

**KLASIFIKASI TWEET TIDAK SENONOH DI TWITTER
DENGAN NAÏVE BAYES CLASSIFIER**

SKRIPSI



disusun oleh

Yurike Pratiwi

17.12.0397

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

**KLASIFIKASI TWEET TIDAK SENONOH DI TWITTER
DENGAN NAÏVE BAYES CLASSIFIER**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Sistem Informasi



disusun oleh

Yurike Pratiwi

17.12.0397

**PROGRAM SARJANA
PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

KLASIFIKASI TWEET TIDAK SENONOH DI TWITTER DENGAN NAÏVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh

Yurike Pratiwi

17.12.0397

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Februari 2022

Dosen Pembimbing,

Ainul Yaqin, M.Kom.

NIK.190302255

PENGESAHAN

SKRIPSI

KLASIFIKASI TWEET TIDAK SENONOH DI TWITTER DENGAN NAÏVE BAYES CLASSIFIER

yang dipersiapkan dan disusun oleh
Yurike Pratiwi

17.12.0397

telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Februari 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Agus Fatkhurohman, M.Kom
NIK. 190302249

Irma Rofni Wulandari, S.Pd., M.Eng
NIK. 190302238

Ainul Yaqin, M.Kom.
NIK. 190302255

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Februari 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, M.Kom.
NIK. 190302096

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, 22 Februari 2022



Yurike Pratiwi
NIIM. 17.12.0397

MOTTO

"Ubahlah hidupmu dari hari ini. Jangan pernah bertaruh pada masa depan, kamu harus bertindak sekarang tanpa menunda-nunda."

– Simone de Beauvoir

"Untuk jadi maju memang banyak tantangan dan hambatan. Kecewa semenit dua menit boleh, tetapi setelah itu harus bangkit lagi."

-Jokowi

"Maka apabila kamu telah selesai (dari sesuatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain, dan hanya kepada Tuhanmulah hendaknya kamu berharap."

-QS. Al-Insyirah:7-8

PERSEMBAHAN

Segala puji dan syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat, taufik, hidayah, serta inayah-Nya kepada penulis sehingga dapat menyelesaikan skripsi ini sesuai dengan target dan mendapatkan hasil yang terbaik.

Dalam kesempatan kali ini, penulis ingin mengucapkan banyak terimakasih kepada semua pihak yang telah membantu penulis dalam menyelesaikan skripsi ini, yaitu:

1. Alm. Bapak saya, Tri Prajoko yang menjadikan motivasi terhadap diri saya untuk terus berjalan kedepan.
2. Ibu saya, Secar Roro Wilis yang selalu mendoakan saya, selalu mendukung dalam segala hal dan selalu memberikan motivasi dan nasihat. Selalu mengingatkan untuk melaksanakan ibadah wajib yaitu sholat 5 waktu tepat pada waktunya. Terimakasih tanpa ibu, saya tidak akan sampai di titik ini, terimakasih sudah mau mengorbankan banyak hal untuk keberhasilan dan kebahagiaan putrimu ini.
3. Terimakasih kepada Kakak saya Danang Pratama dan adik saya Prima Pramudya Wijaya karena sudah memberi semangat selama pembuatan skripsi ini.
4. Terimakasih kepada Padang perwira yudha karena sudah banyak berbagi ilmu dan bantuan selama pengerjaan skripsi ini.
5. Terimakasih kepada bapak Ainul Yaqin M.Kom, selaku dosen pembimbing. Terimakasih atas bimbingan dan dukungannya sehingga dapat menyelesaikan skripsi dengan baik.
6. Terimakasih kepada teman teman 17 Sistem Informasi 06 yang menjadi tempat berbagi ilmu, serta canda dan tawa.
7. Terimakasih kepada teman seperjuangan

(Amanda,aziz,erlangga,faizal,giden,hilal,nuha,sigid,taufiq, dan cahyo) yang tergabung pada “LUCKNUT SQUAD” udah menjadi partner kerja kelompok mulai dari semester 2 hingga semester 6. Terimakasih sudah mau menjadi teman mengeluh dan sudah banyak membantu untuk setiap project. Sehingga saya dapat menyisihkan waktu untuk pengerjaan skripsi ini.

8. Terimakasih kepada mantan sekelas saya,Yoga dari semester 1 sampai sekarang. Terimakasih sudah mengajarkan hal-hal baik.Dan selalu support dalam pengerjaan skripsi ini.
9. Terimakasih kepada sahabat saya,Tri Asih Naibaho yang selalu menjadi tempat keluh kesah dalam mengerjakan skripsi.
10. Terimakasih kepada semua teman-teman dan semua pihak yang tidak dapat disebutkan satu per satu yang sesudah bersedia membantu dan mendukung penulis, sehingga penulis dapat menyelesaikan skripsi ini.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat, taufik, hidayah, serta inayah-Nya, sehingga penulis dapat menyelesaikan skripsi ini sesuai dengan waktu yang diinginkan penulis. Tidak lupa sholawat serta salam penulis haturkan kepada Nabi Agung Muhammad SAW.

Skripsi ini disusun dalam rangka memenuhi salah satu syarat kelulusan jenjang Program Sarjana pada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.

Proses penyusunan hingga selesainya laporan skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak baik secara langsung maupun tidak langsung. Maka dari itu, sebagai rasa hormat penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M, selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Ainul Yaqin, M.Kom, selaku dosen pembimbing penulis yang telah memberikan petunjuk, bimbingan dan nasihatnya dalam proses penulisan skripsi ini.
4. Kedua orang tua dan serta saudara dan saudari yang telah memberikan motivasi, doa, masukan dan semangat.
5. Teman-teman seperjuangan Mahasiswa/i 17-S1 Sistem Informasi-06, yang telah banyak berdiskusi dengan penulis dalam masa pendidikan.

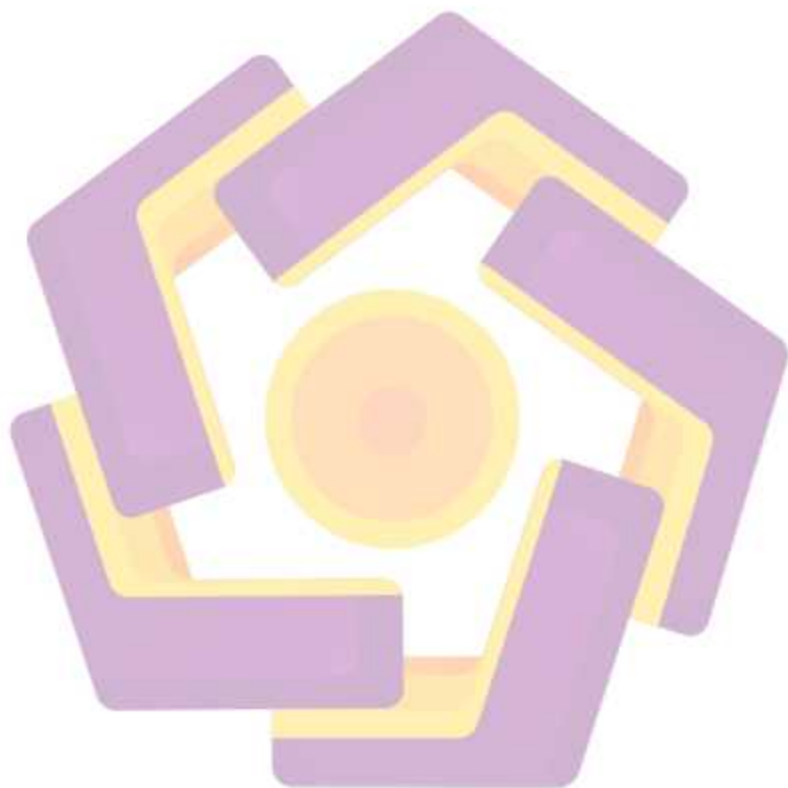
DAFTAR ISI

| | |
|--|------|
| JUDUL | i |
| PERSETUJUAN | ii |
| PERNYATAAN | iv |
| MOTTO | v |
| PERSEMBAHAN | vi |
| KATA PENGANTAR | viii |
| DAFTAR ISI | ix |
| DAFTAR TABEL | xiii |
| DAFTAR GAMBAR | xiv |
| INTISARI | xvii |
| <i>ABSTRACT</i> | xix |
| BAB I PENDAHULUAN | 1 |
| 1.1 Latar Belakang | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah | 3 |
| 1.4 Maksud dan Tujuan Penelitian | 3 |
| 1.5 Manfaat Penelitian | 4 |
| 1.6 Metode Penelitian | 4 |
| 1.6.1 Metode Pengumpulan Data | 5 |
| 1.6.2 Metode Studi Pustaka | 5 |
| 1.6.3 Metode Analisis Data | 5 |
| 1.6.4 Metode Perancangan | 5 |
| 1.6.5 Metode Pengujian | 6 |
| 1.7 Sistematika Penulisan | 7 |
| BAB II LANDASAN TEORI | 9 |

| | | |
|---|--|-----------|
| 2.1 | Kajian Pustaka..... | 9 |
| 2.2 | Landasan Teori..... | 15 |
| | 2.2.1 Pengertian Sistem..... | 15 |
| | 2.2.2 Karakteristik Sistem..... | 15 |
| 2.3 | Konsep Pemodelan Sistem..... | 17 |
| | 2.3.1 Unified Modeling Language (UML)..... | 17 |
| 2.4 | Klasifikasi Teks..... | 22 |
| | 2.4.1 Pengertian Klasifikasi Teks..... | 22 |
| | 2.4.2 Data Mining..... | 23 |
| | 2.4.3 Twitter..... | 24 |
| | 2.4.4 Crawling Data..... | 29 |
| | 2.4.5 API Twitter..... | 29 |
| | 2.4.6 Naïve Bayes Classifier..... | 31 |
| | 2.4.7 Praproses Data..... | 32 |
| 2.5 | Pengujian..... | 33 |
| | 2.5.1 Black Box Testing..... | 33 |
| | 2.5.2 White Box Testing..... | 33 |
| | 2.5.3 Confusion Matrix..... | 34 |
| BAB III METODOLOGI PENELITIAN..... | | 37 |
| 3.1 | Deskripsi Singkat..... | 37 |
| 3.2 | Alur Penelitian..... | 38 |
| | 3.2.1 Pengumpulan Data Twitter..... | 38 |
| | 3.2.2 Pelabelan Data..... | 38 |
| | 3.2.3 Pembagian Data..... | 39 |
| | 3.2.4 Praproses Data..... | 39 |
| 3.3 | Analisis Kebutuhan..... | 42 |
| | 3.3.1 Analisis Kebutuhan Fungsional..... | 42 |
| | 3.3.2 Analisis Kebutuhan Non Fungsional..... | 43 |
| | 3.3.3 Implementasi..... | 31 |
| | 3.3.4 Pengujian..... | 31 |
| | 3.3.5 Pemeliharaan..... | 31 |
| 3.4 | Unified Modeling Language..... | 44 |
| | 3.4.1 <i>Use Case Diagram</i> | 44 |

| | | |
|--|---|-----------|
| 3.4.2 | <i>Activity Diagram</i> | 45 |
| 3.4.3 | <i>Sequence Diagram</i> | 50 |
| 3.4.4 | <i>Class Diagram</i> | 57 |
| 3.5 | Perancangan Sistem | 57 |
| 3.5.1 | Flowchart Naïve Bayes Classifier | 57 |
| 3.5.2 | Flowchart Sistem | 59 |
| 3.6 | Perancangan Tampilan | 61 |
| 3.6.1 | Halaman Login | 61 |
| 3.6.2 | Halaman Dashboard | 61 |
| 3.6.3 | Halaman Dataset | 62 |
| 3.6.4 | Halaman Text Processing | 63 |
| 3.6.5 | Halaman Klasifikasi | 64 |
| 3.6.6 | Halaman Pengujian | 65 |
| 3.7 | Metode Analisis | 66 |
| 3.7.1 | Perhitungan Manual Metode Naïve Bayes | 66 |
| BAB IV HASIL DAN PEMBAHASAN | | 80 |
| 4.1 | Implementasi Rancangan Tampilan | 80 |
| 4.1.1 | Implementasi Halaman Login dan Pembahasan | 80 |
| 4.1.2 | Implementasi Halaman Dashboard dan Pembahasan | 81 |
| 4.1.3 | Implementasi Halaman Dataset dan Pembahasan | 82 |
| 4.1.4 | Implementasi Halaman Text Processing dan Pembahasan | 82 |
| 4.1.5 | Implementasi Halaman Klasifikasi dan Pembahasan | 83 |
| 4.1.6 | Implementasi Halaman Pengujian dan Pembahasan | 84 |
| 4.2 | Hasil Pengujian | 85 |
| 4.2.1 | Black Box Testing | 85 |
| 4.2.2 | White Box Testing | 87 |
| 4.2.3 | Pengujian Algoritma Confusion Matrix | 88 |
| 4.2.4 | Perhitungan Manual Confusion Matrix | 88 |
| 4.2.5 | Confusin Matrix Naïve Bayes Classifier | 89 |
| BAB V PENUTUP | | 91 |
| 5.1 | Kesimpulan | 91 |

| | |
|---------------------|----|
| 5.2 Saran..... | 91 |
| Daftar Pustaka..... | 92 |



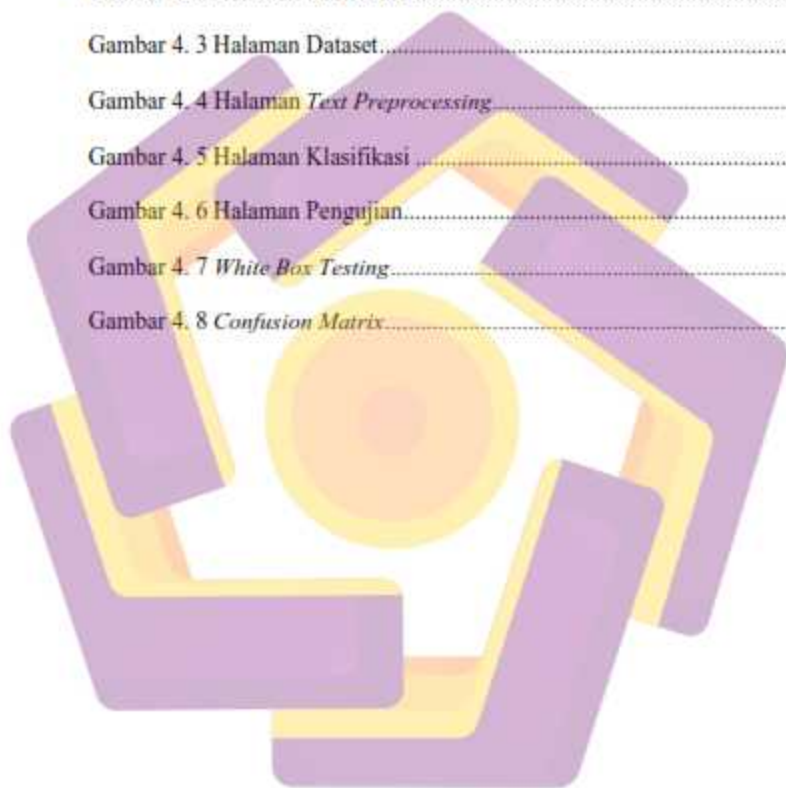
DAFTAR TABEL

| | |
|--|----|
| Tabel 2. 1 Pebandingan Penelitian Sebelumnya | 11 |
| Tabel 2. 2 Use Case Diagram..... | 18 |
| Tabel 3. 6 Data Uji Setelah Preprocessing..... | 77 |
| Tabel 2. 3 Activity Diagram..... | 19 |
| Tabel 2. 4 Sequence Diagram | 20 |
| Tabel 2. 5 Class Diagram | 21 |
| Tabel 2. 6 Confusion Matrix | 34 |
| Tabel 3. 1 Representasi Data..... | 40 |
| Tabel 3. 2 Stopword Removal..... | 41 |
| Tabel 3. 3 Cleansing..... | 42 |
| Tabel 3. 4 Contoh Kasus Data Training..... | 67 |
| Tabel 3. 5 Data Uji | 76 |
| Tabel 3. 7 Data Uji dan Data Latih..... | 77 |
| Tabel 4. 1 Pengujian Black Box..... | 85 |


DAFTAR GAMBAR

| | |
|--|----|
| Gambar 3. 1 Diagram Singkat..... | 37 |
| Gambar 3. 2 Pelabelan Data..... | 39 |
| Gambar 3. 3 <i>Use Case Diagram</i> Aplikasi..... | 45 |
| Gambar 3. 4 <i>Activity Diagram Login</i> | 46 |
| Gambar 3. 5 <i>Activity Diagram Import Data</i> | 46 |
| Gambar 3. 6 <i>Activity Diagram Dataset</i> | 47 |
| Gambar 3. 7 <i>Activity Diagram Preprocessing Data</i> | 47 |
| Gambar 3. 8 <i>Activity Diagram</i> Olah Data NBC dan <i>Confusion Matrix</i> | 48 |
| Gambar 3. 9 <i>Activity Diagram</i> Validasi Aplikasi..... | 49 |
| Gambar 3. 10 <i>Activity Diagram</i> Akurasi..... | 50 |
| Gambar 3. 11 <i>Sequence Diagram Login</i> | 51 |
| Gambar 3. 12 <i>Sequence Diagram Import Data</i> | 52 |
| Gambar 3. 13 <i>Sequence Diagram Dataset</i> | 53 |
| Gambar 3. 14 <i>Sequence Diagram Preprocessing Data</i> | 54 |
| Gambar 3. 15 <i>Sequence Diagram</i> Olah Data NBC dan <i>Confusion Matrix</i> | 55 |
| Gambar 3. 16 <i>Sequence Diagram</i> Validasi Klasifikasi..... | 56 |
| Gambar 3. 17 <i>Sequence Diagram</i> Akurasi..... | 56 |
| Gambar 3. 18 <i>Class Diagram</i> | 57 |
| Gambar 3. 19 <i>Flowchart Naïve Bayes Classifier</i> | 58 |
| Gambar 3. 20 <i>Flowchart</i> Sistem..... | 59 |
| Gambar 3. 21 Rancangan Tampilan Halaman Login..... | 61 |
| Gambar 3. 22 Rancangan Tampilan Halaman Dashboard..... | 62 |
| Gambar 3. 23 Rancangan Tampilan Halaman Dataset..... | 63 |
| Gambar 3. 24 Rancangan Tampilan Halaman Text Preprocessing..... | 64 |
| Gambar 3. 25 Rancangan Tampilan Halaman Klasifikasi..... | 65 |
| Gambar 3. 26 Halaman Pengujian..... | 66 |

| | |
|---|----|
| Gambar 4. 1 Halaman Login | 80 |
| Gambar 4. 2 Halaman Dashboard | 81 |
| Gambar 4. 3 Halaman Dataset | 82 |
| Gambar 4. 4 Halaman <i>Text Preprocessing</i> | 83 |
| Gambar 4. 5 Halaman Klasifikasi | 84 |
| Gambar 4. 6 Halaman Pengujian | 85 |
| Gambar 4. 7 <i>White Box Testing</i> | 87 |
| Gambar 4. 8 <i>Confusion Matrix</i> | 88 |



DAFTAR ISTILAH



| | |
|-------------|---|
| Dataset | Kumpulan data yang akan diolah. Machine Learning Aplikasi dari disiplin ilmu kecerdasan buatan. |
| Text Mining | Menambang data yang berupa teks dimana sumber data biasanya di dapatkan dari dokumen. |
| Literatur | Sumber ataupun acuan yang digunakan dalam berbagai macam aktivitas di dunia Pendidikan ataupun aktivitas lainnya. |

INTISARI

Pada era milenial ini, penggunaan internet di Indonesia dapat dikatakan berkembang pesat. Adanya koneksi internet dapat dimanfaatkan dengan baik sebagai sarana untuk menyalurkan hobi, baik itu jual beli secara *online*, menulis atau menyampaikan pendapat terhadap sesuatu melalui media sosial. Sosial media yang cukup populer di Indonesia salah satunya adalah Twitter. Melalui Twitter pengguna dapat menyampaikan pendapatnya secara bebas pada base. Base sering kali disalahgunakan, dengan mengeluarkan kata-kata tidak senonoh. Mayoritas pengguna twitter, banyak yang dikategorikan masih berusia dibawah umur. Oleh karena itu, perlu diterapkan pengklasifikasian *tweet* yang tidak senonoh yang terdapat di Twitter.

Metode yang digunakan penulis dalam penelitian ini yaitu dengan melakukan pengumpulan data twitter, pelabelan data, pembagian data, dan praproses data. Tujuan dari penelitian ini adalah mengimplementasikan metode *Naive Bayes Classifier* terhadap klasifikasi *tweet* tidak senonoh di twitter. Pada Pengklasifikasian *tweet* ini, penulis menggunakan *Usecase Diagram*, *Activity Diagram*, *Sequence Diagram*, serta *Flowchart* sistem. Dan Menggunakan Bahasa pemrograman Python, *Code Editor Sublime Text*, serta *Microsoft SQL Server* sebagai *database*.

Setelah peneliti membuat sistem Klasifikasi *tweet* tidak senonoh, dapat ditarik kesimpulan bahwa hasil pengujian tingkat akurasi dengan *confusion matrix* dalam pengklasifikasian teks menghasilkan presentase sebesar 99,5%. Dan untuk *Recall* menghasilkan presentase 100%, Presisi menghasilkan presentase 99%, dan *Error Rate* menghasilkan presentase 49,3%.

Kata kunci : Twitter, Klasifikasi *tweet* tidak senonoh, Naive Bayes Classifier.

ABSTRACT

In this millennial era, internet use in Indonesia can be said to be growing rapidly. The existence of an internet connection can be put to good use as a means to channel hobbies, be it buying and selling online, writing or expressing opinions on something through social media. One of the most popular social media in Indonesia is Twitter. Through Twitter, users can express their opinions freely on the base. Base is often misused, by issuing obscene words. The majority of Twitter users, many of whom are categorized as underage. Therefore, it is necessary to apply the classification of inappropriate tweets found on Twitter.

The method used by the author in this study is to collect Twitter data, label data, share data, and preprocess data. The purpose of this study is to implement the Naïve Bayes Classifier method for the classification of indecent tweets on Twitter. In this tweet classification, the author uses Usecase Diagrams, Activity Diagrams, Sequence Diagrams, and system flowcharts. And using the Python programming language, Sublime Text Code Editor, and Microsoft SQL Server as a database.

After the researchers made a classification system for indecent tweets, it can be concluded that the results of testing the level of accuracy with the confusion matrix in text classification produce a percentage of 99.5%. And for Recall it produces a percentage of 100%, Precision produces a percentage of 99%, and Error Rate produces a presentation of 49.3%.

Keywords: *Twitter, Indecent tweet classification, Naive Bayes Classifier*