

METHODE K-NN FOR ANALYS SENTIMENT REVIEW KIDS APPS

Sucitra Sahara¹, Rizqi Agung Permana²

¹Program Studi Sistem Informasi, Universitas Bina Sarana Informatika
Cengkareng, Jakarta Barat, Indonesia

²Program Studi Teknik Informasi, STMIK Antar Bangsa
Tangerang, Indonesia

e-mail: sucitrasahara@gmail.com, rizqiangung@gmail.com

Abstrak

Aplikasi game berbasis android ataupun ios kini sudah merambah ke semua pengguna smartphone mulai dari dewasa remaja sampai anak-anak, maka dari itu para vendor aplikasi maupun pihak pebisnis berlomba menciptakan aplikasi guna meraup keuntungan, mulai kualitas dan performa tinggi sampai kualitas yang masih sering diragukan khususnya pada pengguna aplikasi game anak dibawah umur yang membuat orang tua khawatir dengan yang di konsumsi anaknya, sehingga peneliti melakukan penyeleksian melalui komentar terhadap aplikasi game kategori anak. Berdasarkan pendapat atau komentar masyarakat yang sudah menggunakan aplikasi tersebut pada kolom komentar di website penyelenggara aplikasi. Dari sampel komentar yang telah direview oleh peneliti dan memperoleh data set berupa teks positif dan negatif yang diolah pada pengklasifikasian data dengan menggunakan metode k-Nearest Neighbors (k-NN), dimana metode tersebut salah satu metode algoritma yang paling terkenal dan sering digunakan pada pengenalan pola. Banyak peneliti menyatakan bahwa algoritma k-NN dapat menyelesaikan kinerja data yang sangat baik pada data set yang berbeda terutama pada penyeleksian teks, untuk meningkatkan kinerja klasifikasinya. Pada data set yang telah diuji nilai akurasi yang didapat 78.50% yang diartikan masih dalam batas yang baik pada penggunaan metode k-NN. Pada metode k-NN sesuai dengan konsep text mining yaitu bertujuan untuk mencari pola pada teks, yang dilakukan pada text mining kali ini *text classification*.

Kata Kunci: Review Komentar, Kids Apps, k-Nearest Neighbors, Penyeleksian Teks

Abstract

Now applications based on Android or iOS have penetrated to all smartphone users from teenage parents to children, application vendors and business people are competing to create applications to reap profits, from high quality and high performance to quality that is still often in doubt, especially in children under age that makes parents concerned with the consumption of their children, so researchers conducted research in the selection of applications for children. Based on the opinions or comments of people who have used the application that they write into online media. From the sample of comments that have been reviewed by researchers and obtained data sets in the form of positive and negative texts that are processed in the classification of data using the k-Nearest Neighbors (k-NN) method, where the method is one of the most well-known algorithmic methods and is often used in the introduction pattern. Many researchers claim that the k-NN algorithm can solve very good data performance on different data sets, especially on text selection, to improve the performance of classification. In the data set that has been tested the accuracy value obtained is 78.50% which means that it is still within good limits on the use of the k-NN method. In the k-NN method in accordance with the concept of text mining which aims to find patterns in text, which is done in text mining this time is text classification.

Keywords: Review Comments, Kids Apps, k-Nearest Neighbors, Text Classification

PENDAHULUAN

A. Latar Belakang

Orangtua selalu memberikan yang terbaik untuk anaknya, mulai dari kebutuhan makanan yang sehat, sandang yang layak, pendidikan, bahkan mainan yang terbaik. Bimbingan dan pemantauan orangtua adalah salah satu faktor utama, semakin tumbuh besarnya anak semakin luas pergaulan dan kesibukan anak menjadi tidak terpantau sepenuhnya oleh orangtua, hal ini menjadi kekhawatiran pada gadget yang digunakan oleh anak, seleksi isi konten aplikasi yang di install pada smartphone.

Pesatnya perkembangan internet pada komputasi memungkinkan kita untuk mampu menganalisa sejumlah besar data dan memprediksi minat pelanggan terhadap suatu bentuk produk pada masa depan. Hal ini menjadikan kecenderungan emosional pelanggan dan produk favorit melalui komentar teks online sangat penting untuk dipelajari (Zhang, dkk).

Trend analisis menunjukkan peningkatan penargetan platform Android sebagai sasaran malware sehingga Aplikasi android sangat perlu untuk diprediksi dengan akurat sebelum konsumen melakukan instalasi (Talla, Alpher dan Aydin, 2015).

Beberapa penelitian yang sudah dilakukan dalam klasifikasi sentimen terhadap review yang tersedia secara online diantaranya, analisa sentiment aplikasi smartphone dengan membandingkan metode *Support Vector Machine* (SVM) dan *Naïve Bayes* (Zhang dkk). Kategorisasi teks merupakan solusi yang tepat untuk mengelola informasi yang saat ini berkembang dengan sangat cepat dan melimpah. Kategorisasi teks membuat pengelolaan informasi tersebut menjadi efektif dan efisien. Dengan menggunakan kategorisasi teks, dapat dilakukan penyusunan dokumen menurut kategorinya, penyaringan terhadap email *spam*, melakukan penggalian opini (*opinion mining*) dan analisis sentimen. Algoritma kategorisasi teks saat ini telah banyak berkembang, antara lain: *Support Vector Machines* (SVM), *Naive Bayesian* (NB),

pohon keputusan, *K-Nearest Neighbour* (k-NN). Metode k-Nearest Neighbors (k-NN) adalah salah satu metode nonparametrik yang paling populer diperkenalkan oleh Fix dan Hodges pada tahun 1951 (Tan, 2006). Setiap partikel menyampaikan informasi berupa posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi yang terbaik tersebut (Shuzhou & Bo, 2011). Pada penelitian kali ini, menggunakan pengklasifikasian k-Nearest Neighbors (k-NN) pemilihan fitur akan diterapkan untuk klasifikasi *text* pada pendapat atau pemikiran publik mengenai produk *Kids Apps*.

B. Landasan Teori

1. Internet

Hasil Survei Penetrasi dan Perilaku Pengguna Internet Indonesia 2018 menurut survei yang diadakan oleh APJI (Asosiasi Penyelenggara Internet Indonesia) setiap tahun menunjukkan pengguna internet di indonesia mencapai angka 27.916.716 tahun 2017-2018.



Sumber: Survei APJII (2018)

Gambar II.1. Pengguna Internet Indonesia

2. Text Mining

Adalah proses ekstraksi pola (informasi dan pengetahuan yang berguna) dari sejumlah besar data tak terstruktur. Masukkan untuk pengembangan teks adalah data yang tidak (atau kurang) terstruktur, seperti dokumen Word, PDF, Jurnal Sains dan Teknologi | 128

kutipan teks, sedangkan masukkan untuk pengembangan data adalah data yang terstruktur (Ronen Feldman, 2007).

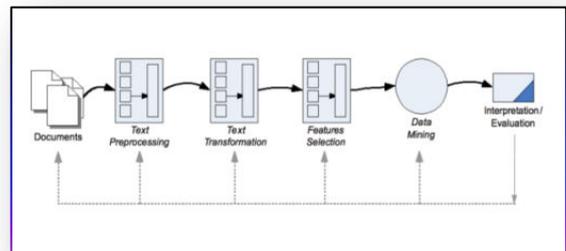
Menurut Moraes (Moraes et al., 2013) langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen adalah:

- a. **Definisikan domain dataset**
Pengumpulan dataset yang melingkupi suatu domain, misalnya dataset review film, dataset review produk, dan lain sebagainya.
- b. **Pre-processing**
Tahap pemrosesan awal yang umumnya dilakukan dengan proses *Tokenization*, *Stopwords removal*, dan *Stemming*.
- c. **Transformation**
Proses representasi angka yang dihitung dari data tekstual. *Binary representation* yang umumnya digunakan dan hanya menghitung kehadiran atau ketidakhadiran sebuah kata di dalam dokumen. Berapa kali sebuah kata muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual. Proses yang umumnya digunakan yaitu TF-IDF, *Binary transformation*, dan *Frequency transformation*.
- d. **Feature Selection**
Pemilihan fitur (*feature selection*) bisa membuat pengklasifikasi lebih efisien/efektif dengan mengurangi jumlah data untuk dianalisa dengan mengidentifikasi fitur yang relevan yang selanjutnya akan diproses. Metode pemilihan fitur yang biasanya digunakan adalah Expert. Knowledge, Minimum Frequency, Information gain, Chi-Square, dan lain sebagainya.
- e. **Classification**
Proses klasifikasi umumnya menggunakan pengklasifikasi seperti

Naïve Bayes, Support Vector Machine, dan lain sebagainya.

f. **Interpretation/Evaluation**

Tahap evaluasi biasanya menghitung akurasi, recall, precision, dan F-1.



Sumber: Ronen Feldman(2007)

Gambar II.2 Tahapan Proses Text Mining

3. Review Produk

Data yang digunakan dalam kebanyakan studi klasifikasi sentimen dikumpulkan dari situs e-commerce seperti www.amazon.com (review produk), www.yelp.com (ulasan restoran), www.CNETdownload.com (review produk) dan www.reviewcentre.com, yang menjadi tuan rumah jutaan ulasan produk oleh konsumen. Selain itu, situs yang tersedia adalah situs review profesional seperti www.dpreview.com, www.zdnet.com dan situs pendapat konsumen tentang topik yang luas dan produk-produk seperti www.consumerreview.com, www.epinions.com, www.bizrate.com (Popescu&Etzioni, 2005; Hu, B.Liu , 2006; Qinliang Mia, 2009; Gambaran Sompraserti, 2010).

4. Android

Android adalah sistem operasi yang biasa disematkan pada gadget, baik itu handphone atau tablet. Jangan heran jika saat ini Android, baik itu tablet atau handphone ini bisa menggantikan peran

dari sebuah komputer jinjing, apabila digunakan untuk kebutuhan entertainment, seperti mendengarkan lagu, menonton video, mengirim email, bermain game, twitter, atau facebook, juga kegiatan hiburan online lainnya. Bedanya dengan tablet atau handphone lebih ringkas dan lebih praktis, karena modelnya memang didisain seperti itu, seperti buku yang bisa ditenteng kemana – mana (Wahadyo, Agus, Sudarma,2012).

Seperti teknologi lainnya, Android muncul tidak langsung canggih seperti saat ini. Teknologinya yang bersifat open source, terus berkembang dan selalu terbuka untuk digunakan dan dikembangkan oleh siapa saja. Mungkin ini yang membuat Android begitu diminati. Android versi beta muncul pertama kali pada November 2007, Android benar-benar dipasarkan dengan versi Android 1.0 pada September 2007 dengan kode nama Apple pie. Android versi ini disematkan pada sebuah handphone HTC Dream. Versi-versi Android hingga tahun 2014, dapat kita lihat seperti: Versi rilis prakomersial (2007–2008): Android alpha, Android beta. Sejarah versi Android menurut level [Antarmuka Pemrograman Aplikasi](#) (API): Android 1.0 (API level 1), Android 1.1 (API level 2), Android 1.5 Cupcake (API level 3), Android 1.6 Donut (API level 4), Android 2.0 Eclair (API level 5), Android 2.0.1 Eclair (API level 6), Android 2.1 Eclair (API level 7), Android 2.2–2.2.3 Froyo (API level 8), Android 2.3–2.3.2 Gingerbread (API level 9), Android 2.3.3–2.3.7 Gingerbread (API level 10), Android 3.0 Honeycomb (API level 11), Android 3.1 Honeycomb (API level 12), Android 3.2 Honeycomb (API level 13), Android 4.0–4.0.2 Ice Cream Sandwich (API level 14), Android 4.0.3–4.0.4 Ice Cream Sandwich (API level 15), Android 4.1 Jelly Bean (API level 16), Android 4.2 Jelly Bean (API level 17), Android 4.3 Jelly Bean (API level 18), Android 4.4 KitKat (API level 19), Android 5.0 Lollipop (API level

21), Android 6.0 Milkshake (API level 24). Distribusi versi Android secara global sejak Desember 2009 Hingga Januari 2015, Android 4.x Jelly Bean adalah versi Android yang paling populer, digunakan oleh sekitar 59% perangkat Android di seluruh dunia.

5. Analisa Sentiment

Analisis Sentimen adalah jenis pengolahan bahasa alami untuk melacak mood masyarakat tentang produk tertentu atau topik. Analisis sentimen, yang juga disebut tambang pendapat, melibatkan dalam membangun sistem untuk mengumpulkan dan meneliti pendapat tentang produk yang dibuat dalam posting blog, komentar, ulasan atau tweet. Analisis Sentimen dapat berguna dalam beberapa cara. Misalnya, dalam pemasaran membantu injudging keberhasilan kampanye iklan atau peluncuran produk baru, menentukan versi produk atau jasa yang populer dan bahkan mengidentifikasi demografi suka atau tidak suka terhadap fitur tertentu (Vinodhini, Chandrasekaran, 2012).

6. Pre-processing

Proses pengubahan bentuk bentuk menjadi data yang terstruktur sesuai kebutuhannya untuk proses dalam data mining, yang biasanya akan menjadi nilai-nilai numerik, proses ini sering disebut dengan *text processing* (Ronen Feldman,2007). Setelah data menjadi data terstruktur dan berupa nilai numerik maka data dapat disajikan sebagai sumber data yang dapat diolah lebih lanjut.

Beberapa proses yang dilakukan adalah sebagai berikut:

a. Tokenizazion

Peneliti menggunakan Tokenize untuk memisahkan kata atau huruf dari tanda baca dan simbol.

b. Stopwards Removal

kata yang dianggap tidak perlu dalam pengolahan data sentimen review, sebagai contoh *if, the, of, or, etc.*

c. Steaming

Proses perubahan bentuk kata menjadi kata dasar. Metode perubahan bentuk kata menjadi kata dasar ini menyesuaikan struktur bahasa yang digunakan dalam proses stemming (Langgeni et al., 2010).

7. Algoritma k-Nearest Neighbors(k-NN)

Algoritma k-Nearest Neighbors (k-NN) adalah penentu klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, tetapi bergantung pada label kategori yang melekat pada dokumen pelatihan mirip dengan dokumen tes. Mengingat tes dokumen, sistem menemukan k tetangga terdekat antara dokumen pelatihan. Rata-kesamaan setiap dokumen tetangga terdekat dokumen uji digunakan sebagai berat kelas dokumen tetangga (Songho tan, 2008).

Ketika *k-values* yang ditetapkan terlalu kecil, maka akan menghasilkan akurasi yang rendah, dikarenakan hasil kategori akan lebih terpengaruh dengan *noise* (Chairina., Rizal., dan Agung. 2008).

Metode k-Nearest Neighbors (k-NN) adalah salah satu metode nonparametrik yang paling populer diperkenalkan oleh Fix dan Hodges pada tahun 1951 (Tan, 2006). Karena hanya ada satu parameter K (jumlah nearest neighbors) yang perlu ditentukan, mudah untuk menerapkan metode k-NN.

8. Validasi dan Evaluasi Text Mining

Dari beberapa metode yang digunakan untuk memvalidasi suatu model berdasarkan data yang didapat, seperti holdout, cross validation, random sub-sampling, dan lain-lain. *Confusion matrix* memberikan keputusan yang diperoleh dalam training dan testing, *Confusion matrix* memberikan penilaian performance klasifikasi berdasarkan objek dengan benar atau salah (Gorunescu, 2011). *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi.

Tabel II.1 Tabel Confution Matrix

Classification	Predicted Class		
	Class = Yes	Class = No	
Observed Class	Class = Yes	<i>a</i> <i>true positive -</i> (TP)	<i>b</i> <i>(false negative -</i> (FN)
	Class = No	<i>c</i> <i>(false positive -</i> (FP)	<i>d</i> <i>(true negative -</i> (TN)

Sumber: Gorunescu (2011)

Keterangan:

True Positive (TP) = Proporsi positif dalam data set yang diklasifikasikan positif.

True Negative (TN) = Proporsi negative dalam data set yang diklasifikasikan negative.

False Positive (FP) = Proporsi negatif dalam data set yang diklasifikasikan positif.

False Negative (FN) = Proporsi negative dalam data set yang diklasifikasikan negative.

Berikut adalah persamaan model Confusion Matrix:

- a. Nilai *Accuracy* adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- b. *Sensitivity* digunakan untuk membandingkan proporsi TP terhadap tupel yang positif, yang dihitung dengan menggunakan persamaan:

$$Sensitivity = \frac{TP}{TP + FN}$$

- c. *Specificity* digunakan untuk membandingkan proporsi TN terhadap tupel yang negatif, yang dihitung dengan menggunakan persamaan:

$$Specificity = \frac{TN}{TN + FP}$$

- d. PPV (*positive predictive value*) adalah proporsi kasus dengan hasil diagnosa positif, yang dihitung dengan menggunakan persamaan:

$$PPV = \frac{TP}{TP + FP}$$

- e. NPV (*negative predictive value*) adalah proporsi kasus dengan hasil diagnosa negatif, yang dihitung dengan menggunakan persamaan:

$$NPV = \frac{TN}{TN + FN}$$

C. Masalah

Pada penelitian kali ini penulis melihat kekhawatiran orangtua pada aplikasi game kategori anak-anak yang ada pada smartphone, orangtua selalu memberikan yang terbaik untuk anaknya, mulai dari kebutuhan makanan, sandang yang layak,

pendidikan, bahkan sampai pemilihan aplikasi game yang baik dan beredukasi. Bimbingan dan pemantauan orangtua pada saat anak menggunakan smartphone adalah suatu keharusan supaya anak terhindar dari hal negatif yang disebabkan dari penggunaan aplikasi game anak pada smartphone berbasis android maupun ios. Dan orang tua masih bingung dalam memilih aplikasi game anak yang baik atau buruk bagi anaknya.

D. Rencana Penyelesaian Masalah

Peneliti melalui seleksi review pengguna aplikasi game untuk anak yang diinstall pada smartphone, dengan memisahkan komentar positif dan negatif. Dan mencari model data yang sesuai dengan pengolahan data set dari review tersebut. Metode k-NN menghasilkan model dimana peneliti menentukan nilai k untuk mencari tingkat keakurasian yang tinggi pada pengujian dataset. Nilai akurasi yang paling tinggi akan digunakan peneliti dalam menentukan apakah nilai tersebut optimal atau tidak, dan ternyata nilai tersebut sudah cukup baik. Maka disimpulkan bahwa struktur algoritma k-NN yang dirancang mencapai ideal dalam pemecahan masalah pada penelitian ini.

METODE PENELITIAN

A. Pengumpulan Data

Data yang digunakan dalam kebanyakan studi klasifikasi sentimen dikumpulkan dari situs e-commerce seperti www.amazon.com (review produk), www.yelp.com (ulasan restoran), www.CNETdownload.com (review produk) dan www.reviewcentre.com, yang menjadi tuan rumah jutaan ulasan produk oleh konsumen. Selain itu, situs yang tersedia adalah situs review profesional seperti www.dpreview.com, www.zdnet.com dan situs pendapat konsumen tentang topik

yang luas dan produk-produk seperti www.consumerreview.com, www.epinions.com, www.bizrate.com (Popescu& Etzioni ,2005 ; Hu,B.Liu ,2006; Qinliang Mia, 2009; Gambaran Sompraserti ,2010).

a. Definiskan domain dataset

Pengumpulan dataset yang melingkupi suatu domain, misalnya dataset review film, dataset review produk, dan lain sebagainya.

b. *Pre-processing*

Tahap pemrosesan awal yang umumnya dilakukan dengan proses *Tokenization*, *Stopwords removal*, dan *Stemming*.

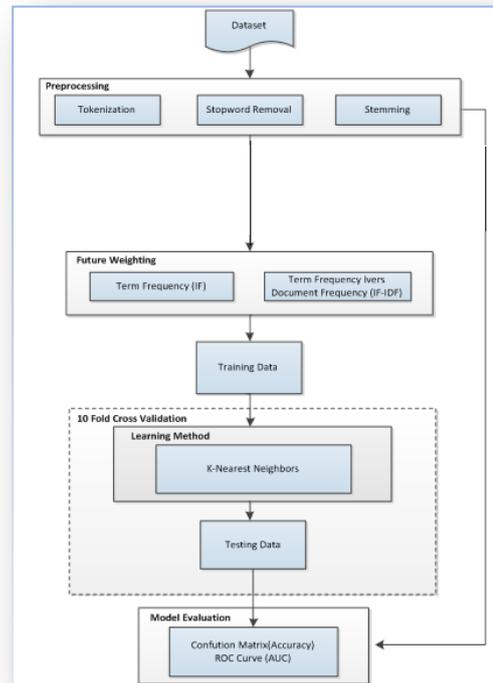
c. *Transformation*

Proses representasi angka yang dihitung dari data tekstual. *Binary representation* yang umumnya digunakan dan hanya menghitung kehadiran atau ketidakhadiran sebuah kata di dalam dokumen. Berapa kali sebuah kata muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual. Proses yang umumnya digunakan yaitu TF-IDF, *Binary transformation*, dan *Frequency transformation*.

d. *Feature Selection*

Pemilihan fitur (*feature selection*) bisa membuat pengklasifikasi lebih efisien/efektif dengan mengurangi jumlah data untuk dianalisa dengan mengidentifikasi fitur yang relevan yang selanjutnya akan diproses. Metode pemilihan fitur yang biasanya digunakan adalah Expert. Knowledge, Minimum Frequency, Information gain, Chi-Square, dan lain sebagainya.

Berikut ini gambaran metode penelitian yang peneliti susun:



Sumber: Peneliti (2018)

Gambar II.3 Metode Penelitian

1. Evaluasi dan Hasil Validasi

Pada evaluasi kali ini penulis mengusulkan penggunaan model dalam kegiatan review aplikasi pada android yaitu model k-Nearest Neighbors (k-NN), yang dilakukan dalam dua tahap penerapan. Algoritma k-NN menghasilkan model dimana peneliti menentukan nilai k untuk mencari tingkat keakurasian yang tinggi pada pengujian dataset. Nilai akurasi yang paling tinggi akan digunakan peneliti dalam menentukan apakah nilai tersebut optimal atau tidak, dan ternyata nilai tersebut sudah cukup baik. Maka disimpulkan bahwa struktur

algoritma yang dirancang mencapai ideal dalam pemecahan masalah.

HASIL DAN PEMBAHASAN PENELITIAN

A. Klasifikasi Text Menggunakan Algoritma k-Nearest Neighbors

Data training yang digunakan dalam pengklasifikasian *text* terdiri atas 100 review positif pada Appstore for Android dan 100 review negatif pada Appstore for ios dan Playstore untuk Android. Data review yang akan diolah masih berupa sekumpulan text yang dipisah dalam bentuk dokumen. Sebelum diklasifikasikan, data tersebut harus melewati proses tahapan agar data dapat diolah dengan baik.

Pada hasil penelitian kali ini, peneliti menjabarkan hasil dari pengujian algoritma dan optimasinya yang menghasilkan nilai akurasi dan AUC.

Data training yang digunakan dalam pengklasifikasian *text* terdiri atas 100 review positif pada Kids apps dan 100 review negatif pada Game Appstore for Android. Data review yang akan diolah masih berupa sekumpulan text yang dipisah dalam bentuk dokumen. Sebelum diklasifikasikan, data tersebut harus melewati proses tahapan agar data dapat diolah dengan baik.

Berikut ini pengelompokan dokumen berdasarkan pengklasifikasian:

Tabel IV.4. Tabel Vector Dokumen Boolean dengan Label Class Hasil Klasifikasi

Dokumen	Great	Good	Annoy	Bad	Class
Positif1.txt	1	1	0	0	Positif
Positif8.txt	1	0	0	0	Positif
Positif12.txt	0	1	0	0	Positif
Positif16.txt	1	1	0	0	Positif
Positif96.txt	1	0	0	0	Positif
Negatif20.txt	0	0	1	0	Negatif
Negatif8.txt	0	0	0	1	Negatif
Negatif12.txt	0	0	0	1	Negatif
Negatif78.txt	0	0	0	1	Negatif
Negatif90.txt	0	0	0	1	Negatif

Sumber: Hasil Penelitian (2018)

B. Penerapan Metode k-NN

Hasil Eksperimen Pengujian Metode k-NN, Hasil Nilai query instance dalam penelitian disini ditentukan dengan cara melakukan uji coba memasukkan nilai k (jumlah tetangga terdekat).

Tabel 3. Eksperimen Metode k-NN

Nilai k	Accuracy	AUC
1	64.50%	0.500
2	64.50%	0.696
3	65.00%	0.734
4	68.00%	0.742
5	70.00%	0.764
6	72.50%	0.790
7	74.00%	0.797
8	74.50%	0.808
9	70.00%	0.816
10	74.50%	0.895

Sumber: Hasil Penelitian (2018)

Hasil menunjukkan pada penerapan metode k-Nearest Neighbors pada tabel IV.5 dengan penentuan nilai k=10 yang nilai akurasi mencapai 74.50% dan AUC 0.895 menunjukkan hasil yang paling tertinggi diantara penentuan nilai k yang lain. Hasil ROC:



Sumber: Hasil Penelitian (2018)

Gambar II.4. Hasil ROC Pengujian k-NN

C. Hasil Analisis Evaluasi Hasil dan Validasi Model

Dari hasil pengujian yang peneliti lakukan dari awal pembahasan, pengukuran akurasi menggunakan confusion matrix dan kurva ROC membuktikan bahwa hasil pengujian algoritma k-Nearest Neighbors (k-NN) cukup tinggi, Nilai akurasi untuk model algoritma k-NN sebesar 78.50% pada k=10, dijabarkan pada tabel 4.

Tabel 4. Pengujian Algoritma k-NN

	Nilai k	Accuracy	AUC
k-NN	10	78.50%	0.825

Sumber: Hasil Penelitian (2018)

D. Confusion Metrix

Data training yang digunakan kami ambil **100** data review positif mengenai aplikasi android dan **100** data review negatif pada aplikasi android. Data review negatif, setelah melalui beberapa tahap pengolahan model k-NN pada, diklasifikasikan untuk review positif yang sesuai prediksi sebanyak **85** data, kemudian **15** data yang diprediksi positif namun masuk kedalam kategori review negatif. Sedangkan untuk data review negatif, yang sesuai dengan prediksi bahwa data tersebut negatif adalah **71**, dan untuk prediksi review negatif yang masuk dalam prediksi review positif adalah **29** data. Hasil akurasi yang muncul adalah **78.50%**.

Accuracy : 78.50% +/- 8.50% (mikro: 78.50%)

	True Positif	True Negatif	Class Precision
Prediksi Positif	85	29	70.25%
Prediksi Negatif	15	71	81.01%
Class Recall	85.00%	64.00%	

Tabel 5. Model Confusion Matrix Model k-NN

Sumber: Hasil Penelitian (2018)

Berdasarkan tabel 5 menunjukkan tingkat akurasi yang cukup baik (78.50%), dan dilihat dari hasil yang muncul pada hasil tabel pengujian bisa dicari nilai accuracy, dijabarkan pada perhitungan berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} =$$

$$\frac{85+71}{85+71+15+29} = \frac{156}{200} = 0.78$$

Keterangan :

TP: True Positif

TN: True Negatif

FP: False Positif

FN: False Negatif

Kesimpulan pengujian adalah algoritma k-NN dapat meningkatkan nilai akurasi yang merupakan solusi yang baik dalam permasalahan pada klasifikasi sentimen review kids apps.

PEMBAHASAN

Penerapan algoritma k-NN menghasilkan nilai akurasi pada klasifikasi review kids apps untuk mengidentifikasi antara review komentar positif dan review komentar negatif, dengan memiliki model klasifikasi teks pada review, pembaca dapat dengan mudah mengidentifikasi mana review yang positif maupun yang negatif. Dari data review yang sudah ada, dipisahkan menjadi kata-kata, lalu diberikan bobot pada masing-masing kata tersebut. Dapat dilihat kata mana saja yang berhubungan dengan sentimen yang sering muncul dan mempunyai bobot paling tinggi. Dengan demikian dapat diketahui review tersebut termasuk review positif atau review negatif.

Dalam penelitian ini, hasil pengujian model akan dibahas melalui *confusion matrix* untuk menunjukkan model yang terbaik. Tanpa menggunakan metode pemilihan fitur, k-Nearest Neighbors sendiri sudah menghasilkan akurasi yang cukup tinggi sebesar 78.50% dan nilai AUC 0.875.

Peneliti menyediakan aplikasi berbasis web untuk menguji model menggunakan dataset yang berbeda dan belum diklasifikasikan sesuai dengan kelasnya. Diaplikasikan dengan menggunakan bahasa pemrograman PHP berbasis Web.

E. Implikasi Penelitian

Implikasi penelitian ini mencakup beberapa aspek, di antaranya:

1. Implikasi terhadap aspek sistem Hasil evaluasi menunjukkan penerapan Algoritma k-Nearest Neighbors (k-NN) merupakan metode yang cukup baik dalam mengklasifikasi teks review Kids apps. Metode ini dapat membantu para calon pengguna android dan iOS dalam menentukan aplikasi apa saja yang layak mereka install, supaya tidak lagi asal menginstall aplikasi yang banyak tersedia diberberapa media online, yang dampak buruknya bisa mengakibatkan memory dan kualitas mobile pengguna melemah kualitasnya, dan mengurangi waktu dalam membaca review dan komentar mengenai aplikasi pada android.
2. Implikasi Aspek Manajerial Membantu para pengembang dan vendor sistem yang berkaitan dengan dunia aplikasi android, baik dari sumber sosial media atau dari situs resmi para pengusaha dibidang aplikasi android, agar menggunakan aplikasi RapidMiner dalam membangun suatu sistem.
3. Implikasi terhadap aspek penelitian lanjutan Penelitian selanjutnya bisa menggunakan metode pemilihan fitur ataupun dataset dari domain yang berbeda, seperti review hotel, review restoran, dan banyak lainnya yang bisa dicari dalam bidang pengembangannya.

KESIMPULAN

Klasifikasi text dengan data berupa review kids apps, salah satu pengklasifikasian yang dapat digunakan adalah k-Nearest Neighbors (k-NN). Hal ini dikarenakan k-NN merupakan metode yang dapat sesuai dengan klasifikasi data dan mudah dipahami. Dari data review yang sudah ada, dipisahkan menjadi kata-kata, lalu diberikan bobot pada masing-masing kata tersebut. Dapat dilihat kata mana saja yang berhubungan dengan sentimen yang sering muncul dan mempunyai bobot paling tinggi. Dengan demikian dapat diketahui review tersebut termasuk review positif atau review negatif. k-NN juga sering digunakan pada beberapa peneliti dalam klasifikasi teks dan memiliki performa yang baik. Dari pengolahan data yang sudah dilakukan. Data review yang peneliti olah dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi k-NN pada data review pruduk pada kids apps mencapai 78.50% dan nilai AUC sebesar 0.875.

UCAPAN TERIMA KASIH

Alhamdulillah puji syukur kepada Allah swt, karena kehendak dan ridhaNya peneliti dapat menyelesaikan penelitian ini. Peneliti sadari paper ini tidak akan selesai tanpa doa, dukungan dan dorongan dari berbagai pihak. Adapun dalam kesempatan ini peneliti ingin mengucapkan banyak terima kasih kepada penerbit jurnal Jurnal Sains dan Teknologi Ganesha.

DAFTAR PUSTAKA

Feldman, Ronen and Sanger, James. 2007. *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, New York.

Gorunescu. 2011. *Data Mining Concepts, Models and Techniques*. Romania: Springer-Verlag Berlin Heidelberg

Popescu, A. M., Etzioni, O.: Extracting Product Features and Opinions from Reviews, In Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, 2005, 339–346.

Songbo Tan, Jin Zhang, “An empirical study of sentiment analysis for chinese documents”, *Expert Systems with Applications* 34 (2008) 2622–2629.

Vinodhini.G,Chandrasekaran.RM.2012. *Sentiment Analysis and Opinion Mining: A Survey*. International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X,Vol 2.

Wadyono, Agus dan Sudarma S. 2012. *Tip Trik Android untuk Pengguna Tablet & Handphone*. Media Kita, Jakarta.

Juansyah, Andi. “Pembangunan aplikasi child tracker berbasis assisted – global positioning system (a-gps) dengan platform android”, *Jurnal Komputa* (2015) ISSN:2089-9033 Vol 1.