37
4. 8	Prosedur Inisialisasi Parameter Lapisan <i>Gated Recurrent Unit</i>	37
4. 9	Prosedur Normalisasi Data	38
4. 10	Prosedur Umpan Maju <i>Gated Recurrent Unit</i>	39
4. 11	Prosedur Perhitungan <i>Cross-entropy Error</i>	39

4. 12	Prosedur Umpan Mundur <i>Gated Recurrent Unit</i>	40
4. 13	Prosedur Memperbarui Bobot dan Bias	41

DAFTAR LAMPIRAN

Nomor	Judul Lampiran
1.	Sample Data Latih
2.	Sample Data Uji
3.	<i>Source Code</i> Program
4.	Tampilan Menu Developer
5.	Tampilan Menu Bantuan
6.	Tampilan Menu Ekstraksi CNN
7.	Tampilan Menu Pelatihan GRU
8.	Tampilan Menu Aplikasi Klasifikasi
9.	Tampilan Data Latih
10.	Tampilan Data Uji
11.	Tampilan Bobot CNN
12.	Tampilan Representasi Matriks Hasil CNN
13.	Tampilan Bobot Optimal GRU
14.	Tampilan Test Data Uji