PENERAPAN METODE KLASIFIKASI K-NEAREST NEIGHBOR PADA TWEET PROSTITUSI DI TWITTER TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Teknik Pada
Jurusan Teknik Informatika

Oleh:

RAHMAT RAMADHAN
11451105853

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU

PEKANBARU

2021

LEMBAR PERSETUJUAN

PENERAPAN METODE KLASIFIKASI K-NEAREST NEIGHBOR PADA TWEET PROSTITUSI DI TWITTER

LAPORAN TUGAS AKHIR MAHASISWA
JURUSAN TEKNIK INFORMATIKA
UIN SUSKA RIAU

TUGAS AKHIR

Oleh

RAHMAT RAMADHAN
NIM. 11451105853

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir
di Pekanbaru, pada tanggal 07 Desember 2021

Pembimbing I,	Pembimbing II,
Yusra, S.T., M.T	**Muhammad Fikry, S.T., M.Sc**
NIP. 19840123 201503 2001	**NIP. 19801018 200710 1002**

LEMBAR PENGESAHAN

PENERAPAN METODE KLASIFIKASI K-NEAREST NEIGHBOR PADA TWEET PROSTITUSI DI TWITTER

LAPORAN TUGAS AKHIR MAHASISWA

JURUSAN TEKNIK INFORMATIKA

UIN SUSKA RIAU

Oleh

RAHMAT RAMADHAN

NIM. 11451105853

Telah dipertahankan di depan sidang dewan penguji

sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik

pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 07 Desember 2021

Mengesahkan,

Dekan,	Ketua Jurusan,
Dr. Hartono, M.Pd.	**Iwan Iskandar, ST.M.T**
NIP. 19640301 199203 1 003	**NIP. 19821216 201503 1 003**

DEWAN PENGUJI

Ketua	: Dr. Elin Haerani, S.T., M.Kom.
Pembimbing I	: Yusra, S.T., M.T
Pembimbing II	: Muhammad Fikry, S.T., M.Sc
Penguji I	: Surya Agustian, ST, M.Kom
Penguji II	: Fadhila Syafria, S.T.M.Kom

LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 07 Desember 2021

Yang membuat pernyataan,

Rahmat Ramadhan

11451105853

LEMBAR PERSEMBAHAN

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Alhamdulillahirabbil'alamin…

Rasa Syukur yang tak terhingga pada Allah SWT yang telah memberikan nikmat-Nya, hingga diri ini masih diberikan kesempatan untuk menyelesaikan Tugas Akhir ini

Waktu berlalu, detik terlewati, begitu banyak rintangan yang datang silih berganti tak membuat hati harus menyerah lalu berhenti. Terkadang, ada kalanya hidup terasa begitu berat, namun lagi-lagi menyerah bukan pilihan yang harus diambil

Demi sebuah perjuangan,

Demi sebuah cita-cita,

Kupersembahkan Tugas Akhir ini untuk orang-orang tersayang, yang telah memberikan dukungan dan doa

Teruntuk Ayahanda dan Ibunda terkasih, tercinta, dan terhormat

Saya sadar bahwa apa-apa yang telah saya raih saat ini masih jauh dari kata cukup untuk membalas segala kebaikan yang pernah diberikan. Terima kasih atas dukungan dan doa yang selalu dilangitkan, karya ini kupersembahkan untuk Ayahanda dan Ibunda sebagai wujud rasa Terima Kasih atas segala pengorbanan dan jerih payah kalian sehingga saya mampu untuk menggapai apa yang telah dicita-citakan.

Terima kasih.

ABSTRAK

Dengan berkembangnya media sosial, kemudahan berkicau juga bermanfaat bagi para pekerja seks komersial (PSK) yang sering ditemui dengan cara menjual diri untuk mencari calon pengguna jasanya. Kegiatan prostitusi online ini tentunya harus diberantas mengingat dampaknya yang begitu besar. Berbagai upaya telah dilakukan pemerintah untuk melakukan pembenahan terhadap akun prostitusi online yang tersebar di media sosial Twitter. Namun, dari waktu ke waktu akun prostitusi meningkat. Pada penelitian ini dilakukan penerapan K-Nearest Neighbor untuk klasifikasi tweet prostitusi di Twitter. Pengelompokan tweet menggunakan metode K-Nearest Neighbor menggunakan 2000 data. Pada penelitian ini menggunakan uji K-Nearest Neighbor 90:10% memiliki akurasi tertinggi 97,50% dengan nilai k 9, pengujian 80:20% memiliki tingkat akurasi 94,25% dengan nilai k 9, dan 70:30 memiliki tingkat akurasi 94,50%.

Kata kunci: *Confusion Matrix,* Klasifikasi, *K-Nearest Neighbor, tweet,* Twitter.

ABSTRAK

With the development of social media, the convenience of Twitter is also beneficial for commercial sex workers (CSWs) who are often found by selling themselves to find potential users of their services. This online prostitution activity, of course, must be eradicated considering the impact it has had on it is so large. The government has made many efforts to make improvements to online prostitution accounts spread on Twitter social media. However, over time the accounts of prostitution are increasing. In this study, a comparison was made between Naïve Bayes and K-Nearest Neighbor for the classification of prostitution tweets on Twitter. Classifying tweets using the K-Nearest Neighbor using 2000 data. In this study using the K-Nearest Neighbor test 90:10% has the highest accuracy of 97.50% with a k value of 9, the 80:20% test has an accuracy rate of 94.25% with a k value of 9, and the 70:30% test has an accuracy rate of 94.50%.

Keywords: *Confusion Matrix, Classification*, *K-Nearest Neighbor, , tweet,* Twitter.

DAFTAR ISI

DAFTAR GAMBAR

Gambar **Halaman**

DAFTAR TABEL

Tabel **Halaman**

DAFTAR RUMUS

Rumus **Halaman**

BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi membawa sebuah perubahan dalam masyarakat. Lahirnya media sosial menjadikan pola prilaku masyarakat mengalami pergeseran baik budaya dan etika. Media sosial adalah media online, dengan penggunannya bisa dengan mudah berbagi informasi. Dampak positif media sosial memudahkan kita berinteraksi dengan banyak orang, sedangkan dampak negatif media sosial menjauhkan orang yang sudah dekat, membuat orang kecanduan internet, menimbulkan konflik, dan mudah mempercayai berita-berita yang tidak benar(Negoro & Atmadja, 2014). Menurut (Azeharie, 2014) beberapa sebab mengapa manusia membutuhkankomunikasi dengan manusia lain dikatakan oleh Harold D.Laswell dalam buku Hafied Cangara bahwa antara lain manusia ingin mengontrol lingkungan melalui berbagai informasi yang didapatnya dari proses komunikasi dalam lingkungan. Sebab lain manusia membutuhkan komunikasi adalah agar manusia dapat beradaptasi dengan lingkungannya.

Besarnya minat penggunaan media sosial di Indonesia dapat di lihat dari situs (cnnindonesia.com, 2020) dipaparkan bahwa penggunaan media sosial di Indonesia sudah mencapai 160juta atau 59% dari total jumlah penduduk. Berdasarkan data tersebut, waktu rata-rata penggunaan media sosial di Indonesia mencapai 3 jam 26 menit per hari dalam menggunakan media sosial.Pengguna Twitter, berdasarkan data PT Bakrie Telkom, memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna. Menurut (Paramastri & Gumilar, 2019) Terdapat beragam motif pengguna internet dalam mengakses berbagai media, termasuk media sosial. Salah satu motif utama adalah adanya fasilitas pendukung yang memungkinkan mereka mengakses dan melakukan interaksi di dalam media sosial.

Besarnya minat masyarakat Indonesia terhadap media sosial Twitter dapat menjadi salah satu alasan mengapa banyak pula kalangan yang menggunakan media sosial tersebut untuk berbagai kepentingan. Bagi sebagian masyarakat Twitter merupakan tempat untuk berkomunikasi dan berbagi informasi, namun bagi beberapa pihak Twitter dapat menjadi ajang promosi, pemberitaan, dan bisnis(Puspita & Gumelar, 2014).Dilihat dari beberapa fitur, salah satu fitur yang sangat menarik dalam media sosial twitter yaitu fitur follower. Fitur ini bisa memberikan gambaran bagi pengguna bahwa bagi akun pengguna twitter yang mempunyai follower terbanyak menunjukkan akun tersebut dipersepsi mempunyai kredibilitas yang memengaruhi daya tarik tertentu (Nurhadi, 2017). Motif penggunaan Twitter ini merupakan wadah yang pantas untuk menyalurkan aspirasi, melihat karakter, motif perkembangan pergaulan, motif hiburan referensi.

Namun, penggunaan media sosial Twitter juga disalah gunakan oleh pihak-pihak yang mencoba mencari keuntungan pekerja seks komersial (PSK) dengan cara menjual jasa untuk memuaskan kebutuhan seksual pelanggan. Prostitusi Online menjadi salah satu bentuk kejahatan yang berkembang karena kemajuan ilmu pengetahuan dan teknologi, semakin berkembangnya teknologi menyebabkan semakin merebaknya bisnis prostitusi online karena dapat memanfaatkan sarana internet dalam berintraksi dan penawaran prostitusi. Modus yang digunakan dengan menawarkan dan memasang foto-foto Pekerja Seks Komersial lengkap dengan data diri dan info kontak yang setiap saat dapat dihubungi oleh konsumen baik maupun telepon seluler (Negoro and Atmadja, 2014). Pihak Twitter selaku penyedia layanan media sosial pun telah melakukan upaya pencegahan dengan menyediakan fitur Report. Jadi masyarakat dapat melaporkan akun yang terindikasi melakukan tindak prostitusi online. Tetapi cara ini juga dirasa kurang efektif mengingat budaya masyarakat yang terkesan acuh terhadap akun-akun semacam ini (Azhar, 2017). Hal ini tentunya berdampak buruk bagi masyarakat terutama generasi muda karena merekalah pengguna terbanyak dari media sosial.

Penelitian pertama kali terkait akun prostitusi *online* dengan judul

Klasifikasi Akun Prostitusi Berdasarkan Skoring Tweet (Azhar, 2017). Terdapat banyak metode yang dapat digunakan untuk mengklasifikasi di antaranya *Support Vector Machine, K-Nearst Neighbor, Naïve bayes* dan *Modified K-Nearest Neigbor*. Diharapkan dengan adanya penelitian ini dapat di bandingkan metode mana yang memiliki akurasi terbaik untuk melakukan klasifikasi pada *tweet* di Twitter. Metode dengan akurasi terbaik dapat digunakan untuk membangun sebuah aplikasi klasifikasi *tweet* Twitter yang digunakan untuk memberantas akun-akun prostitusi yang tersebar di Media Sosial khususnya di Twitter.

Penelitian yang terkait dengan metode yang menunjukan bahwa tingkat akurasi algoritma dalam pengklasifikasian relative tinggi. Hal ini dibuktikan dengan yang dilakukan oleh Rahmad Robi Waliyansah Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN) dengan akurasi 82,7%, sedangkan dalam penelitian ClaudioFresta Suharno Klasifikasi Teks Bahasa Indonesia pada Dokumen Pengaduan Sambat Online menggunakan Metode K-Nearest Neighbors (K-NN) dan ChiSquaredengan akurasi 78%, sedangkan dalam penelitian Yusra Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode NaïveBayes Classifier dan K-Nearest Neighbor dengan akurasi tertinggi naive bayes 87%, dalam penelitan yang dilakukan oleh Oki Arifin dalam Analisa perbandingan tingkat performansi metode *Support Vector Machine* dan *Naïve Bayes Classifier* Untuk Klasifikasi Jalur Minat SMA dengan akurasi SVM 97% dan Naïve Bayes 92 %, pada penelitan Elly Indrayuni Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes dengan akurasi 90%, dalam penelitian M.Azman Maricar dalam Perbandingan Akurasi Naive Bayes dan K-NN pada klasfikasi untuk meramalkan status pekerjaan Alumni ITB STIKOM Bali dengan akurasi Naive bayes 83% dan Knn 82%.

Dalam penelitian Firman Tempola Perbandingan klasifikasi antara Knn dan Naive Bayes pada penentuan status gunung berapi dengan k-fold cross validation dengan akurasi K-NN 78% dan Naive Bayes 91%, sedangkan dalam penelitan

Viona Novalia dalam penelitian Perbandingan Metode Klasifikasi *Naïve Bayes* dan *K-Nearest Neighbor* (Studi Kasus : Status Kerja Penduduk di Kabupaten Kutai Kartanegara Tahun 2018) dengan tingkat akurasi Naïve Bayes 90% dan K-NN 94%, pada penelitian Aida Indriani Analisa Perbandingan Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* terhadap Klasifikasi Data dengan akurasi KNN80% dan 73% Naïve Bayes, sedangkan dalam penelitian M. Ali Fauzi pada Klasifikasi Teks Pengaduan Pada Sambat Online Menggunakan Metode NGram dan Neighbor Weighted K-Nearest Neighbor (NW-KNN) dengan akurasi 75%.

Permasalahan-permasalahan yang telah diteliti oleh peneliti sebelumnya terkait dengan klasifikasi, maka penelitian ini dilakukan untuk menyelesaikan permasalahan agar dapat mengklasifikasi dan membandingkan akurasi metode. Metode yang di gunakan dalam penelitian adalah menggunakan metode *K-Nearest Neigbor* untuk menghasilkan akurasi terbaik.

Adapun dipilihnya algoritma *K-Nearest Neigbor* sebagai pengklasifikasian teks karena tingkat akurasi algoritma ini dalam pengklasifikasian relatif tinggi dalam penelitan yang dilakukan oleh Suharno dalam klasifikasi teks bahasa Indonesia pada dokumen pengaduan dengan akurasi K-Nearst Neighbor 90%. Berdasarkan latar belakang permasalahan di atas, maka akan dibangun suatu aplikasi berdasarkan penelitian yaitu Penerapan Metode K Nearest Neighbor pada *tweet* prostitusi di Twitter.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka pada penelitian ini diperoleh rumusan masalah "Bagaimana Penerapan Metode *K Nearest Neighbor* dalam mengklasifikasi *tweet* prostitusi di Twitter?".

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan dari Twitter
2. Data *tweet* yang digunakan berjumlah 2000 *tweet* dengan pembagian 1000 *tweet* yang terindikasi *tweet* Prostitusi dan 1000 *tweet* bukan prostitusi.
3. Klasifikasi *tweet* prostitusi di Twitter menggunakan 2 kelas yaitu prostitusi dan bukan prostitusi.

1.4 Tujuan Penelitian

Tujuan penelitian yang ingin dicapai dalam penelitian ini adalah:

1. Menerapkan metode *K-Nearest Neighbor* pada Twitter untuk mengklasifikasi *tweet* prostitusi.
2. Mengetahui akurasi metode *K-Nearest Neighbor* dalam pengklasifikasian *tweet* di Twitter.

1.5 Sistematika Penulisan

Laporan tugas akhir ini terdiri dari enam bab, dengan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan dasar-dasar dari penulisan tugas akhir yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitan, serta sistematika penulisan laporan tugas akhir.

BAB II LANDASAN TEORI

Bab ini menjelaskan teori-teori yang berhubungan dengan spesifikasi pembahasan penelitian yang akan diangkat baik buku, jurnal, prostitusi, metode *K Nearest Neighbor* yang digunakan sebagai landasan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang metodologi penelitian, identifikasi masalah, teknik pengumpulan data, analisa algoritma dan alat bantu penelitian.

BAB IV ANALISA PERANCANGAN

Bab ini menjelaskan tentang analisa data, analisa proses dan perancangan aplikasi.

BAB V IMPLEMENTASI DAN PENGUJIAN

Pada bab ini menjelaskan mengenai implementasi dan perhitungan tingkat akurasi metode *K Nearest Neighbor* pada pengklasifikasian *tweet* prostitusi di Twitter.

BAB VI PENUTUP

Bab ini berisi tentang kesimpulan dan saran-saran berkaitan dengan penelitian yang telah dilakukan.

BAB II
LANDASAN TEORI

2.1 Metode Klasifikasi

Menurut (Imanda et al., 2018) Klasifikasi adalah sebuah metode mengelompokkan data. Klasifikasi juga diartikan sebagai pengelompokan data atau objek baru ke dalam kelas atau kategori berdasarkan variabel-variabel tertentu. Klasifikasi merupakan *data mining* yang melihat atribut dari kelompok data yang sudah didefenisikan sebelumnya. Atribut- atribut tersebut digunakan sebagai variabel dalam penentuan kelas suatu objek baru. Tujuan klasifikasi yaitu menentukan kelas dari suatu objek yang belum diketahui kelasnya dengan akurat.

Terdapat beberapa tahapan dalam klasifikasi yaitu pembangunan model, penerapan model, dan evaluasi. Pembangunan model dilakukan berdasarkan data latih yang telah memiliki atribut dan kelas data. Kemudian data-data tersebut diadaptasi untuk menentukan kelas dari data atau objek yang baru. Setelah menerapkan model, maka tahap berikutkanya yaitu evaluasi. Tahap ini meliputi bagaimana tingkat akurasi sesuai dengan yang diinginkan, maka model yang telah dibangun tersebut dapat dilanjutkan untuk analisa data lebih lanjut.

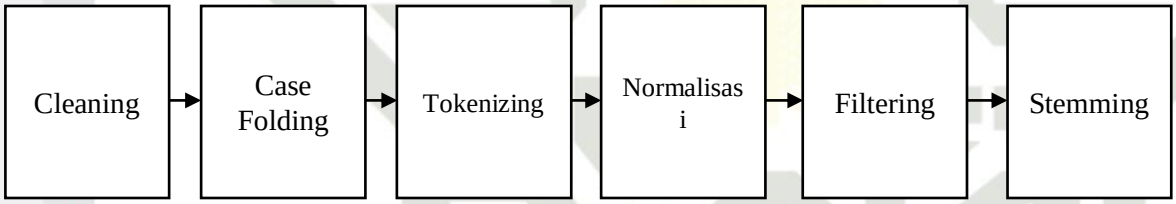
Proses klasifikasi terdiri dari dua fase, yaitu fase *learning* dan fase *testing*. Fase *Learning* yaitu fase di mana sebagian data yang kelas datanya telah diketahui dijadikan model untuk aplikasi yang akan dibangun. Sedangkan fase *testing* yaitu fase di mana model yang sudah terbentuk diuji dengan sebagian data lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi, maka model ini dapat digunakan untuk memprediksi kelas data yang belum diketahui.

2.2 Text *Preprocessing*

Text Preprocessing merupakan tahapan awal dari proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Sebuah teks yang ada harus dipisahkan, hal ini dapat dilakukan dalam beberapa tingkatan yang berbeda.

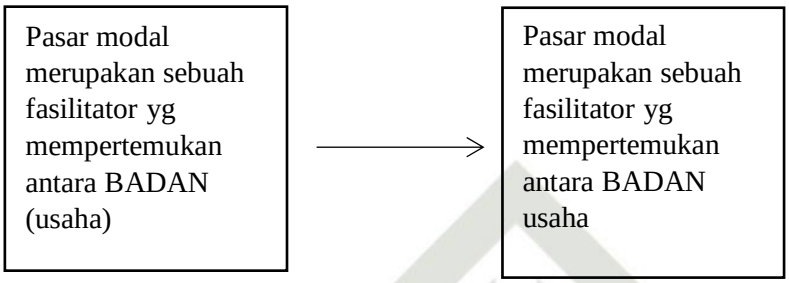
Suatu dokumen dapat dipecah menjadi bab, sub-bab, paragraf, kalimat dan pada akhirnya menjadi potongan kata/*token.* Selain itu pada tahapan ini keberadaan digit angka, huruf kapital, atau karakter- karakter yang lainnya dihilangkan dan dirubah.

Pada tahapan *text perpocessing* memiliki beberapa proses yaitu *cleaning,case foding, tokenizing,* dan *filtering, stemming* (Tifani wulandari, 2018).



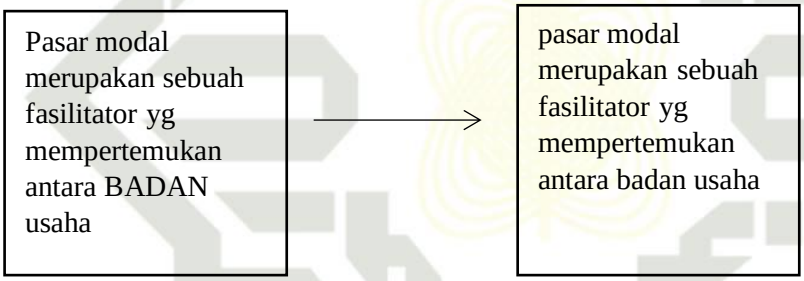
Gambar 2. 1 Pre-*Processing*

1. *Cleaning* adalah proses yang digunakan untuk melakukan pembersihan text dari karakter-karakter selain huruf, tanda baca dan tag yang tidak digunakan nantinya di dalam proses pengklasifikasian. Kata yang perlu dihilangkan seperti *username* dan *mention,* dan sebagainya.



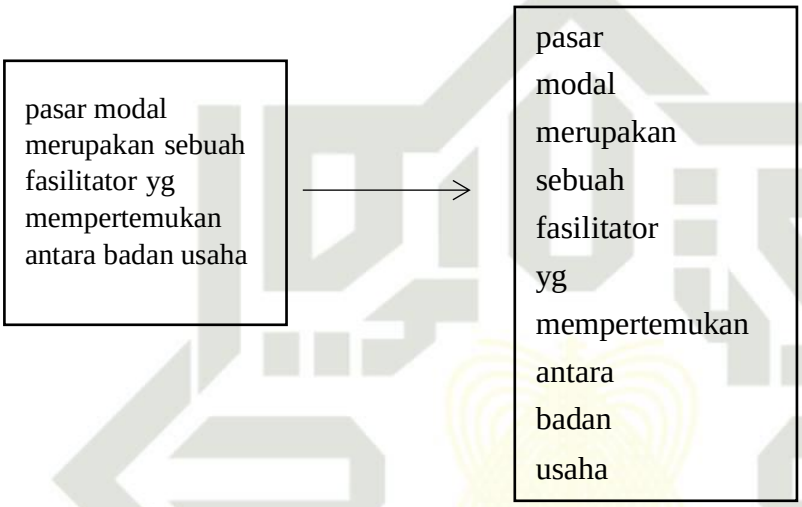
Gambar 2. 2 Proses *Cleaning*

2. *Case Folding* pada tahap ini semua karakter alphabet akan dihapus dengan tujuan mengurangi *noise* sehingga karakter khusus, URL, *email*, angka dan simbol akan dihapus.



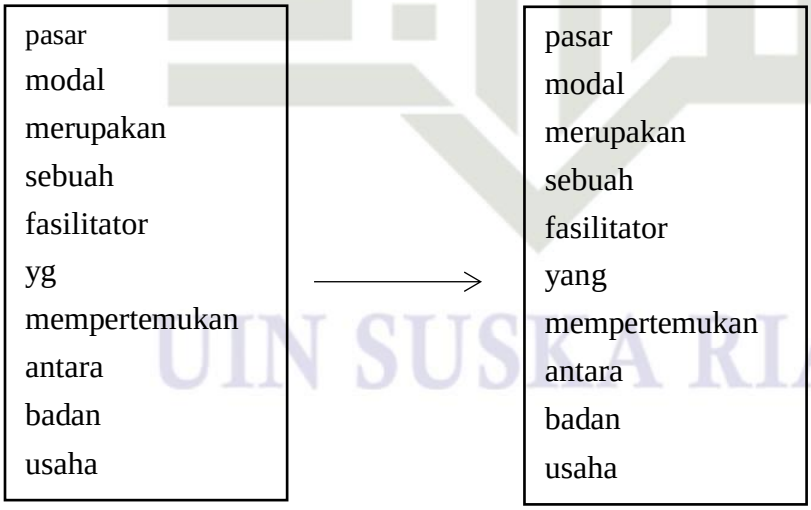
Gambar 2. 3 Proses *Case Folding*

3. *Tokenizing* proses pemecahan sekumpulan karakter dalam suatu teks pada *tweet* ke dalam satuan kata. Proses untuk membagi teks yang dapat berupa kalimat, paragraph atau dokumen, menjadi *token-token* atau bagian tertentu.



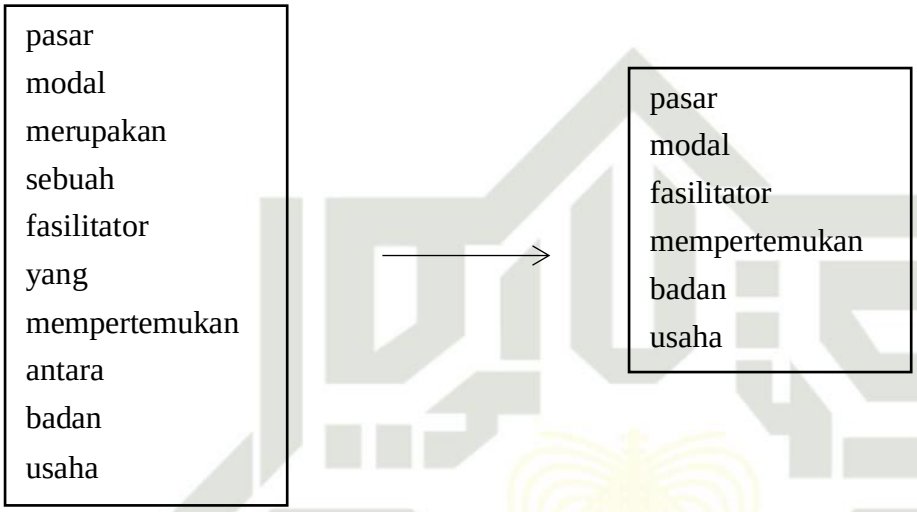
Gambar 2. 4 Proses *Tokenizing*

4. Normalisasi merupakan proses ejaan kata terhadap kata yang tidak standar, misalnya “temanni” menjadi “temani”. Terhadap singkatan yang tidak baku seperti, “bhng” menjadi “bohong”Gambar 2. 5



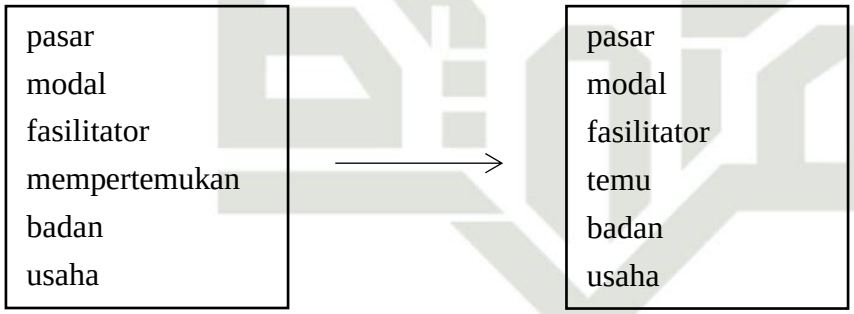
Gambar 2. 5 Proses Normalisasi

5. *Filtering* merupakan proses pengambilan kata-kata yang penting dari hasil *tokenizing*. *Filtering* juga merupakan proses memperbaiki kata-kata yang tidak dibutuhkan.



Gambar 2. 6 Proses *Filtering*

6. *Stemming* merupakan salah satu proses untuk menemukan kata dasar dari sebuah kata dengan menghilangkan semua imbuhan



Gambar 2. 7 Proses *Stemming*

2.3 Pembobotan dan Seleksi Fitur

Tahap selanjutnya setelah melakukan *text preprocessing* adalah melakukan pemilihan fitur agar mendapatkan hasil klasifikasi yang lebih maksimal. Untuk mengubah data tersebut menjadi numerik yaitu menggunakan pembobotan *Term Frequency Inverse Documents Frequensy* (TF-IDF).

Term Frequency Inverse Documents Frequensy merupakan tahapan yang digunakan untuk menentukan seberapa jauh keterhubungan kata(*term*) terhadap dokumen dengan memberikan bobot setiap kata. Metode TF-IDF ini menggabungkan dua konsep yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut. Dalam perhitungan bobot menggunakan TF-IDF, dihitung terlebih dahulu nilai TF perkata dengan bobot masing-masing kata adalah 1. Sedangkan nilai IDF diformulasikan pada persamaan.

Rumus dalam menentukan pembobotan dengan TF-IDF adalah sebagai berikut:

W_{ij}= tf Xidf

Idf = log $(\frac{n}{dfi})$

Dengan : i = 1,2,….,p (Jumlah Variabel)

J = 1,2,…,N (Jumlah Data) .....(2. 1)

Dimana *wij* adalah bobot dari kata *i* pada artikel ke *j*, *N* merupakan jumlah seluruh dokumen, *tfij* adalah jumlah kemunculan kata i pada dokumen *j*, *dfj* adalah jumlah artikel *j* yang mengandung kata *i.* TF-IDF dilakukan agar data dapat di analisis dengan menggunakan *K-Nearest Neigbor* .

2.4 *Chi Square Feature Selection*

Metode CHAID merupakan suatu metode pohon klasifikasi yang pertama kali dikenalkan oleh Dr. G. V. Kass tahun 1980 pada buku Applied Statistics dalam sebuah artikel yang berjudul "An Exploratory Technique for Investigating Large Quantities of Categorical Data". CHAID merupakan suatu teknik iteratif yang menguji variabel-variabel independen secara individual yang digunakan dalam klasifikasi dan menyusunnya pada tingkat signifikansi statistik chi-square terhadapvariabel dependennya.(Aritonang and et al, 2016).

Seleksi fitur (feature selection) dilakukan untuk mereduksi fitur-fitur yang tidak relevan dalam proses klasifikasi oleh *K-Nearest Neighbor*. Seleksi fitur Chi Square menggunakan teori statistika untuk menguji independensi sebuah term dengan kategorinya. Salah satu tujuan penggunaan seleksi fitur adalah untuk menghilangkan fitur pengganggu dalam klasifikasi. Dalam seleksi fitur Chi Square berdasarkan teori statistika, dua peristiwa di antaranya adalah, kemunculandari fitur dan kemunculan dari kategori, yang kemudian setiap nilai term diurutkandari yang tertinggi.(Anisah et al., 2016) Uji Chi Square dalam statistika diterapkan untuk menguji independensi dari dua peristiwa.

Rumus :

$$X^2 = \sum \frac{(O-E)2}{E} = \text{.....} \quad 2.2$$

O : Nilai Observasi (pengamatan)

E : Nilai Expected (Harapan)

Df : (b-1)(k-1).....2. 3

B : Jumlah Baris

K : Jumlah Kolom

2.5 K-Nearest Neighbor

K-Nearest Neighbor merupakan metode untuk melakukan klasifikasi terhadap objek sesuai dengan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Data pembelajaran diproyeksikan ke ruang berdimensi banyak, dimana tiap dimensi merepresentasikan fitur dari data.

Algoritma ini mengklasifikasikan data baru yang belum diketahui kelasnya atau memprediksi kelas dengan data sejumlah k yang letaknya terdekat dari data baru tersebut. Nilai k umumnya ditentukan dalam jumlah ganjil untuk menghindari munculnya jumlah jarak yang sama dalam pengklasifikasian. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak *Euclidean* (Srianto & Mulyanto, 2016).

Untuk mendefinisikan jarak antara dua titik yaitu titik pada data *training* (X) dan titik pada data *testing* (Y) maka dipakai rumus *Euclidean* (Iriantoro et al., 2018).

Persamaan *Euclidean* adalah:

$$d_{(x,y)} = \sqrt{\sum_{k-1}^{n} (x_1 - y_2)^2} \quad \text{..... (2. 4)}$$

Keterangan:

X1 = data latih

Y2 = data uji

k = variabel data

d = jarak

n = dimensi data

Langkah-langkah untuk menghitung metode algoritma *K-Nearest Neighbor*:

1. Menentukan parameter k (jumlah tetangga paling dekat)
2. Menghitung kuadrat jarak *eucliden (query instance)* masing-masing objek terhadap data sampel yang diberikan.
3. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *eucliden* terkecil.
4. Mengumpulkan kategori k (klasifikasi *nearest neighbor*)
5. Dengan memakai kategori *nearest neighbor* yang paling mayoritasmaka dapat diprediksi nilai *queri instance* yang telah dihitung.

Berikut kelebihan dan kelemahan algoritma *K-Nearest Neighbor*:

Kelebihan:

Algoritma *K-Nearest Neighbor* memiliki beberapa kelebihan yaitu cukup tangguh terhadap *training* data yang *noisy* dan efektif apabila data latihnya besar.

Kelemahan:

1. *K-Nearest Neighbor* perlu menentukan nilai dari parameter k (jumlah dari tetangga terdekat).
2. Pembelajaran berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil yang terbaik.

Biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap sampel uji pada keseluruhan data latih.

2.6 *White box*

White box juga dikenal sebagai pengujian struktural, pengujian *transparent* box, pengujian berbasis logika atau pengujian berbasis kode. Kata *white box*, mengacu pada sebuah test case, perangkat lunak yang sedang diuji dianggapsebagai kotak, sedangkan white mengacu pada bahwa kotak itu terlihat jelas bagian dalamnya.

Berdasarkan (Irawan, 2017) *White box* adalah pengujian yang didasarkan pada pengecekan terhadap detail perancangan, menggunakan struktur kontrol dari desain program secara prosedur untuk membagi pengujian kedalam beberapa *test case*.

Kasus yang sering menggunakan *white box* testing akan diuji dengan beberapa tahapan yaitu:

1. Pengujian seluruh keputusan yang menggunakan logika.
2. Pengujian keseluruh loop yang ada sesuai batasan-batasannya.
3. Pengujian pada struktur data yang sifatnya internal dan yang terjamin validitasnya.

Persyaratan dalam pengujian *white box* testing

Berikut ini terdapat beberpa persyaratan dalam pengujian *white box* testing, terdiri atas:

1. Mendefenisikan semua alur logika
2. Membangun kasus untuk digunakan dalam pengujian
3. Mengevaluasi semua hasil pengujian
4. Melakukan pengujian secara menyeluruh.

2.7 *Confusion Matrix*

Confusion Matrix merupakan metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Presisi atau *confidence* adalah proporsi kasus yang diprediksi positif yang juga positif benar pada data yang sebenarnya. *Recall* atau *Sensitivity* adalah proporsi kasus positif yang sebenarnya yang diprediksi positif secara benar.

		Kelas Prediksi	
		True	**False**
Kelas Sebenarnya	**True**	**TP**	**FN**
	False	**FP**	**TN**

Tabel 2. 1 Confusion Matrix

Rumus untuk mencari nilai akurasi yaitu:

$$\boldsymbol{Akurasi} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \boldsymbol{X}\ \mathbf{100\%} \quad \text{..... (2.5)}$$

Keterangan:

1. True Positif (TP)
   True positif adalah merupakan jumlah dokumen dari kelas *true* yang benar diklasifikasikan sebagai kelas *true*.
2. True Negatif (TN)
   True negatif adalah merupakan jumlah dokumen yang berasal dari kelas *true* salah yang diklasifikasikan sebagai kelas *false*.
3. False Positif (FP)
   False positif adalah merupakan jumlah dokumen yang berasal dari kelas *false* yang salah yang diklasifikasikan sebagai kelas *false*.
4. False Negative (FN)
   False negatif adalah merupakan jumlah dokumen yang berasal dari kelas *true* yang salah yang diklasifikasikan sebagai kelas *false*.

2.8 Penelitian terkait

No	Judul	Penulis dan tahun	Hasil
1	Klasifikasi Teks Bahasa Indonesia pada Dokumen Pengaduan Sambat Online menggunakan Metode K-Nearest Neighbors (K-NN) dan ChiSquare	Claudio Fresta Suharno , M. Ali Fauzi , Rizal Setya Perdana (2017)	Dari penelitian ini menunjukkan bahwa hasil yang didapatkan dengan menggunakan seleksi fitur lebih baik daripada tanpa adanya proses seleksi fitur. Precision dan recall terbaik didapatkan pada k = 15 dengan seleksi fitur sebesar 25%. Sedangkan hasil dari F-Measure terbaik didapatkan dengan nilai 78% pada k = 15 dan k = 5 dengan seleksi fitur sebesar 25%.
2	Perbandingan Akurasi Klasifikasi Citra Kayu Jati Menggunakan Metode Naive Bayes dan k-Nearest Neighbor (k-NN)	Rahmat Robi Waliyansyah, Citra Fitriyah (2019)	Klasifikasi citra kayu jati yang berasal dari sulawesi dengan Metode Naive Bayes paling baik dengan persentase tingkat akurasi sebesar 82,7%.
3	Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbor	Yusra , Dhita Olivita , Yelfi Vitriani (2016)	Akurasi yang didapat pada pengujian 10-fold cross validation untuk seratus data tugas akhir, metode Naïve Bayes Classifier dapat melakukan klasifikasi terhadap tugas akhir lebih baik daripada K-Nearest Neighbor dengan akurasi sebesar 87%.

4	Analisa perbandingan tingkat performansi metode *Support Vector Machine* dan *Naïve Bayes Classifier* Untuk Klasifikasi Jalur Minat SMA	Oki Arifin, Theopilus Bayu Sasongko (2018)	Pada pengujian kedua, implementasi pengujian komparasi metode SVM dengan Naïve Bayes Classifier dilakukan pada dataset penjurusan xyz yang berjumlah 280 data dengan 12 atribut penentu dengan akurasi SVM 97% dam Naïve Bayes 92%
5	Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali	M. Azman Maricar , Dian Pramana (2019)	Berdasarkan ketentuan rentang nilai MAPE, baik Naïve Bayes 83% dan K-Nearest Neighbor 82% dengan nilai K=9 memiliki arti bahwa metode tersebut baik dalam kasus ini, namun Naïve Bayes sedikit lebih baik. .
6	Perbandingan klasifikasi antara Knn dan Naive Bayes pada penentuan status gunung berapi dengan k-fold cross validation.	Firman Tempola , Miftah Muhammad , Amal Khairan (2018)	Berdasarkan pengujian dari dua metode machine learning yang telah diterapkan pada sistem tersebut, diperoleh rata-rata akurasi sistem ketika menggunakan k-nn sebesar 78%, sedangkan Naïve Bayes 81%.

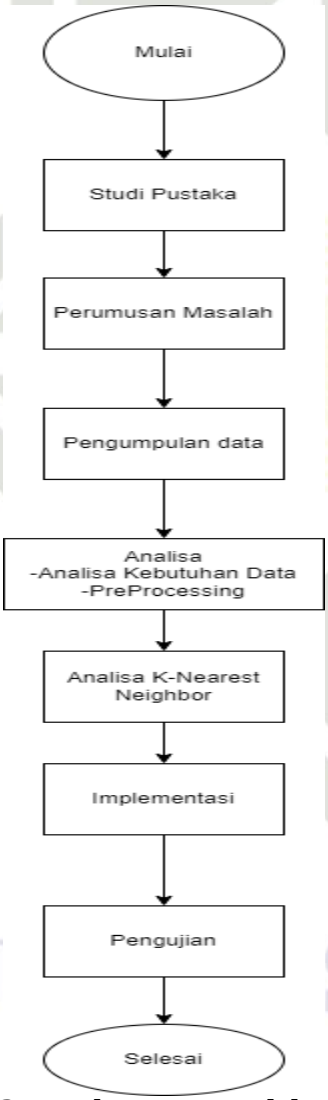
7	Perbandingan Metode Klasifikasi *Naïve Bayes* dan *K-Nearest Neighbor* (Studi Kasus : Status Kerja Penduduk di Kabupaten Kutai Kartanegara Tahun 2018)	Viona Novalia, Rito Goejantoro dan Sifriyani (2018)	Pengklasifikasian status kerja penduduk dengan metode naive Bayes menghasilkan akurasi sebesar 90,08% dan pada metode k-nearest neighbor menghasilkan akurasi sebesar 94,66%.
8	Klasifikasi Teks Bahasa Bali dengan Metode *Supervised Learning Naïve Bayes Classifier*.	Ida Bagus Widnyana Putra, Made Sudarma, I Nyoman Satya Kumara. (2016)	Pengukuran efektivitas klasifikasi terhadap 15 data uji diperoleh nilai 80% untuk metode k-NN dan sebesar 73% untuk metode NBC dengan menggunakan pengukuran efektivitas Confusion matrix .
9	Analisa Perbandingan Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* terhadap Klasifikasi Data	Aida Indriani (2020)	Model klasifikasi yang telah dibuat kemudian diuji dengan mengunakan teknik confusion matrix dengan masing-masing hasil adalah nilai presisi sebesar 83%, nilai recall sebesar 80%, dan terakhir sebesar 80% yang didapatkan pada pengujian akurasi.

10	Klasifikasi Teks Pengaduan Pada Sambat Online Menggunakan Metode NGram dan Neighbor Weighted K-Nearest Neighbor (NW-KNN)	Annisya Aprilia Prasanti , M. Ali Fauzi , M. Tanzil Furqon. (2018)	Hasil pengujian dalam penelitian ini menunjukkan bahwa penggunaan metode NW-KNN dengan nilai tetangga k = 3 dan metode N-Gram dengan Unigram memiliki nilai f-measure tertinggi sebesar 75.25%.
8	Klasifikasi Teks Bahasa Bali dengan Metode *Supervised Learning Naïve Bayes Classifier*.	Ida Bagus Widnyana Putra, Made Sudarma, I Nyoman Satya Kumara. (2016)	Pengukuran efektivitas klasifikasi terhadap 15 data uji diperoleh nilai 80% untuk metode k-NN dan sebesar 73% untuk metode NBC dengan menggunakan pengukuran efektivitas Confusion matrix .
9	Analisa Perbandingan Metode *Naïve Bayes Classifier* dan *K-Nearest Neighbor* terhadap Klasifikasi Data	Aida Indriani (2020)	Model klasifikasi yang telah dibuat kemudian diuji dengan mengunakan teknik confusion matrix dengan masing-masing hasil adalah nilai presisi sebesar 83%, nilai recall sebesar 80%, dan terakhir sebesar 80% yang didapatkan pada pengujian akurasi.
10	Klasifikasi Teks Pengaduan Pada Sambat Online Menggunakan Metode NGram dan Neighbor Weighted K-Nearest Neighbor (NW-KNN)	Annisya Aprilia Prasanti , M. Ali Fauzi , M. Tanzil Furqon. (2018)	Hasil pengujian dalam penelitian ini menunjukkan bahwa penggunaan metode NW-KNN dengan nilai tetangga k = 3 dan metode N-Gram dengan Unigram memiliki nilai f-measure tertinggi sebesar 75.25%.

BAB III

METODOLOGI PENELITIAN

Metodologi penelitian merupakan sebuah proses pelaksanaan penelitian yang terdiri dari langkah-langkah dan juga menerapkan prinsip metode ilmiah. Adapun langkah-;langkah yang penulis lakukan selama melakukan penelitian ini dapat dilihat dari gambar.



Gambar 3. 1 Tahapan Metodologi Penelitian

3.1 Studi Pustaka

Studi Pustaka merupakan cara yang dilakukan untuk menemukan dan mengumpulkan permasalahan, data dan informasi yang ada pada klasifikasi *tweet* di Twitter berdasarkan hasil penelitian terkait sebelumnya, fakta yang berhubungan dengan penelitian serta membaca dan mempelajari buku-buku, jurnal-jurnal, penelitian atau referensi yang lain berhubungan dengan *tweet* di Twitter.

3.2 Perumusan Masalah

Pada tahapan perumusan masalah ini dilakukan setelah mendapatkan berbagai informasi dari jurnal dan buku yang terkait dengan penelitian yang akan dilakukan. Setelah hasil didapatkan pada tahapan studi pustaka, maka dapat dirumuskan permasalahan penelitian ini yaitu Perbandingan Metode Klasifikasi Naïve Bayes dan K Nearest Neighbor pada Tweet Prostitusi di Twitter.

3.3 Pengumpulan Data

Pada tahap ini merupakan salah satu tahapan pengambilan data. Data yang digunakan 2000 data yang nantinya dibagi sebagai data latih dan data uji. Pengumpulan data latih dan data uji didapatkan dari akun yang menawarkan prostitusi, dari masing- masing akun diambil *tweet* yang menawarkan prostitusi. Pembagian data latih dan data uji berdasarkan *tweet* yang diambil yaitu 70% data latih : 30% data uji, 80% data latih : 20% data uji, dan 90% data latih : 10% data uji. Pada tahap ini data di ambil dengan menggunakan aplikasi dengan cara manual.

Dalam penelitian ini, data di ambil dari Twitter. Pengambilan data dari Twitter cukup mudah dilakukan karena Twitter sudah menyediakan API (*Application Programing Interface*) yang ditujukan kepada pengembang aplikasi untuk mempermudah pengambilan data dari Twitter.

3.4 Analisa

Tahapan analisa ini akan dilakukan terhadap data yang sudah didapatkan. Pada penelitian ini terdapat 5 tahap yaitu analisa kebutuhan data, analisa *preprecessing,* pelabelan, implementasi metode *K-Nearest Neigbor*.

a. Analisa kebutuhan data

Pada tahapan analisa dan perancangan ini berisi tentang tahapan analisa klasifikasi *tweet* prostitusi di Twitter beserta perancangan pada aplikasi yang akan dibangun pada tahap Implementasi.

b. Analisa *Preprocessing*

Pada tahapan *preprocessing* ini dilakukan terhadap dokumen berupa teks yang digunakan untuk mendapatkan data dalam bentuk yang terstruktur sehingga mempermudah dalam proses perhitungan yang akan dilakukan selanjutnya. Pada tahap *preprocessing* yang akan dilakukan selanjutnya. Pada tahap *preprocessing* memiliki beberapa tahapan yaitu terdiri dari *case folding, cleaning, tokenizing,* normalisasi, *filtering, stemming.*

a. *Cleaning*

   Proses penghilang karakter atau tanda baca yang tidak diperlukan dari teks tersebut, seperti menghilangkan *mention, URL, email*.

b. *Case folding*

   Merupakan proses untuk mengubah seluruh huruf di dalam dokumen menjadi huruf kecil.

c. *Tokenizing*

   Proses untuk memecah kalimat menjadi beberapa bagian atau kata.

d. Normalisasi

   Proses pengambilan kata-kata yang tidak baku kedalam bahasa baku dalam kamus bahasa Indonesia.

e. *Filtering*

Merupakan proses untuk menghilangkan kata-kata yang dianggap tidak memiliki makna selama proses klasifikasi. Dengan menggunakan stopword.

I. *Stemming*

Merupakan proses pengubahan kata menjadi kata dasarnya. Algoritma stemming digunakan adalah ECS (*Enhanced Confix Stripping).*

c. Pelabelan

Pelabelan data yaitu proses memberikan label atau kelas pada *tweet*. Pelabelan *tweet* dibuat berdasarkan rule, apabila *tweet* berisi kalimat ajakan untuk menggunakan jasa prostitusi maka pelabelannya adalah prostitusi, dan jika *tweet.* Berisi kalimat yang tidak menawarkan jasa prostitusi maka hasil pelabelannya adalah bukan prostitusi. Proses pelabelan menggunakan seorang Ahli Bahasa yang bernama Yudi Audina, S.Pd telah menyelesaikan proses SI(Strata I) Pendidikan Bahasa Indonesia.

d. Proses *K-Nearest Neighbor*

Algoritma *K-Nearest Neighbor* merupakan algoritma supervised learning dengan proses belajar berdasarkan nilai dari variabel target yang terasosiasi dengan nilai dari variabel prediktor. Dalam algoritma K-nn semua data yang memiliki harus memiliki label, sehingga ketika ada data baru yang telah ada dan diambil data yang paling mirip dan melihat label dari data tersebut. Adapun langkah-langkat dari algoritma K-nn tersebut.

a. Menentukan parameter k (jumlah tetangga paling dekat)
b. Menghitung kuadrat jarak *eucliden (query instance)* masing-masing objek terhadap data sampel yang diberikan.
c. Kemudian mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak *eucliden* terkecil.
d. Mengumpulkan kategori Y (klasifikasi *nearest neighbor*).

e. Dengan memakai kategori *nearest neighbor* yang paling mayoritas maka dapat diprediksi nilai *queri instance* yang telah dihitung.

3.5 Implementasi dan Pengujian

Pada tahapan implementasi dan pengujian dilakukan setelah proses analisa dan perancangan pada aplikasi yang akan dibangun.

1. Implementasi

Implementasi dilakukan dalam bentuk pengkodean pada aplikasi yang telah dirancang pada tahap sebelumnya. Lingkungan pada implementasi terdiri dari perangkat keras (*hardware*) dan perangkat lunak (*software*). Penjelasan dari bagian implementasi sebagai berikut:

1. Perangkat keras (*hardware*)

Processor	: intel inside core i7
Ram	:8,00 GB
Harddisk	: 1 Terrabyte (TB)

2. Perangkat lunak (*software*)

Sistem Operasi	: Windows 10
Bahasa pemrograman	: PHP
Tools	: Sublime Text 3, *Microsoft Visio*
Web Server	: Apache
Web Browser	: *Mozila Firefox, Chrome*
Database	: *MySQl*

2. Pengujian

Pada tahap pengujian akan dilakukan pengujian *whitebox*, pengujian akurasi aplikasi dengan menggunakan *confusion matrix*. Kinerja dari kedua metode tersebut akan dibandingkan, sehingga dapat diketahui metode yang paling efektif dalam melakukan klasifikasi *tweet* prostitusi di Twitter dengan beberapa mekanisme pembagian data latih dan data uji. Hasil dari setiap metode kemudian divalidasi dengan *confusion matrix*.

Dataset yang digunakan pada penelitian yaitu sebanyak 2000 tweet. Terdapat tiga pembagian data latih dan data uji yaitu 90% data latih dengan 10% data uji, 80% data latih dengan 20% data uji dan 70% data latih dengan 30% data uji.

3.6 Kesimpulan dan Saran

Pada tahapan ini berisi kesimpulan dan saran yang akan ditarik dari hasil penelitian. Kesimpulan berisi mengenai hasil dari penelitian yang sudah dilakukan seta untuk saran yaitu berisi saran-saran untuk perkembangan aplikasi kedepannya agar menjadi lebih baik.

DAFTAR PUSTAKA

Amisah, S., Honggowibowo, A. S., & Pujiastuti, A. (2016). Klasifikasi Teks Menggunakan Chi Square Feature Selection Untuk Menentukan Komik Berdasarkan Periode, Materi Dan Fisikdengan Algoritma Naivebayes. *Compiler*, *5*(2), 59–66. https://doi.org/10.28989/compiler.v5i2.171

Azeharie, S. (2014). Analisis Penggunaan Twitter Sebagai Media Komunikasi Selebritis Di Jakarta. *Jurnal Komunikasi Untar*, *6*(2), 83–98.

Azhar, Y. (2017). Klasifikasi Akun Prostitusi Berdasarkan Skoring Tweet. *Jurnal Ilmiah Networking Engineering Research Operation (NERO)*, *3*(1), 15–19.

Imanda, A. C., Hidayat, N., & Furqon, M. T. (2018). Klasifikasi Kelompok Varietas Unggul Padi Menggunakan Modified K- Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *2*(8), 2392–2399.

Irawan, Y. (2017). Pengujian Sistem Informasi Pengelolaan Pelatihan Kerja UPT BLK Kabupaten Kudus dengan Metode Whitebox Testing. *Sentra Penelitian Engineering Dan Edukasi*, *9*(3), 59–63.

Iriantoro, D. N. D., Dewi, C., & Fitriani, D. (2018). Klasifikasi pada Penyakit Dental Caries Menggunakan Gabungan K-Nearest Neighbor dan Algoritme Genetika. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *2*(8), 2926–2933.

Mentari, N. D., Fauzi, M. A., & Muflikhah, L. (2018). Analisis Sentimen Kurikulum 2013 Pada Sosial Media Twitter Menggunakan Metode K-Nearest Neighbor dan Feature Selection Query Expansion Ranking. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya, *2*(8), 2739–2743.

Negoro, P. A., & Atmadja, I. G. O. (2014). Analisis Terhadap Prostitusi Online

Ditinjau Dari Hukum Pidana Positif Di Indonesia. Jurnal Universitas Sebelas Maret, *3*(1), 71. file:///C:/Users/WIN10-PC/Documents/SKRIPSI/JURNAL SKRIPSII/Teori Pasal.pdf

Nurhadi, Z. F. (2017). Model Komunikasi Sosial Remaja Melalui Media Twitter. *Jurnal ASPIKOM*, *3*(3), 539. https://doi.org/10.24329/aspikom.v3i3.154

Paramastri, N. A., & Gumilar, G. (2019). Penggunaan Twitter Sebagai Medium Distribusi Berita dan News Gathering Oleh Tirto.Id. *Jurnal Kajian Jurnalisme*, *3*(1), 18. https://doi.org/10.24198/jkj.v3i1.22450

Puspita, R. S. D., & Gumelar, G. (2014). Pengaruh Empati Terhadap Perilaku Prososial Dalam Berbagi Ulang Informasi Atau Retweet Kegiatan Sosial Di Jejaring Sosial Twitter. *JPPP - Jurnal Penelitian Dan Pengukuran Psikologi*, *3*(1), 1–7. https://doi.org/10.21009/jppp.031.01

Srianto, D., & Mulyanto, E. (2016). Perbandingan K-Nearest Neighbor Dan Naive Bayes. *Jurnal Techno.COM*, *15*(3), 241–245.

Wulandari, T. (2018). Klasifikasi Jenis Emosi dari Tweet Berbahasa Indonesia Menggunakan Metode *Support Vector Machine*.