



**PENERAPAN ALGORITMA NAÏVE BAYES DAN LOGISTIC
REGRESSION DALAM KLASIFIKASI DATA DEWEY DECIMAL
CLASIFICATION (DDC) INDONESIA**

LAPORAN PENELITIAN

**SELVI ANGRAINI
191410170**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA
PALEMBANG
2023**



**PENERAPAN ALGORITMA NAÏVE BAYES DAN LOGISTIC
REGRESSION DALAM KLASIFIKASI DATA DEWEY DECIMAL
CLASIFICATION (DDC) INDONESIA**

**Laporan Penelitian Ini Diajukan Sebagai Syarat Memperoleh Gelar
Sarjana Komputer**

**SELVI ANGRAINI
191410170**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS SAINS TEKNOLOGI
UNIVERSITAS BINA DARMA
PALEMBANG
2023**

HALAMAN PENGESAHAN

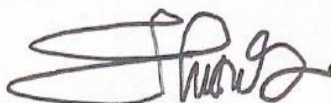
PENERAPAN ALGORITMA NAÏVE BAYES DAN LOGISTIC
REGRESSION DALAM KLASIFIKASI DATA DEWEY DECIMAL
CLASIFICATION (DDC) INDONESIA

SELVI ANGRAINI
191410170

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi

Pembimbing

Palembang, 09 September 2023
Fakultas Sains Teknologi
Universitas Bina Darma
Dekan,



Ilman Zuhri Yadi, M.M., M.Kom.




Universitas Bina Darma
Fakultas Sains Teknologi

Dr. Tata Sutabri, S.Kom., MMSI, MKM.

HALAMAN PERSETUJUAN

Penelitian Berjudul "PENERAPAN ALGORITMA NAÏVE BAYES DAN LOGISTIC REGRESSION DALAM KLASIFIKASI DATA DEWEY DECIMAL CLASSIFICATION (DDC) INDONESIA" Oleh SELVI ANGRAINI telah dipertahankan didepan komisi penguji pada hari Sabtu tanggal 09 SEPTEMBER 2023.

Komisi Penguji

1. Ketua : Ilman Zuhri Yadi, M.M., M.Kom.


(.....)

2. Anggota : Mohamad Farozi, M.Kom


(.....)

3. Anggota : Irman Effendy, M.Kom.


(.....)

Mengetahui,

Program Studi Sistem Informasi

Universitas Bina Darma

Ketua,


Universitas Bina Darma
Fakultas Sains dan Teknologi

Nita Rosa Damayanti, M.Kom., Ph. D.

SURAT PERNYATAAN

Saya yang bertanda tangan bawah ini :

Nama : SELVI ANGRAINI

Nim : 191410170

Dengan ini menyatakan bahwa :

1. Karya tulis saya (Riset) adalah asli dan belum pernah diajukan untuk mendapat gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses secara daring;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpanan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang- undangan yang berlaku ;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, 09 September 2023

Yang membuat pernyataan,



SELVI ANGRAINI
191410157

ABSTRAK

Dalam Perpustakaan, terdapat istilah kode Dewey Decimal Classification (DDC) yaitu berupa salah satu sistem klasifikasi buku yang digunakan dengan menerapkan kode-kode DDC pada buku-buku yang akan diklasifikasi. Pada penelitian ini, klasifikasi DDC yang dilakukan menggunakan algoritma Naïve Bayes dan Logistic Regression. Namun, dalam penelitian ini hanya dilakukan pengujian dan pelatihan kedua algoritma tersebut untuk mengetahui sejauh mana kemampuan algoritma naïve bayes dan logistic regression dalam melakukan klasifikasi DDC. Penelitian ini melakukan proses klasifikasi berdasarkan judul buku, kata kunci dan nama penulis buku sehingga menghasilkan nilai akurasi klasifikasi. Model Naïve Bayes ini menghasilkan nilai akurasi yang cukup baik dari hasil pengujian berdasarkan “Judul buku dan keyword” dari masing-masing kode pengujian. Pada ke-4 dataset yang diuji berdasarkan “Judul buku + Keyword”, kode pengujian yang menghasilkan nilai akurasi tertinggi adalah Naïve Bayes + CV + TFIDF pada dataset keempat dengan nilai precision 0.78, recall 0.77, F1-score 0,76 dan akurasi 0.77.

Kata Kunci: Naïve Bayes, Sistem Klasifikasi, DDC

KATA PENGANTAR

Puji syukur senantiasa kita haturkan kepada Allah SWT berkat kehadiran, rahmatnya, hidayahnya dan karunianya kita dapat menyelesaikan laporan penelitian yang berjudul “Penerapan Algoritma Naïve Bayes dan Logistic Regression dalam Klasifikasi Data *Dewey Decimal Classification* (DDC) Indonesia”. Laporan penelitian ini disusun untuk memenuhi salah satu syarat memperoleh gelar sarjana (S1) Program Studi Sistem Informasi, di Fakultas Sains Teknologi Universitas Bina Darma Palembang.

Dengan segala hormat dan kesadaran hati penulis mengucapkan terima kasih yang sebesar-besarnya dan penghargaan yang tulus penulis sampaikan kepada kedua orang tua tercinta, yaitu Ayahanda H. Umiri dan Ibunda Hj Hotamun Aryawati yang telah menjaga, mendidik, mengajari, merawat, menesehati, memberi semangat, dan tidak pernah lelah berdoa demi anaknya. Melalui kesempatan ini, penulis juga ingin mengucapkan rasa terima kasih yang tulus kepada segala rekan yang sudah ikut membantu dan memberikan dukungan dalam penyelesaian laporan penelitian ini. Rasa terima kasih penulis haturkan kepada:

1. Prof. Dr. Sunda Ariana, M.Pd., M.M selaku Rektor Universitas Bina Darma Palembang.
2. Dr. Tata Sutabri, S.Kom., MMSI., MKM. selaku Dekan Fakultas Sains Teknologi.
3. Nita Rosa Damayanti, M.Kom., Ph.D selaku Ketua Program Studi Sistem Informasi.
4. Bapak Ilman Zuhri Yadi, M.M., M.Kom. dan Ibu Dr. Yesi Novaria Kunang, S.T., M.Kom. selaku pembimbing saya yang senantiasa membimbing, mendukung dan memberikan pengetahuan untuk belajar menjadi lebih baik.
5. Bapak Mohamad Farozi, M.Kom. dan Bapak Irman Effendy, M.Kom sebagai Dosen penguji.
6. Semua Dosen pengajar Program Studi Sistem Informasi yang telah memberikan ilmu, pengalaman dan wawasan selama saya menempuh studi di universitas ini.

7. Kakakku tersayang (Yuli Karyawati, Marlana, Ria Agustina) keponakan tersayangku (Dzakia Marwah, M. Fatria Agam, Dzakira Aptani, Sahir Romadon serta keluarga besarku) terima kasih untuk kasih sayang, motivasi, dukungan, perhatian dan doa yang selalu dipanjatkan selama ini untuk keberhasilanku.
8. Untuk keluarga perpustakaan Universitas Bina Darma terima kasih untuk semua mendukung, nasehat, motivasi, support dan doa yang selalu mendorong penulis selama menempuh penelitian di perpustakaan ini.
9. Untuk Ceker Squad (Dio, Subhan, Irfan, Aji, Rizki, Rido, Nia, Tia, Wulan, Winda) dan TeamGirls (Chairani, Sherly, Kusuma) terima kasih sudah menjadi teman terbaik yang selalu memberikan dukungan dan selalu ada disetiap proses penulis lalui.
10. Semua pihak yang telah membantu penulis selama penelitian dan penulisan skripsi ini dan tidak dapat disebutkan penulis satu persatu hanya ucapan terima kasih yang dapat penulis berikan.

Laporan penelitian ini masih jauh dari kata sempurna, namun penulis berharap Semoga laporan penelitian ini dapat memberikan inspirasi dan sumbangan pemikiran bagi para pembaca yang ingin mengeksplorasi lebih jauh mengenai topik yang sama. Akhir kata, penulis menyadari bahwa masih banyak kekurangan dalam laporan penelitian ini. Kritik dan saran yang membangun sangat kami harapkan guna perbaikan dan pengembangan penelitian di masa yang akan datang. Semoga Allah SWT senantiasa meridhai segala upaya dan hasil dari penulisan laporan penelitian ini ini. Aamiin

Palembang, 31 Agustus 2023

Penulis
Selvi Angraini

DAFTAR ISI

COVER LUAR.....	i
COVER SAMPUL	ii
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 <i>Latar Belakang</i>	1
1.2 <i>Rumusan Masalah</i>	3
1.3 <i>Tujuan Penelitian</i>	3
1.4 <i>Batasan Masalah</i>	3
1.5 <i>Manfaat Penelitian</i>	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 <i>Landasan Teori</i>	4
2.1.1 <i>Sistem Klasifikasi</i>	4
2.1.2 <i>Dewey Decimal Clasification (DDC)</i>	4
2.1.3 <i>Text Mining</i>	6
2.1.4 <i>Naïve Bayes</i>	6
2.1.5 <i>Machine Learning</i>	7
2.1.6 <i>Precision, Recall dan F1-score dan Accuracy</i>	8
2.1.7 <i>Logistic Regression dan Count Vectorize</i>	9
2.2 <i>Penelitian Sebelumnya</i>	9
BAB III METODOLOGI PENELITIAN.....	11
3.1 <i>Flowchart Alur Penelitian</i>	11
3.2 <i>Pengumpulan Data</i>	12
3.3 <i>Pre-processing</i>	12
3.4 <i>Pembagian Data</i>	12
3.5 <i>Data Tranning</i>	12
3.6 <i>Pelatihan model Naïve Bayes</i>	12
3.7 <i>Data Testing</i>	12

3.8	<i>Evaluasi model</i>	13
BAB IV		14
4.1	<i>Pengumpulan Data</i>	14
4.1.1	<i>Dataset Pertama</i>	14
4.1.2	<i>Dataset Kedua</i>	15
4.1.3	<i>Dataset Ketiga</i>	15
4.1.4	<i>Dataset Keempat</i>	16
4.2	<i>Pre-processing</i>	16
4.3	<i>Pembagian Data</i>	25
4.4	<i>Evaluasi Model</i>	25
4.4.2	<i>Test 2 berdasarkan Judul buku dan keyword</i>	27
4.4.3	<i>Test 3 berdasarkan Judul buku dan nama penulis</i>	28
4.5	<i>Hasil Evaluasi</i>	30
BAB V PENUTUP		32
5.1	<i>Kesimpulan</i>	32
5.2	<i>Saran</i>	32
DAFTAR PUSTAKA		33
LAMPIRAN		36

DAFTAR GAMBAR

Gambar 3.1 Alur Metodologi Penelitian.....	11
Gambar 4.1 Proses Case folding Dataset pertama.....	17
Gambar 4.2 Proses case folding Dataset 2.....	17
Gambar 4.3 Proses case folding Dataset 3.....	18
Gambar 4.4 Proses case folding Dataset 4.....	18
Gambar 4.5 Proses Normalize Dataset pertama.....	19
Gambar 4.6 Proses Normalize Dataset kedua.....	19
Gambar 4.7 Proses Normalize Dataset ketiga.....	20
Gambar 4.8 Proses Normalize Dataset keempat.....	20
Gambar 4.9 Proses Stopword Dataset pertama.....	21
Gambar 4.10 Proses Stopword Dataset kedua.....	21
Gambar 4.11 Proses Stopword Dataset ketiga.....	22
Gambar 4.12 Proses Stopword Dataset keempat.....	22
Gambar 4.13 Proses Stemming Dataset pertama.....	23
Gambar 4.14 Proses Stemming Dataset kedua.....	23
Gambar 4.15 Proses Stemming Dataset ketiga.....	24
Gambar 4.16 Proses Stemming Dataset keempat.....	24

DAFTAR TABEL

Tabel 2.1 Kode Dewey Decimal Clasification.....	5
Tabel 2.2 Contoh penomoran kode DDC.....	6
Tabel 4.1 Dataset pertama.....	14
Tabel 4.2 Dataset kedua.....	15
Tabel 4.3 Dataset ketiga.....	15
Tabel 4.4 Dataset keempat.....	16
Tabel 4.5 Test Dataset 1 berdasarkan Judul buku.....	26
Tabel 4.6 Test Dataset 2 berdasarkan Judul buku.....	26
Tabel 4.7 Test Dataset 3 berdasarkan Judul buku.....	26
Tabel 4.8 Test Dataset 4 berdasarkan Judul buku.....	26
Tabel 4.9 Test Dataset 1 Judul dan keword.....	27
Tabel 4.10 Test Dataset 2 Judul+keyword.....	27
Tabel 4.11 Test Dataset 3 Judul+keyword.....	28
Tabel 4.12 Test Dataset 4 Judul+keyword.....	28
Tabel 4.13 Test Dataset 1 Judul buku dan nama penulis.....	29
Tabel 4.14 Test Dataset 2 Judul buku dan nama penulis.....	29
Tabel 4.15 Test Dataset 3 Judul buku dan nama penluis.....	29
Tabel 4.16 Test Dataset 4 Judul buku dan nama penulis.....	29
Tabel 4.17 Hasil tertinggi berdasarkan judul buku + keyword.....	30

DAFTAR LAMPIRAN

LEMBAR KONSULTAS.....	36
SK PEMBIMBING.....	37
LEMBAR REVISI.....	38
LEMBAR PENGAJUAN JUDUL.....	39
LEMBAR VALIDASI JILID.....	40

