**Analisis Sentimen Aplikasi Go-Jek Menggunakan Metode SVM (Support Vektor Machine) dan NBC (Naive Bayes Classifier)**

**(Studi Kasus : Komentar pada Play Store)**

**SKRIPSI**



Disusun oleh :

MOCH TRI ANJASMORO

NIM : 161552019149984

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS WIDYAGAMA MALANG**

**2020**

# BAB I

# PENDAHULUAN

## Pendahuluan

Transportasi digunakan oleh masyarakat untuk mempermudah aktifitas sehari – hari. Mulai dari aktifitas bekerja, sekolah, belanja, silaturahmi dan juga mengantar barang. Dengan banyaknya masyarakat dalam kegiatan sehari-hari, sangat sulit untuk dilakukan terutama bagi masyrakat yang tidak dapat mengendarai motor atau mobil sendiri. Lalulintas yang semakin macet akan menambah kesulitan masyarakat dalam melakukan kegiatannya. Untuk itu dibutuhkan bantuan jasa orang lain yang berupa jasa kurir, belanja dan antar. Jasa ini biasa dilakukan oleh pekerja ojek motor atau tukang ojek.

Ketergantungan masyarakat terhadap ojek tidak diimbangi dengan palayanan yang kurang memuaskan, mulai dari tarif yang tidak seragam, sepeda motor tidak standart, tukang ojek tidak membawa helm untuk penumpang dan juga penumpang sering menunggu atau mencari keberadaan ojek jika berada disuatu tempat yang jauh dari jalan umum.

Kemajuan teknologi dalam bidang transportasi memberi sebuah keuntungan bagi penumpang maupun tukang ojek. Go-Jek adalah pelopor Ojek online di Indonesia dan menjadi yang terbesar untuk saat ini, dan bernaung di bawah perusahaan PT. Go-Jek Indonesia. Perusahaan ini bergerak di bidang jasa layanan transportasi sebagai perantara yang menghubungkan antara para pengendara ojek dengan pelanggan. Go-Jek merupakan model transportasi yang harus dipesan melalui via www.GoJek.com melalui Smartphone untuk Aplikasi Go-Jek, iOS dan android [1].

Go-Jek meraih Decacorn dan jadi perusahaan pertama yang menyandang status ini. Menurut data CB Insight yang dikutip Jumat (5/4/2019), Go-Jek memiliki nilai valuasi lebih dari 10 miliar dolar Amerika Serikat (AS) dan menduduki peringkat nomor 19 di dunia. Decacorn adalah status tingkat lanjut yang diberikan kepada startup dengan valuasi lebih dari 10 miliar dolar. CB Insights mencatat terdapat belasan perusahaan rintisan yang memiliki status ini [2].

Pengguna jasa ojek online memberi pendapat dan opini melalui berbagai media. Salah satunya melalui ulasan google play. Biasanya, ulasan pengguna berisi dua bagian, yaitu nilai rating dan komentar secara tekstual. Nilai rating menunjukkan evaluasi keseluruhan pengalaman pengguna menggunakan skala numerik, namun komentar tekstual mampu bercerita lebih mendalam. Setelah beberapa bulan, mungkin ada lebih dari sepuluh ribu komentar tekstual dari aplikasi baru yang diluncurkan di market [3].

Berdasarkan penjelasan diatas yang ada, maka diperlukan sebuah solusi berupa analisis terhadap saran atau keluhan yang diterima oleh Go-Jek sehingga dapat diketahui informasi sentimen terhadap aplikasi yang telah diluncurkan. Sebelumnya telah dilakukan penelitian klasifikasi keluhan pada akun facebook group iRaise Helpdesk menggunakan metode SVM berhasil mendapatkan nilai akurasi tertinggi sebesar 95.67% didapat hasil perbandingan klasifikasi data antara prediksi manual dengan prediksi model, maka hasil ketepatan prediksi yang dibangun model sebesar 995 data, dan tingkat kesalahan prediksi sebanyak 45 data [4].

Pada penelitian ini akan membandingan hasil metode SVM (Support Vector Machine) dan NBC (Naive Bayes Classifier) untuk mengklasifikasikan komentar pada play store terhadap aplikasi Gojek.

## Rumusan masalah

Berdasarkan latar belakang masalah maka dapat dirumuskan bagaimana tingkat akurasi metode Support Vektor Machine dan Naive Bases Classifier dalam mengklasifikasi sentimen kedalam kelas positif dan kelas negatif.

## Tujuan penelitian

Membangun suatu sistem yang dapat menghasilkan informasi tentang perbandingan metode Support Vektor Machine dan Naive Bases Classifier aplikasi Go-jek berdasarkan hasil sentiment analysis terhadap komentar yang terdapat pada google play.

## Batasan masalah

Agar penelitian yang dilakukan lebih terarah dan mencapai sasaran yang ditentukan maka diperlukan suata pembatasan masalah atau ruang lingkup kajian sebagai berikut :

1. Data sentimen mengenai aplikasi Go-jek diperoleh dari komentar aplikasi Go-Jek di playstore.
2. Data sentimen diklasifikasikan kedalam dua kelas yaitu positif dan negatif.
3. Metode yang digunakan adalah SVM dan NBC.
4. Sistem aplikasi ini berjalan diconsole python.
5. Bahasa yang digunakan adalah python.

# BAB II

# LANDASAN TEORI

## Penelitian terdahulu

Pada penelitian klasifikasi keluhan menggunakan metode SVM (Support Vector Machine) studi kasus akun facebook group iRaise Helpdesk menggunakan dataset 1040 data, didapat hasil perbandingan klasifikasi data antara prediksi manual dengan prediksi model, maka hasil ketepatan prediksi yang dibangun model sebesar 995 data, dan tingkat kesalahan prediksi sebanyak 45 data. Ketepatan klasifikasi sebanyak 995 data, dan klasifikasi error sebanyak 45 data [5].

Media sosial masyarakat mengeluarkan beragam opini tentang pelayanan dari transportasi ini dengan jumlah yang banyak, sehingga terdapat kesulitan untuk menentukan opini yang bersifat positif, negatif ataupun netral. Penelitian terdahulu mengenai analisis sentimen pada Twitter menggunakan metode Naïve Bayes dengan data sebanyak 3000 data yang terdiri dari 1000 kalimat netral, 1000 kalimat positif dan 1000 kalimat negatif dengan hasil akurasi hingga 88%. Sedangkan penelitian yang akan dilakukan yaitu membuat sistem yang mampu mengklasifikasikan sentimen ke dalam sentimen positif, netral atau negatif serta menyalurkan opini tersebut ke setiap jasa yang bersangkutan dengan opini yang muncul. Hasil yang didapatkan dari akurasi naïve bayes memperoleh ketepatan 80% [6].

## Landasan teori

* 1. **SVM (Support Vektor Machine)**

SVM adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. SVM memiliki prinsip dasar linier classifier yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Fungsi kernel yang digunakan untuk memetakan dimensi awal (dimensi yang lebih rendah) himpunan data ke dimensi baru (dimensi yang relatif lebih tinggi) [4].

Macam – macam fungsi kernel :

1. Kernel linier :

K(u , v) = uvT

1. kernel polinomial :

K(u , v) = (1 + u vT)d, d ≥ 2

1. kernel RBF :

K(u , v) = exp(-y ║u – v ║2), y > 0

1. Kernel Sigmoid :

K(u , v) = exp

* 1. **NBC (Naive Bases Classifier)**

Naïve bayes classifier adalah penggolongan menggunakan statistik sederhana berdasarkan teorema bayes yang mengasumsikan bahwa keberadaan atau ketiadaan dari suatu kelas dengan fitur lainnya (Lorosae, dkk. 2018). Tahap ini adalah proses pengklasifikasian yang akan diproses menggunakan metode Naive Bayes Classifier untuk menentukan data mana yang termasuk dalam opini positif dan negatif [7]. Berikut rumusnya

P(H│X) =

Keterangan

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

* 1. **Python**

Python merupakan bahasa pemrograman tingkat tinggi. Hal ini disebabkan karena kode yang dituliskan akan di compile menjadi byte code dan dieksekusi sehingga Python cocok digunakan untuk scripting language, aplikasi web dan lain sebagainya. Hal lain yang menjadikan bahasa ini menjadi bahasa pemgrograman tingkat tinggi adalah Python dapat di extend kedalam bahasa C dan C++ serta bahasa pemgrograman ini memiliki struktur konstruksi yang kuat (blok kode, fungsi, class, modules, dan packages) dan serta konsisten menggunakan konsep Object Oriented Programming (OOP) (Kuhlman, 2015).

* 1. ***Pre-processing***

*Pre-processing* dilakukan untuk memproses awal data dengan tujuan komentar untuk diklasifikasikan sebagai bersih dari tanda baca dan kata-kata yang tidak penting [8]. Beberapa langkah yang dilakukan pre-processing:

1. *Cleaning* : digunakan untuk membersihkan komentar dari setiap tanda baca untuk menyederhanakan proses selanjutnya.
2. *Case folding* : mengubah semua huruf besar menjadi huruf kecil.
3. *Stop word removal* : digunakan untuk menghapus kata – kata yang tidak memiliki makna.
4. *Tokenization* : memisahkan kata berdasarkan spasi yang terdapat dalam kalimat.

Tabel 1 Tahap *pre-processing*

|  |  |
| --- | --- |
| Proses data | Data |
| Komentar | Disini saya bukan mau komplain masalah aplikasi, tapi mengenai sikap para driver yg kalo ambil orderan pilih-pilih, maksudnya apa????? Cuma karena saya order pake promoan terus di kesampingkann??? Udah ngerasa okeee luuu sebagai mode transportasi disini?????? |
| *cleaning* | Disini saya bukan mau komplain masalah aplikasi tapi mengenai sikap para driver yg kalo ambil orderan pilih-pilih maksudnya apa Cuma karena saya order pake promoan terus di kesampingkann Udah ngerasa okeee luuu sebagai mode transportasi disini |
| *Case folding* | disini saya bukan mau komplain masalah aplikasi tapi mengenai sikap para driver yg kalo ambil orderan pilih-pilih maksudnya apa cuma karena saya order pake promoan terus di kesampingkann udah ngerasa okeee luuu sebagai mode transportasi disini |
| *Stopword removal* | komplain masalah pilih-pilih kesampingkann |
| *tokenization* | “komplain”, “masalah”, “pilih-pilih”, “kesampingkann” |

* 1. **Pembobotan kata (*Term Weighting*)**

Pembobotan kata memberikan bobot pada fitur kata berdasarkan frekuensi kemunculan kata. Fitur kata yang telah diberi bobot dapat digunakan untuk proses klasifikasi [9]. Tahap pembobotan kata yaitu:

1. Term Frequency (tf)

TF (Term Frequency) adalah frekuensi dari kemunculan sebuah term dalam dokumen yang bersangkutan.

1. Document Frequency (df)

Document Frequency (df) adalah jumlah dokumen yang mengandung term.

1. Inverse Document Frequency (idf)

IDF (Inverse Document Frequency) merupakan sebuah perhitungan dari bagaimana term didistribusikan secara luas ada koleksi dokumen yang bersangkutan. Rumus idf

D = jumlah dokumen

df = dokumen frekuensi

1. Term Frequency-Inverse Document Frequency (tf-idf)

Term Frequency-Inverse Document Frequency (tf-idf) adalah hasil perkalian dari pembobotan term frequency dan inverse document frequency dari suatu term.

D = jumlah dokumen

df = dokumen frekuensi

Tf = term frekuensi

* 1. **Analisis Sentimen**

Analisis Sentimen adalah penambangan kontekstual teks yang mengidentifikasi dan mengekstrak informasi subjektif dalam sumber, dan membantu para pembisnis untuk memahami sentimen sosial dari merek, produk atau layanan mereka saat memantau percakapan online.

# BAB III

# METODE PENELITIAN

## Metode Pengambilan Data

Metode pengambilan data yang digunakan untuk mengumpulkan data-data pendukung yang dibutuhkan dalam proses pembuatan aplikasi. Data didapat dari Play store pada komentar aplikasi go-jek. Beberapa cara yang dapat digunakan adalah melalui studi literatur, yaitu dengan mengumpulkan dan mempelajari beberapa referensi dari berbagai sumber yang berkaitan dengan judul penelitian yang dilakukan. Selain itu, untuk dataset yang digunakan diambil melalui Webscrapper dengan proses scrapping. Dataset yang diambil berupa file csv.

## Metode Pengembangan Perangkat Lunak

Pada penelitian ini, penulis menggunakan metode Waterfall. Metode waterfall sering juga disebut model sekuensial linier (sequencial linier) atau alur hidup klasik (classic life cycle). Metode ini menyediakan pendekatan alur hidup perangkat secara sekuensial atau terurut dimulai dari analisis kebutuhan (requirement analysis), desain sistem (design), pengodean (coding), pengujian (testing) dan tahap pendukung (implementation). Gambar 1.1 merupakan pengembangan perangkat lunak metode Waterfall.



Gambar 1. 1

Tahapan metode waterfall adalah sebagai berikut:

1. Requirement Analysis

Ditahap ini penulis akan melakukan analisa kebutuhan yang akan digunakan untuk pengembangan aplikasi. Analisa kebutuhan dilakukan untuk melengkapi data yang diperlukan dalam pengembangan aplikasi sehingga dapat menghasilkan aplikasi yang baik.

1. System Design

Desain sistem adalah persiapan rancang bangun implementasi yang menggambarkan bagaimana suatu sistem dibentuk yang berupa penggambaran, perencanaan, dan pembuatan sketsa atau pengaturan dari beberapa elemen yang terpisah kedalam satu kesatuan yang utuh dan berfungsi, menyangkut didalamnya konfigurasi komponen perangkat lunak dan perangkat keras dari suatu sistem.

1. Testing

Seluruh unit yang dikembangkan dalam tahap coding diintegrasikan kedalam sistem setelah pengujian yang dilakukan masing-masing unit. Setelah integrasi seluruh sistem diuji untuk mengecek setiap kegagalan maupun kesalahan.

1. Implementation

Tahap implementasi adalah tahap dimana setelah melakukan desain program yang diterjemahkan kedalam kode program dengan menggunakan bahasa pemrograman python. Pada tahap ini nantinya akan dilakukan pengujian terhadap sistem implementasi analisis clustering dan sentiment data komentar aplikasi gojek di playstore menggunakan metode SVM dan NBC.

## Metode Pengolahan Data

Data yang didapat dari Webscrapping komentar merupakan data yang langsung diambil dari komentar pengguna langsung. Data tersebut perlu dilakukan pengolahan agar menjadi data yang mudah digunakan dalam proses sentiment analysis. komentar akan mengalami penyeleksian kata-kata komentar sehingg komentar menjadi lebih ringkas. Beberapa komponen dari komentar akan dihapus untuk menyeleksi komentar. Proses ini dapat disebut dengan pre-processing. Setelah melalui pre-processing, data yang berupa teks akan diubah dalam bentuk angka melalui perhitungan TF/IDF. Nilai TF/IDF ini yang akan menjadi masukan untuk algoritma SVM dan NBC. Algoritma SVM dan NBC akan menghasilkan klasifikasi untuk data yang dimasukkan.

## Analisa Data

Komponen yang diperlukan untuk melakukan penelitian ini salah satunya adalah data. Data yang diperlukan sebagai objek pengujian sentimen analisis yaitu tweet dari komentar aplikasi gojek pada playstore yang dijadikan dalam format dokumen .csv.

## Analisa Kebutuhan

Dalam membangun sistem aplikasi Analisis Sentimen Twitter Untuk Penilaian Opini Kualitas aplikasi gojek Menggunakan NBC dan SVM terdapat beberapa analisa kebutuhan antara lain meliputi kebutuhan perangkat lunak (software), perangkat keras (hardware), kebutuhan pengguna sistem (brainware), dan file.

## Perancangan Sistem

Dalam merancang sistem ini. Penulis membangi menjadi beberapa tahapan proses yang harus direncanakan. Berikut adalah tahapan-tahapan yang harus dirancang :

1. Text Pre-Processing

Pada tahapan ini teks perlu dilakukan sebuah ekstraksi sehingga teks berubah menjadi data numerik yang dapat diolah. Prose-proses tersebut biasa disebut text pre-processing. Gambar 1.2 menunjukkan bagaimana proses Text Pre-Processing dilakukan pada teks masukan.



Gambar 1. 2

Pada gambar 1.2. Teks masukan yang digunakan adalah tweet bahasa Indonesia. Proses pertama yang dilakukan dalam melakukan case folding yaitu mengubah semua huruf dalam dokumen menjadi huruf kecil, hanya karakter huruf a-z dan karakter angka 0-9 yang diterima dan tahap tokenizing akan pemotongan teks masukan berdasarkan tiap kata yang menyusunnya. Setelah diperoleh daftar kata yang membentuk teks masukan, dilakukan filter terhadap daftar kata tersebut berdasarkan stop-word list. Selanjutnya dilakukan proses stemming yaitu mencari kata dasar dari setiap kata yang sudah terpisah. Stemming yang dilakukan mengunakan algoritma Nazief dan Adriani. Dataset tweet yang telah mengalami preprocessing akan masuk ke proses selanjutnya, yaitu perhitungan TF IDF. Proses ini akan sangat menentukan training dan testing yang menjadi proses selanjutnya karena training dan testing akan menggunakan dataset yang telah diubah kedalam TF IDF, yaitu data yang berupa angka dari ektraksi kalimat dalam tweet. TF IDF merupakan salah sau cara yang digunakan untuk menghitung bobot setiap kata dalam setimen analysis. Perhitungan ini dilakukan dengan cara menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap kata dalam setiap tweet. Rumus yang digunakan dalam menghitung TF IDF sebagai berikut.

D = jumlah dokumen

df = dokumen frekuensi

Tf = term frekuensi

Setelah didapatkan bobot dari masing-maisng kata dalam setiap komentar, maka akan dilakukan proses perhitungan menggunakan metode SVM dan NBC. Langkah- langkah yang bisa dilakukan yaitu :

SVM

Teori SVM diawali dengan mengelompokkan kasus linier yang mampu untuk dipisahkan (dikelompokan) dan dibagi sesuai dengan kelas ataupun hubungan sebab akibatnya. Dalam teori ini, fungsi yang digunakan adalah fungsi linier yang dapat didefinisikan sebagai berikut:

g(x) := sgn(f(x))              (2.1)

dengan              (2,2)

atau      (2.3)

dimana x, w ∈and b ∈

Permasalahan dalam klasifikasi ini dapat dirumuskan dalam satu set parameter (w, b) sehingga f(xi) =<w, x > +b = yi untuk semua i. Teknik ini berfungsi untuk menemukan fungi pemisah ( clasifier / hyperplane ) yang terbaik diantara banyak fungsi yang tidak terbatas jumlahnya yang memungkinkan untuk memisahkan dua buah objek. Hyperplan yang paling baik adalah klasifier yang terdapat di tengah atau diantara dua set objek dar dua kelas klasifikasi, yang didasarkan atas pemaksimalan ekuivalen dari margin atau jarak diantara kedua set yang berbeda kelasnya.

NBC

* Membaca Data Training
* Menghitung Jumlah dan Probabilitas, namun jika data numerik maka
  + Menghitung nilai *mean* dan Dtandar Deviasi dari masing-masing parameter yang merupakan numerik. Adapun persamaan untuk mencari nilai rata-rata hitung (*mean*) adalah seperti dalam persamaan NBC berikut ini:

Atau

Keterangan :

**μ** : nilai rata-rata hitung (mean)

**xi** : nilai x ke-i

**n** : jumlah sampel

Sedangkan persamaan untuk menghitung nilai Nilai Simpangan Baku (Standar Deviasi) dirumuskan sebagai berikut :

Keterangan :

**σ** :standar deviasi

**xi** : nilai x ke-i

**μ** : nilai rata-rata hitung (mean)

**n** : jumlah sampel

* + Menghitung nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
* Mendapatkan nilai dalam tabel mean, Standar Deviasi dan Probabilitas dengan rumus :

P(H│X) =

Keterangan

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

* Menghasilkan Solusi

## Implementasi Sistem

Implementasi sistem dilakukan dengan mengacu perancangan sistem berdasarkan metode waterfall.

## Validasi dan Evaluasi

Validasi yang digunakan dalam penelitian ini adalah k-fold cross validation. Dalam k-fold cross validation, data awal dipartisi secara acak menjadi sejumlah k subset (fold), yaitu D1, D2, …, Dk, yang masing-masingnya berukuran sama. Proses training dan testing dilakukan sebanyak k kali. Dalam iterasi ke-i, data subset Di digunakan sebagai dataset testing, sedangkan data lainnya digunakan dalam proses training untuk mendapatkan suatu model klasifikasi yang nantinya akan digunakan dalam proses testing. Evaluasi performansi dilakukan untuk menguji hasil dari klasifikasi dengan mengukur nilai performansi dari sistem yang telah dibuat. Parameter pengujian yang digunakan untuk evaluasi yaitu akurasi yang perhitungannya diperoleh dari tabel coincidence matrix (matrik klasifikasi atau tabel kontigensi).

## Analisa hasil dan Kesimpulan

Pada tahap terakhir yaitu berupa analisa hasil penelitian dan penarikan kesimpulan. Setelah melakukan pengujian aplikasi maka penulis akan melakukan analisa hasil penelitian dan membuat kesimpulan. Salah satu poin kesimpulan yang direncanakan adalah perbandingan penilaian secara manual dan penilaian sistem.

# BAB IV

# Hasil Dan Pembahasan

## SVM

Pada metode SVM dalam pengujian ini menggunakan 3 kernel yaitu kernel linear, kernel polynomial, kernel RBF.

## Kernel Linear

Pada table 1 terdapat score accuracy, recall dan precision metode svm kernel linear. Berdasarkan table 1 score accuracy, recall dan linear paling tinggi adalah 0,87129 dengan data testing sebanyak 10% dan 90% data training. Score accuracy, recall dan precision metode svm kernel linear paling terendah adalah 0,81354 dengan data testing sebanyak 90% dan 10% data training.

Table 1 SVM Kernel Linear

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | | SVM Kernel Linear | | |
| Testing | Training | Accuracy | Recall | Precision |
| 10% | 90% | 0.87129 | 0.87129 | 0.87129 |
| 20% | 80% | 0.86567 | 0.86567 | 0.86567 |
| 30% | 70% | 0.83056 | 0.83056 | 0.83056 |
| 40% | 60% | 0.85287 | 0.85287 | 0.85287 |
| 50% | 50% | 0.84431 | 0.84431 | 0.84431 |
| 60% | 40% | 0.84359 | 0.84359 | 0.84359 |
| 70% | 30% | 0.84451 | 0.84451 | 0.84451 |
| 80% | 20% | 0.82896 | 0.82896 | 0.82896 |
| 90% | 10% | 0.81354 | 0.81354 | 0.81354 |

## SVM Kernel Polynomial

Pada table 2 terdapat score accuracy, recall dan precision metode svm kernel polynomial. Berdasarkan table 2 score accuracy, recall dan linear paling tinggi adalah 0,49813 dengan data testing sebanyak 80% dan 20% data training. Score accuracy, recall dan precision metode svm kernel polynomial paling terendah adalah 0,47525 dengan data testing sebanyak 10% dan 90% data training.

Table 2 SVM Kernel Polynomial

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | | SVM Kernel Polynomial | | |
| Testing | Training | Accuracy | Recall | Precision |
| 10% | 90% | 0.47525 | 0.47525 | 0.47525 |
| 20% | 80% | 0.48259 | 0.48259 | 0.48259 |
| 30% | 70% | 0.48173 | 0.48173 | 0.48173 |
| 40% | 60% | 0.49127 | 0.49127 | 0.49127 |
| 50% | 50% | 0.48703 | 0.48703 | 0.48703 |
| 60% | 40% | 0.48752 | 0.48752 | 0.48752 |
| 70% | 30% | 0.49643 | 0.49643 | 0.49643 |
| 80% | 20% | 0.49813 | 0.49813 | 0.49813 |
| 90% | 10% | 0.49723 | 0.49723 | 0.49723 |

## SVM Kernel RBF

Pada table 3 terdapat score accuracy, recall dan precision metode svm kernel RBF. Berdasarkan table 3 score accuracy, recall dan linear paling tinggi adalah 0,49813 dengan data testing sebanyak 80% dan 20% data training. Score accuracy, recall dan precision metode svm kernel RBF paling terendah adalah 0,47525 dengan data testing sebanyak 10% dan 90% data training.

Table 3 SVM Kernel RBF

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | | SVM Kernel RBF | | |
| Testing | Training | Accuracy | Recall | Precision |
| 10% | 90% | 0.47525 | 0.47525 | 0.47525 |
| 20% | 80% | 0.48259 | 0.48259 | 0.48259 |
| 30% | 70% | 0.48173 | 0.48173 | 0.48173 |
| 40% | 60% | 0.49127 | 0.49127 | 0.49127 |
| 50% | 50% | 0.48703 | 0.48703 | 0.48703 |
| 60% | 40% | 0.48752 | 0.48752 | 0.48752 |
| 70% | 30% | 0.49643 | 0.49643 | 0.49643 |
| 80% | 20% | 0.49813 | 0.49813 | 0.49813 |
| 90% | 10% | 0.49723 | 0.49723 | 0.49723 |

## NBC

Pada table 4 terdapat score accuracy, recall dan precision metode NBC. Berdasarkan table 4 score accuracy, recall dan linear paling tinggi adalah 0,86139dengan data testing sebanyak 10% dan 90% data training. Score accuracy, recall dan precision metode NBC paling terendah adalah 0,76471 dengan data testing sebanyak 90% dan 10% data training.

Table 4 NBC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data | | NBC | | |
| Testing | Training | Accuracy | Recall | Precision |
| 10% | 90% | 0.86139 | 0.86139 | 0.86139 |
| 20% | 80% | 0.81592 | 0.81592 | 0.81592 |
| 30% | 70% | 0.8206 | 0.8206 | 0.8206 |
| 40% | 60% | 0.81047 | 0.81047 | 0.81047 |
| 50% | 50% | 0.81238 | 0.81238 | 0.81238 |
| 60% | 40% | 0.797 | 0.797 | 0.797 |
| 70% | 30% | 0.8117 | 0.8117 | 0.8117 |
| 80% | 20% | 0.8015 | 0.8015 | 0.8015 |
| 90% | 10% | 0.76471 | 0.76471 | 0.76471 |

## Pembahasan

Berdasarkan table 1 sampai dengan table 4 score accuracy, recall dan precision paling tinggi adalah 0,87129 dengan data testing sebanyak 10% dan 90% data training menggunakan metode SVM kernel linear. Score accuracy, recall dan precision yang paling rendah adalah 0,47525 dengan data testing sebanyak 10% dan 90% data training menggunakan metode SVM kernel RBF, polynomial dan sigmoid.

Berdasarkan banyak data yang menjadi data training dan data testing, data testing 10% dan training 90% memiliki score accuracy, recall dan precision yang paling tertinggi memiliki score 0,87129 dan yang paling rendah data testing 90% dan data training 10% memiliki score 0,47525.

Dari data table diatas untuk mendapatkan score accuracy, recall dan precision yang tinggi dipengaruhi oleh metode yang dipakai dan data testing dan data training yang digunakan.

# BAB V JADWAL PENELITIAN

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NO | KEGIATAN | FEBRUARI | | | | MARET | | | | APRIL | | | | MEI | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Pengumpulan Data |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Identifikasi Masalah |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Analisa Kebutuhan Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Perancangan Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Coding |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Testing |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Revisi Rancangan dan Code  Program |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Implementasi Sistem |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | Pembimbingan Penulisan  Naskah Skripsi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 | Penulisan Akhir Laporan |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

# REFERENSI

[1] M. Damis dan T. Mulianti, “PENGGUNA OJEK ONLINE DI KALANGAN MAHASISWA UNIVERSITAS SAM RATULANGI MANADO,” no. 21, hlm. 19, 2018.

[2] Dipna Videlia Putsanra, “Go-Jek Jadi Startup Pertama yang Raih Decacorn di Indonesia,” *Go-Jek Jadi Startup Pertama yang Raih Decacorn di Indonesia*, 05-Apr-2019. [Daring]. Tersedia pada: https://tirto.id/go-jek-jadi-startup-pertama-yang-raih-decacorn-di-indonesia-dleL.

[3] L. B. Ilmawan dan E. Winarko, “Aplikasi Mobile untuk Analisis Sentimen pada Google Play,” vol. 9, no. 1, hlm. 12, 2015.

[4] M. Affandes, “Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus : Akun Facebook Group iRaise Helpdesk),” hlm. 7, 2017.

[5] M. Affandes, “Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) (Studi Kasus : Akun Facebook Group iRaise Helpdesk),” hlm. 7, 2017.

[6] D. G. Nugroho, Y. H. Chrisnanto, dan A. Wahana, “ANALISIS SENTIMEN PADA JASA OJEK ONLINE MENGGUNAKAN METODE NAÏVE BAYES,” hlm. 6, 2016.

[7] Y. S. Mahardhika dan E. Zuliarso, “ANALISIS SENTIMEN TERHADAP PEMERINTAHAN JOKO WIDODO PADA MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAIVES BAYES CLASSIFIER,” hlm. 5, 2018.

[8] F. S. Fitri, M. N. S. Si, dan C. Setianingsih, “Sentiment Analysis on the Level of Customer Satisfaction to Data Cellular Services Using the Naive Bayes Classifier Algorithm,” dalam *2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)*, Bali, 2018, hlm. 201–206, doi: 10.1109/IOTAIS.2018.8600870.

[9] U. Rofqoh, R. S. Perdana, dan M. A. Fauzi, “Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features,” hlm. 9.