



**PENGEMBANGAN APLIKASI & PENGUMPULAN BASIS DATA
TEKS DENGAN MONGODB UNTUK ANALISIS *BIG DATA*
ARTIKEL BERITA *ONLINE***

SKRIPSI

**ALDO JULIANDRA
161410326**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOPMUTER
UNIVERSITA BINADARMA
PALEMBANG
2020**



**PENGEMBANGAN APLIKASI & PENGUMPULAN BASIS DATA
TEKS DENGAN MONGODB UNTUK ANALISIS *BIG DATA*
ARTIKEL BERITA *ONLINE***

**ALDO JULIANDRA
161410326**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana
Komputer**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOPMUTER
UNIVERSITA BINADARMA
PALEMBANG
2020**

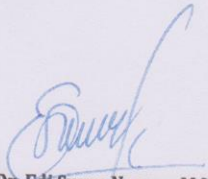
HALAMAN PENGESAHAN

**PENGEMBANGAN APLIKASI & PENGUMPULAN BASIS DATA
TEKS DENGAN MONGODB UNTUK ANALISIS BIG DATA
ARTIKEL BERITA ONLINE**

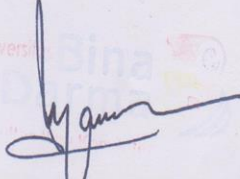
**ALDO JULIANDRA
161410326**

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

Pembimbing


Dr. Edi Surya Negara, M.Kom.

Palembang, 16 September 2020
Fakultas Ilmu Komputer
Universitas Bina Darma
Dekan,


Dedy Syamsuar, S.Kom., M.I.T., Ph.D.

HALAMAN PERSETUJUAN

Skripsi Berjudul "PENGEMBANGAN APLIKASI & PENGUMPULAN BASIS DATA TEKS DENGAN MONGODB UNTUK ANALISIS BIG DATA ARTIKEL BERITA ONLINE" Oleh "Aldo Juliandra", telah dipertahankan di depan komisi penguji pada hari Rabu tanggal 16 September 2020.

Komisi Penguji

1. Ketua : Dr. Edi Surya Negara, M.Kom.
2. Anggota : Dedy Syamsuar, S.Kom., M.I.T., Ph.D.
3. Anggota : Ilman Zuhri Yadi, M.M., M.Kom.



Mengetahui,
Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Bina Darma
Ketua,

Universitas Bina Darma
Fakultas Ilmu Komputer

Dr. Edi Surya Negara, M.Kom.

SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Aldo Juliandra

NIM : 161410326

Dengan ini menyatakan bahwa :

1. Karya tulis saya (Skripsi) adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma ataupun perguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftarrujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiaris checker serta di unggah ke internet, sehingga dapat diakses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan 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.

Pelambang, Agustus 2020

Yang membuat pernyataan,



Aldo Juliandra

161410326

MOTTO DAN PERSEMBAHAN

MOTTO :

Terus Mencoba dan Pantang Menyerah.

PERSEMBAHAN :

Alhamdulillah Atas Rahmat dan Hidayah-Nya Allah SWT Akhirnya saya mampu menyelesaikan skripsi ini dengan baik dan benar. Karya Sederhana ini Kupersembahkan kepada :

- Allah SWT.
- Kedua Orang Tua Tercinta, Terimakasih atas semangat dan support yang diberikan, serta tak pernah putus mendoakan anaknya untuk sukses.
- Pembimbing Bapak Dr. Edi Surya Negara, M.Kom.
- Teman-teman seangkatan Pejuang Skripsi.
- Rekan-Rekan Perjuangan Satria Biruku "HIMSIF".
- Rekan-Rekan Bina Darma Programmer.
- Almamater Universitas Bina Darma.

ABSTRAK

Perkembangan teknologi saat ini menjadikan data menjadi sangat berharga karena semakin hari data mengalami pertumbuhan yang sangat pesat, data memiliki bentuk yang kompleks seperti kata, huruf, angka, grafik tabel, gambar, serta lambang. Semakin banyaknya data yang diperoleh dari berbagai sumber informasi yang ada setiap harinya menjadikan semakin banyaknya data yang telah dihasilkan fenomena ini disebut dengan *Big Data*. Permasalahan yang biasa di temukan yaitu banyaknya data yang di hasilkan oleh berbagi sumber salah satunya berita *online* serta media sosial akan tetapi tidak ada sumber basis data dan *big data analytic* pada sumber berita *online* tersebut. Penelitian ini menggunakan data dari twitter karena Twitter sebagai salah satu potensi sumber data informasi, Twitter telah mengalami pertumbuhan sangat pesat dan dengan cepat meraih popularitas di seluruh dunia, terbukti dengan banyaknya pengguna Twitter mengirimkan kicauan per harinya. Tujuan dari penelitian ini adalah melakukan data analytic terutama pencarian data yang sedang *trends* yang sedang terjadi pada masyarakat melalui data dari Twitter pada saat melakukan *update* status atau mengunggah *tweet* berupa *hashtag* sehingga bisa diketahui apa yang sedang di bicarakan di tengah masyarakat. Penelitian ini mengadopsi metode deskriptif yaitu *crawling, analyzing, storage, visualyizing* serta menggunakan MongoDB sebagai fasilitas penyimpananya. Hasil studi ini mendapatkan hasil *crawling data* sebanyak 9089 data *tweets* menjadi 2933 data *tweets* yang telah diproses pada tahap *praprosesing* data yang menghasilkan data bebas dari *noise*, data ini kemudian divisualisasikan walaupun tidak dijelaskan secara detail tahap proses nya dan menghasilkan sebuah informasi topik modeling yang sedang trends. Di harapkan data informasi tersebut dapat menjadi sebuah referensi sebagai sumber data yang dapat diolah dalam keperluan analisis penentuan topik untuk berita yang sedang terpopuler pada saat ini yang di tujukan pada media berita *online*.

Kata Kunci : Berita *Online*, *Big Data*, Twitter, MongoDB.

ABSTRACT

Current technological developments make data very valuable because data is experiencing rapid growth every day, data has complex forms such as words, letters, numbers, table graphs, pictures, and symbols. The more information from online news that is obtained from various information sources that exist every day, the more data that has been generated this phenomenon is called big data, The increasing number of data that has been generated by this phenomenon is called big data. The problem commonly found is the large amount of data generated by sharing online news sources but the absence of big data analytics on these online news sources. This research uses data from Twitter because Twitter as one of the potential sources of information data, Twitter has experienced rapid growth and is rapidly gaining popularity worldwide, as evidenced by the large number of Twitter users sending tweets per day. The purpose of this study is to search for data on trends that are happening in the community through data from Twitter when updating status or uploading tweets in the form of hashtags so that you can find out what is being discussed in the community. This study adopts descriptive methods, namely Crawling, Analyzing, Storage, Visualizing and using MongoDB as a storage facility. The results of this study get 9089 data crawling data tweets into 2933 tweets data that have been processed in the pre-processing stage of data which results in noise-free data, this data is then visualized even though the process stage is not explained in detail and produces information on modeling topics that are trending. . It is hoped that the information data can become a reference as a source that can be processed in the analysis of determining the topic for the most popular news at this time aimed at online news media.

Keywords : Online News, Big Data, Twitter, MongoDB.

KATA PENGANTAR



Puji syukur kehadiran Tuhan Yang Maha Esa karena berkat rahmat dan karunia-Nya yaitu skripsi ini dapat diselesaikan dengan cukup baik untuk memenuhi salah satu syarat agar mendapatkan gelar Sarjana Komputer di Program Studi Sistem Informasi, Fakultas Ilmu Komputer, di Universitas Bina Darma Palembang.

Dalam penulisan penelitian skripsi ini, tentunya masih sangat jauh dari kata sempurna. Hal ini dikarenakan keterbatasannya pengetahuan yang dimiliki. Oleh karena itu dalam rangka melengkapi kesempurnaan dari penulisan skripsi ini diharapkan adanya saran dan kritik yang diberikan bersifat membangun.

Pada kesempatan yang baik ini, tak lupa penulis menghaturkan terima kasih kepada semua pihak yang telah memberikan bimbingan, pengarahan, nasehat dan pemikiran dalam penulisan skripsi penelitian ini, terutama kepada :

1. Dr. Sunda Ariana, M.Pd., M.M. selaku Rektor Universitas Bina Darma Palembang.
2. Dedi Syamsuar, S.Kom., M.I.T, Ph.D selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Edi Surya Negara, M.Kom. selaku Ketua Program Studi Sistem Informasi sekaligus pembimbing dan mentor terbaik untuk penulis.
4. Kepada Bapak Dedi Syamsuar, M.I.T, Ph.D dan Bapak Ilman Zuhriyadi, M.M., M.Kom sebagai Penguji.
5. Orang Tua Tercinta, Keluarga besar, Keluarga kecil, saudara - saudaraku, dan seluruh teman serta sahabat-sahabatku yang selalu memberikan dorongan dan masukan serta bantuan baik moral maupun materil yang tak ternilai harganya.
6. Kepada seluruh dosen dan mahasiswa Universitas Bina Darma yang telah membantu atas terlaksananya skripsi ini.

7. Kepada teman-teman seperjuangan Program Studi Sistem Informasi angkatan 2016.
8. Kepada Rekan-rekan kepengurusan HIMSIF Universitas Bina Darma.
9. Kepada Rekan-rekan BDPRO (Bina Darma Programmer).

Palembang, 31 Agustus 2020

Penulis

DAFTAR ISI

Halaman

HALAMAN PENGESAHAN.....	i
HALAMAN PERSETUJUAN	i
HALAMAN PERNYATAAN	i
MOTTO DAN PERSEMBAHAN.....	iv
ABSTRAK.....	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR	xi
DAFTAR KODE PROGRAM.....	xiii
BAB I.....	1
PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
2.1. Rumusan Masalah	2
3.1. Batasan Masalah	2
4.1. Tujuan	2
5.1. Manfaat Penelitian	3
6.1. Sistematika Penulisan.....	3
BAB II.....	4
TINJAUAN PUSTAKA	4
2.1. Landasan Teori.....	4
2.1.1. Big Data	4
2.1.2. Text mining.....	5
2.1.3. Media Sosial.....	5
2.1.4. Twitter	8
2.1.5. Basis Data	10
2.1.6. MongoDB	11
2.1.7. Python.....	12
2.1.8. API	12

2.1.9.	Visual Studio Code.....	12
2.1.10.	Metode Deskriptif	13
2.1.11.	Preprosesing data.....	14
2.1.12.	Steaming.....	14
2.1.13.	Stopword	14
2.2.	Penelitian Sebelumnya.....	15
BAB III		17
METODOLOGI PENELITIAN		17
3.1.	Metode Penelitian.....	17
3.1.1.	Waktu dan Tempat.....	17
3.1.2.	Alat dan Bahan.....	17
3.1.3.	Data Tweet	18
3.1.4.	Tahapan Penelitian	18
3.1.5.	Crawling Data.....	20
3.1.6.	Storing data.....	23
3.1.7.	Analyzing Data	25
3.1.8.	Visualyzing Data.....	26
BAB IV		27
HASIL DAN PEMBAHASAN		27
4.1 Hasil.....		27
4.1.1. Implementasi Crawling Data		27
4.1.2. Implementasi Storing Data.....		31
4.1.3. Implementasi Analyzing Data		33
4.1.4. Implementasi Visualyzing Data		53
4.2 Pembahasan		54
BAB V		56
KESIMPULAN DAN SARAN.....		56
5.1.	Kesimpulan	56
5.2.	Saran.....	56
DAFTAR PUSTAKA		57

DAFTAR GAMBAR

Gambar	Halaman
BAB I PENDAHULUAN	1
BAB II TINJAUAN PUSTAKA	6
BAB III ANALISIS DAN PERANCANGAN	17
Gambar 3.1 Flowcharts Bagian Tahap-tahap penelitian.....	19
Gambar 3.2 Proses Autentikasi API Twitter	20
Gambar 3.3 Alur Crawling Data	22
Gambar 3.4 Perintah Pada MongoDB	23
Gambar 3.5 Alur Proses Storing Data	24
Gambar 3.6 Bentuk Data Preprocessing.....	25
Gambar 3.7 Alur Proses Analyzing	26
BAB IV HASIL DAN PEMBAHASAN	28
Gambar 4.1 Keys & Token Twitter Developer.....	27
Gambar 4.2 Hasil Perintah <i>Crawling Data</i>	30
Gambar 4.3 Hasil Crawling Sebanyak 9089 Data.....	30
Gambar 4.4 Membuat Nama Database dan Collection.....	31
Gambar 4.5 Import Data Json ke MongoDB.....	31
Gambar 4.6 File Data Json diMongoDB.	32
Gambar 4.7 Fungsi Pretty di MongoDB.....	32
Gambar 4.8 Export Data Json CMD.	33
Gambar 4.9 Hasil Export Json.	34
Gambar 4.10 Convert json ke Csv.....	36
Gambar 4.11 Save Convert Json ke CSV.	36
Gambar 4.12 Hasil dari Convert.....	37
Gambar 4.13 Upload Indonesia.csv kedalam jupyter.	37
Gambar 4.14 Data Indonesia.csv berhasil upload di Jupyter.....	38
Gambar 4.15 Implementasi modul Jupyter.	39
Gambar 4.16 Implementasi Hasil Data.....	40
Gambar 4.17 Implementasi data frame.	40

Gambar 4. 18 Hasil id_str & text.....	41
Gambar 4. 19 Hasil menghapus nama user @.....	42
Gambar 4. 20 Menghapus angka-angka 0-9, ge, rt, dan simbol #.	43
Gambar 4. 21 Hapus Duplikat.....	44
Gambar 4. 22 <i>Stopwords</i>	44
Gambar 4. 23 Tokenize.....	45
Gambar 4. 24 Remove happy emoticon.	45
Gambar 4. 25 <i>Remove sad emoticon</i>	46
Gambar 4.26 Remove seluruh emoticon, stopword, steaming, tokenize.....	47
Gambar 4. 27 Kolom baru tweet_clean.	48
Gambar 4. 28 Sort nilai true.	49
Gambar 4. 29 kolom tweet.....	49
Gambar 4. 30 Output 2933 data.....	50
Gambar 4. 31 Implementasi export csv.	50
Gambar 4. 32 File output csv di jupyter.....	51
Gambar 4. 33 File output_cleaning_indonesia.csv di jupyter.	51
Gambar 4. 34 File output_cleaning_indonesia.csv di excel.....	52
Gambar 4. 35 Visualisasi Topic Modelling.....	53

DAFTAR KODE PROGRAM

Kode Program	Halaman
BAB IV HASIL DAN PEMBAHASAN	28
Kode Program 4. 1 Consumer key dan consumer secret	28
Kode Program 4. 2 Crawling Data Python.	28
Kode Program 4. 3 Import File Json ke MongoDB	31
Kode Program 4. 4 Cek Isi Data di Mongodb.....	31
Kode Program 4. 5 Cek Isi Data MongoDB Menggunakan Pretty	32
Kode Program 4. 6 Export File Json Dari MongoDB.....	33
Kode Program 4. 7 Convert Tools Json to Csv	34
Kode Program 4. 8 Modul.....	38
Kode Program 4. 9 Load Data Csv	39
Kode Program 4. 10 Merubah Data Frame	40
Kode Program 4. 11 Remove User dan @	41
Kode Program 4. 12 Menghapus Angka-Angka 0-9, GE, Huruf RT, dan Simbol #	42
Kode Program 4. 13 Remove Duplikat Tweets	43
Kode Program 4. 14 Stopword	44
Kode Program 4. 15 Modul Toknize.....	45
Kode Program 4. 16 Remove Happy Emoticon.....	45
Kode Program 4. 17 Remove Sad Emoticon.....	46
Kode Program 4. 18 Mengapus Kembali Ge, Rt, Https, Hastag, Stopwords, Tokenaize, Steaming, Remove Emoticon.....	46
Kode Program 4. 19 Membuat Kolom Tweet_clean	47
Kode Program 4. 20 Membuat Sort Nilai True	48
Kode Program 4. 21 Menghapus Collom Id_str, Text, Remove_user, remove_http, tweet_clean.....	49
Kode Program 4. 22 Menghapus Data Kosong.....	50
Kode Program 4. 23 Export Csv	50