



SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI

**ANALISIS SENTIMEN HASIL CRAWLING PADA MEDIA
ONLINE MENGGUNAKAN ELASTICSEARCH DAN
ALGORITMA NAIVE BAYES**

TUGAS AKHIR

MUHAMAD FADRI WIJAYA

0110216072

PROGRAM STUDI TEKNIK INFORMATIKA

DEPOK

FEBRUARI 2020



SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI

**ANALISIS SENTIMEN HASIL CRAWLING PADA MEDIA
ONLINE MENGGUNAKAN ELASTICSEARCH DAN
ALGORITMA NAIVE BAYES**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
sarjana komputer**

**Muhammad Fadri Wijaya
0110216072**

**PROGRAM STUDI TEKNIK INFORMATIKA
DEPOK
FEBRUARI 2020**

HALAMAN PERNYATAAN ORISINALITAS

**Skripsi/Tugas Akhir ini adalah hasil karya penulis,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.**



Nama : Muhammad Fadri Wijaya

NIM : 0110216072

Tanda Tangan :

Tanggal : 24 Februari 2020

HALAMAN PENGESAHAN

Skripsi/Tugas Akhir ini diajukan oleh :

Nama : Muhammad Fadri Wijaya

NIM : 0110216072

Program Studi : Teknik Informatika

Judul Skripsi : Analisis Sentimen Hasil Crawling pada Media Online
Menggunakan elasticsearch dan Algoritma Naive Bayes

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika Sekolah Tinggi Teknologi Terpadu Nurul Fikri

DEWAN PENGUJI

Pembimbing I

(Sirojul Munir, S.Si, M.Kom)

Penguji I **S T T - N F** Penguji II

(Ahmad Rio Adriansyah, S.Mi, M.Mi) (Hilmy Abidzar Tawakal, ST., M.Kom)

Ditetapkan di : Depok

Tanggal : 24 Februari 2020

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Tugas Akhir ini. Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana komputer Program Studi Teknik Informatika pada Sekolah Tinggi Teknologi Terpadu Nurul Fikri. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak sangatlah sulit bagi penulis untuk menyelesaikan Tugas Akhir ini. Oleh karena itu dalam kesempatan ini, penulis mengucapkan banyak terima kasih kepada:

1. Bapak Lukman Rosyidi, ST, M.M., M.T selaku Ketua Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
2. Bapak Sirojul Munir, S.Si, M.Si., selaku dosen pembimbing yang telah memberikan banyak bantuan, bimbingan, serta arahan dalam pengerjaan tugas akhir ini.
3. Seluruh dosen di Sekolah Tinggi Teknologi Terpadu Nurul Fikri, atas ilmu yang bermanfaat dan bimbingannya selama perkuliahan hingga saat ini.
4. Seluruh dosen penguji yang telah memberikan masukan bermanfaat bagi penelitian ilmiah ini.
5. Kedua orang tua dan keluarga, yang selalu memberikan dukungan dan doa hingga saat ini.
6. Teman-teman seperjuangan di Sekolah Tinggi Teknologi Terpadu Nurul Fikri.

Akhir kata, penulis mohon maaf yang sebesar - besarnya apabila ada kesalahan ataupun kekeliruan didalam penulisan tugas akhir ini. Semoga Allah SWT membalas segala kebaikan semua pihak yang telah membantu dan semoga tugas akhir ini membawa manfaat bagi yang membacanya.

Depok, 24 Februari 2020

Muhammad Fadri Wijaya

**ALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Sekolah Tinggi Teknologi Terpadu Nurul Fikri, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Fadri Wijaya

NIM : 0110216072

Program Studi : Teknik Informatika

Jenis karya : Skripsi / Tugas Akhir

demikian demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada STT-NF **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty - Free Right*)** atas karya ilmiah saya yang berjudul :

**Analisis Sentimen Hasil Crawling pada Media Online Menggunakan elasticsearch dan
Algoritma Naive Bayes**

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini STT-NF berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

STT - NF

Dibuat di : Depok

Pada tanggal : 24 Februari 2020

Yang menyatakan



(Muhammad Fadri Wijaya)

ABSTRAK

Nama : Muhammad Fadri Wijaya

NIM : 0110216072

Program Studi : Teknik Informatika

Judul : Analisis Sentimen Hasil Crawling pada Media Online Menggunakan elasticsearch dan Algoritma Naive Bayes

Sentiment analysis dan opinion mining adalah area yang telah mengalami pertumbuhan yang cukup besar selama dekade terakhir. Bidang penelitian ini mencoba untuk menentukan perasaan, pendapat, emosi, antara lain, orang-orang terhadap sesuatu atau seseorang. Untuk melakukan ini, teknik bahasa alami dan algoritma pembelajaran mesin digunakan. Artikel ini membahas masalah penggalian sentimen dan opini dari kumpulan berita yang diambil dari portal berita. Tujuan pertama dari penelitian ini adalah menghasilkan mesin yang mampu mengklasifikasikan sentimen suatu informasi yang didapatkan dari portal berita yang diambil menggunakan teknik scraping berdasarkan kata kunci yang ditambahkan oleh user melalui aplikasi dashboard. Text preprocessing yang digunakan yaitu tokenize words, remove stopwords dan stemming. Berita dari hasil text processing melalui proses matching sentiment. Yang mengklasifikasikan setiap kata yang termasuk sentiment positif atau negatif, setelah dihitung berapa total positif dan negatif diimplementasikan algoritma Naive Bayes Classifier untuk menghitung probabilitas sentiment, Confusion Matrix sebagai metode pengujian. Serta menggunakan Elastic search sebagai database untuk menampung data hasil scraping dan Opinion words. Hasil dari data yang telah diklasifikasi akan divisualisasikan di dashboard dalam bentuk chart, yang berisi nilai sentiment positif dan negatif dari berita yang sebelumnya telah diolah berdasarkan kata kunci yang tersimpan didatabase.

Kata kunci: Analisis Sentimen, Visualisasi Data, Crawling, Naive Bayes, Elasticsearch

ABSTRACT

Name : Muhammad Fadri Wijaya

NIM : 0110216072

Study Program : Informatics Engineering

Title : Sentiment Analysis of Crawling Results on Online Media Using Elasticsearch and Naive Bayes Algorithm

Sentiment analysis and opinion mining are areas that have experienced considerable growth over the last decade. This field of research tries to determine the feelings, opinions, emotions, among other things, of people towards something or someone. To do this, natural language techniques and machine learning algorithms are used. This article discusses the issue of sentiment and opinion from a collection of news taken from news portals. The first objective of this research is to produce a machine that is able to classify information obtained from news portals that are retrieved using a scraping technique based on keywords added by users through the application dashboard. Text preprocessing used is tokenize words, remove stopwords and stemming. News from text processing results through a sentiment matching process. Which classifies each word that includes positive or negative sentiments, after calculating how many total positives and negatives the Naïve Bayes Classifier algorithm implements to calculate sentiment probability, Confusion Matrix as a test method. And using Elastic search as a database to accommodate scraping data and Opinion words. The results of the classified data will be visualized on the dashboard in the form of a graph, which contains the positive and negative sentiment values of the news that have previously been processed based on keywords stored in the database.

Keywords: Sentiment analysis, Data Visualization, Crawling, Naïve Bayes, Elasticsearch

DAFTAR ISI

DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xi
KATA PENGANTAR.....	i
BAB I PENDAHULUAN.....	v
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	4
BAB II LANDASAN TEORI.....	6
2.1 Tinjauan Pustaka.....	6
2.1.1 Media Online, Crawling dan Sentimen Analisis.....	6
2.1.2 Metodologi Pengembangan Scrum.....	9
2.1.3 Infrastruktur Pengembangan.....	13
2.1.4 Algoritma Naïve Bayes.....	17
2.1.5 Blackbox Testing.....	21
2.2 Penelitian Terkait.....	22
2.3 Posisi Penelitian.....	22
BAB III METODE PENELITIAN.....	24
3.1 Tahapan Penelitian.....	24
3.2 Rancangan.....	25
3.2.1 Pemecahan Masalah.....	25
3.2.2 Jenis Penelitian.....	26
3.2.3 Metode Pengumpulan Data.....	27
3.2.4 Lingkungan Pengembangan.....	27
BAB IV ANALISIS DAN PERANCANGAN.....	28
4.1 Analisis Sistem.....	28
4.1.1 Product Backlog.....	28

4.1.2	User Story	29
4.2	Perancangan Sistem.....	30
4.2.1	Data Model	30
4.2.2	Rancangan Alur Sistem	31
4.2.3	Deployment	34
4.3	Rancangan Pengujian	36
4.3.1	Blackbox Testing.....	36
BAB V	38
IMPLEMENTASI DAN EVALUASI	38
5.1	Tools & Aplikasi.....	38
5.1.1	Webhose	38
5.1.2	Python.....	39
5.1.3	Elasticsearch.....	39
5.1.4	Crontab	40
5.1.5	Mysql Connector	40
5.2	Implementasi Aplikasi.....	41
5.2.1	Collecting Data.....	41
5.2.2	Preprocessing Text	42
5.2.3	Sentiment Analisis Naïve Bayes	46
5.2.4	Crontab	49
5.2.5	Mysql Connector	49
5.3	Implementasi Alur Sistem.....	52
5.4	Implementasi Scrum.....	55
5.5	Rekomendasi & Tindak Lanjut	56
5.6	Hasil Pengujian	57
5.6.1	Blackbox Testing.....	57
5.6.2	Confusion Matrix.....	58
5.6.3	Klasifikasi Naïve Bayes	61
BAB VI	62
KESIMPULAN DAN SARAN	62
6.1	Kesimpulan.....	62
6.2	Saran.....	62

DAFTAR PUSTAKA	63
----------------------	----

DAFTAR GAMBAR

Gambar 1 : Gambar Flow Scrum	10
Gambar 2 : Implementasi Naïve Bayes Classifier NLTK	15
Gambar 3 : Tahapan penelitian	24
Gambar 4 : Flowchart pemecahan masalah	25
Gambar 5 : Data Model.....	30
Gambar 6 : flowchart keseluruhan proses crawling dan pemrosesan sentimen.....	31
Gambar 7 : Flowchart Proses Crawling Data	31
Gambar 8 : Flowchart Proses Preprocessing Text	32
Gambar 9 : Flowchart Proses Sentimen Naïve Bayes Classifier	34
Gambar 10 : Deployment Diagram Aplikasi Crawling	35
Gambar 11 : Halaman Webhose.io	38
Gambar 12 : Tampilan Jupyter Notebook IDE Python.....	39
Gambar 13: Tampilan List Index Elasticsearch.....	39
Gambar 14: Tampilan halaman repository python-crontab	40
Gambar 15: Tampilan Halaman repositori Mysql-Connector	40
Gambar 16: Fungsi crawling berita.....	52
Gambar 17: Hasil Crawling Berita dari keyword yang diset.....	52
Gambar 18 :Fungsi untuk Preprocessing text	53
Gambar 19: Hasil Sentimen dari proses matching dan penghitungan polaritas perkalamat	53
.....	
Gambar 20 : Hasil Sentimen total polaritas keseluruhan.....	53
Gambar 21: Proses training dan testing menggunakan NLTK naive bayes	54
Gambar 22: Hasil Klasifikasi dari pemodelan yang sudah ditraining	54
Gambar 23 : Gambar Rekomendasi & Tindak Lanjut.....	56
Gambar 24 : Implementasi Training Menggunakan NLTK Naïve Bayes Classifier...	58
Gambar 25: Hasil Confusion Matrix.....	59
Gambar 26 : Gambar Hasil Crawling & Sentimen Berita	60

STT - NF

DAFTAR TABEL

Table 1 : Confusion Matrix.....	16
Table 2 : Penelitian Terkait.....	22
Table 3: Posisi Penelitian.....	23
Table 4: User Story	29
Table 5 : Blackbox Testing.....	36
Table 6: Source Code Collecting Data.....	41
Table 7 : Source Code Preprocessing Text.....	42
Table 8 : Source Code Analysis Naive Bayes	46
Table 9 : Source Code Crontab.....	49
Table 10 : Source Code Mysql-Connector.....	49
Table 11 : Hasil BlackBox Testing.....	57
Table 12: Hasil Confusion Matrix	59



STT - NF