

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK MENDETEKSI
KOMENTAR PELECEHAN SEKSUAL TERHADAP TOKOH PUBLIK
DI INSTAGRAM**

SKRIPSI



disusun oleh

Dian Indriastuti

15.11.9140

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK MENDETEKSI
KOMENTAR PELECEHAN SEKSUAL TERHADAP TOKOH PUBLIK
DI INSTAGRAM**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Dian Indriastuti

15.11.9140

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2019**

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK MENDETEKSI
KOMENTAR PELECEHAN SEKSUAL TERHADAP TOKOH PUBLIK
DI INSTAGRAM**

yang dipersiapkan dan disusun oleh

Dian Indriastuti

15.11.9140

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 7 Oktober 2019

Dosen Pembimbing,



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

PENGESAHAN
SKRIPSI
IMPLEMENTASI NAÏVE BAYES CLASSIFIER UNTUK MENDETEKSI
KOMENTAR PELECEHAN SEKSUAL TERHADAP TOKOH PUBLIK
DI INSTAGRAM

yang dipersiapkan dan disusun oleh

Dian Indriastuti

15.11.9140

telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Oktober 2019

Susunan Dewan Penguji

Nama Penguji

Bety Wulan Sari, M.Kom.
NIK. 190302254

Acihmah Sidauruk, M.Kom.
NIK. 190302238

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

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
tanggal 13 November 2019



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, 29 Oktober 2019



Dian Indriastuti

15.11.9140

MOTTO

“Balas dendam terbaik adalah dengan memperbaiki dirimu.” – Ali Bin Abi Thalib

“Jika kau tak suka sesuatu, ubahlah. Jika tak bisa, maka ubahlah cara pandangmu tentangnya.” – Maya Angelou

“Nikmat paling indah adalah takut pada Tuhan.” – SYP

“Time keep ticking, a well time spent is a must.” – DW

“Your time is yours. My time is mine. But everytime is God’s” – DI

“Setiap baik dan buruk yang terjadi pasti ada maksudnya. Tapi selalu kembali pada Tuhan adalah jawabannya.” – DI

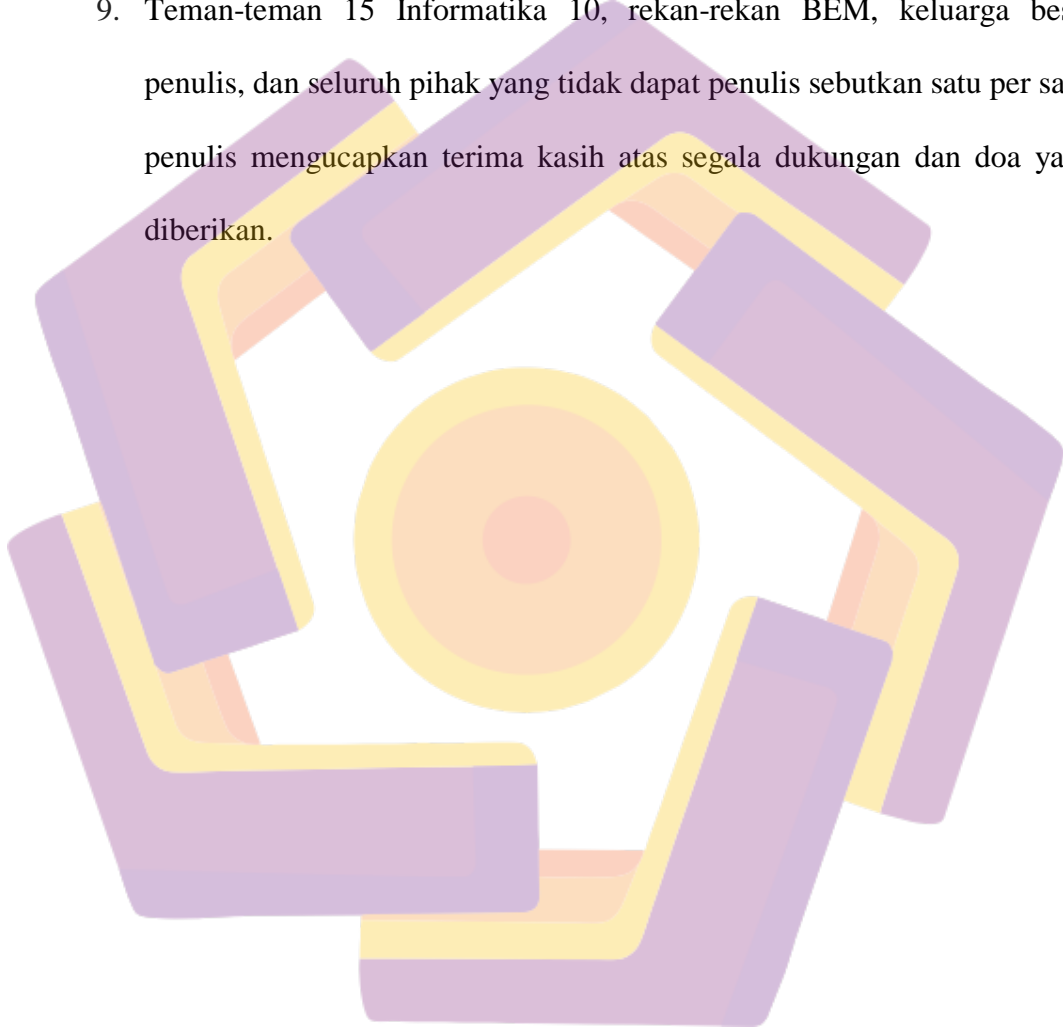
PERSEMBAHAN

Alhamdulillah rabbil 'alamin segala puji bagi Allah SWT atas curahan rahmat, anugerah, cinta, dan kasih sayang-Nya. Segala syukur penulis ucapkan kepada Allah SWT telah memberikan kesempatan, kemudahan, ilmu, sehat, serta manusia-manusia terbaik di sekeliling penulis yang sangat membantu dalam penyelesaian skripsi ini.

Dengan bangga dan penuh syukur, penulis persembahkan karya yang sederhana ini kepada mereka:

1. Ibunda tercinta, Sri Hastuti, terima kasih telah menjadi Ibu yang luar biasa. Cinta, kasih, serta sayangmu tak terhitung. Segala sabar dan ikhlas yang tercurah tak akan tergantikan. Segala kebaikan tak akan pernah sanggup terbalaskan.
2. Ayahanda tercinta, Suwarno, terima kasih telah menjadi Ayah dengan sabarnya tiada tertandingi. Cinta dan kasihmu mengajarkan banyak hal yang tak terlupakan.
3. Rani Widya Prananingrum, terima kasih telah menjadi sosok kakak yang kuat, mengajarkan pentingnya komunikasi dalam kehidupan
4. Azaria Puspitasari, terima kasih telah menjadi sosok adik yang mengajarkan selalu bersabar dalam kehidupan.
5. Ibu Krisnawati selaku pembimbing, terima kasih atas segala kesempatann, waktu, bimbingan, serta nasehatnya dalam penyelesaian skripsi ini.
6. Abinda Dwi Nur Ahmad, terima kasih telah menjadi partner belajar, berdiskusi dan berdebat yang luar biasa.

7. Beta Priyoko, terima kasih atas waktu serta bimbingan program yang sangat membantu dalam skripsi ini.
8. Mitha, Shoffi, Yuli, Kharis, Syifa, Afan, Nur, Gray, Santoso, Rizki dan seluruh sahabat terbaik yang mendukung dan menjadi tempat sambat berlabuh, tanpanya skripsi ini akan terasa sepi.
9. Teman-teman 15 Informatika 10, rekan-rekan BEM, keluarga besar penulis, dan seluruh pihak yang tidak dapat penulis sebutkan satu per satu, penulis mengucapkan terima kasih atas segala dukungan dan doa yang diberikan.



KATA PENGANTAR

Bismillahirrahmanirrahim,

Alhamdulillah *rabbil 'alamin*, segala puji bagi Allah SWT atas curahan rahmat, anugerah, cinta, dan kasih sayang-Nya, serta sholawat beriring salam kepada Nabi Muhammad SAW sehingga penulis diberikan kelancaran dalam menyelesaikan skripsi yang berjudul **"Implementasi Naïve Bayes Classifier untuk Mendeteksi Komentar Pelecehan Seksual terhadap Tokoh Publik di Instagram"** dengan baik.

Skripsi ini disusun sebagai salah satu syarat kelulusan program S1 Informatika Universitas AMIKOM Yogyakarta. Dalam penyusunan skripsi ini tidak lepas dari arahan, bimbingan, dan dukungan dari berbagai pihak secara langsung maupun tidak langsung. Oleh karena itu penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta sekaligus pembimbing yang telah memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.
3. Bapak Sudarmawan, M.T. selaku ketua jurusan S1-Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Bety Wulan Sari, M.Kom. dan Ibu Acihmah Sidauruk, M.Kom. selaku dosen penguji yang memberikan wawasan baru serta masukan untuk perkembangan dan perbaikan skripsi ini.

5. Para dosen yang telah membagi pengetahuan dan ilmu yang berharga selama perkuliahan.
6. Para penulis sumber bacaan, jurnal, dan makalah yang penulis jadikan referensi dalam penyusunan skripsi ini.
7. Keluarga, sahabat dan teman-teman yang telah berpartisipasi dan memberikan doa serta dukungan dalam penyusunan skripsi secara langsung maupun tidak langsung.

Penulis menyadari dalam skripsi ini masih jauh dari kata sempurna. Oleh karenanya, penulis akan terbuka terhadap saran serta kritik yang membangun sebagai bahan perbaikan ke depan.

Akhir kata, semoga karya yang telah disusun ini dapat memberikan manfaat serta sumbangsih untuk menambah ilmu pengetahuan pembaca.

Terima kasih.

Yogyakarta, 29 Oktober 2019
Penulis,

Dian Indriastuti
15.11.9140

DAFTAR ISI

| | |
|---------------------------------------|-------|
| HALAMAN JUDUL | I |
| PERSETUJUAN..... | II |
| PENGESAHAN | III |
| PERNYATAAN..... | IV |
| MOTTO..... | V |
| PERSEMBAHAN | VI |
| KATA PENGANTAR..... | VIII |
| DAFTAR ISI | X |
| DAFTAR TABEL | XIII |
| DAFTAR GAMBAR..... | XV |
| INTISARI..... | XVII |
| ABSTRACT | XVIII |
| BAB I PENDAHULUAN | 1 |
| 1.1 LATAR BELAKANG | 1 |
| 1.2 RUMUSAN MASALAH..... | 3 |
| 1.3 BATASAN MASALAH | 4 |
| 1.4 MAKSUD DAN TUJUAN PENELITIAN..... | 4 |
| 1.5 METODOLOGI PENELITIAN..... | 5 |
| 1.5.1 Metode Pengumpulan Data | 5 |
| 1.5.2 Metode Analisis..... | 5 |
| 1.5.3 Metode Perancangan | 6 |
| 1.5.4 Metode Pengujian..... | 6 |
| 1.6 SISTEMATIKA PENULISAN..... | 6 |
| BAB II LANDASAN TEORI | 8 |

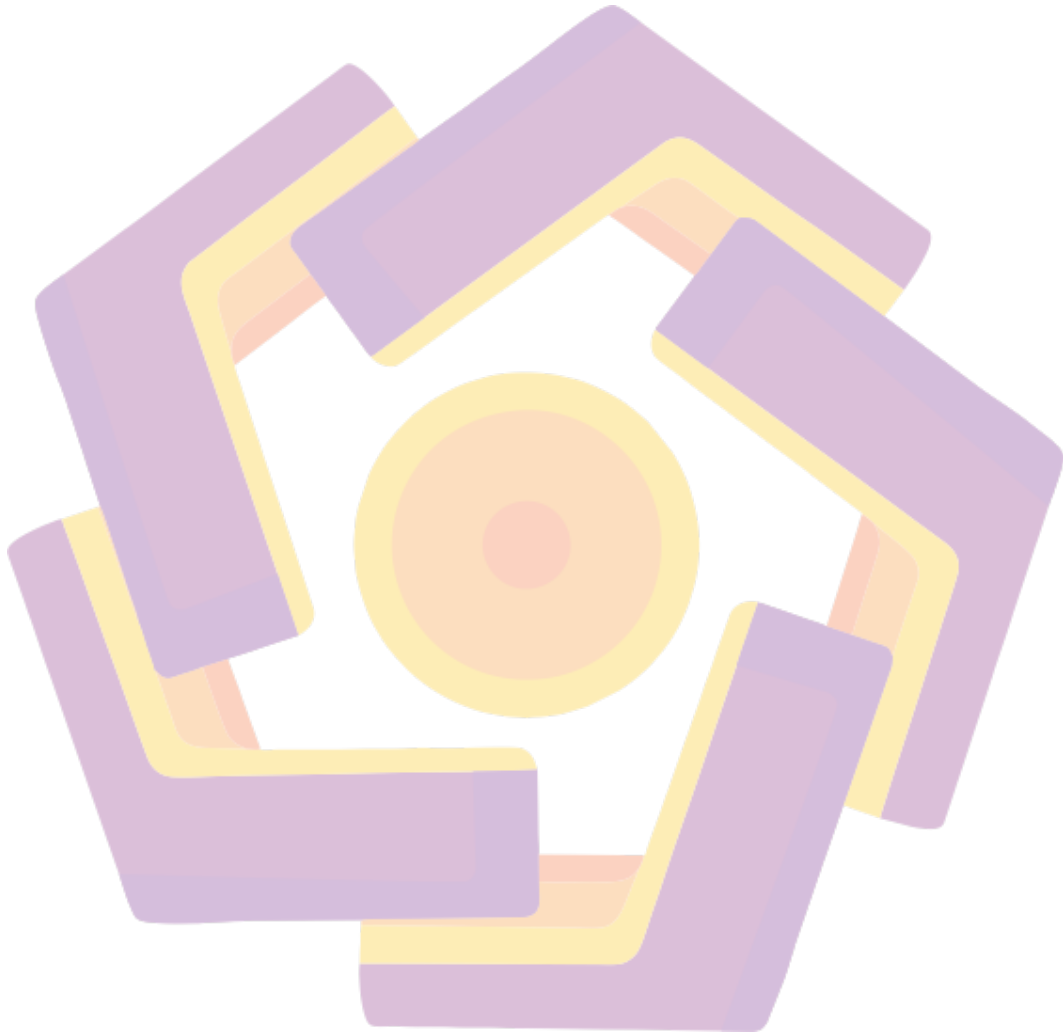
| | | |
|---|--|-----------|
| 2.1 | KAJIAN PUSTAKA | 8 |
| 2.2 | DASAR TEORI | 11 |
| 2.2.1 | Instagram | 11 |
| 2.2.2 | Tokoh Publik | 12 |
| 2.2.3 | Pelecehan Seksual | 13 |
| 2.2.4 | Data Mining | 15 |
| 2.2.5 | Text Mining | 16 |
| 2.2.6 | Klasifikasi | 18 |
| 2.2.7 | Naïve Bayes Classifier | 18 |
| 2.2.8 | Confusion Matrix | 21 |
| 2.2.9 | K-Fold Cross Validation | 23 |
| 2.3 | SYSTEM DEVELOPMENT LIFE CYCLE (SDLC) | 24 |
| 2.3.1 | Langkah-langkah Metode SDLC | 25 |
| 2.4 | ANALISIS | 27 |
| 2.4.1 | Analisis Kebutuhan | 27 |
| 2.5 | PERANCANGAN | 28 |
| 2.5.1 | Data Flow Diagram (DFD) | 28 |
| 2.5.2 | Flowchart | 30 |
| 2.6 | PENGUJIAN | 33 |
| BAB III ANALISIS DAN PERANCANGAN | | 36 |
| 3.1 | ANALISIS MASALAH | 36 |
| 3.1.1 | Identifikasi Masalah | 36 |
| 3.1.2 | Hasil Analisis | 37 |
| 3.2 | ANALISIS SISTEM | 39 |
| 3.2.1 | Analisis Kebutuhan | 39 |
| 3.3 | ANALISIS ALGORITMA | 40 |
| 3.3.1 | Analisis Kebutuhan Data | 40 |
| 3.3.2 | Pelabelan Data | 41 |
| 3.3.3 | Preprocessing | 42 |
| 3.3.4 | Klasifikasi | 44 |
| 3.4 | PERANCANGAN SISTEM | 49 |
| 3.4.1 | DFD | 49 |

| | | |
|---|---|----|
| 3.4.2 | Perancangan Antarmuka..... | 55 |
| BAB IV IMPLEMENTASI DAN PEMBAHASAN..... | | 59 |
| 4.1 | IMPLEMENTASI | 59 |
| 4.1.1 | Pengumpulan Dataset Training | 59 |
| 4.1.2 | Implementasi Antar Muka..... | 60 |
| 4.2 | PEMBAHASAN SOURCE CODE | 65 |
| 4.2.1 | Source Code Pengambilan Data | 66 |
| 4.2.2 | Source Code Training Data | 66 |
| 4.2.3 | Source Code Preprocessing Data..... | 67 |
| 4.2.4 | Source Code Klasifikasi | 68 |
| 4.3 | PENGUJIAN SISTEM..... | 69 |
| 4.3.1 | White Box Testing | 69 |
| 4.3.2 | Black Box Testing | 70 |
| 4.4 | HASIL DAN ANALISIS PENGUJIAN MODEL | 72 |
| 4.4.1 | Pengujian Model Menggunakan 10 Fold Cross Validation dan Confusion Matrix..... | 72 |
| 4.4.2 | Hasil Pengujian..... | 73 |
| BAB V PENUTUP | | 87 |
| 5.1 | KESIMPULAN | 87 |
| 5.2 | SARAN | 87 |
| DAFTAR PUSTAKA..... | | 87 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2. 1 Tabel Referensi Penelitian..... | 10 |
| Tabel 2. 2 Tabel Referensi Penelitian (Lanjutan)..... | 11 |
| Tabel 2. 3 Tabel Confusion Matrix | 22 |
| Tabel 2. 4 Simbol-simbol Data Flow Diagram (DFD)..... | 29 |
| Tabel 2. 5 Tabel Flowchart..... | 31 |
| Tabel 2. 6 Tabel Flowchart (Lanjutan)..... | 32 |
| Tabel 3. 1 Contoh Komentar | 40 |
| Tabel 3. 2 Contoh Pelabelan Komentar..... | 41 |
| Tabel 3. 3 Contoh Pelabelan Komentar (Lanjutan)..... | 42 |
| Tabel 3. 4 Contoh Case Folding Komentar | 43 |
| Tabel 3. 5 Contoh Tokenizing Komentar | 43 |
| Tabel 3. 6 Contoh Tokenizing Komentar (Lanjutan) | 44 |
| Tabel 3. 7 Contoh Stemming Komentar | 44 |
| Tabel 3. 8 Jumlah Dokumen..... | 46 |
| Tabel 3. 9 Token Frekuensi | 46 |
| Tabel 3. 10 Token Frekuensi (Lanjutan) | 47 |
| Tabel 4. 1 White box Testing | 70 |
| Tabel 4. 2 Black Box Testing..... | 70 |
| Tabel 4. 3 Black Box Testing (Lanjutan)..... | 71 |
| Tabel 4. 4 Hasil Pengujian Confusion Matrix Fold 1..... | 74 |
| Tabel 4. 5 Hasil Pengujian Confusion Matrix Fold 2..... | 75 |
| Tabel 4. 6 Hasil Pengujian Confusion Matrix Fold 2 (Lanjutan)..... | 76 |
| Tabel 4. 7 Hasil Pengujian Confusion Matrix Fold 3..... | 77 |
| Tabel 4. 8 Hasil Pengujian Confusion Matrix Fold 4..... | 78 |
| Tabel 4. 9 Hasil Pengujian Confusion Matrix Fold 5..... | 79 |
| Tabel 4. 10 Hasil Pengujian Confusion Matrix Fold 6..... | 80 |
| Tabel 4. 11 Hasil Pengujian Confusion Matrix Fold 7..... | 81 |
| Tabel 4. 12 Hasil Pengujian Confusion Matrix Fold 7 (Lanjutan)..... | 82 |
| Tabel 4. 13 Hasil Pengujian Confusion Matrix Fold 8..... | 83 |
| Tabel 4. 14 Hasil Pengujian Confusion Matrix Fold 9..... | 84 |

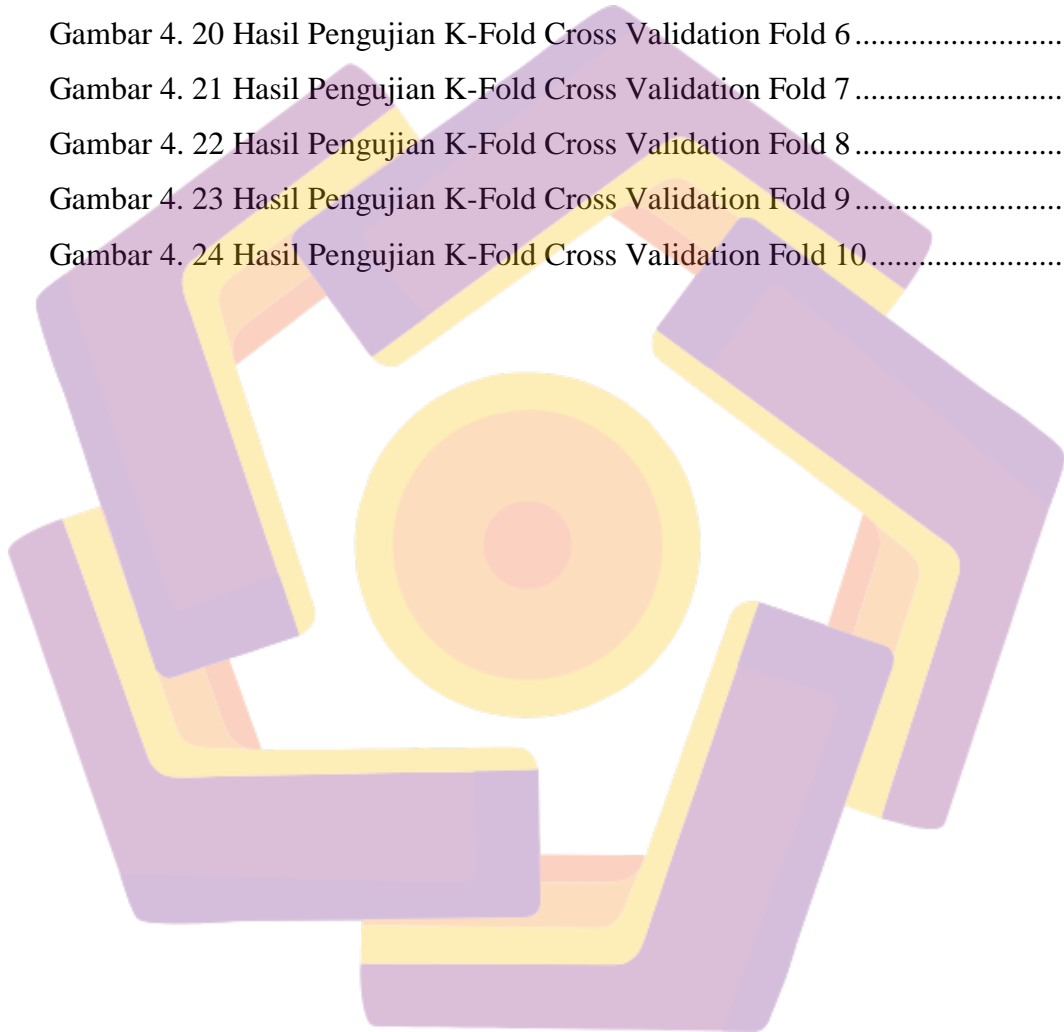
Tabel 4. 15 Hasil Pengujian Confusion Matrix Fold 10.....85
Tabel 4. 16 Hasil Pengujian K-Fold Cross Validation dan Confusion Matrix
Seluruh Fold86



DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2. 1 Skenario K Fold Cross Validation..... | 24 |
| Gambar 2. 2 Tahapan SDLC | 25 |
| Gambar 3. 1 Alur Proses Klasifikasi | 38 |
| Gambar 3. 2 Flowchart Preprocessing..... | 42 |
| Gambar 3. 3 Flowchart Naive Bayes Classifier | 45 |
| Gambar 3. 4 Diagram Konteks | 49 |
| Gambar 3. 5 DFD Level 1 | 50 |
| Gambar 3. 6 DFD Level 2 Proses Training | 50 |
| Gambar 3. 7 DFD Level 2 Proses Klasifikasi | 51 |
| Gambar 3. 8 DFD Level 2 Proses Evaluasi | 52 |
| Gambar 3. 9 DFD Level 3 Proses Training | 53 |
| Gambar 3. 10 DFD Level 3 Proses Klasifikasi | 54 |
| Gambar 3. 11 DFD Level 3 Proses Evaluasi | 55 |
| Gambar 3. 12 UI Halaman Utama | 56 |
| Gambar 3. 13 UI Navigation Bar | 56 |
| Gambar 3. 14 UI Menu Training | 57 |
| Gambar 3. 15 UI Menu Classification..... | 58 |
| Gambar 3. 16 UI Halaman Evaluation | 58 |
| Gambar 4. 1 Data Komentar Harrasment | 60 |
| Gambar 4. 2 Data Komentar Not Harrasment | 60 |
| Gambar 4. 3 Implementasi Halaman Utama | 61 |
| Gambar 4. 4 Implementasi Navigation Bar | 62 |
| Gambar 4. 5 Implementasi Halaman Training | 62 |
| Gambar 4. 6 Halaman Training Result..... | 63 |
| Gambar 4. 7 Implementasi Halaman Classification | 64 |
| Gambar 4. 8 Halaman Classification Result..... | 64 |
| Gambar 4. 9 Implementasi Halaman Evaluation..... | 65 |
| Gambar 4. 10 Halaman Evaluation Result | 65 |
| Gambar 4. 11 Source Code Scrapping Data | 66 |
| Gambar 4. 12 Source Code Training Data | 67 |

| | |
|--|----|
| Gambar 4. 13 Source Code Preprocessing Data..... | 68 |
| Gambar 4. 14 Source Code Klasifikasi | 69 |
| Gambar 4. 15 Hasil Pengujian K-Fold Cross Validation Fold 1 | 74 |
| Gambar 4. 16 Hasil Pengujian K-Fold Cross Validation Fold 2 | 75 |
| Gambar 4. 17 Hasil Pengujian K-Fold Cross Validation Fold 3 | 76 |
| Gambar 4. 18 Hasil Pengujian K-Fold Cross Validation Fold 4 | 77 |
| Gambar 4. 19 Hasil Pengujian K-Fold Cross Validation Fold 5 | 79 |
| Gambar 4. 20 Hasil Pengujian K-Fold Cross Validation Fold 6 | 80 |
| Gambar 4. 21 Hasil Pengujian K-Fold Cross Validation Fold 7 | 81 |
| Gambar 4. 22 Hasil Pengujian K-Fold Cross Validation Fold 8 | 82 |
| Gambar 4. 23 Hasil Pengujian K-Fold Cross Validation Fold 9 | 83 |
| Gambar 4. 24 Hasil Pengujian K-Fold Cross Validation Fold 10 | 85 |



INTISARI

Beragamnya pengguna Instagram, beragam pula tujuannya. Bagi tokoh publik, Instagram adalah salah satu media yang digunakan untuk lebih dekat dengan penggemar dan masyarakat. Postingan apapun dapat menjadi sorotan publik, tidak selalu komentar positif yang mereka dapatkan namun tidak jarang pula ada komentar-komentar negatif, seperti *bullying*, bahkan asusila yang bisa jadi termasuk dalam pelecehan seksual. Meskipun korban perempuan lebih banyak, tapi laki-laki juga ada yang menjadi korban. Oleh karenanya, deteksi komentar pelecehan seksual dibutuhkan.

Penelitian ini menggunakan metode *Naïve Bayes Classifier* untuk mengklasifikasikan komentar-komentar pada publik figur di Instagram. Komentar tersebut akan digolongkan ke dalam dua kelas yaitu kelas *harassment* untuk komentar yang tergolong pelecehan seksual dan kelas *not harassment* untuk komentar yang tidak tergolong pelecehan seksual.

Dataset yang digunakan untuk proses *training* sebanyak 1000 data, dengan komposisi 500 komentar *harassment* dan 500 komentar *not harassment*. Penelitian ini diawali dengan proses *preprocessing*, dilanjutkan proses *training* dan klasifikasi. Berdasarkan hasil evaluasi yang menggunakan *10 fold cross validation* dan *confusion matrix*, didapatkan hasil akurasi rata-rata sebesar 82%, dengan *error rate* 18%, *precision* 85%, dan *recall* sebesar 80%. Dari *10 fold*, akurasi tertinggi terdapat pada *fold* 3 dengan hasil akurasi 88%, sedangkan akurasi terendah yaitu pada *fold* 8 sebesar 73%. Dari hasil tersebut, *Naïve Bayes Classifier* dapat diterapkan untuk mengklasifikasikan komentar pelecehan seksual di Instagram.

Kata kunci: Klasifikasi, Naïve Bayes Classifier, Pelecehan Seksual, Instagram, 10-Fold Cross Validation, Confusion Matrix

ABSTRACT

The diversity of Instagram users, also various their intentions. For the public figures, Instagram is one of the media that used to get closer to their fans and public. Any post can be in the public spotlight, not always positive comments that they got but not infrequently also negative comments, such as bullying, and even immoral that may be included in sexual harassment comment. Although there are more female victims, there are also male victims. Therefore, detection of sexual harassment comments is needed.

This study uses the Naïve Bayes Classifier method to classify comments on public figures on Instagram. The comments will be classified into two classes, namely the harassment class for comments which are classified as sexual harassment and not harassment class for comments which are not classified as sexual harassment.

The data that used for the training process amounted to 1000 data, with the contribution of 500 harassment comments and 500 not harassment comments. The research begins with the preprocessing data, followed by the training and classification process. Based on the evaluation results that used 10 fold cross validation and confusion matrix, showed an average accuracy of 82%, with an error rate of 18%, precision of 85%, and recall of 80%. Out of 10 times, the highest accuracy in fold 3 with the accuracy result of 88%, while the lowest accuracy in fold 8 of 73%. From these results, Naïve Bayes Classifier can be applied to classify sexy harassment comments on Instagram.

Keywords: *Classification, Naïve Bayes Classifier, Sexual Harassment, Instagram, 10-Fold Cross Validation, Confusion Matrix*