

**ANALISIS SENTIMEN PENGGUNA *E-MONEY* PADA
TWITTER MENGGUNAKAN *ALGORITMA C4.5* DAN
*NAÏVE BAYES***

SKRIPSI



Oleh :

ZELI ERIKA SHOLIKHA

NPM. 1634010012

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2020**

**ANALISIS SENTIMEN PENGGUNA *E-MONEY* PADA
TWITTER MENGGUNAKAN *ALGORITMA C4.5* DAN
*NAÏVE BAYES***

SKRIPSI

Diajukan Untuk Memenuhi Persyaratan Dalam Memperoleh Gelar Sarjana
Komputer Program Studi Teknik Informatika



Oleh :

ZELI ERIKA SHOLIKHA

NPM. 1634010012

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2020**

LEMBAR PENGESAHAN SKRIPSI

Judul : Sistem Informasi Penjualan Pakaian Berbasis *Web* pada DEFIRZA
Collection Surabaya

Oleh : Deny Alif Firmansyah

NPM : 1434010198

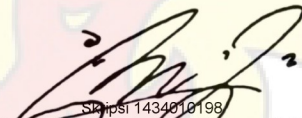
Telah Diseminarkan Dalam Ujian Skripsi
Pada Tanggal : 20 Juli 2020

Menyetujui:


Dosen Pembimbing

Dosen Penguji


1.


Budi Nugroho, S.Kom, M.Kom
NPT. 3 8009 050 205 1


1.


Henni Endah Wahanani, S.T, M.Kom
NPT. 3 7809 13 0342 1

2.


Firza Prima Aditiawan, S.Kom, MTI
NPT. 3 8605 13 03441

2.


Mohammad Idhom, S.P, S.Kom, M.T
NPT. 3 8303 10 0285 1

Mengetahui:



Dekan
Fakultas Ilmu Komputer,

Dr. Ir. Ni Ketut Sari, MT
NIP. 19650731 199203 2 001

Koordinator Program Studi
Informatika,


Budi Nugroho, S.Kom, M.Kom
NPT. 3 8009 050 205 1

SURAT PERNYATAAN ANTI PLAGIAT

Saya, mahasiswa Teknik Informatika UPN “Veteran” Jawa Timur, yang bertandatangan di bawah ini:

Nama : ZELI ERIKA SHOLIKHA

NPM : 1634010012

Menyatakan bahwa Judul Skripsi/ Tugas Akhir yang Saya ajukan dan akan dikerjakan, yang berjudul:

“ANALISIS SENTIMEN PENGGUNA E—MONEY PADA TWITTER MENGUNAKAN ALGORITMA C4.5 DAN NAÏVE BAYES ”

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk dan atau *software* yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di UPN “Veteran” Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata di kemudian hari pernyataan ini terbukti tidak benar, maka Saya siap menerima segala konsekuensinya.

Surabaya, 15 Juli 2020

Hormat Saya,



ZELI ERIKA SHOLIKHA

NPM. 1634010012

ANALISIS SENTIMEN PENGGUNA *E-MONEY* PADA TWITTER MENGGUNAKAN
ALGORITMA C4.5 DAN *NAÏVE BAYES*

Nama Mahasiswa : Zeli Erika Sholikha

NPM : 1634010012

Program Studi : Teknik Informatika

Dosen Pembimbing : Eva Yulia Puspaningrum,S.Kom.M.Kom

Wahyu Syaifullah JS.S.Kom.M.Kom

ABSTRAK

Twitter merupakan *microblogging service* dimana pengguna dapat memperbarui status tentang berbagai informasi dan opini terhadap berbagai topik atau isu-isu yang terjadi. Pengguna twitter menyebarkan konten yang berbeda-beda sesuai dengan pola kebiasaan pengguna, dari informasi tersebut terdapat data yang dapat diolah menjadi analisis sentimen.

Penelitian dilakukan untuk membangun perangkat lunak yang dapat mengumpulkan pesan twitter yang bersifat opini terhadap e-money seperti OVO, Dana dan Go-Pay dan di analisis sehingga mendapatkan kategori (positif, negatif dan undefined) dari pesan twitter tersebut. Dalam penelitian ini akan dilakukan analisis sentimen dengan menggunakan dua algoritma, yaitu *Naïve Bayes* dan *Decision Tree (C4.5)* Dan dibandingkan dengan metode *voting by majority*. Langkah pertama dalam penelitian ini yaitu mengumpulkan pesan twitter yang mengandung *keyword* mengenai aplikasi e-money (OVO, Dana dan Go-Pay) dengan menggunakan API Twitter (proses *crawling*). Langkah selanjutnya kedua yaitu tahap preprocessing lalu tahap ketiga pembobotan kata dengan *information gain*. Pengolahan data yang mempersiapkan data untuk dapat dianalisis dengan algoritma *naïve bayes* dan *Decision Tree (C4.5)*. Langkah terakhir yaitu tahap analisis pesan twitter yang akan dibandingkan melalui metode *voting by majority* dan akan dikategorikan ke dalam kategori positif, negatif dan undefined.

Pada penelitian ini berhasil mendapatkan accuracy sebesar 98,1%, recall sebesar 100% dan yang terakhir precision 98%.

Kata Kunci : *e-money*, Twitter, Sentimen, Algoritma *naïve bayes*, Algoritma *Decision Tree (C4.5)*, *voting by majority*.

KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa karena berkat rahmat dan karuniaNya, penulis dapat menyelesaikan laporan penelitian skripsi yang berjudul “ANALISIS SENTIMEN PENGGUNA *E-MONEY* PADA TWITTER MENGGUNAKAN *ALGORITMA C4.5* DAN *NAÏVE BAYES* ” sebagai bentuk pertanggungjawaban penulis terhadap pelaksanaan penelitian skripsi dengan baik. Isi dari laporan ini adalah pemaparan terkait sistem rekomendasi yang dapat memberikan daftar referensi penelitian pendahulu dan rekomendasi dosen pembimbing tugas akhir yang sesuai topik tugas akhir mahasiswa dengan inputan berupa judul terbaru.

Adapun penelitian ini dilakukan sebagai persyaratan dalam memperoleh gelar Sarjana Komputer Program Studi Teknik Informatika di Universitas Pembangunan Nasional “Veteran” Jawa Timur. Penulis menyadari bahwa laporan penelitian tugas akhir atau skripsi ini masih jauh dari kesempurnaan. Atas segala kekurangan dan ketidaksempurnaan penelitian ini, penulis sangat mengharapkan masukan, kritik, dan saran yang bersifat membangun kearah perbaikan dan penyempurnaan dalam proses penelitian dan pembuatan laporan. Akhir kata penulis mengucapkan terimakasih kepada semua pihak yang berperan dalam penelitian skripsi ini. Penulis berharap semoga laporan penelitian skripsi ini dapat bermanfaat bagi semua pihak.

Surabaya, 15 Juli 2020

Penulis

UCAPAN TERIMA KASIH

Dalam penyusunan dan penulisan skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan

ini penulis menyampaikan terima kasih kepada terutama kepada Allah SWT yang selalu ada dan memberi kelancaran, kemudahan, kekuatan, dan sumber segala pencerahan dan tempat pertama mengadu keluh kesah dan rasa gelisah penulis.

Tidak lupa penulis berterima kasih kepada yang terhormat:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Ni Ketut Sari, M.T. selaku Dekan Fakultas Ilmu Komputer.
3. Bapak Budi Nugroho, S.kom., M.kom., selaku Ketua Program Studi Teknik Informatika. Beserta Staff Dosen Teknik Informatika UPN “Veteran” Jawa Timur yang telah memberi dan membekali penulis dengan berbagai ilmu selama mengikuti perkuliahan sampai akhir penulisan skripsi.
4. Ibu Eva Yulia Puspaningrum, S.Kom. M.Kom dan bapak Wahyu Syaifullah JS.S.Kom. M.Kom selaku Dosen Pembimbing yang selalu memberikan dukungan, arahan dan doa dalam proses penyelesaian penelitian skripsi ini.

Kedua orang tua dan kakak tercinta, Moh.Zaini, Lilik Pujiati yang selalu mendoakan keberhasilan saya dan telah memberikan fasilitas untuk memudahkan saya dalam melakukan semua hal yang berkaitan dengan perkuliahan. Terimakasih atas semua bantuannya, semoga Tuhan Yang Maha

Esa memberikan balasan yang setimpal kepada semua pihak yang telah membantu penulis dengan dukungan secara finansial, emosional, doa dan waktu yang telah diluangkan, sehingga penulis telah berhasil melaksanakan penelitian dan proses penyusunan laporan penelitian skripsi dengan baik dan tepat.

Akhir kata, penulis menyadari dalam penulisan laporan penelitian skripsi ini masih jauh dari kata sempurna serta masih terdapat banyak kekurangan. Untuk itu penulis mengharapkan adanya masukan kritik dan saran yang bersifat membangun demi perbaikan proses penelitian atau penyusunan laporan yang mungkin akan dilakukan lagi di kemudian hari. Penulis berharap, semoga laporan penelitian skripsi ini memberikan dampak yang positif bagi pembaca dan khususnya bagi penulis juga.

Surabaya, 15 Juli 2020

Penulis

DAFTAR ISI

ANALISIS SENTIMEN PENGGUNA <i>E-MONEY</i> PADA TWITTER MENGGUNAKAN <i>ALGORITMA C4.5</i> DAN <i>NAÏVE BAYES</i>	i
ANALISIS SENTIMEN PENGGUNA <i>E-MONEY</i> PADA TWITTER MENGGUNAKAN <i>ALGORITMA C4.5</i> DAN <i>NAÏVE BAYES</i>	ii
LEMBAR PENGESAHAN SKRIPSI	i
SURAT PERNYATAAN ANTI PLAGIAT	Error! Bookmark not defined.
ABSTRAK	iii
KATA PENGANTAR	iv
UCAPAN TERIMA KASIH	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	xi
DAFTAR CODE.....	xii
DAFTAR PERSAMAAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan	4
1.5 Manfaat	5
BAB II TINJAUAN PUSTAKA	6
2.1. Penelitian Pendahulu.....	6
2.2. Landasan Teori	7
2.3.6. Konsep Algoritma <i>C4.5 (Decision tree)</i>	14
2.3.6. Konsep Naïve Bayes Classifier (NBC).....	25
2.3.7. Karakteristik <i>Naïve Bayes</i>	27
2.3.8. Metode Voting-by-majority	27
BAB III METODE PENELITIAN	29
3.1 Analisis Data.....	29
3.2. Analisis Sistem	29
BAB IV HASIL DAN PEMBAHASAN	52
4.1. Implementasi Data	52
4.2. Implementasi Antarmuka.....	52

1.3. Pembahasan	58
1.4. Analisis Uji Coba.....	67
BAB V KESIMPULAN DAN SARAN	79
5.1. Kesimpulan	79
5.2. Saran	79
DAFTAR PUSTAKA.....	80
LAMPIRAN	82
Hasil Crawling Yang Telah Diberi Label	82
Hasil Text Preprocessing	87
Hasil Klasifikasi Algoritma C4.5.....	89
Hasil Klasifikasi Algoritma Naive Bayes	102
Hasil Data Uji Setelah Di Klasifikasi Algoritma C4.5 Dan Naive Bayes Beserta Metode Voting By Majority	104

DAFTAR GAMBAR

Gambar 2. 1 Tahap Preprocessing	11
Gambar 2. 2 Contoh Case Folding dan Tokenizing	12
Gambar 2. 3 Contoh Filtering.....	13
Gambar 2. 4 Hasil Proses Stemming	14
Gambar 2. 5 Hasil Perhitungan Nilai Gain Simpul Akar	20
Gambar 2. 6 Hasil Perhitungan Nilai Gain Cabang Simpul Error Nilai 0	21
Gambar 2. 7 Hasil Perhitungan Nilai Gain Cabang Simpul Cepat Nilai 1.....	22
Gambar 2. 8 Proses Pengecekan Kalimat Opini.....	24
Gambar 3. 1 Skema Aliran Data pada Sistem.....	31
Gambar 3. 2 Diagram Alir Preprocessing	33
Gambar 3. 3 Diagram Alir Case folding dan Tokenizing	34
Gambar 3. 4 Diagram alir dari proses Filtering	36
Gambar 3. 5 Diagram Alir proses stemming	38
Gambar 3. 6 Pembentukan Pohon Keputusan Cabang akar.....	43
Gambar 3. 7 Pembentukan Pohon Keputusan Cabang Tidak	44
Gambar 3. 8 Pembentukan Pohon Keputusan Cabang bagus	45
Gambar 3. 9 Pembentukan Pohon Keputusan Cabang Mudah	47
Gambar 3. 10 Pembentukan Pohon Keputusan Cabang error.....	48
Gambar 4. 1 Tampilan menu Home	53
Gambar 4. 2 Tampilan menu crawling data	54
Gambar 4. 3 tampilan menu preprocessing.....	54
Gambar 4. 4 Tampilan menu Algoritma C4.5.....	55
Gambar 4. 5 Tampilan menu algoritma naive bayes.....	56

Gambar 4. 6 Tampilan menu Pengujian.....	57
Gambar 4. 7 Tampilan hasil data testing.....	57

DAFTAR TABEL

Tabel 2. 1 Contoh Training Data.....	17
Tabel 2. 2 Nilai Training Data.....	17
Tabel 2. 3 Perhitungan Information Gain Untuk Simpul Akar	19
Tabel 2. 4 Perhitungan Information Gain Untuk Cabang Simpul error	20
Tabel 2. 5 Perhitungan Information Gain Untuk Cabang Simpul.....	21
Tabel 3. 1 Contoh Tahap Case Folding.....	35
Tabel 3. 2 Contoh Tahap Tokenizing.....	35
Tabel 3. 3 Contoh Tahap Filtering	37
Tabel 3. 4 Contoh Tahap Stemming	40
Tabel 3. 5 Data Latih.....	41
Tabel 3. 6 Perhitungan Information Gain Simpul Akar	42
Tabel 3. 7 Perhitungan Information Gain Simpul Tidak.....	43
Tabel 3. 8 Perhitungan Information Gain Simpul bagus.....	44
Tabel 3. 9 Perhitungan Information Gain Simpul Mudah.....	46
Tabel 3. 10 Perhitungan Information Gain Simpul Tanpa.....	47
Tabel 3. 13 Perhitungan Probabilitas Data Latih	49
Tabel 3. 14 Metode voting by majority.....	50
Tabel 4. 1 Daftar Fuction dan Kegunaan untuk proses Stemming	61
Tabel 4. 2 Tabel Akurasi Sistem.....	Error! Bookmark not defined.
Tabel 4. 3 hasil akurasi dari semua metode	78

DAFTAR CODE

Kode 4. 1 Potongan Sourcecode fungsi koneksi dengan token	59
Kode 4. 2 Potongan Sourcecode case folding dan tokenizing	59
Kode 4. 3 Potongan Sourcecode Filtering.....	60
Kode 4. 4 Potongan sourcecode Stemming	61
Kode 4. 5 Potongan sourcecode Information Gain	63
Kode 4. 6 Potongan sourcecode C4.5	63
Kode 4. 7 Potongan sourcecode Naive Bayes.....	65
Kode 4. 8 Fungsi Naive Bayes	66
Kode 4. 9 Fungsi Voting by majority	67

DAFTAR PERSAMAAN

Persamaan 2.1 Perhitungan Gain	14
Persamaan 2.2 Perhitungan Nilai Entropy	15
Persamaan 2.3 Perhitungan Entropy Training Data.....	18
Persamaan 2.4 VMAP (1)	25
Persamaan 2.5 VMAP (2)	26
Persamaan 2.6 VMAP (3)	26
Persamaan 2.7 Probabilitas (1).....	26
Persamaan 2.8 Probabilitas (2).....	26