

**ANALISIS SENTIMEN MASYARAKAT TERHADAP
HACKER BJORKA DALAM KASUS KEBOCORAN
DATA BPJS KETENAGAKERJAAN MENGGUNAKAN
ALGORITMA NAÏVE BAYES DAN C4.5**

SKRIPSI

Oleh:

Eka Nur A'ini

201910225052



PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS BHAYANGKARA JAKARTA RAYA

2023

LEMBAR PERSETUJUAN PEMBIMBING
(DIGUNAKAN UNTUK TUGAS AKHIR)

Judul Tugas Akhir : Analisis Sentimen Masyarakat Terhadap *Hacker*
Bjorka dalam Kasus Kebocoran Data BPJS
Ketenagakerjaan Algoritma *Naive Bayes* dan C4.5

Nama Mahasiswa : EKA NUR A'INI
Nomor Pokok Mahasiswa : 201910225052
Program Studi/Fakultas : Informatika/Ilmu Komputer
Tanggal Lulus Ujian Skripsi : 04 Agustus 2023

Pembimbing I

Pembimbing II

Wowon Priatna, S.T., M.Ti.
NIDN: 0429118007

Siti Setiawati, S.Pd., M.Pd.
NIDN: 0313107904

LEMBAR PENGESAHAN

Judul Skripsi : ANALISIS SENTIMEN MASYARAKAT
TERHADAP HACKER BJORKA DALAM KASUS
KEBOCORAN DATA BPJS KETENAGAKERJAAN
MENGUNAKAN ALGORITMA *NAÏVE BAYES*
DAN C4.5

Nama Mahasiswa : EKA NUR A'INI
Nomor Induk Mahasiswa : 201910225052
Program Studi/Fakultas : Informatika/Illmu Komputer
Tanggal Lulus Ujian Skripsi : 04 Agustus 2023

Jakarta, 09 Agustus 2023

MENGESAHKAN

Ketua Tim Penguji : Dr. Dra. Tyastuti Sri Lestari, M.M
NIDN: 0327036701

Penguji II : Ajif Yunizar Pratama Yusuf, S.Si., M.Eng
NIDN: 0328068603

Penguji III : Wowon Priatna, S.T., M.Ti
NIDN: 0429118007

MENGETAHUI,

Ketua
Program Studi Informatika


Ahmad Fathurrozi, S.E., M.M.S.I
NIP. 2012486

Dekan
Fakultas Ilmu Komputer


Dr. Dra. Tyastuti Sri Lestari, M.M
NIP. 1408206



LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Eka Nur A'ini
NPM : 201910225052
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Analisis Sentimen Masyarakat terhadap Hacker Bjorka
dalam Kasus Kebocoran Data BPJS Ketenagakerjaan
Menggunakan Algoritma *Naïve Bayes* dan C4.5.

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Jakarta, 11 Agustus 2023
Penulis



Eka Nur A'ini
201910225052

ABSTRAK

Eka Nur A'ini. 201910225052. Analisis Sentimen Masyarakat terhadap *Hacker Bjorka* dalam Kasus Kebocoran Data BPJS Ketenagakerjaan Menggunakan Algoritma *Naïve Bayes* dan C4.5. Fakultas Ilmu Komputer. Bekasi: Universitas Bhayangkara Jakarta Raya. 2023.

Pada pertengahan Maret 2023, Indonesia dibuat gempar lagi oleh kehadiran *hacker* dengan nama Bjorka. Bjorka tidak melakukan aksinya satu atau dua kali, namun sering kali Bjorka membuat gempar seluruh masyarakat Indonesia. Pasalnya Bjorka kembali melakukan aksi pembobolan data, data yang dibobol oleh Bjorka ini merupakan data BPJS Ketenagakerjaan milik masyarakat Indonesia sebanyak 19 juta data dan diperjualbelikan. Sejak ramainya kasus Bjorka hal itu memicu banyak masyarakat yang mengkritik di berbagai media sosial salah satu diantaranya media sosial *Facebook* sehingga kritik atau opini tersebut dapat dimanfaatkan untuk melakukan analisis sentimen. Berdasarkan hal tersebut, diperlukan sebuah metode yang dapat secara otomatis melakukan klasifikasi opini ke dalam kategori positif dan negatif melalui proses analisis sentimen. Proses analisis sentimen dilakukan dengan proses *data preprocessing*, pembobotan kata menggunakan metode TF-IDF, penerapan algoritma, dan pembahasan atas hasil klasifikasi. Metode klasifikasi data yang digunakan dalam penelitian ini adalah *Naive Bayes* dan C4.5. Data tersebut akan diproses menggunakan *text mining* dan klasifikasi menggunakan algoritma *Naive Bayes* dan C4.5. Berdasarkan pengujian yang telah dilakukan, hasil klasifikasi terbaik diperoleh oleh *Naive Bayes* dengan akurasi sebesar 70% dan 68% untuk algoritma C4.5. Algoritma *Naive Bayes* mampu memprediksi ketepatan data sebesar 70% untuk sentimen positif dan negatif.

Kata Kunci: Bjorka, Analisis Sentimen, Pembobolan Data, Opini, *Naïve Bayes*, C4.5.

ABSTRACT

Eka Nur A'ini. 201910225052. *Analysis of Public Sentiment towards Hacker Bjorka in the Case of Data Breach of BPJS Ketenagakerjaan Using Naive Bayes and C.45 Algorithms.* Faculty of Computer Science. Bekasi: Bhayangkara University, Jakarta Raya. 2023.

In mid-March 2023, Indonesia was once again shaken by the presence of a hacker named Bjorka. Bjorka didn't carry out their actions just once or twice, but often managed to stir up the entire Indonesian society. This is because Bjorka once again engaged in data breaches, with the data breached by Bjorka consisting of 19 million records belonging to the Indonesian social security agency (BPJS Ketenagakerjaan) and being traded. Since the Bjorka case gained widespread attention, it triggered a lot of public criticism on various social media platforms, including Facebook. Given this situation, a method is required to automatically classify opinions into positive and negative categories through sentiment analysis. The process of sentiment analysis involves data preprocessing, word weighting using the TF-IDF method, algorithm implementation, and discussion of the classification results. The data classification method employed in this study is Naive Bayes and C4.5. The data will be processed using text mining and classified using the Naive Bayes and C4.5 algorithms. Based on the conducted testing, the best classification results were obtained using Naive Bayes with an accuracy of 70%, while the C4.5 algorithm achieved 68%. The Naive Bayes algorithm demonstrated a prediction accuracy of 70% for both positive and negative sentiments.

Keywords: *Bjorka, Sentiment Analysis, Data Breach, Opinion, Naive Bayes, C4.5.*

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK**

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Eka Nur A'ini
NPM : 201910225052
Program Studi : Informatika
Fakultas : Ilmu Komputer
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalty-Free Right)**, atas karya ilmiah saya yang berjudul :

Analisis Sentimen Masyarakat terhadap Hacker Bjorka dalam Kasus Kebocoran Data BPJS Ketenagakerjaan Menggunakan Algoritma *Naïve Bayes* dan C4.5

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 11 Agustus 2023
Yang Menyatakan



Eka Nur A'ini
201910225052

KATA PENGANTAR

Puji syukur peneliti panjatkan kepada Tuhan Yang Maha Esa Allah SWT yang telah melimpahkan karunia-Nya sehingga peneliti dapat menyelesaikan skripsi yang disusun sebagai salah satu syarat untuk mencapai gelar sarjana Strata satu (S1) pada program studi Informatika Universitas Bhayangkara Jakarta Raya.

Penelitian ini tidaklah dapat terwujud dengan baik tanpa adanya bimbingan, dorongan, saran-saran dan bantuan dari berbagai pihak. Pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah S.W.T yang telah meridhai penulis selaku hamba-Nya untuk menulis proposal studi independen ini hingga selesai.
2. Bapak Inspektur Jenderal Polisi Irjen Pol. (Purn) Dr, Drs. H. Bambang Karsono, S.H.,M.M. selaku Rektor Universitas Bhayangkara Jakarta Raya.
3. Ibu Dr. Tyastuti Sri Lestari, M.M. selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
4. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. selaku Ketua Program Studi Informatika.
5. Bapak Wowon Priatna, S.T., M.Ti selaku dosen pembimbing satu penyusunan laporan skripsi.
6. Ibu Siti Setiawati, S.Pd., M.Pd. selaku dosen pembimbing dua penyusunan laporan skripsi.

7. Ibu Hafizah, S.S., M.Pd. selaku validator instrumen dalam penentuan palabelan data dalam pembuatan laporan skripsi.
8. Kedua orang tua saya yang selalu memberikan dukungan dan semangat dan selalu medoakan saya setiap harinya dalam agar diberikan kemudahan dalam menyelesaikan skripsi ini.
9. Dhea Putri Aprilyana dan Nur Fadila yang selalu membantu, memberikan dukungan dan semangat dalam menyelesaikan skripsi ini.
10. Teman-teman ANGK yang selalu memberikan dukungan dan semangat dalam menyelesaikan skripsi ini.
11. Teman-teman seperjuangan pada program studi Informatika Universitas Bhayangkara Jakarta Raya yang selalu memberikan dukungan dan selalu membantu dalam menyelesaikan skripsi ini.

Peneliti menyadari bahwa penelitian ini jauh dari kesempurnaan, untuk itu peneliti mohon maaf apabila terdapat kekurangan dan kesalahan baik dari isi maupun tulisan. Oleh karena itu, peneliti mengharapkan kritik dan saran yang membangun

Bekasi, 10 Juli 2023

Eka Nur A'ini
201910225052

DAFTAR ISI

Halaman

LEMBAR PERSETUJUAN PEMBIMBING.....	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN BUKAN PLAGIASI.....	iii
ABSTRAK.....	iv
ABSTRACT.....	v
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	iv
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	5
1.4 Batasan Masalah.....	5

1.5	Tujuan Penelitian.....	6
1.6	Manfaat Penelitian.....	6
1.7	Sistematika Penulisan	7
BAB II TINJAUAN PUSTAKA		9
2.1	<i>Table Related Reserch</i>	9
2.2	Penelitian Terdahulu	10
2.3	<i>Data Science</i>	12
2.4	Data Mining	12
2.5	Analisis Sentimen	14
2.5.1	Sentimen	15
2.5	Statistika	15
2.6	Data	16
2.6	<i>Knowledge Discovery in Database</i>	17
2.7	Pembobotan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF).....	18
2.8	Algoritma <i>Naïve Bayes</i>	21
2.9	Algoritma C4.5.....	23
2.10	Confusion Matrix	25
2.11	BPJS Ketenagakerjaan.....	27

2.12	Media sosial	28
BAB III METODOLOGI PENELITIAN.....		31
3.1	Objek Penelitian dan Sumber Data	31
3.2	Kerangka Penelitian.....	31
3.2	Uraian Metodologi.....	32
3.2.1	<i>Data Selection</i>	32
3.3.1.1	<i>Scrapping Data</i>	33
3.3.1.2	Pelabelan Data	33
3.2.2	<i>Preprocessing</i>	34
3.2.2.1	<i>Case Folding</i>	34
3.2.2.2	<i>Tokenizing</i>	36
3.2.2.3	<i>Filtering / Stopword Removing</i>	39
3.2.2.4	<i>Stemming</i>	40
3.2.3	<i>Transformation</i>	44
3.3.3.1	<i>Splitting Data</i>	44
3.3.3.2	<i>TF-IDF Vectorizing</i>	44
3.2.4	<i>Data Mining</i>	44
3.2.4.1	Algoritma <i>Naïve Bayes</i>	44

3.2.4.2	Algoritma C.45	46
3.2.5	<i>Evaluation</i>	48
3.2.6	<i>Knowledge Presentation</i>	48
3.3	Instrumen Penelitian	48
BAB IV	HASIL DAN PEMBAHASAN	62
4.1	Hasil	62
4.2	Pembahasan	63
4.2.1	<i>Data Selection</i>	63
4.2.1.1	Pengambilan Data	63
4.2.1.2	Pelabelan Data	67
4.2.2	<i>Preprocessing</i>	69
4.2.2.1	<i>Case Folding</i>	69
4.2.2.2	<i>Tokenizing</i>	72
4.2.2.3	<i>Filtering / Stopword Removing</i>	76
4.2.2.4	<i>Stemming</i>	79
4.2.3	<i>Transformation</i>	82
4.2.3.1	Mengubah Dataset Menjadi <i>Numpy Array</i>	82
4.2.3.2	<i>Splitting Dataset</i>	83

4.2.3.3	Pembobotan <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF)	84
4.2.4	<i>Data Mining</i>	103
4.2.4.1	Implementasi Algoritma <i>Naïve Bayes</i>	103
4.2.4.1	Implementasi Algoritma C.45.....	111
4.2.5	<i>Evaluation</i>	125
4.2.5.1	<i>Evaluation Naïve Bayes Algorithm</i>	125
4.2.5.2	<i>Evaluation C.45 Algorithm</i>	130
4.2.6	<i>Knowladge Presentation</i>	134
BAB V	PENUTUP	137
5.1	Kesimpulan	137
5.2	Saran.....	138
DAFTAR PUSTAKA		139
LAMPIRAN		147

DAFTAR TABEL

Tabel 3. 1 Hasil <i>Labelling</i> Manual	33
Tabel 3. 2 Aturan Tahap Case Folding	35
Tabel 3. 3 Ilustrasi Tahapan <i>Case Folding</i>	35
Tabel 3. 4 Aturan Tahap <i>Tokenizing</i>	38
Tabel 3. 5 Ilustrasi Tahapan <i>Tokenizing</i>	38
Tabel 3. 6 Ilustrasi Tahapan <i>Filtering / Stopword Removing</i>	40
Tabel 3. 7 Ilustrasi Tahapan <i>Stemming</i>	43
Tabel 3. 8 Pengelompokan Data.....	48
Tabel 3. 9 Tabel Validasi Opini	53
Tabel 4. 1 Perbandingan <i>Confusion Matrix Algoritma Naïve Bayes dan C.45</i>	62
Tabel 4. 2 Pelabelan Data	68
Tabel 4. 3 Tahapan <i>Case Folding</i> Manual	70
Tabel 4. 4 Tahapan <i>Tokenizing</i> Manual	73
Tabel 4. 5 Tahapan <i>Filtering / Stopword Removing</i> Manual	76
Tabel 4. 6 Tahapan <i>Stemming</i> Manual.....	79
Tabel 4. 7 Pembuatan <i>Word Vector</i> Manual.....	86
Tabel 4. 8 Perhitungan TF dan DF	89
Tabel 4. 9 Perhitungan IDF	93
Tabel 4. 10 Perhitungan Bobot.....	97
Tabel 4. 11 Sampel <i>Data Train</i>	104

Tabel 4. 12 Sampel <i>Data Test</i>	104
Tabel 4. 13 Penentuan Prediksi dan Aktual.....	105
Tabel 4. 14 Perhitungan <i>Prior Probability</i>	107
Tabel 4. 15 Sampel <i>Data Test</i>	108
Tabel 4. 16 Hasil Prediksi <i>Data Test</i>	110
Tabel 4. 17 Contoh <i>Training Data</i>	111
Tabel 4. 18 Nilai <i>Training Data</i>	112
Tabel 4. 19 Perhitungan <i>Information Gain</i> Untuk Simpul Akar.....	116
Tabel 4. 20 Nilai <i>Training Data</i>	117
Tabel 4. 21 Perhitungan <i>Gain</i> Untuk Simpul Akhir.....	121
Tabel 4. 22 Nilai <i>Training Data</i>	122
Tabel 4.23 Perhitungan <i>Gain</i> Untuk Simpul Akhir	123
Tabel 4. 24 <i>Confusion Matrix Naïve Bayes</i>	126
Tabel 4. 25 <i>Confusion Matrix C.45</i>	130

DAFTAR GAMBAR

Gambar 2. 1 Proses <i>Knowledge Discovery in Databases</i> [15].....	18
Gambar 3. 1 <i>Official Fanpage</i> BPJS Ketenagakerjaan.....	31
Gambar 3. 2 Diagram Alir Tahapan Penelitian.....	32
Gambar 3. 3 Tahapan <i>Case Folding</i>	34
Gambar 3. 4 Tahapan <i>Tokenizing</i>	37
Gambar 3. 5 Tahapan <i>Filterring / Stopword</i>	39
Gambar 3. 6 Tahapan <i>Stemming</i>	41
Gambar 3. 7 Tahapan <i>Naïve Bayes</i>	45
Gambar 3. 8 Tahapan C.45	47
Gambar 4. 1 Logo <i>Instan Data Scraper</i>	64
Gambar 4. 2 Ekstensi <i>Instan Data Scraper</i>	64
Gambar 4. 3 Situs Halaman yang di <i>Scrapping</i>	65
Gambar 4. 4 Komentar Masyarakat Kasus Kebocoran Data BPJS.....	65
Gambar 4. 5 Pengambilan Data Menggunakan <i>Instan Data Scraper</i>	66
Gambar 4. 6 Menyimpan Hasil <i>Scrapping</i> Data	67
Gambar 4. 7 Hasil <i>Scrapping</i> Data dan <i>Labelling</i> Data.....	69
Gambar 4. 8 <i>Syntax</i> Tahapan <i>Case Folding</i>	72
Gambar 4. 9 Hasil <i>Syntax</i> Tahapan <i>Case Folding</i>	72
Gambar 4. 10 <i>Syntax</i> Tahapan <i>Tokenizing</i>	75
Gambar 4. 11 Hasil Tahapan <i>Tokenizing</i>	75
Gambar 4. 12 <i>Syntax</i> Tahapan <i>Filtering / Stopword Removing</i>	78

Gambar 4. 13 Hasil <i>Syntax Filtering/Stopword Removing</i>	79
Gambar 4. 14 <i>Syntax Tahapan Stemming</i>	81
Gambar 4. 15 Hasil <i>Syntax Stemming</i>	81
Gambar 4. 16 <i>Syntax Tahapan Merubah Data Menjadi Numpy Array</i>	82
Gambar 4.17 Hasil <i>Syntax Tahapan Merubah Data Menjadi Numpy Array Kolom</i> <i>Komentar</i>	83
Gambar 4. 18 Hasil <i>Syntax Tahapan Merubah Data Menjadi Numpy Array Kolom</i> <i>Sentimen</i>	83
Gambar 4. 19 Tahapan <i>Splitting Data</i> dan Hasil <i>Splitting Data</i>	84
Gambar 4. 20 <i>Syntax Tahapan Pembobotan TF IDF</i>	102
Gambar 4. 21 Hasil <i>Syntax Pembobotan Kata</i>	103
Gambar 4. 22 <i>Syntax Tahapan Modelling Naïve Bayes</i>	110
Gambar 4.23 Hasil Perhitungan <i>Gain Simpul Akar</i>	117
Gambar 4. 24 Hasil Perhitungan <i>Gain Simpul Akar</i>	122
Gambar 4. 25 Hasil Perhitungan <i>Gain Simpul Akar</i>	124
Gambar 4. 26 <i>Syntax Menampilkan Confusion Matrix Naïve Bayes</i>	128
Gambar 4. 27 Hasil <i>Confusion Matrix Naïve Bayes</i>	129
Gambar 4. 28 Tahapan dan Hasil Uji Model <i>Naïve Bayes</i>	129
Gambar 4. 29 <i>Syntax Menampilkan Confusion Matrix C.45</i>	133
Gambar 4. 30 Hasil <i>Confusion Matrix C.45</i>	133
Gambar 4. 31 <i>Syntax Tahapan dan Hasil Uji Model C.45</i>	134
Gambar 4. 32 <i>Pie Chart Hasil Labelling Data</i>	134

Gambar 4. 33 Hasil *Wordcloud*..... 135



DAFTAR LAMPIRAN

Lampiran I	Uji Plagiarisme
Lampiran II	Kartu Konsultasi Bimbingan Skripsi
Lampiran III	Validasi Instrumen
Lampiran IV	Surat Rekomendasi Dosen Pembimbing

