

**KLASIKASI KATEGORI BERITA MENGGUNAKAN
XLNET PRETRAINED MODEL**

SKRIPSI



disusun oleh

Faizal Fakhri Irfani

17.11.1500

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**KLASIKASI KATEGORI BERITA MENGGUNAKAN
XLNET PRETRAINED MODEL**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Falzal Fakhri Irfani

17.11.1500

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

**KLASIKASI KATEGORI BERITA MENGGUNAKAN
XLNET PRETRAINED MODEL**

yang dipersiapkan dan disusun oleh

Fatzal Fakhri Irfani

17.11.1500

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 7 Juli 2021

Dosen Pembimbing,

Mardhiya Hayaty, S.T., M.Kom
NIK. 190302108

PENGESAHAN

SKRIPSI

**KLASIKASI KATEGORI BERITA MENGGUNAKAN
XLNET PRETRAINED MODEL**

yang dipersiapkan dan disusun oleh

Falzal Fakhri Irfani

17.11.1500

telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Juli 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Balta, M.Kom

NIK. 190302290

Norhikmah, M.Kom

NIK. 190302245

Mardhiya Havaty, S.T., M.Kom.

NIK. 190302108

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Juli 2021

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta,S.Kom., M.Kom.

NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terlapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terlapat karya atau pendapat yang pernah ditulis dan atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 7 Juli 2021



Faizul Fakhri Irfani
NIM. 17.11.1500

PERSEMBAHAN

Dengan mengucapkan syukur Alhamdulillah saya persembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung dalam proses pembuatan skripsi.

1. Kedua orang tua, kakak dan adik saya, yang selalu mendoakan dan memberikan semangat serta motivasi tiada henti.
2. Dosen pembimbing saya Ibu Mardhiya Hayaty, S.T., M.Kom. yang telah membimbing saya dari awal sampai akhir pembuatan skripsi
3. Dosen-dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak ilmu selama kuliah.
4. Teman-teman satu perbimbingan yang memberikan dukungan dan motivasi selama proses pengerjaan skripsi.
5. Teman-teman kelas 17-IF-09 yang telah menemani dan selalu memberikan semangat untuk menyelesaikan skripsi.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shawat serta salam juga tidak lupa penulis panjatkan kepada junjungan kita Nabi Muhammad SAW yang telah memberikan teladan mulia dalam menuntun umatnya sehingga penulis dapat menyelesaikan skripsi dengan judul **“KLASIKASI KATEGORI BERITA MENGGUNAKAN XLNET PRETRAINED MODEL”** dengan sebaik – baiknya sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta

Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

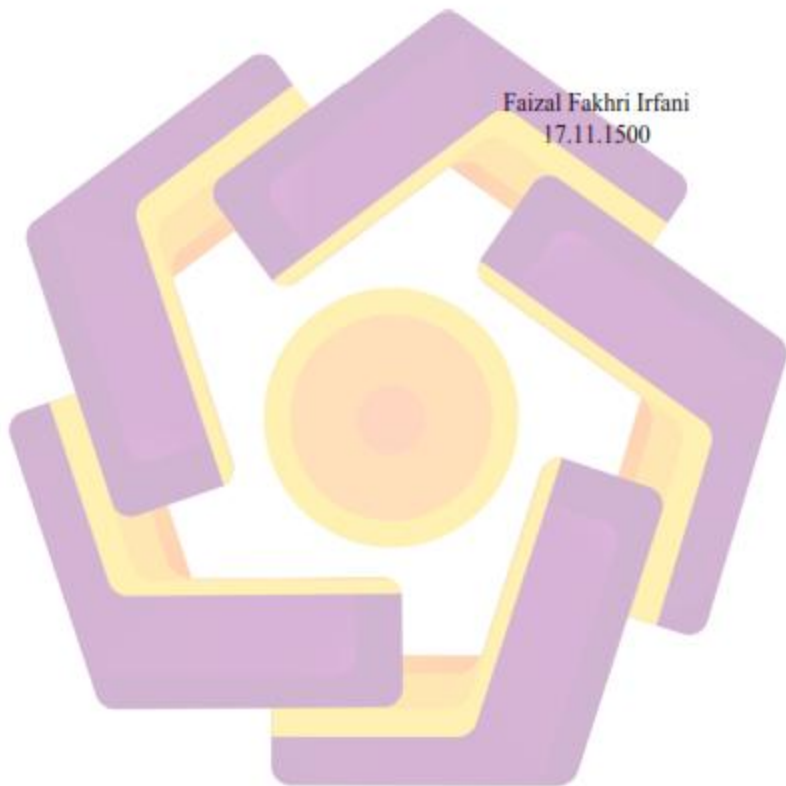
1. Prof. Dr. M. Suyanto, MM selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Mardhiya Hayaty, S.T., M.Kom. selaku dosen pembimbing yang telah banyak memberikan nasehat dan arahan bagi penulis agar penulis menyelesaikan skripsi ini dengan baik.
3. Para Dosen dan Staff Universitas AMIKOM Yogyakarta yang telah banyak memberikan ilmu pengetahuan, pengalaman, dan bantuannya selama penulis kuliah hingga terselesaikannya skripsi ini.
4. Keluarga dan teman-teman yang telah memberikan dukungan dan do'a kepada penulis.

Penulis menyadari masih banyak kekurangan yang masih harus diperbaiki dari penulisan skripsi ini. Penulis menerima kritik dan saran yang membangun

untuk kesempurnaan untuk kesempurnaan skripsi ini. Semoga tulisan ini bisa bermanfaat dan mendorong kita untuk melakukan penelitian yang lebih baik.

Yogyakarta, 7 Juli 2021


Faizal Fakhri Irfani
17.11.1500



DAFTAR ISI

JUDUL	i
PERSETUJUAN.....	ii
PENGESAHAN	iii
PERNYATAAN.....	iv
MOTTO	iv
PERSEMBAHAN.....	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Maksud dan Tujuan Penelitian.....	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI.....	5
2.1 Tinjauan Pustaka	5
2.2 <i>Natural Language Processing</i>	7
2.3 <i>Data Mining</i>	8

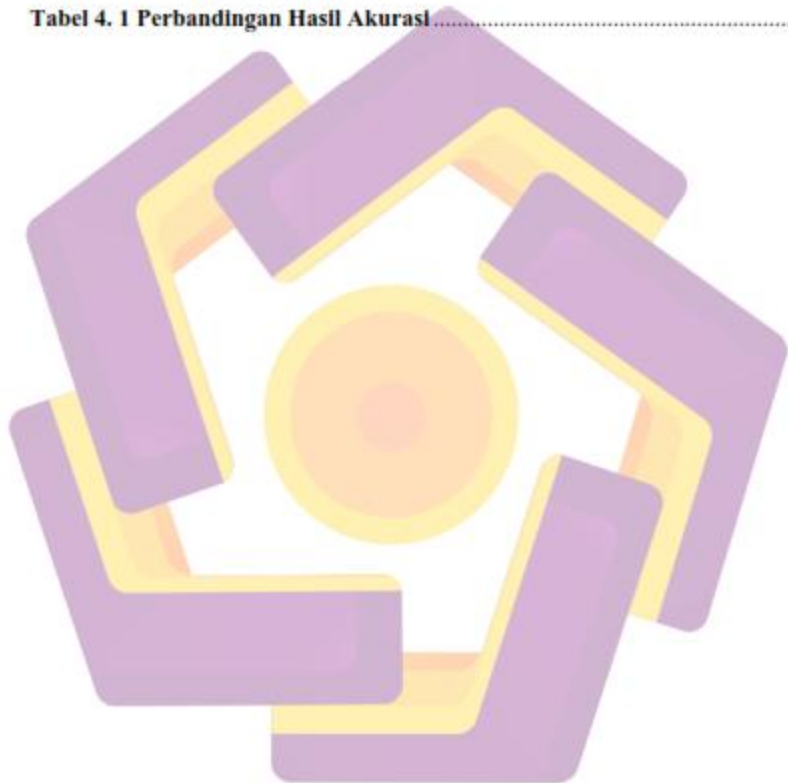
2.4	<i>Text Mining</i>	8
2.5	<i>Text Classification</i>	9
2.6	<i>Deep Learning</i>	9
2.7	<i>Web Scraping</i>	11
2.8	<i>Preprocessing</i>	11
2.9	<i>Word Embedding</i>	12
2.10	XLNet	13
2.10.1	<i>Auto-regressive Language Model</i>	13
2.10.2	<i>Permutation Language Modeling</i>	14
2.10.3	<i>Two-Stream Self-Attention</i>	16
2.10.4	<i>Partial Prediction</i>	19
2.10.5	Integrasi Transformer-XL	20
2.10.6	<i>Pretraining XLNet</i>	21
2.10.7	Arsitektur Model	21
2.10.8	<i>Fine-tuning XLNet</i>	21
2.11	<i>AdaMax Optimizer</i>	22
2.12	<i>Epoch, Batch Size dan Learning Rate</i>	23
2.13	<i>Logistic Regression</i>	23
2.14	<i>Confusion Matrix</i>	24
BAB III METODE PENELITIAN		26
3.1	Tahapan Penelitian	26
3.2	Alat Penelitian	26
3.3	Penentuan Kategori dan Portal Berita	27
3.4	Pengumpulan Data	27
3.5	<i>Preprocessing Data</i>	27



3.6	Pembagian <i>Dataset</i>	28
3.7	Melakukan <i>Training</i>	28
3.8	Pengujian	28
3.9	Penghitungan Akurasi dan Evaluasi	29
BAB IV HASIL DAN PEMBAHASAN		30
4.1	Penentuan Kategori dan Portal Berita	30
4.2	Pengumpulan <i>Data</i>	30
4.3	Melakukan <i>Preprocessing Data</i>	33
4.4	Melakukan Pembagian <i>Dataset</i>	37
4.5	Melakukan <i>Training</i>	37
4.6	Melakukan <i>Testing</i>	43
4.7	Menghitung Akurasi dan Evaluasi	45
BAB V PENUTUP		48
5.1	Kesimpulan	48
5.2	Saran	48
DAFTAR PUSTAKA		49

DAFTAR TABEL

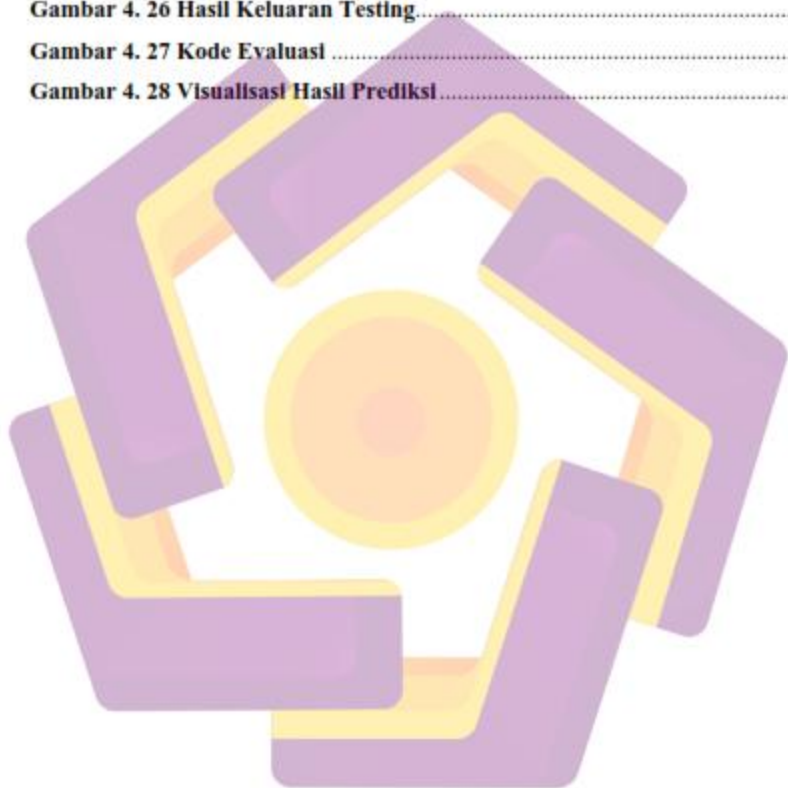
Tabel 2. 1 Tabel Perbandingan Penelitian	6
Tabel 2. 2 Tabel Arsitektur Model XLNet	21
Tabel 2. 3 Tabel Confusion Matrix	25
Tabel 4. 1 Perbandingan Hasil Akurasi	47



DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Deep Learning.....	11
Gambar 2. 2 Ilustrasi Auto-regressive Language Model	14
Gambar 2. 3 Ilustrasi Permutation Language Modeling	15
Gambar 2. 4 Daftar Permutasi	15
Gambar 2. 5 Ilustrasi pola permutasi X3.....	16
Gambar 2. 6 Ilustrasi representasi konten	17
Gambar 2. 7 Ilustrasi representasi kueri	18
Gambar 2. 8 Ilustrasi Two-Stream Self-Attention	18
Gambar 3. 1 Flowchart Tahapan Penelitian	26
Gambar 4. 1 Kode Generate Url.....	31
Gambar 4. 2 Function untuk mendapatkan judul dan isi berita.....	31
Gambar 4. 3 Kode untuk mendapatkan semua isi dari berita.....	31
Gambar 4. 4 Kode untuk pelabelan kategori	32
Gambar 4. 5 Isi Dataset	32
Gambar 4. 6 Kode Konversi Label.....	33
Gambar 4. 7 Kode Tokenizer XLNet	33
Gambar 4. 8 Kode Percobaan Preprocessing	34
Gambar 4. 9 Hasil Preprocessing	34
Gambar 4. 10 Kode Preprocessing Content	35
Gambar 4. 11 Input Ids	36
Gambar 4. 12 Attention Mask	36
Gambar 4. 13 Kode Split Dataset	37
Gambar 4. 14 Kode Konversi Data ke Format Torch Tensor	37
Gambar 4. 15 Kode Inisialisasi Dataloader	38
Gambar 4. 16 Kode Impor Model XLNet	38
Gambar 4. 17 Kode Optimizer Adamax	38
Gambar 4. 18 Kode Inisialisasi Scheduler	39
Gambar 4. 19 Kode Deklarasi Class flat_accuracy	39
Gambar 4. 20 Kode Training	40

Gambar 4. 21 Kode Valldasi	41
Gambar 4. 22 Hasil Keluran Training dan Validation.....	42
Gambar 4. 23 Visuallsasi Model Loss	42
Gambar 4. 24 Visuallsasi Model Accuracy.....	43
Gambar 4. 25 Kode Testing	44
Gambar 4. 26 Hasil Keluaran Testing.....	44
Gambar 4. 27 Kode Evaluasi	45
Gambar 4. 28 Visuallsasi Hasil Prediksi.....	46



INTISARI

Klasifikasi kategori berita adalah tugas mengategorikan artikel berita ke dalam suatu kategori. Pengelompokan berita umumnya dilakukan secara manual. Pengelompokan berita ini akan lebih efisien apabila pengklasifikasian berita diproses secara otomatis dengan komputer menggunakan metode tertentu hingga mengeluarkan hasil kategori terkait berita yang sudah dimasukkan. Salah satu cara yang dapat mengatasi permasalahan tersebut menggunakan teknologi pendekatan *deep learning*.

Dalam penelitian ini menggunakan menggunakan pendekatan *deep learning* dengan arsitektur XLNet sebagai metode untuk *word embedding* dan *logistic regression* untuk memprediksi probabilitas kategori dari artikel berita. Penelitian ini menggunakan konsep *transfer learning* dengan menggunakan model XLNet yang sudah dilatih yang kemudian dikombinasikan dengan dataset baru untuk keperluan klasifikasi kategori berita. Untuk mendapatkan model yang optimal dilakukan *tuning hyperparameter* seperti *learning rate* dan penggunaan beberapa *optimizer*.

Dalam Penelitian ini menggunakan dataset sebesar 10000 artikel berita dengan pembagian data latih sebesar 90%, data validasi sebesar 15% dari data latih dan data uji sebesar 10%. Hasil penelitian menunjukkan bahwa klasifikasi kategori berita menggunakan XLNet memperoleh akurasi tertinggi sebesar 97.6%.

Kata Kunci: klasifikasi, *deep learning*, XLNet, *transfer learning*, akurasi

ABSTRACT

News category classification is the task of categorizing news articles into a category. News grouping is generally done manually. This news grouping will be more efficient if the news classification is processed automatically with a computer using certain methods to issue category results related to the news that has been entered.

One way to solve this problem is to use deep learning approach technology. This study uses a deep learning approach with XLNet architecture as a method for word embedding and logistic regression to predict the category probability of news articles. This study uses the concept of transfer learning by using the XLNet model that has been trained which is then combined with a new dataset for the purpose of classifying news categories. To get the optimal model, hyperparameter tuning is done such as learning rate and the use of several optimizers.

In this study using a dataset of 10000 news articles with a distribution of training data by 90%, validation data by 15% from training data and test data by 10%. The results showed that the classification of news categories using XLNet obtained the highest accuracy of 97.6%.

Keywords: *classification, deep learning, XLNet, transfer learning, accuracy*