

PERBANDINGAN METODE *SUPPORT VECTOR MACHINE* DAN *DECISION TREE* UNTUK ANALISIS SENTIMEN *REVIEW* KOMENTAR PADA APLIKASI TRANSPORTASI ONLINE

Khoirul Abbi Rokhman¹⁾, Berlilana²⁾ Primandani Arsi³⁾

^{1), 2), 3)} *Informatika, Universitas Amikom Purwokerto*

email : rokhmanabi@gmail.com¹⁾, berli_amikom@yahoo.com²⁾, ukhti.prima@amikompurwokerto.ac.id²⁾

Abstraksi

Secara umum transportasi digunakan untuk memudahkan manusia melakukan aktivitas sehari-hari. Saat ini, transportasi tidak hanya terdapat secara konvensional saja namun berkembang dengan adanya transportasi berbasis *online* yang harganya lebih terjangkau dan lebih praktis. Gojek Adalah salah satu aplikasi transportasi online yang memiliki pengguna bisa dikatakan banyak di Indonesia. Namun dalam system ini pasti memiliki banyak kekurangan yang dirasakan penggunanya. Dengan menganalisis kekurangan dari aplikasi perusahaan dapat mengetahui kekurangan dari aplikasi dan bagaimana cara memperbaikinya. Tujuan dari penelitian ini yaitu untuk melakukan analisis sentiment dengan menggunakan data ulasan yang terdapat pada *Google Play* guna mengetahui perbandingan keakurasian antara metode *Support Vector Machine* untuk mengklasifikasikan ulasan dari dua kategori yaitu ulasan positif dan negative. Kemudian dibandingkan dengan metode *Decision Tree*. Melalui klasifikasi diperoleh hasil akurasi sebesar 90.20% untuk metode *Support Vector Machine* sedangkan 89.80% untuk metode *Decision Tree*. Jadi bisa disimpulkan untuk metode *Support Vector Machine* nilai akurasinya lebih tinggi dibandingkan metode *Decision Tree*.

Kata Kunci :

Text Mining, Support Vector Machine, Decision Tree.

Abstract

In general, transportation is used to make it easier for humans to carry out their daily activities. At present, transportation is not only available conventionally but is developing with online-based transportation which is more affordable and more practical. Gojek is an online transportation application that has many users in Indonesia. But in this system there must be many disadvantages that are felt by its users. By analyzing the deficiencies of company applications, we can find out the deficiencies of the application and how to fix them. The purpose of this study is to conduct a sentiment analysis using review data found on Google Play to determine the accuracy comparison between the Support Vector Machine method to classify reviews from two categories, namely positive and negative reviews. Then compared with the Decision Tree method. Through classification, the results obtained are 90.20% accuracy for the Support Vector Machine method while 89.80% for the Decision Tree method. So it can be concluded for the Support Vector Machine method the accuracy value is higher than the Decision Tree method.

Keywords :

Text Mining, Support Vector Machine, Decision Tree

Pendahuluan

Dalam era globalisasi sekarang ini, persaingan bisnis yang dihadapi perusahaan semakin ketat, sehingga menuntut manajemen perusahaan untuk lebih cermat dalam menentukan strategi bersaingnya agar dapat memenangkan persaingan yang dihadapinya. Bisnis jasa semakin beragam guna memenuhi kebutuhan masyarakat. Transportasi merupakan sarana yang sangat penting untuk menunjang aktivitas atau mobilitas manusia setiap hari sehingga harus dipersiapkan dengan baik dan aman karena sangat berpengaruh dalam aktivitas perekonomian, pengiriman barang atau jasa

transportasi penumpang dan lain sebagainya. Di era modern saat ini masyarakat memiliki berbagai macam aktivitas dan untuk memenuhi aktivitas tersebut dibutuhkan transportasi sebagai alat penunjang dalam menjalankan aktivitasnya.

Saat ini banyak sekali aplikasi yang menawarkan jasa transportasi online di Indonesia bahkan bisa terus berkembang. Di antaranya, Aplikasi Transportasi Online yang sedang populer antara lain Gojek, Grab, Nujek, Oke Jack, Indo-Jek. Gojek merupakan salah satu aplikasi transportasi online terpopuler dan sering digunakan oleh banyak konsumen di Indonesia. Gojek pertama kali berdiri

pada tahun 2010. Gojek merupakan layanan transportasi online yang sangat populer bahkan bisa dikatakan berada di posisi pertama. Berdasarkan survei pada 19 April 2020 di Google Play, pengguna yang telah mengunduh aplikasi Gojek lebih dari 50 juta, sehingga bisa dikatakan Gojek lebih populer di era sekarang..

Ulasan pengguna sering kali digunakan sebagai alat yang efektif dan efisien dalam mencari informasi tentang suatu produk atau aplikasi. Bahwa studi terbaru menemukan bahwa hampir 50% pengguna internet mengandalkan rekomendasi dari mulut ke mulut (opini) sebelum menggunakan suatu produk, hal ini dikarenakan review dari pengguna lain dapat memberikan informasi terbaru mengenai produk tersebut berdasarkan perspektif pengguna lain. yang telah menggunakan produk. [1]

Pengguna yang tidak puas dengan layanan atau produk yang ditawarkan suatu perusahaan biasanya akan menuliskan keluhannya di kolom komentar pada aplikasi di situs *Google Play*. Di sisi lain, ada juga pelanggan yang puas mengungkapkan sikap positifnya terhadap suatu produk di kolom komentar pada aplikasi di situs *Google Play*. Disadari atau tidak, opini pelanggan tertulis, sedikit atau banyak, akan dapat memengaruhi pelanggan potensial. Namun, memantau dan mengatur opini publik juga tidak mudah. Ada terlalu banyak opini yang diposting di situs *Google Play* untuk diproses secara manual. Oleh karena itu, diperlukan suatu metode atau teknik khusus yang dapat secara otomatis mengkategorikan review tersebut, baik positif maupun negatif.

Setelah data terkumpul, dengan menggunakan text mining dapat diketahui percakapan apa saja yang sering dibicarakan oleh pengguna atau konsumen. Salah satu analisis penambangan teks adalah analisis sentimen yang dapat diterapkan pada review produk dari perusahaan yang terdaftar di *Google Play*. Analisis sentimen diterapkan untuk mengklasifikasikan umpan balik positif dan negatif dari konsumen sehingga dapat mempercepat dan mempermudah tugas perusahaan dalam meninjau kekurangan produk mereka. Jika ditemukan sentiment negatif, perusahaan dapat segera mengambil tindakan untuk mengatasinya. Metode yang dapat digunakan dalam analisis sentimen tinjauan pengguna adalah metode *Support Vector Machine* dan metode *Decision Tree*.

Support Vector Machine (SVM) adalah metode klasifikasi data mining. Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai metode klasifikasi yang efisien untuk masalah nonlinier. Selain itu, SVM juga merupakan teknik untuk menemukan hyperplane yang dapat memisahkan dua kumpulan data dari dua kelas yang berbeda. SVM memiliki kelebihan diantaranya menentukan jarak menggunakan *Support Vector Machine* sehingga proses komputasi menjadi cepat. Kemudian metode SVM yaitu metode

Decision Tree merupakan alat pendukung keputusan yang menggunakan model keputusan yang berbentuk seperti pohon. *Decision Tree* mendefinisikan berbagai alternatif yang mungkin untuk memecahkan masalah, dan ada juga kemungkinan faktor yang dapat mempengaruhi alternatif ini bersama dengan perkiraan akhir jika ada alternatif.

Setelah melakukan klasifikasi, penulis mencoba menggali dan menggali seluas-luasnya informasi apa saja yang terdapat pada setiap klasifikasi sentimen positif dan negatif yang dianggap penting untuk digunakan untuk berbagai keperluan. Tujuan dari penelitian ini adalah untuk mengetahui hasil perbandingan ketepatan klasifikasi persepsi pengguna Gojek dengan menggunakan metode *Support Vector Machine* dan metode *Decision Tree*. Harapannya dengan membandingkan metode *Support Vector Machine* (SVM) dan *Decision Tree*, penelitian ini dapat memperoleh hasil teks review yang lebih baik dan akurat. Dengan informasi yang diperoleh lebih baik dan lebih akurat, informasi ini nantinya dapat dijadikan acuan dalam menjaga popularitas, kualitas dan guna memperbaiki kekurangan serta mengevaluasi ke arah yang lebih baik.

Tinjauan Pustaka

Beberapa Landasan Teori yang digunakan

1. Transportasi Online

Transportasi Online adalah pelayanan jasa transportasi yang berbasis internet dalam setiap kegiatan transaksinya, mulai dari pemesanan, pemantauan jalur, pembayaran dan penilaian terhadap pelayanan jasa itu sendiri[2].

2. Gojek

Gojek adalah perusahaan penyedia jasa yang bergerak dalam bidang teknologi dan transportasi berbasis online. Selain melayani antar jemput penumpang (GO-RIDE), Gojek juga melayani pengiriman barang (GO-SEND), pemesanan makanan (GO-FOOD) dan beberapa layanan jasa lainnya. Selain itu fasilitas yang didapatkan oleh Driver sendiri bukan hanya dari sistem bagi hasil yang menguntungkan mereka juga mengatakan bahwa pendapatannya meningkat semenjak bergabung sebagai mitra, mereka juga mendapatkan santunan kesehatan dan kecelakaan, serta mendapatkan akses ke lebih banyak pelanggan aplikasi ojek online Gojek[3].

3. Text Mining

Text Mining adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis [4]. Selain itu *Text mining* atau *text analytics* adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan data mining dimana data mining mengolah data yang sifatnya terstruktur[5].

4. Support Vector Machine

Support Vector Machine (SVM) diperkenalkan oleh Vapnik pada tahun 1992 sebagai suatu teknik klasifikasi yang efisien untuk masalah nonlinear. *Support Vector Machine (SVM)* juga dikenal sebagai teknik pembelajaran mesin (*machine learning*) paling mutakhir setelah pembelajaran mesin sebelumnya yang dikenal sebagai *Neural Network (NN)*. Baik SVM maupun NN tersebut telah berhasil digunakan dalam pengenalan pola. Pembelajaran dilakukan dengan menggunakan pasangan data input dan data output berupa sasaran yang diinginkan. Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari hyperplane terbaik yang berfungsi sebagai pemisah dua buah kelas pada input space. SVM berusaha menemukan fungsi pemisah (*hyperplane*) dengan memaksimalkan jarak antar kelas. Dengan cara ini, SVM dapat menjamin kemampuan generalisasi yang tinggi untuk data-data yang akan datang[6].

5. Decision Tree

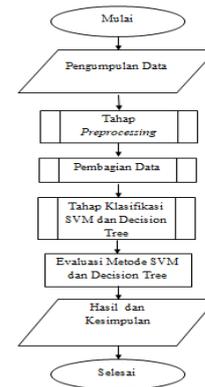
Decision Tree adalah top-down pohon rekursif dari algoritma induksi, yang menggunakan ukuran seleksi atribut untuk memilih atribut yang diuji. Algoritma *decision tree* mencoba untuk meningkatkan akurasi dengan menghapus cabang-cabang pohon yang mencerminkan noise dalam data [7].

Penelitian yang berjudul analisis sentiment pada Twitter untuk mengenai pengguna transportasi umum dalam kota besar di Indonesia. Penelitian ini menggunakan metode Support Vector Machine. Dalam menggunakan metode Support Vector Machine memperoleh hasil akurasi sebesar 78,12%. Variabel yang berpengaruh terhadap akurasi adalah jumlah data, perbandingan jumlah data latihan dan uji, serta perbandingan jumlah data positif dan negatif yang digunakan[8].

Selain itu penelitian tentang analisis sentiment online pengguna E-commerce. Metode yang digunakan pada penelitian ini Support Vector Machine dan *Maximum Entropy (Maxent)* dalam mengklasifikasikan teks ulasan berbahasa Indonesia. Klasifikasi dengan metode SVM diperoleh tingkat akurasi sebesar 91,95%. Sedangkan dengan metode *Maximum Entropy (Maxent)* memberikan tingkat akurasi yang lebih tinggi yaitu sebesar 92,98%[9].

Metode Penelitian

Penelitian ini dilakukan menggunakan langkah-langkah seperti pada Gambar 1.



Gambar 1. Konsep Penelitian

Berdasarkan dari gambar konsep penelitian di atas maka dijelaskan sebagai berikut :

1. Tahap Pengumpulan Data

Pada tahap pengumpulan data pada penelitian ini yaitu mengambil data sekunder yang diambil dari ulasan aplikasi Gojek di *google Play*.

2. Tahap Preprocessing

Dalam tahap ini, data yang telah diperoleh akan diolah untuk menemukan data yang diperlukan dan membuang data-data yang tidak diperlukan untuk proses analisis. Teknik pengolahan data pada tahap *preprocessing* meliputi :

- Case Folding yaitu data tersebut akan diubah menjadi sama (lowercase).
- Tokenizing yaitu teks pada data-data yang telah diperoleh akan dipisah untuk tiap katanya
- Replace Tokens yaitu proses perbaikan kata yang salah.
- Filtering yaitu proses mengeliminasi kata-kata yang tidak memiliki pengaruh atau tidak informative.
- Stemming merupakan proses menghilangkan semua imbuhan yang ada pada kata tersebut menjadi kata dasar.

3. Pembagian Data

Dalam penelitian ini data yang telah di peroleh akan dibagi menjadi data *training* dan data *testing*.

- Data *training* yaitu data yang akan di olah dengan metode yang akan digunakan, yang hasilnya akan digunakan sebagai prediksi untuk data *testing*.
- Data *testing* merupakan data yang akan diuji dan diprediksi.

4. Tahap Klasifikasi

Dalam tahap klasifikasi penelitian ini menggunakan beberapa cara yaitu :

- Klasifikasi menggunakan metode *Support Vector Machine (SVM)*, mencari *hyperplane* (batas keputusan) terbaik yang memisahkan tiap ulasan kedalam dua kelas yaitu ulasan positif dan ulasan negatif dengan menggunakan algoritma *Support Vector*

Machine. Dan selanjutnya data tersebut akan digunakan untuk prediksi terhadap data *testing*.

b. Klasifikasi dengan metode *decision tree*, mencari *hyperplane* (batas keputusan) terbaik yang memisahkan tiap ulasan kedalam dua kelas yaitu ulasan positif dan ulasan negatif dengan menggunakan algoritma *Decision Tree*. Dan selanjutnya data tersebut akan digunakan untuk prediksi terhadap data *testing*.

5. Evaluasi Metode SVM dan *Decision Tree*

Dalam pengujian dari kedua algoritma yang digunakan akan diuji dengan menggunakan $fold = 2$ s/d 10, pengujian dilakukan untuk mendapatkan akurasi terbaik dalam proses pengujiannya. Setelah itu kedua metode tersebut akan dibandingkan manakah metode yang mendapatkan nilai akurasi terbaik. Dari metode SVM akan menggunakan 3 kernel type yaitu dot, radial dan polynomial.

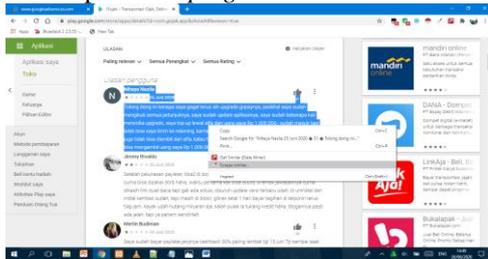
6. Hasil dan Kesimpulan

Dalam tahap ini dilakukan penalaran kesimpulan yang diperoleh berdasarkan hasil dari perbandingan klasifikasi dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Decision Tree*, dan perhitungan akurasi dengan menggunakan tabel *confision matrix* manakah dari kedua metode itu yang terbaik.

Hasil dan Pembahasan

1. Pengumpulan Data

Proses pengambilan data ulasan aplikasi Gojek yang ada pada *Google Play* menggunakan Proses *Scraping*. Dengan cara mengblok nama user, tanggal dan kalimat ulasan. Kemudian klik kanan lalu pilih *Scraping Similar*.



Gambar 2. Proses *Scraping*

2. Preprocessing

Setelah melakukan proses pengumpulan data maka kita akan melakukan pembersihan data. Tujuannya untuk mengurangi dimensi-dimensi kata yang tidak pengaruh terhadap proses analisis. Berikut tahapan *Preprocessing* :

a) *Case Folding*

Case Folding adalah tahap merubah kalimat menjadi huruf kecil (*lowercase*). Tujuan dari proses ini yaitu untuk memudahkan pada proses selanjutnya. Proses ini menggunakan operator *Transform Cases*. Berikut sebagian contoh hasil dari proses *Case Folding*.

Tabel 1. Hasil *Case Folding*

| No | Sebelum <i>Case Folding</i> | Sedudah <i>Case Folding</i> |
|----|---|---|
| 1 | Tolong dong ini kenapa saya gagal terus sih upgrade gopaynya, padahal saya sudah mengikuti semua petunjuknya, saya sudah update aplikasinya, saya sudah beberapa kali mencoba upgrade, saya top up lewat alfa dan uang saya Rp 1.000.000,- sudah masuk tapi tidak bisa saya kirim ke rekening, karna belum bisa di upgrade, bagaimana ini? uang saya juga tidak bisa diambil dari alfa, kalau terus seperti ini, saya akan laporkan ke POLISI agar bisa mengambil uang saya Rp 1.000.000,- tersebut. | tolong dong ini kenapa saya gagal terus sih upgrade gopaynya, padahal saya sudah mengikuti semua petunjuknya, saya sudah update aplikasinya, saya sudah beberapa kali mencoba upgrade, saya top up lewat alfa dan uang saya rp 1.000.000,- sudah masuk tapi tidak bisa saya kirim ke rekening, karna belum bisa di upgrade, bagaimana ini? uang saya juga tidak bisa diambil dari alfa, kalau terus seperti ini, saya akan laporkan ke polisi agar bisa mengambil uang saya rp 1.000.000,- tersebut. |
| 2 | Lengkap Setelah pelunasan paylater, tiba2 di blokir tanpa ada pemberitahuan, padahal saldo limitnya cuma bisa dipakai 30rb haha. waktu pertama kali bisa 500rb, di email jawabannya cuma dikasih link buat baca tapi gak ada solusi, disuruh update versi terbaru udah, di uninstal dan instal kembali sudah, tapi masih di blokir, giliran telat 1 hari bayar tagihan di telponin terus tiap jam, kayak udah hutang milyaran aja, kalah pulak la tukang kredit haha. Slogannya pasti ada jalan, tapi ya paham sendirilah | lengkap setelah pelunasan paylater, tiba2 di blokir tanpa ada pemberitahuan, padahal saldo limitnya cuma bisa dipakai 30rb haha. waktu pertama kali bisa 500rb, di email jawabannya cuma dikasih link buat baca tapi gak ada solusi, disuruh update versi terbaru udah, di uninstal dan instal kembali sudah, tapi masih di blokir, giliran telat 1 hari bayar tagihan di telponin terus tiap jam, kayak udah hutang milyaran aja, kalah pulak la tukang kredit haha. slogannya pasti ada jalan, tapi ya paham sendirilah |
| 3 | Saya sudah bayar paylater, janjinya cashback 30% paling lambat tgl 15 juni Tp sampai saat ini belum ada jawaban, kalau di email cm blg suruh tunggu. Bnr jg kata org kalau kayak hutang ratusan juta, Pas janji gini ga ada wujudnya, tolong lah jangan bohongi customer kalau | saya sudah bayar paylater, janjinya cashback 30% paling lambat tgl 15 juni tp sampai saat ini belum ada jawaban, kalau di email cm blg suruh tunggu. bnr jg kata org kalau kayak hutang ratusan juta, pas janji gini ga ada wujudnya, tolong lah jangan bohongi customer kalau |

| No | Sebelum Case Folding | Sedudah Case Folding |
|----|--|--|
| | memang ga bisa kasi cashback jangan janjikan apa2, toh kami juga punya tanggung jawab utk bayar,bukan tidak mau bayar. | janjikan apa2, toh kami juga punya tanggung jawab utk bayar,bukan tidak mau bayar. |

b).Tokenizing

Tokenizing adalah tahap memecahkan teks yang dapat berupa kalimat, paragraf atau dokumen menjadi token-token atau bagian-bagian tertentu berdasarkan tiap kata menyusunnya. Tokenizing dilakukan untuk mendapatkan suatu potongan kata yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Dari data ulasan yang didapat menghasilkan 3.312 kata.

c).Replace Tokens

Replace Tokens berguna untuk melakukan perbaikan kata-kata yang disingkat ataupun memperbaiki kata yang mengalami salah ejaan yang memiliki maksud yang sama.

d).Filtering

Filtering proses ini berfungsi untuk melakukan penyaringan kata pada dokumen atau pengurangan kata didalam corpus yang disebut *stopwords*. Proses ini sendiri merupakan tahap mengeliminasi kata-kata yang tidak memiliki pengaruh namun sering muncul dalam dokumen. Dari 3.312 yang dihasilkan pada proses *Tokenizing* disaring pada proses *filtering* dan memperoleh kata sejumlah 278 kata. Kata-kata tersebut merupakan kata yang dirasa cukup penting dan berpengaruh.

e).Stemming

Stemming merupakan proses menghilangkan semua imbuhan yang ada pada kata tersebut menjadi kata dasar. Sebagai contoh imbuhan yaitu : meng-, me-, kan-, di-, pe-, peng-, dan lain-lain. *Stemming* bertujuan untuk mengurangi variasi kata yang memiliki kata dasar sama.

3.Tahap Klasifikasi

Setelah melakukan proses *Preprocessing* proses selanjutnya proses klasifikasi. Pada uji coba yang dilakukan akan menggunakan nilai *folds* 2 sampai 10. Tujuannya agar melihat performa terbaik yang diperoleh dari masing-masing algoritma. Algoritma yang digunakan pada penelitian ini adalah *Support Vector Machine* dan *Decision Tree*. Alasan menggunakan metode SVM karena kemampuannya yang meminimalkan error. Uji coba menggunakan 3 *karnel type* yaitu dot, radial dan polynomial. Dari uji coba *fold* 2 sampai 10 menghasilkan nilai berikut :

Tabel 2.Hasil eksperimen SVM

| Nilai K-Folds | Karnel Type | | |
|---------------|-------------|---------------|------------|
| | Dot | Radial | Polynomial |
| 2 | 78.80% | 90.00% | 89.20% |
| 3 | 73.60% | 90.20% | 88.00% |
| 4 | 70.40% | 90.00% | 83.20% |
| 5 | 75.60% | 90.00% | 83.40% |
| 6 | 69.80% | 90.01% | 85.23% |
| 7 | 69.62% | 89.60% | 82.97% |
| 8 | 67.39% | 90.00% | 88.40% |
| 9 | 71.16% | 90.01% | 88.01% |
| 10 | 68.60% | 90.00% | 86.80% |

Pada Tabel 2 diperlihatkan hasil eksperimen yang dilakukan dengan menggunakan metode SVM dengan menghasilkan nilai akurasi tertinggi pada Kaenel type = radial dan terdapat pada k-Fold 3 sebesar 90.20 %. Dari hasil tersebut menghasilkan bobot dimasing-masing kata. Pembobotan pada tiap kata yang didapat. Pada pembobotan kata tersebut memiliki nilai yang berbeda. Semakin besar jumlah kemunculan suatu kata maka semakin besar pula bobotnya atau akan membrikan nilai kesesuaian yang semakin besar. Tabel 3 berikut adalah beberapa contoh hasil pembobotan kata menggunakan metode SVM :

Tabel 3. Hasil Pembobotan Kata

| Kata | Nilai Bobot |
|-----------|-------------|
| Ahli | -2.609 |
| Ajar | -4.409 |
| Akomodasi | -2.597 |
| Akses | -3.681 |
| Aksi | -2.607 |
| Aktivitas | -3.677 |
| Akun | -3.145 |
| Akurasi | -1.855 |
| Alih | -2.612 |
| Apartemen | -2.599 |
| Arti | -2.575 |
| Arus | -2.607 |
| Asia | 22.338 |
| Asik | -2.604 |
| Asli | -5.192 |
| Bagus | 32.938 |
| Baik | -9.003 |

Metode kedua yang digunakan untuk membandingkan hasil akurasi dari algoritma SVM adalah algoritma *decision tree*. Dari uji coba yang dilakukan melalui nilai *k-fold* 2 sampai 10.

Tabel 4. Hasil Eksperimen Decision Tree

| Nilai K-Folds | Hasil |
|---------------|---------------|
| 2 | 89.60% |
| 3 | 89.30% |
| 4 | 89.40% |
| 5 | 89.80% |
| 6 | 89.60% |
| 7 | 89.40% |
| 8 | 89.21% |
| 9 | 89.15% |
| 10 | 89.25% |

Dari hasil diatas bisa dilihat hasil akurasi terbaik berada pada nilai *k-fold* 5 dengan nilai

89.80%. Nantinya hasil ini akan dibandingkan dengan metode sebelumnya yaitu metode SVM.

4. Evaluasi

Berdasarkan eksperimen pengujian yang telah dilakukan maka dihasilkan model dengan tingkat akurasi tertinggi pada setiap metode. Pengujian menggunakan operator *cross validation* dengan menggunakan nilai *k-fold* 2 sampai 10. Pada eksperimen menggunakan metode *Support Vector Machine* menghasilkan nilai 90.20% pada nilai *k-fold* 3.

| | true NEGATIVE | true POSITIF | class precision |
|---------------|---------------|--------------|-----------------|
| pred NEGATIVE | 449 | 48 | 90.34% |
| pred POSITIF | 1 | 2 | 66.67% |
| class recall | 99.78% | 4.00% | |

Gambar 3. Hasil Akurasi SVM

Sedangkan untuk hasil eksperimen terbaik pada metode kedua adalah metode *decision tree* terdapat pada nilai *k-fold* 5 dengan nilai 89.80%.

| | true NEGATIVE | true POSITIF | class precision |
|---------------|---------------|--------------|-----------------|
| pred NEGATIVE | 446 | 47 | 90.47% |
| pred POSITIF | 4 | 3 | 42.86% |
| class recall | 99.11% | 6.00% | |

Gambar 4. Hasil Akurasi *decision tree*

Perbandingan hasil eksperimen dengan nilai akurasi terbaik pengujian menggunakan nilai *k-fold* 3 untuk SVM dan *k-fold* 5 untuk *decision tree* dapat dilihat pada Tabel 5 dibawah ini:

Tabel 5. Perbandingan Algoritma

| Prediksi | SVM | | Decision Tree | |
|----------|----------------|---------|----------------|---------|
| | Negatif | Positif | Negatif | Positif |
| Negatif | 449 | 48 | 446 | 47 |
| Positif | 1 | 2 | 4 | 3 |
| | Akurasi 90.20% | | Akurasi 89.80% | |

Berdasarkan Tabel 5 menggambarkan seberapa besar presentasi data yang diprediksi oleh metode SVM dan akan dibandingkan dengan metode *Decision Tree*. *True Negatif* merupakan banyaknya data observasi dengan kategori negative yang mampu diprediksi negatif atau diprediksi dengan tepat oleh metode tersebut. *True Positif* yaitu banyaknya data observasi dengan kategori positif yang mampu diprediksi positif atau diprediksi dengan tepat oleh metode tersebut. *False negatif* merupakan banyaknya data observasi yang berkategori negatif namun terdapat kesalahan prediksi. Sedangkan *False positif* merupakan banyaknya data observasi yang berkategori positif namun terdapat kesalahan prediksi.

Pada metode SVM diperoleh hasil bahwa pada kelas positif terdapat 50 ulasan positif yang diuji, terdapat 48 ulasan yang telah diklasifikasikan dengan benar dan terdapat kesalahan prediksi sejumlah 2 ulasan yang masuk ulasan negatif. Kemudian pada ulasan negatif dengan total 450 ulasan yang diuji terdapat 449 ulasan yang diklasifikasikan dengan benar sebagai ulasan negatif dan 1 yang masuk ulasan

positif. Selanjutnya dari nilai tersebut diperoleh akurasi sebesar 90.20%, yang berarti dari 500 data yang diujikan terdapat 497 ulasan yang benar dalam pengklasifikasiannya oleh model SVM.

Kemudian pada Metode *decision tree* diperoleh prediksi bahwa pada kelas positif terdapat 50 ulasan positif yang diuji, terdapat 47 ulasan yang telah diklasifikasikan dengan benar dan kesalahan prediksi sejumlah 3 ulasan yang masuk kedalam ulasan negatif. Kemudian pada ulasan negative dengan total 450 ulasan yang diuji terdapat 446 ulasan yang diklasifikasikan dengan benar sebagai ulasan negative dan 4 yang masuk ulasan positif. Selanjutnya dari nilai *confusion matrix* tersebut diperoleh akurasi sebesar 89.80%, yang berarti dari 500 data yang diujikan terdapat 493 ulasan yang benar dalam pengklasifikasiannya oleh model *Decision Tree*.

Adapun proses perhitungan nilai akurasi dilakukan dengan menggunakan rumus untuk perhitungan nilai akurasi pada metode SVM sebagai berikut :

$$\text{Akurasi} = \frac{TN+TP}{TN+FN+FP+TP}$$

$$\text{Akurasi} = \frac{449+2}{449+1+48+2}$$

$$\text{Akurasi} = \frac{451}{500} = 0,902 = 90.20\%$$

Sedangkan untuk perhitungan nilai akurasi pada metode *decision tree* sebagai berikut :

$$\text{Akurasi} = \frac{TN+TP}{TN+FN+FP+TP}$$

$$\text{Akurasi} = \frac{446+3}{446+4+47+3}$$

$$\text{Akurasi} = \frac{449}{500} = 0.898 = 89.80\%$$

5. Penarikan Kesimpulan

Dari hasil eksperimen didapatkan hasil akurasi terbaik yaitu untuk metode *Support Vector Machine* sebesar 90.20%. Dengan hasil pada kelas positif terdapat 50 ulasan positif yang diuji, terdapat 48 ulasan yang telah diklasifikasikan dengan benar dan terdapat kesalahan prediksi sejumlah 2 ulasan yang masuk ulasan negatif. Kemudian pada ulasan negatif dengan total 450 ulasan yang diuji terdapat 449 ulasan yang diklasifikasikan dengan benar sebagai ulasan negatif dan 1 yang masuk ulasan positif. Sedangkan pada metode *Decision Tree* mendapatkan nilai akurasi 89.80 dengan memperoleh prediksi bahwa pada kelas positif terdapat 50 ulasan positif yang diuji, terdapat 47 ulasan yang telah diklasifikasikan dengan benar dan kesalahan prediksi sejumlah 3 ulasan yang masuk kedalam ulasan negatif. Kemudian pada ulasan negative dengan total 450 ulasan yang diuji terdapat 446 ulasan yang diklasifikasikan dengan benar sebagai ulasan negative dan 4 yang masuk ulasan positif.

Jadi dapat disimpulkan bahwa metode *Support Vector Machine* untuk nilai akurasinya

lebih tinggi dibandingkan metode *Decision Tree* dengan nilai akurasi yang didapat yaitu 90.20%. Nilai akurasi ini didapatkan di nilai *k-fold 3* dengan *karnel type radial*.

Kesimpulan dan Saran

Berdasarkan penelitian yang sudah peneliti lakukan dengan membandingkan metode SVM dan metode *Decision Tree* terhadap data ulasan komentar pengguna aplikasi Gojek pada *Google Play* dengan tujuan utama dari penelitian ini yaitu untuk mengetahui tingkat akurasi yang dihasilkan. Dalam proses penelitian ini menggunakan RapidMiner 9.6. Dari pengolahan data tersebut diperoleh hasil berupa hasil dari perbandingan metode *Support Vector Machine* dan *Decision Tree*. Metode *Support Vector Machine* nilai akurasinya lebih tinggi dengan nilai akurasi 90.20% yang terdapat pada nilai *k-fold 3 karnel type radial*.

Analisis sentiment terhadap ulasan pada *Google Play* dengan menggunakan metode *Support Vector Machine* dan *Decision Tree* pada penelitian ini masih memiliki banyak kekurangan. Berikut beberapa saran yang dapat digunakan sebagai proses penelitian-penelitian selanjutnya yang akan mengambil penelitian hamper serupa dengan penelitian ini:

- a) Dapat melakukan penelitian dengan menggunakan gabungan algoritma klasifikasi seperti dengan algoritma asosiasi atau algoritma clustering.
- b) Dapat melakukan penelitian dua atau lebih dataset ulasan pada aplikasi transportasi online seperti Grab, Nujek, dan lain sebagainya.

Daftar Pustaka

- [1] Fanani. F. "*Klasifikasi Review Software Pada Google Play Menggunakan Pendekatan Analisis Sentimen*", Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta, 2017
- [2] Wijaya, A., "*Aspek hukum bisnis transportasi jalan online*", Jakarta : Gramedia, 2016
- [3] Yola, Melfa, and Budianto. D, "*Analisis kepuasan konsumen terhadap kualitas pelayanan dan harga produk pada supermarket dengan menggunakan metode Importance Performance Analysis (IPA)*." *Jurnal Optimasi Sistem Industri* 12(12). (2013): 301-309, hlm 302, 2013
- [4] Feldman. R., & Sanger J., "*The Text Mining Handbook Advance Approaches In Analyzing Unstructured Data*". New York: Cambridge University Press, 2007
- [5] Jamil. H. N., "*Analisis Sentimen Pada Online Review Menggunakan Kombinasi Metode Lexicon Based and Naïve Bayes Classifier*", Program Studi Statistika FMIPA UII Yogyakarta, 2017
- [6] Han. J., Kamber. M., & Pei. J., "*Data Mining: Concepts and Techniques Third edition*", Waltham: Elsevier, 2011
- [7] Suyanto, "*Data Mining Untuk Klasifikasi dan Klasterisasi Data*", Bandung : Informatika Bandung, 2017

- [8] Novantirani. A., Sabariah. M. K., Effendy. V., "*Analisis Sentimen Pada Twitter untuk Mengenai Penggunaan Transportasi Umum Darat Dalam Kota dengan Metode Support Vector Machine (SVM)*". Program Studi Teknik Informatika, Fakultas Informatika, Universitas Telkom, Bandung, 2015
- [9] Praptiwi. D. Y., "*Analisis Sentimen Online Review Pengguna E-Commerce Menggunakan Metode Support Vector Machine Dan Maximum Entropy*". Program Studi Statistika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Islam Indonesia, 2018