

**IMPLEMENTASI METODE NAÏVE BAYES UNTUK KLASIFIKASI
TINDAKAN PROSTITUSI ONLINE PADA TWITTER**

SKRIPSI



disusun oleh
Lukman Hakim
16.11.0120

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

**IMPLEMENTASI METODE NAÏVE BAYES UNTUK KLASIFIKASI
TINDAKAN PROSTITUSI ONLINE PADA TWITTER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh
Lukman Hakim
16.11.0120

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI METODE NAÏVE BAYES UNTUK KLASIFIKASI
TINDAKAN PROSTITUSI ONLINE PADA TWITTER**

yang dipersiapkan dan disusun oleh

Lukman Khakim

16.11.0120

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 5 Juni 2020

Dosen Pembimbing,

Arif Dwi Laksito, M.Kom.
NIK. 190302150

PENGESAHAN

SKRIPSI

IMPLEMENTASI METODE NAÏVE BAYES UNTUK KLASIFIKASI TINDAKAN PROSITUSI ONLINE PADA TWITTER

yang dipersiapkan dan disusun oleh
Lukman Khakim

16.11.0120

telah dipertahankan di depan Dewan Penguji
pada tanggal 19 Juni 2020

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ainul Yaqin, M. Kom.
NIK. 190302255

Akhmad Dahlan, M. Kom.
NIK. 190302174

Arif Dwi Laksito, M. Kom.
NIK. 190302150

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 Juni 2020

DEKAN FAKULTAS ILMU KOMPUTER

Krisnawati, S.Si, M.T.
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 19 Juni 2020



Lukman Khakim

NIM. 16.11.0120

MOTTO

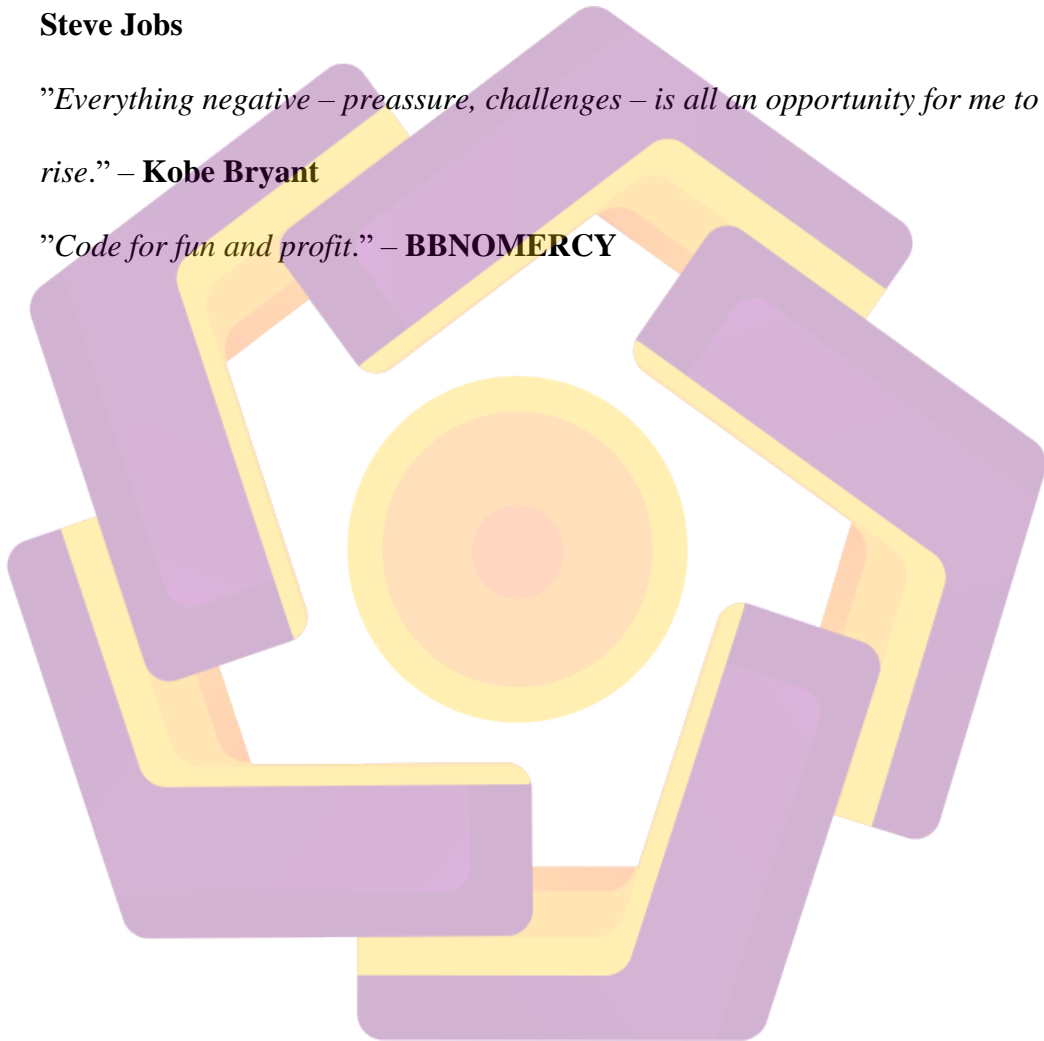
"Life is what happens when you're busy making other plans." - **John Lennon**

"Your time is limited, so don't waste it living someone else's life. Don't be trapped by dogma – which is living with the results of other people's thinking." –

Steve Jobs

"Everything negative – pressure, challenges – is all an opportunity for me to rise." – **Kobe Bryant**

"Code for fun and profit." – **BBNOMERCY**



PERSEMBAHAN

Saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung dalam proses pembuatan skripsi. Berkat rahmat Tuhan Yang Maha Esa yang memberikan segala nikmat dan kasih sayang-Nya hingga detik ini.

1. Kedua orang tua yang dengan ikhlas tanpa mengharapkan imbalan apapun dalam memberikan dukungan, do'a dan restunya tanpa henti untuk penulis.
2. Kakak dari penulis yang selalu memberikan dukungan dalam proses pengerjaan skripsi ini.
3. Bapak Arif Dwi Laksito selaku dosen pembimbing yang dengan kesabarannya membimbing dan mengajari banyak hal untuk penulis.
4. Febrilia Endah selaku teman yang tanpa henti memberikan dorongan untuk segera menyusun skripsi ini hingga selesai.
5. Teman-teman dekat penulis yang selalu memberikan semangat dalam menjalani hidup dari awal hingga saat ini.
6. Semua dosen dan karyawan Universitas Amikom Yogyakarta

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya dan shalawat serta salam juga tidak lupa penulis panjatkan kepada junjungan nabi kita Nabi Muhammad SAW yang telah memberikan teladan mulia dan menuntun umatnya sehingga penulis dapat menyelesaikan skripsi ini.

Skripsi yang berjudul “Impelementasi Metode Naïve Bayes untuk Melakukan Klasifikasi Tindakan Prostitusi Online pada Twitter” ini disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas Amikom Yogyakarta.

Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada:

1. Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta
2. Bapak Arif Dwi Laksito, M. Kom selaku dosen pembimbing yang telah banyak memberikan arahan bagi penulis agar penulis dapat menyelesaikan skripsi ini dengan baik.
3. Kedua orang tua, saudara, dan keluarga besar yang telah memberikan dukungan, doa, dan kasih sayangnya kepada penulis.
4. Seluruh dosen dan staff Universitas Amikom Yogyakarta yang telah membantu dan membimbing selama proses perkuliahan.
5. Teman-teman kelas 16 S1IF 02 dan teman-teman dekat.

6. Seluruh teman-teman dan keluarga besar Universitas Amikom Yogyakarta dan semua pihak yang tidak dapat saya sebutkan satu persatu, terima kasih banyak atas segala bantuannya dalam menyelesaikan skripsi ini.

Penulis menyadari skripsi ini masih ada kekurangan, maka dari itu kritik dan saran yang membangun serta teguran dari semua pihak, penulis menerima dengan lapang dada untuk kesempurnaan karya selanjutnya. Akhirnya kepada Allah SWT semoga skripsi yang sederhana ini dapat memberikan manfaat. Khususnya bagi penulis dan pembaca yang yang budiman pada umumnya.

Yogyakarta, 19 Juni 2020

Lukman Khakim

16.11.0120

DAFTAR ISI

JUDUL	I
PERSETUJUAN.....	II
PENGESAHAN	III
MOTTO	V
PERSEMBAHAN	VI
KATA PENGANTAR.....	VII
DAFTAR ISI	VIII
DAFTAR TABEL	XI
DAFTAR GAMBAR.....	XII
INTISARI.....	XIVII
ABSTRACT	XIV
BAB I PENDAHULUAN	1
1.1 LATAR BELAKANG	1
1.2 RUMUSAN MASALAH.....	4
1.3 BATASAN MASALAH	4
1.4 MAKSUD DAN TUJUAN PENELITIAN.....	5
1.5 MANFAAT PENELITIAN	5
1.6 METODE PENELITIAN.....	6
1.6.1 Metode Pengumpulan Data	6
1.6.2 Metode Analisis.....	6
1.6.3 Metode Perancangan.....	6
1.6.4 Metode Implementasi	7
1.6.5 Metode Testing.....	7
1.7 SISTEMATIKA PENULISAN.....	7
BAB II LANDASAN TEORI.....	9

2.1	KAJIAN PUSTAKA	9
2.2	DASAR TEORI	14
2.2.1	Pengertian Prostitusi	14
2.2.2	Pengertian Twitter	15
2.2.3	Text Mining	15
2.2.4	Text Preprocessing	16
2.2.5	Naive Bayes	18
2.2.6	Term Frequency-Inverse Document Frequency (TF-IDF)	20
2.2.7	Confusion Matrix	21
2.2.8	Precision	22
2.2.9	Recall	22
2.2.10	F1-Score	23
2.2.11	K-Fold Cross Validation	23
2.2.12	Python	24
BAB III METODE PENELITIAN		26
3.1	IDENTIFIKASI MASALAH	26
3.2	ANALISIS MASALAH	26
3.3	HASIL ANALISIS	27
3.4	ANALISIS KEBUTUHAN	27
3.4.1	Analisis Kebutuhan Fungsional	27
3.4.2	Analisis Kebutuhan Non-Fungsional	28
3.5	DESKRIPSI SISTEM	28
3.6	KLASIFIKASI ALGORITMA NAIVE BAYES	30
3.6.1	Pengumpulan Dataset	30
3.6.2	Cleaning Data	34
3.6.3	Case Folding	35
3.6.4	Tokenization	36
3.6.5	Stemming	37
3.6.6	TF-IDF	38
3.6.7	Naive Bayes	41
3.7	PERANCANGAN SISTEM	47
3.7.1	Diagram Pengembangan Sistem	47

3.7.2	Diagram Alur Preprocessing	48
3.7.3	Proses Klasifikasi Naive Bayes	50
BAB IV HASIL DAN PEMBAHASAN.....		52
4.1	DESKRIPSI IMPLEMENTASI.....	52
4.2	IMPLEMENTASI PENGUMPULAN DATA TRAINING	52
4.3	IMPLEMENTASI PELABELAN DATA	54
4.4	IMPLEMENTASI PERANCANGAN FUNGSI DAN SISTEM	55
4.4.1	Pengaksesan Data Training	55
4.4.2	Implementasi Data Cleaning dan Case Folding	56
4.4.3	Implementasi Tokenizing	57
4.4.4	Implementasi Stemming.....	58
4.4.5	Implementasi Pembobotan TF-IDF.....	58
4.4.6	Implementasi Klasifikasi Naive Bayes.....	58
4.4.7	Implementasi Pengukuran Akurasi Model Niave Bayes.....	60
4.4.8	Implementasi Klasifikasi Data Baru.....	61
4.5	EVALUASI PROGRAM.....	63
BAB V PENUTUP		69
5.1	KESIMPULAN	69
5.2	SARAN.....	70
DAFTAR PUSTAKA.....		74

DAFTAR TABEL

Table 2. 1 Tabel Penelitian Sebelumnya	11
Table 2. 2 Confusion matrix prediksi kelas	21
Table 3. 1 Contoh Data Hasil Crawling	30
Table 3. 2 Pelabelan data	32
Table 3. 3 Cleaning Data	34
Table 3. 4 Case folding	36
Table 3. 5 Tokenization	36
Table 3. 6 Stemming	37
Table 3. 7 Pengelompokan feature list	38
Table 3. 8 TF-IDF Feature True	39
Table 3. 9 TF-IDF Feature False	40
Table 3. 10 Data untuk pengujian naïve bayes	43
Table 3. 11 Pengecekan kata terhadap features model	44
Table 3. 12 Penentuan prior kelas	44
Table 3. 13 Hasil dan prosentase prediksi	45
Table 4. 1 Confussion Matrix Model Naive Bayes	63
Table 4. 2 Hasil 10 Fold Cross Validation	65
Table 4. 3 Confusion Matrix pada K-Fold Cross Validation	66
Table 4. 4 Evaluasi Pembagian Data Training	67

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Sistem	29
Gambar 3. 2 Diagram Pengembangan Sistem	47
Gambar 3. 3 Diagram Alur Preprocessing	48
Gambar 3. 4 Diagram Alur Proses Kalsifikasi Naïve Bayes	50
Gambar 4. 1 Script Crawling Data Twitter	53
Gambar 4. 2 Sample data hasil crawling	54
Gambar 4. 3 Pelabelan data training	55
Gambar 4. 4 Pengaksesan data training	55
Gambar 4. 5 Implementasi Fungsi Data Cleaning	56
Gambar 4. 6 Implementasi cleaning dataset	56
Gambar 4. 7 Implementasi Stopwords Removal	57
Gambar 4. 8 Implementasi fungsi <i>tokenizing</i>	57
Gambar 4. 9 Implementasi tokenizing	58
Gambar 4. 10 Implementasi Stemming	58
Gambar 4. 11 Implementasi Pembobotan TF-IDF	58
Gambar 4. 12 Implementasi Model Klasifikasi Naive Bayes	59
Gambar 4. 13 Implementasi Pengukuran Akurasi Model Naive Bayes	60
Gambar 4. 14 Implementasi klasifikasi data baru	61
Gambar 4. 15 Implementasi fungsi realtime crawling	62
Gambar 4. 16 Proses realtime klasifikasi	62
Gambar 4. 17 Implementasi K-Fold Cross Validation	64
Gambar 4. 18 Evaluasi K-Fold Cross Validation	66
Gambar 4. 19 Evaluasi Pembagian Data Training	67
Gambar 4. 20 Pengujian akurasi terhadap train size	68

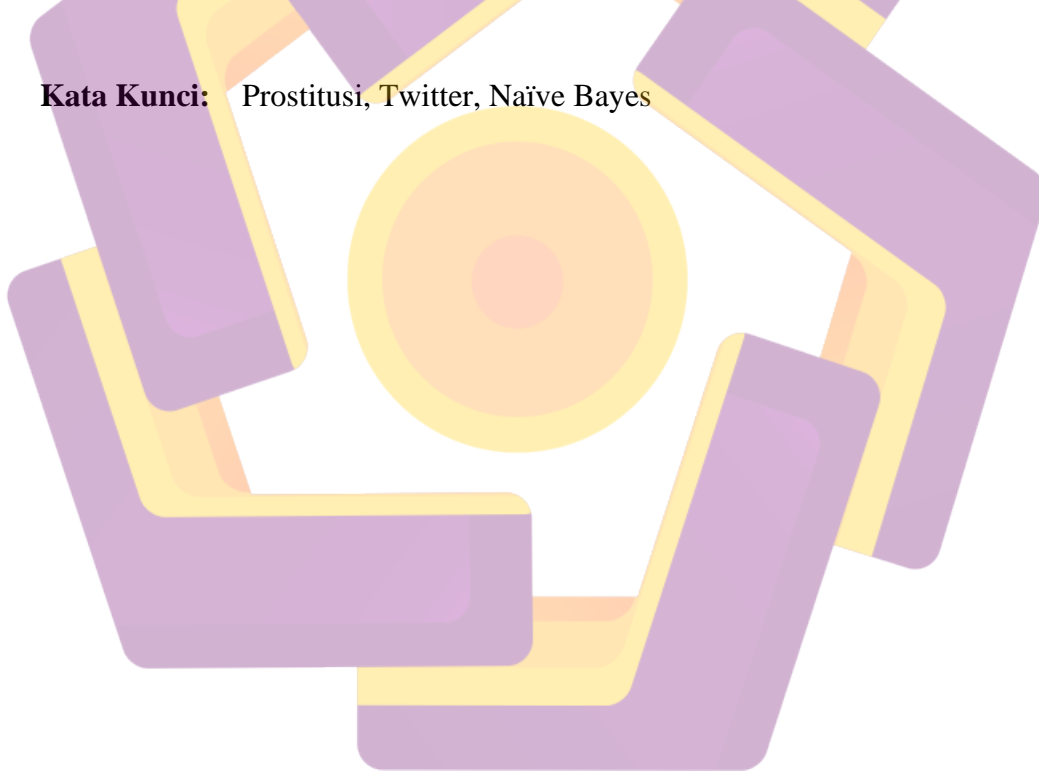
INTISARI

Kegiatan yang melanggar aturan pada media sosial menjadi salah satu dampak negatif dari penggunaan internet saat ini. Twitter menjadi tempat paling terbuka untuk melakukan hal tersebut karena kemudahan dalam penggunaannya. Salah satu kegiatan negatif yang tersebar pada Twitter saat ini adalah tindakan prostitusi. Para pelaku menjajakan dirinya dengan mengirim tweet yang mengacu dalam tindakan prositusi.

Dengan teknik klasifikasi teks menggunakan metode Naïve Bayes dapat dilakukan prediksi terhadap tweet yang mengarah pada tindakan prostitusi dengan melakukan klasifikasi tweet dari dataset yang telah didapatkan. Implementasi metode Naïve Bayes diharapkan mampu melakukan prediksi dengan tingkat akurasi yang baik untuk melakukan klasifikasi tweet pada Twitter.

Hasil penelitian dari 40000 dataset yang dibagi menjadi 80% data training dan 30% data testing diperoleh tingkat akurasi metode Naïve Bayes sebesar 98,41% dengan nilai precision sebesar 98%, nilai recall sebesar 98% dan nilai f1-measure sebesar 98%.

Kata Kunci: Prostitusi, Twitter, Naïve Bayes



ABSTRACT

Activities that violate the rules on social media are one of the negative impacts of using the internet today. Twitter is the most open place to do this because of its ease of use. One of the negative activities that are spreading on Twitter at this time is the act of prostitution. The perpetrators peddled themselves by sending tweets that referred to acts of prostitution.

With the text classification technique using the Naïve Bayes method, predictions of tweets that lead to prostitution can be made by classifying tweets from the dataset that has been obtained. The implementation of the Naïve Bayes method is expected to be able to make predictions with a good level of accuracy for classifying tweets on Twitter.

The results of research from 40000 datasets that were divided into 80% training data and 30% testing data obtained the accuracy rate of the Naïve Bayes method of 98.41% with a precision value of 98%, a recall value of 98% and an f1-measure value of 98%.

Keyword: *Prostitution, Twitter, Naïve Bayes*

