IMPLEMENTASI PARTICLE SWARM OPTIMIZATION PADA ANALYSIS SENTIMENT REVIEW APPSTORE FOR ANDROID MENGGUNAKAN K-NEAREST NEIGHBORS

TESIS

Disusun oleh :

SUCITRA SAHARA
14000814

PROGRAM PASCA SARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
NUSA MANDIRI
JAKARTA
2015

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama	: Sucitra Sahara
NIM	: 14000814
Program Studi	: Magister Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: Manajemen Information System (MIS)

Dengan ini menyatakan bahwa tesis yang telah saya buat dengan judul "Implementasi K-Nearest Neighbors Pada Analysis Centiment Review Appstore for Android dengan Menerapkan Naive Bayes".

Adalah hasil karya sendiri, dan semua sumber baik yang kutipan maupun yang dirujuk telah saya nyatakan dengan benar dan tesis ini belum pernah diterbitkan atau dipublikasikan dimanapun dan dalam bentuk apapun.

Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila dikemudian hari ternyata saya memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tesis yang telah saya buat adalah hasil karya milik seseorang atau badan tertentu, saya bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari Program Pascasarjana Magister Ilmu Komputer Sekolah tinggi Manajemen Informatika dan komputer Nusa Mandiri dicabut/dibatalkan.

Jakarta, 15 Februari 2015
Yang menyatakan,

Sucitra Sahara

HALAMAN PENGESAHAN

Tesis ini diajukan oleh:

Nama	: Sucitra Sahara
NIM	: 14000814
Program Studi	: Magister Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: Manajemen Information System (MIS)
Judul Tesis	: Implementasi K-Nearest Neighbors Pada Analysis Centiment Review Appstore for Android dengan Menerapkan Naive Bayes.

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Magister ilmu komputer (M. Kom) pada program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).

Jakarta, 15 Maret 2015
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur

Prof. Dr. Ir. Kaman Nainggolan, MS

DEWAN PENGUJI

Penguji I	: Dr. Windu Gata, M.Kom	.....
Penguji II	: Dr. Sfenrianto, M.Kom	.....
Penguji III/ Pembimbing	: Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd	.....

KATA PENGANTAR

Alhamdulillah, ucapan syukur penulis panjatkan Kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya, sehingga peneliti dapat menyelesaikan penulisan tesis ini dengan baik dan tepat pada waktunya. Sholawat serta salam pada pucuk pimpinan alam revolusi Islam, Nabi Besar Muhammad SAW. Dimana tesis ini penulis sajikan dalam bentuk buku yang sederhana. Adapun judul tesis, yang penulis ambil sebagai berikut **"**Implementasi Particle Swarm Optimization pada *Analysis Sentiment Review* Appstore *for* Android Menggunakan K-Nearest Neighbors**".**

Tujuan dari penelitian tesis ini adalah sebagai salah satu syarat kelulusan program Pascasarjana dan untuk memperoleh gelar Magister Ilmu Komputer pada program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) Jakarta.

Penulis juga lakukan mencari dan menganalisa berbagai macam sumber referensi, baik dalam bentuk jurnal ilmiah, buku-buku literatur, *internet*, dll yang terkait dengan pembahasan pada tesis ini. Penulis menyadari bahwa tanpa adanya dukungan dan bimbingan dari semua pihak maka tesis ini tidak dapat terselesaikan tepat pada waktunya. Oleh karena itu ijinkanlah penulis mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd selaku Pembimbing dan Ketua program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri).
2. Orang tua tercinta, untuk Ayah dan Mamah yang selalu mendoakan langkahku tanpa henti.
3. Untuk Nenek tersayang, adik-adikku, sepupu, keponakan, dan semua saudara yang selalu memberikan support.
4. Seluruh staf, karyawan Bina Sarana Informatika dan staf Program Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta.

5. Rekan-rekan mahasiswa Pascasarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, yang selalu membagi ilmu serta dukungannya.

Penulis menyadari bahwa tesis ini masih banyak kekurangan disana sini, untuk itu penulis mohon kritik dan saran yang bersifat membangun demi kesempurnaan penulisan karya ilmiah di masa mendatang.

Akhir kata semoga tesis ini dapat bermanfaat bagi penulis khususnya dan bagi para pembaca pada umumnya.

Jakarta, 16 Maret 2015

Sucitra Sahara
Penulis

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama	: Sucitra Sahara
NIM	: 14000814
Program Studi	: Magister Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: Manajemen Information System (MIS)
Jenis Karya	: Tesis

Demi pengembangan ilmu pengetahuan, dengan ini menyetujui untuk memberikan ijin kepada pihak Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri) **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalti-Free Right)*** atas karya ilmiah kami yang berjudul: "Implementasi K-Nearest Neighbors Pada Analysis Centiment Review Appstore for Android dengan Menerapkan Naive Bayes"

Dengan **Hak Bebas Royalti Non-Eksklusif** ini pihak STMIK Nusa Mandiri berhak menyimpan, mengalih-media atau bentukkan, mengelolanya dalam pangkalan data (*database*), mendistribusikannya dan menampilkan atau mempublikasikannya di internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta karya ilmiah tersebut.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak STMIK Nusa Mandiri, segala bentuk tuntutan hukumyang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 15 Februari 2015
Yang menyatakan,

Sucitra Sahara

ABSTRAK

Nama	: Sucitra Sahara
NIM	: 14000814
Program Studi	: Magister Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: Manajemen Information System (MIS)
Judul Tesis	: Implementasi Particle Swarm Optimization pada *Analysis Sentiment Review* Appstore *for* Android Menggunakan K-Nearest Neighbors

Pesatnya aplikasi berbasis android yang menjamur memungkinkan para vendor maupun pihak pebisnis berlomba lomba menciptakan bermacam aplikasi, mulai kualitas dan performa tinggi sampai kualitas yang masih sering diragukan, sehingga peneliti mengadakan penyeleksian terhadap aplikasi untuk android berdasarkan opini atau komentar masyarakat yang telah menggunakan aplikasi tersebut dan dituangkan kedalam media online. Dari banyak komentar yang telah direview memperoleh data set berupa teks positif dan negatif yang akan peneliti buat pengklasifikasian data dengan menggunakan k-Nearest Neighbors(k-NN), k-NN merupakan salah satu algoritma yang paling populer untuk pengenalan pola. Banyak peneliti telah menemukan bahwa algoritma KNN dapat menyelesaikan kinerja yang sangat baik pada data set yang berbeda terutama pada penyeleksian teks, Particle Swarm Optimization(PSO) yang dikombinasikan dengan k-NN berkepentingan untuk meningkatkan kinerja klasifikasinya. Sebelum digunakan optimasi dengan PSO pada data set akurasi yang didapat 75.50% dan setelah dikombinasikan antara k-NN dan PSO akurasinya adalah 88.50%.

Penggunaan PSO dan k-NN sesuai dengan konsep text mining yaitu bertujuan untuk mencari pola-pola yang ada pada teks, kegiatan yang dilakukan oleh text mining disini adalah *text classification*.

Kata kunci: *Review* komentar, k-Nearest Neighbors, Particle Swarm Optimization, *Text classification*

ABSTRACT

Nama	: Sucitra Sahara
NIM	: 14000814
Program Studi	: Magister Ilmu Komputer
Jenjang	: Strata Dua (S2)
Konsentrasi	: Manajemen Information System (MIS)
Judul Tesis	: Implementasi Particle Swarm Optimization pada *Analysis Sentiment Review* Appstore *for* Android Menggunakan K-Nearest Neighbors

The rapid application for android-based the mushrooming allows the seller nor the parties competing businesses competing to create a variety of applications for, start quality and high performance until quality that still often doubtful, so the researchers held a screening of applications for android against opinion or for based on the comments society the been use the application for and poured into online media. From many comments the been data were reviewed to obtain the set of positive and negative text form that would researchers create data classification with use k-nearest neighbor (k-NN), k-NN algorithm is praying one that paled passable fame for pattern recognition. Researchers found that the knn algorithm may have finish boarding costs sets of data collection on a lot of very good thing distinct in terms of selecting text, Particle Swarm Optimization (PSO) the combined by k-NN concerned for improve classification immersed boarding costs. Before pada sets used optimization PSO pada with data accuracy the obtained 75.50% and the combined taxable income between k-nn and pso accuracy is 88.50%.

Use PSO and k-NN with concept that text mining aims to find patterns of text on, activities performed text mining posted here is the text classification.

Keywords: Review comments, k-Nearest Neighbors, Particle Swarm Optimization, Text Classification

DAFTAR ISI

Halaman

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR LAMPIRAN

Halaman

BAB I
PENDAHULUAN

1.1. Latar Belakang Penulisan

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

Sistem operasi Android memiliki pangsa pasar tertinggi pada tahun 2014, sehingga sistem ini menjadi sistem operasi mobile yang paling banyak digunakan di dunia. Fakta ini membuat pengguna Android menjadi target terbesar untuk disusupi malware. 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 apikasi smartphone dengan membandingkan methode *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 Bayessian* (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).

Particle Swarm Optimization (PSO) merupakan teknik komputasi evolusioner yang mampu menghasilkan solusi secara global optimal dalam ruang

pencarian melalui interaksi individu dalam segerombolan partikel. 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). PSO banyak digunakan untuk memecahkan masalah optimasi serta pada seleksi fitur (Liu, et al., 2011).

Pada penelitian, pengklasifikasian k-Nearest Neighbors dengan optimasi *Particle Swarn Optimization* (PSO) sebagai metode pemilihan fitur akan diterapkan untuk klasifikasi *text* pada pendapat atau opini public mengenai *review* produk *Appstore for Android*.

1.2. Permasalahan Penelitian

Pembahasan masalah penelitian akan di uraikan dalam tiga kategori diataranya yaitu:

1.2.1 Identifikasi Masalah

Klasifikasi pada metode k-Nearest Neighbor (k-NN) berbasis mesin *query* indek lokasi terdekat yang sangat sederhana dan sangat efisien untuk klasifikasi teks. Namun, k-Nearest *Neighbor* memiliki kekurangan yaitu menimbulkan *overhead* yang tinggi dan lemah terhadap data yang dinamis, k-NN memiliki keterbatasan seperti: besar kompleksitas perhitungan, sepenuhnya tergantung pada training set, dan tidak ada perbedaan berat badan antara masing-masing kelas. Untuk mengatasi hal ini, metode baru untuk meningkatkan klasifikasi kinerja k-NN menggunakan *Particle Swarm Optimization* yang dapat digunakan dalam seleksi atribut agar k-NN mendapatkan hasil yang lebih optimal. Dibuktuikan dalam penelitian ini, untuk mengetahui terpecahkan atau tidak masalah tersebut.

1.2.2 Batasan Masalah

Seperti yang sudah dibahas pada identifikasi masalah diatas, mengenai metode yang akan digunakan dalam penelitian, penulis akan membatasi masalah yaitu meneliti data analisi sentiment review Appstore pada Android pada website http://www.amazon.com dengan menerapkan seleksi fitur Particle Swarm Optimization (PSO) dengan menggunakan Algoritma k-Nearest Neighbor.

1.2.3 Rumusan Masalah

Dengan menerapkan metode seleksi fitur *Particle Swarm Optimization algoritm* dan penggunaan klasifikasi algoritma k-Nearest Neighbor, apakah nilai pada akurasi pada analisa sentimen review aplikasi Android akan berpengaruh besar?

1.3. Manfaat dan Tujuan Penelitian

Penjabaran mengenai manfaat dan tujuan dalam penelitian ini.

1.3.1 Manfaat Penelitian

a. Manfaat praktis dari hasil penelitian ini adalah dapat digunakan oleh para pengguna maupun vendor dapat mengetahui bahwa aplikasi tersebut baik digunakan atau tidak, bagi pengguna untuk kemudahan dalam menilai dan memilih aplikasi pada android, sedangkan untuk para developer aplikasi dapat mengetahui opini mengenai aplikasi yang telah dibuat dan bahan pertimbangan untuk pembuatan aplikasi android yang lebih baik dari sebelumnya.

b. Manfaat kebijakan dari hasil penelitian ini adalah dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan dalam memilih aplikasi yang baik dan tidak baik.

c. Manfaat teoritis dari penelitian ini yaitu diharapkan dapat memberikan sumbangan untuk pengembangan teori yang berkaitan dengan penerapan *Particle swarm optimization* pada untuk meningkatkan akurasi penentuan sentiment analisis review appstore for Android.

1.3.2 Tujuan Penelitian

Peneliti mempunyai tujuan memperoleh dan menganalisa penerapan metode *Particle Swarm Optimization* (PSO) untuk menganalisa sentiment pada review aplikasi Android dan pengklasifikasi k-Nearest *Neighbor*, agar diperoleh keputusan berdasarkan hasil yang didapat, keputusan tersebut berupa respon positif atau negative dalam menentukan produk tersebut baik atau tidak, sehingga pengguna maupun konsumen produk Appstore for Android dengan mudah menentukan kosumsi produk tersebut.

1.4. Ruang Lingkup Penelitian

Ruang lingkup pembahasan dalam penelitian ini dibatasi pada penerapan model Algoritma k-Nearest Neighbors (k-NN) dengan *Particle swarm optimization* (PSO) yang digunakan untuk seleksi atribut dalam analisis sentimen review yang akan ditetapkan, untuk klasifikasi berdasarkan dari data www.amazon.com/review, *Software* yang digunakan adalah Rapid Miner dimana *software* tersebut memiliki sistem yang komprehensif untuk analisa data serta banyak digunakan karena kemampuan, fleksibilitas dan kemudahan dalam penggunaannya.

1.5. Hipotesis

Diduga k-Nearest Neighbors (k-NN) dapat diterapkan dalam penentuan sentiment analisis review produk, dan *Particle swarm optimization* (PSO) dapat digunakan dalam seleksi atribut yang sesuai pada k-Nearest Neighbors untuk meningkatkan akurasi penentuan review negative dan positif.

1.6. Sistematika Penulisan

Berikut sistematika penulisan tesis yang telah dibuat:

Bab I: Pendahuluan

Membahas mengenai latar belakang penulisan, masalah pada pengklasifikasian teks analisa dari suatu komentar, pemecahan masalah dan tujuan penelitan.

Bab II: Landasan Teori

Membahas teori yang melandasi penelitian, dan peneliti menyajikan beberapa studi kasus dan contoh penggunaan algoritma.

Bab III: Metode Penelitian

Membahas metode pengumpulan data dan eksperimen. Menguji K-Nearest Neighbors untuk meningkatkan akurasi dalam mengklasifikasi komentar pada Review Appstore for Android.

Bab IV: Hasil dan Pembahasan

Menampilkan hasil dari eksperimen, sebelum diterapkannya model dan sesudah diterapkan model.

Bab V: Penutup

Membahas kesimpulan dan kekurangan penelitian, serta kelebihan dari model yang telah digunakan.

BAB II

LANDASAN TEORI DAN KERANGKA PENELITIAN

2.1. Tinjauan Pustaka

Tinjauan pustaka dilakukan dengan menggunakan referensi dari buku-buku, jurnal ataupun artikel yang didapatkan melalui media internet sebagai acuan penulisan ini, berikut adalah pengertian-pengertian mengenai penulisan yang akan dibahas.

2.1.1. Review Data Set

Review text adalah teks yang ditujukan untuk meninjau suatu karya, baik film, buku dan sebagainya, untuk mengetahui kualitas, kelebihan serta kekurangan yang dimiliki oleh karya tersebut. Tujuan koomunikatif dari *Review text* adalah *to criticise an art work, event for a public audience* (melakukan kritik terhadap peristiwa atau karya seni ataupun lainnya untuk khalayak umum).

2.1.2. Analisa Sentimen (*Sentiment Analysis*)

Analisis sentimen adalah “Riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual dilakukan untuk melihat pendapat terhadap sebuah masalah, atau untuk identifikasi kecenderungan hal di pasar. Saat ini pendapat masyarakat menjadi sumber yang penting dalam pengambilan keputusan akan suatu produk.” (Ipmawati, Kusrini, dan Luthfi, 2017).

Analisis sentimen adalah “proses yang bertujuan untuk menentukan isi dari dataset yang berbentuk teks bersifat positif, negatif atau netral. Saat ini, pendapat khalayak umum menjadi sumber yang penting dalam pengambilan keputusan seseorang akan suatu produk”. (Chandani dan Romi Satria Wahono,

2015).

Menurut (Sipayung, et al., 2016) "*Sentiment Analysis* adalah merupakan perpaduan dari data mining dan *text mining*, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi".

Menurut (Moraes, Valiati, & Gavião Neto, 2013) langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen diantaranya:

1. Definisikan Domain Dataset

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

2. Pre-processing

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

3. 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*.

4. *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 Fequency, Information Gain, Chi-Square,* dan lain sebagainya.

5. *Clasfication*

Proses klasifikasi umumnya menggunkan pengklasifikasi seperti *Naive Bayes, Support Vector Machine,* dan lain sebagainya

6. *Interpretation/Evaluation*

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

2.1.3. Seleksi Fitur (*Feature Selection*)

Feature Selection atau seleksi fitur adalah sebuah proses yang biasa digunakan pada *Machine Learning* dimana sekumpulan dari fitur yang dimiliki oleh data digunakan untuk pembelajaran algoritma. *Feature Selection* menurut (Nugroho dan Wibowo, 2017) telah menjadi bidang penelitian aktif dalam pengenalan pola, statistik, dan Data Mining. Seleksi fitur adalah salah satu faktor yang paling penting yang dapat mempengaruhi tingkat akurasi klasifikasi karena jika dataset berisi sejumlah fitur, dimensi dataset akan menjadi besar hal ini membuat rendahnya nilai akurasi klasifikasi. Seleksi fitur adalah kemampuannya untuk mendeteksi hubungan nonlinier antar variabel, hal ini memungkinkan pengambilan relevansi dan redundansi fitur secara bersamaan (Mira et al., 2018).

Masalah dalam seleksi fitur adalah pengurangan dimensi, dimana

awalanya semua atribut diperlukan untuk memperoleh akurasi yang maksimal. Empat alasan utama untuk melakukan pengurangan dimensi menurut (Nugroho dan Wibowo, 2017).

1. *Decreasing the learning cost* atau penurunan biaya pembelajaran.
2. *Increasing the learning performance* atau meningkatkan kinerjapembelajaran.
3. *Reducing irrelevant dimensions* atau mengurangi dimensi yang tidak relevan.
4. *Reducing redundant dimensions* atau mengurangi dimensi yang berlebihan.
5. Adapun tujuan seleksi fitur menurut (Wahyuni, 2016) adalah "Mengurangi fitur data yang berdimensi tinggi", untuk tujuan dari kegiatan data mining dan fitur metode seleksi dapat diklasifikasikan ke dalam tiga kategori utama yaitu:

1. Metode filter

Metode Filter adalah memilih atribut yang relevan sebelum pindah ke tahap pembelajaran berikutnya, atribut yang dianggap paling penting yang dipilih untuk pembelajar, sedangkan sisanya dikecualikan.

2. Metode *wrapper*

Metode *wrapper* menilai sekelompok variabel dengan menggunakan klasifikasi yang sama atau algoritma regresi digunakan untuk memprediksi nilai dari variabel target.

3. Metode *embedded*

Untuk metode *embedded*, proses seleksi atribut terletak di dalam algoritma pembelajaran, sehingga pemilihan set optimal atribut secara langsung dibuat selama fase generasi model.

Ide utama dari *Feature Selection* adalah memilih subset dari fitur yang ada tanpa transformasi karena tidak semua fitur/atribut relevan dengan masalah.

Bahkan beberapa dari fitur atau atribut tersebut mengganggu dan mengurangi akurasi. *Noisy Features* atau fitur yang tidak terpakai tersebut harus dihapus untuk meningkatkan akurasi. Selain itu dengan fitur atau atribut yang sangatbanyak akan memperlambat proses komputasi.

2.1.4 *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, 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*

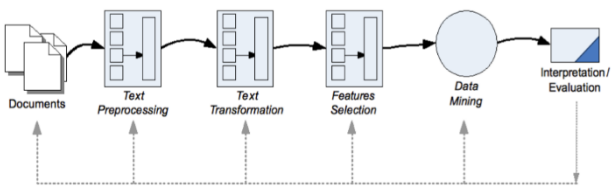
Proses klasifikasi umumnya menggunakan pengklasifikasi seperti Naïve Bayes, Support Vector Machine, dan lain sebagainya.

f. *Interpretation/Evaluation*

Tahap evaluasi biasanya menghitung akurasi, recall, precision.

1. 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.CNET download.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; Gamgaran Somprasertsi, 2010).



Sumber: Ronen Feldman(2007)

Gambar II.1 Tahapan Proses Text Mining

2. 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).

3. *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. *Steamming*

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

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

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, karenakan 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.

5. Penelian Terkait

Dari beberapa peneliti terkait menggunakan model yang berbeda-beda, dan pengklasifikasian mengunakan k-NN memiliki akurasi yang dinilai optimal. Kemudian untuk kesimpulan yang didapat, Fitur Selection PSO mampu mengangkat performa dari metode k-NN.

Judul	Classifier and Future Selection	Accuracy	Hasil
Noisy data elimination using mutual k-nearest neighbor for classification mining (Liu dan Zang, 2011)	MKNNC, K-NN	MKNNC= 82.2 % K-NN= 78.31%	Diambil nilai akurasi tertinggi dari banyak eksperimen. Disimpulkan Akurasi pada MKNNC lebih besar dibanding dengan k-NN.
A direct boosting algorithm for the k-nearest neighbor classifier via local warping of the distance metric (Neo dan Ventura, 2012)	Classifier k-NN via local warping of the distance metric, Bossted k-NN	Accuracy Defference: Boosted k-NN= 0.01 k-NN=0.08	Peneliti mengadakan eksperimen keberbagai dataset salah satunya data resonan
A novel hybrid system for feature selection based on an improvedgravitational search algorithm and k-NN method (Xiang dan Han, 2012)	GSA, k-NN	83.9%	Peneliti banyak melakukan eksperimen terhadap beberapa banyak dataset, yang diambil contoh dari pengujian k-NN dan GA, ternyata menghasilkan akurasi yang cukup optimal.
An Improved k-Nearest Neighbor Classification Using Genetic Algorithm (Saguna, Tanushkodi, 2010)	GA, k-NN	67.59%	Dari dataset yang peneliti uji, diambil sampel pada dataset HIV. Akurasinya belum begitu optimal
Tony Bellotti and Jonathan Crook (2007)	*Support Vector Machine* (SVM), Logistic Regression (LR), Linear Discriminant Analysis (LDA) dan k-Nearest Neighbours(kNN)	-	SVM dengan model liniear dan gaussian mengungguli ketiga metode lainnya
Implementasi Particle Swarm pada Analysis Sentiment Review Appstore For Android menggunakan K-Nearest Neighbors	PSO, k-NN	?	Peneliti melakukan pengujian terhadap sentiment analisis review aplikasi pada android.

2.1.4. Validasi dan Evaluasi Algoritma Text Mining

Validasi merupakan proses mengevaluasi akurasi dari suatu model. Dalam mengevaluasi model klasifikasi berdasaran perhitungan objek data testing mana yang diprediksi benar dan tidak benar.

True Positive (TP) : Proporsi positif dalam data set yang diklasifikasikan Positif

True Negative (TN) : Proporsi negatif dalam data set yang diklasifikasikan Negatif

False Positive (FP) : Proporsi negatif dalam data set yang diklasifikasikan Positif

False Negative (FN) : Proporsi negatif dalam data set yang diklasifikasikan negatif

Berikut adalah persamaan model *Confution Matrix*:

1. Nilai *Accuracy* adalah proporsi jumlah prediksi yang benar. Dapat dihitungdengan menggunakan persamaan:

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

2. *Sensitivity* digunakan untuk membandingkan proporsi TP terhadap tupelyang positif, yang dihitung dengan menggunakan persamaan:

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

3. *Specificity* digunakan untuk membandingan proporsi TN terhadap tupelyang negatif, yang dihitung dengan menggunakan persamaan:

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

4. PPV (*positive predictive value*) adalah proporsi kasus dengan hasil diagnosapositif, yang dihitung dengan menggunakan persamaan:

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

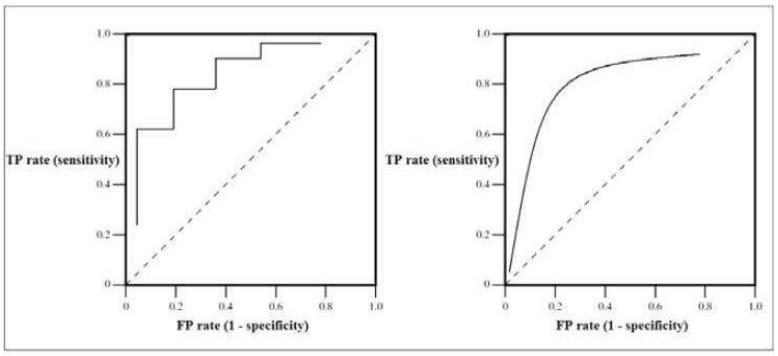
5. NPV (*negative predictive value*) adalah proporsi kasus dengan hasildiagnosa negatif, yang dihitung dengan menggunakan persamaan:

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

Cross validation set pelatihan dan validasi harus *crossover* berturut-turut sehingga setiap data memiliki kesempatan tervalidasi(Witten, Frank, dan Hall, 2011a).

Kurva ROC (*Receiver Operating Characteristic*) digunakan untuk mengevaluasi akurasi classifier dan untuk membandingkan klasifikasi yang berbeda model (Vercellis, 2009). Kurva ROC digunakan untuk mengukur AUC (*Area Under Curve*). Kurva ROC membagi hasil positif dalam sumbu y dan hasil negatif dalam sumbu x (Witten, Frank, dan Hall, 2011b). Sehingga semakin besar area yang berada dibawah kurva. semakin baik pula hasil prediksi.

Permasalahan dalam klasifikasi kurva ROC dapat digunakan untuk menguji dan menilai hasil kinerja pengklasifikasian secara visual dan yang digunakan untuk mengekspresikan *confusion matrix*. Kurva ROC merupakan grafik dua dimensi dengan *false positive* sebagai garis *horizontal* dan *true positive* sebagai garis *vertikal* (Vecellis, 2009).



Gambar 2.3. Kurva ROC

Sumber : Gorunescu (2011)

Gambar 2.3 Grafik ROC (*discrete dan continous*) Pada gambar 2.3 Garis diagonal membagi ruang ROC, yaitu :

1. poin diatas garis diagonal merupakan hasil klasifikasi yang baik.
2. poin dibawah garis diagonal merupakan hasil klasifikasi yang buruk.

Dapat disimpulkan bahwa, satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas

didalam grafik.

Berikut panduan untuk mengklasifikasikan keakuratan diagnosa menggunakan AUC (Gorunescu, 2011) :

1. 0.90-1.00 = *excellent classification*;
2. 0.80-0.90 = *good classification*;
3. 0.70-0.80 = *fair classification*;
4. 0.60-0.70 = *poor classification*;5. 0.50-0.60 = *failure*.

2.1.5. RapidMiner

RapidMiner merupakan perangkat lunak yang bersifat terbuka (*open source*).RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, *text mining* dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriftif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. RapidMiner memiliki kurang lebih

500 operator data mining, termasuk operator untuk input, output, data preprocessing dan visualisasi. RapidMiner merupakan *software* yang berdiri sendiri untuk analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. RapidMiner ditulis dengan menggunakan bahasa java sehingga dapat bekerja di semua sistem operasi (Aprilia dkk : 2013)

RapidMiner sebelumnya bernama YALE (*Yet Another Learning Environmet*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh RalfKlinkenberg, Ingo Mierswa, dan Simon Fischer di *Artificial Intelligence Unit* dari *University of Dortmund*, RapidMiner didistribusikan di bawah licensi AGPL (*GNU Affero General Public License*) versi 3. Hingga saat ini telah ribuan

aplikasi yang dikembangkan menggunakan RapidMiner di lebih dari 40 negara. RapidMiner sebagai *software open source* untuk data mining tidak perlu diragukan lagi karena software ini terkemuka di dunia. RapidMiner menempati peringkat pertama sebagai Software data minig pada polling oleh Kdnuggets, sebuah portal data mining pada 2010-2011.

RapidMiner menyediakan GUI (*Graphic User Interface*) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML (*Extensible Markup Language*) yang mendefinisikan proses analitis keinginan pengguna untuk diterapkan ke data. File ini kemudian dibaca oleh RapidMiner untuk menjalakan analisis secara otomatis.

RapidMiner memiliki beberapa sifat sebagai berikut:

1. Ditulis dengan bahsa pemrograman Java sehingga dapat dijalankan di berbagaisistem operasi.
2. Proses penemuan pengetahuan dimodelkan sebagai operator trees.
3. Representasi XML internal untuk memastikan format standar pertukaran data.
4. Bahasa scripting memungkinkan untuk eksperimen skala besar dan otomatisasieksperimen.
5. Konsep multi-layer untuk menjamin tampilan data yang efisien dan menjaminpenangan data.
6. Memiliki GUI, *command line mode*, dan Jawa API yang dapat dipanggil dariprogram lain.

Beberapa fitur dari *RapidMiner*, antara lain:

1. Banyaknya algoritma data mining, seperti *decision tree* dan *self-*

organizationmap.

2. Bentuk grafis yang canggih, seperti tumpang tindih diagram histogram, *treechart* dan 3D *Scatter plots*.

3. Banyaknya variasi plugin, seperti *text plugin* utuk melakukan analisis text.

4. Menyediakan prosedur data mining dan *machine learning* termasuk: ETL (*extraction, transformation loading*), data *preprocessing*, visualisasi, *modelling* dan evaluasi.

5. Proses *data mining* tersusun atar operator-operator yang *nestable*, dideskripsikan dengan XML, dan dibuat dengan GUI.

2.2. Tinjauan Studi

Berikut adalah beberapa penelitian terdahulu yang terkait dengan topic penelitian yang penulis jadikan rujukan. Secara garis besar tinjauan studi dalam tesis ini meliputi

Judul	Classifier and Future Selection	Accuracy	Hasil
Noisy data elimination using mutual k-nearest neighbor for classification mining (Liu dan Zang, 2011)	MKNNC, K-NN	MKNNC= 82.2 % K-NN= 78.31%	Diambil nilai akurasi tertinggi dari banyak eksperimen. Disimpulkan Akurasi pada MKNNC lebih besar dibanding dengan k-NN.
A direct boosting algorithm for the k-nearest neighbor classifier via local warping of the distance metric (Neo dan Ventura, 2012)	Classifier k-NN via local warping of the distance metric, Bossted k-NN	Accuracy Defference: Boosted k-NN= 0.01 k-NN=0.08	Peneliti mengadakan eksperimen keberbagai dataset salah satunya data resonan
A novel hybrid system for feature selection based on an improvedgravitational search algorithm and k-NN method (Xiang dan Han, 2012)	GSA, k-NN	83.9%	Peneliti banyak melakukan eksperimen terhadap beberapa banyak dataset, yang diambil contoh dari pengujian k-NN dan GA, ternyata menghasilkan akurasi yang cukup optimal.
An Improved k-Nearest Neighbor Classification Using Genetic Algorithm (Saguna, Tanushkodi, 2010)	GA, k-NN	67.59%	Dari dataset yang peneliti uji, diambil sampel pada dataset HIV. Akurasinya belum begitu optimal
Tony Bellotti and Jonathan Crook (2007)	*Support Vector Machine* (SVM), Logistic Regression (LR), Linear Discriminant Analysis (LDA) dan k-Nearest Neighbours(kNN)	-	SVM dengan model liniear dan gaussian mengungguli ketiga metode lainnya
Implementasi Particle Swarm pada Analysis Sentiment Review Appstore For Android menggunakan K-Nearest Neighbors	PSO, k-NN	?	Peneliti melakukan pengujian terhadap sentiment analisis review aplikasi pada android.

2.2. Kerangka Pemikiran

Penerapan algoritma k-NN menghasilkan nilai akurasi pada klasifikasi review appstore for android 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. Peneliti menyediakan aplikasi berbasis web untuk menguji model menggunakan dataset yang berbeda dan belum diklasifikasikan sesuai dengan kelasnya. Diaplikasikan dengan menggunakan bahasa pemgrograman PHP berbasis Web.

BAB III

METODOLOGI PENELITIAN

3.1. Perancangan Penelitian

Dalam penelitian ada kegiatan penyelidikan (*investigation*), yaitu mencari fakta secara teliti dan teratur menurut kaidah tertentu untuk menjawab suatu pertanyaan. Jadi penyelidikan dilakukan untuk menjelaskan sesuatu. Secara umum metode penelitian diartikan sebagai cara ilmiah untuk mendapatkan data dengan tujuan dan kegunaan tertentu.

Metode penelitian yang penulis lakukan adalah metode penelitian eksperimen, dengan tahapan sebagai berikut:

1. **Pengumpulan data set**, untuk kemudian diseleksi dari data yang tidak sesuai.
2. **Pengolahan data awal**, dipilih berdasarkan kesesuaian data dengan metode yang paling baik dari beberapa metode pengklasifikasian teks yang sudah digunakan oleh beberapa peneliti sebelumnya. Model yang digunakan adalah algoritma k-Nearest Neighbors (k-NN).
3. **Metode Yang Diusulkan**, untuk meningkatkan akurasi dari Algoritma k-Nearest Neighbors (k-NN), maka dilakukan penambahan dengan menggabungkan metode pemilihan fitur filter dan wrapper, yaitu Information gain dan Genetic algorithm.
4. **Eksperimen dan Pengujian Metode**, dalam eksperimen data penelitian, penulis menggunakan RapidMiner 5 untuk mengolah data. Sedangkan

untuk pengujian metode, penulis membuat aplikasi menggunakan bahasa pemrograman PHP dan HTML.

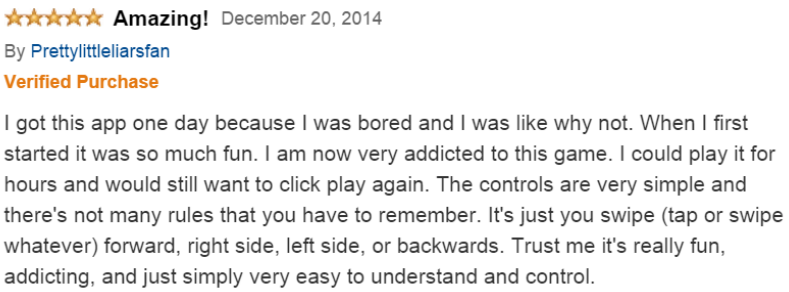
5. **Evaluasi dan Validasi Hasil Evaluasi**, dilakukan untuk mengetahui akurasi dari model algoritma k-Nearest Neighbors (k-NN). Proses validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah Cross Validation.

3.2.Pengumpulan Data

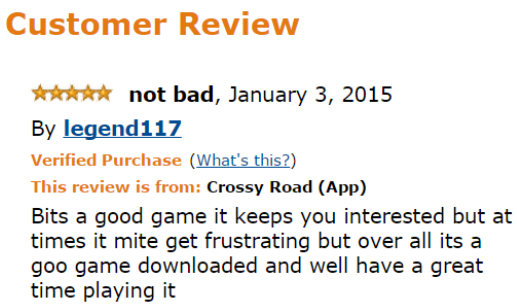
Data yang digunakan dalam kebanyakan studi klasifikasi sentimen dikumpulkan dari situs e-commerce seperti www.amazon.com (review produk), www.yelp.com (ulasan restoran), www.CNET download.com (review produk) dan www.reviewcentre.com, yang menjadi tuan rumah jutaan ulasan produk oleh konsumen. Selain 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; Gamgaran Somprasertsi ,2010).

Peneliti menggunakan data *review* aplikasi Android pada layanan Appstore for Android. Mengolah data yang diambil pada *Customer Reviews* pada aplikasi dari situs http://www.amazon.com.

Contoh review komentar positif :



Contoh review komentar negatif :



3.3.Pengolahan Data Awal

Dari banyaknya data review, peneliti mengambil sample data sebanyak 100 *review* positif dan 100 review negatif sebagai data training berdasarkan pengambilan sampling data secara *simple random*.

Text processing bertujuan untuk mempersiapkan dokumen teks yang tidak terstruktur menjadi data terstruktur yang siap digunakan untuk proses selanjutnya (Chandani,Wahono, dan Purwanto.2015).

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:

1. *Tokenizazion*

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

2. *Stopwards Removal*

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

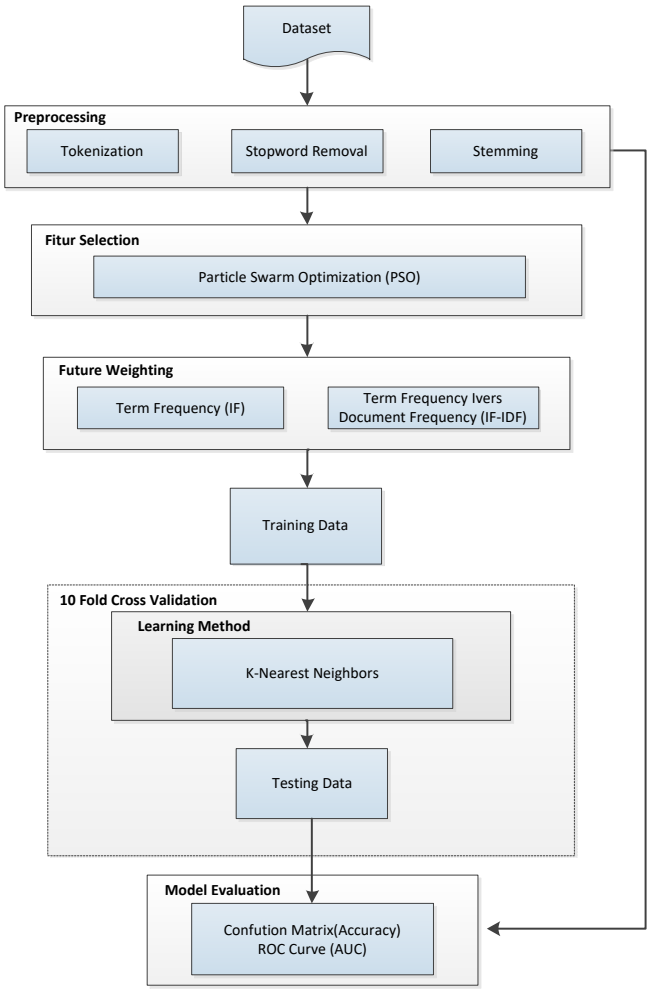
3. *Steamming*

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

Sedangkan untuk tahap *transformation* dengan melakukan pembobotan TF- IDF pada masing-masing kata. Di mana prosesnya 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.

3.4. Metode yang Diusulkan

Peneliti mengusulkan menggunakan 1 algoritma klasifikasi k-Nearest Neighbors (k-NN) dan seleksi fitur menggunakan Particle Swarm Optimization (PSO) untuk meningkatkan akurasi pada k-NN. Kerena dengan menggunakan usulan tersebut dapat memenuhi nilai akurasi yang baik. PSO yang secara teknik optimasi sederhana untuk menerapkan dan memodifikasi beberapa parameter, sedangkan untuk k-NN sebuah metode untuk melakukan klasifikasi terhadap atribut atau objek berdasarkan data set yang jaraknya paling dekat dengan atribut tersebut. Pada Gambar III.1 merupakan detail dari model usulan.



Sumber: Peneliti

Gambar III.1 Model Usulan

Model tersebut menjabarkan bahwa dataset melewati tahap preprocessing terlebih dahulu untuk menyeleksi kata-kata dan huruf yang baik untuk nantinya akan diklasifikasi. Tahap evaluasi digunakan 10 fold cross validation, dan pengukuran akurasi diukur dengan confusion matrix.

Hasil yang dibandingkan disini adalah akurasi k-Nearest Neighbors (k-NN) sebelum menggunakan metode fiture selection dengan akurasi k-Nearest Neighbors (k-NN) setelah menggunakan metode fiture selection Particle Swarm Optimization (PSO).

3.5. Eksperimen dan Hasil Pengujian

Peneliti melakukan proses eksperimen menggunakan aplikasi Rapid Miner 5. Sedangkan untuk pengujian model dilakukan menggunakan dataset review aplikasi android pada appstore for android pada situs http://amazon.com/review. Untuk implementasi model selain data training pada aplikasi Rapid Miner, peneliti membuat aplikasi pengolah review appstore for andorid menggunakan bahasa pemrograman PHP dan HTML. Spesifikasi komputer yang peneliti gunakan ada pada tabel III.1.

Tabel III.1 Spesifikasi Komputer yang Digunakan

Processor	Intel Core i3
Memori	4 GB
Hard Disk	320 GB
Sistem Operasi	Microsoft Windows 2010
Aplikasi Text Mining	Rapid Miner Versi 5.3
Software	Adobe Dreamweaver
Bahasa Pemrograman	PHP, HTML

Sumber: Peneliti

3.6.Evaluasi dan Validasi Hasil

Pada evaluasi kali ini penulis mengusulkan penggunaan model dalam kegiatan review aplikasi pada android yaitu model k-Nearest Neighbors (k-NN) dan k-Nearest Neighbors (k-NN) berbasis Particle Swarm Optimization (PSO), 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, namun peneliti mencoba meninggikan tingkat akurasi dengan menggunakan Particle Swarm Optimization (PSO), pada optimasi PSO harus menentukan nilai population size yang tepat, barulah peneliti bisa menentukan nilai akurasi yang tertinggi. Maka disimpulkan bahwa struktur algoritma yang dirancang mencapai ideal dalam pemecahan masalah.

BAB IV
HASIL PENELITIAN DAN PEMBAHASAN

4.1. Implementasi Metodologi

Berdasarkan metodologi penelitian yang telah dipaparkan pada BAB III, berikut implementasi metodologi yang dilakukan dalam penelitian ini.

4.1.1. Pengumpulan Data

Data training yang digunakan dalam pengkasifikasian *text* terdiri atas 100 review positif pada Appstore for Android dan 100 review negatif pada 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.

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

Penerapan algoritma k-NN menghasilkan nilai akurasi pada klasifikasi review appstore for android 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 74.50% dan nilai AUC 0.825.

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

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

Implikasi penelitian ini mencakup beberapa aspek, di antaranya:

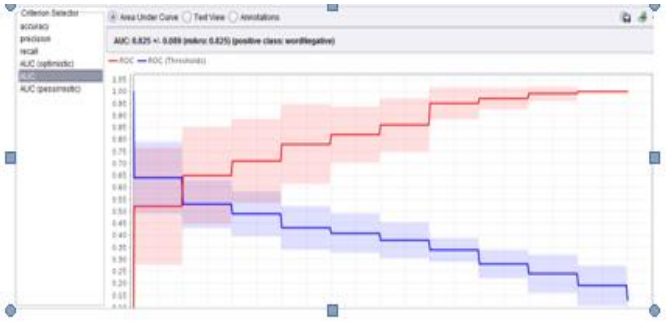
1. Implikasi terhadap aspek sistem Hasil evaluasi menunjukkan penerapan Algoritma k-Nearest Neighbors (k-NN) merupakan metode yang cukup baik dalam mengklasifikasi teks review Appstore for Android. Metode ini dapat membantu para calon pengguna android dalam menentukan aplikasi apa saja yang layak mereka 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 andorid.

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.

4.1.2. Hasil Pengujian

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



Sumber: Peneliti

Gambar 2. Hasil ROC Pengujian k-NN

6. 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 alogoritma k-Nearest Neighbors (k-NN) cukup tinggi, Nilai akurasi untuk model algoritma k-NN sebesar 74.50% pada k=10,dijabarkan pada tabel 4.

Tabel 4. Pengujian Algoritma k-NN

	Nilai k	Accuracy	AUC
k-NN	10	74.50%	0.825

Sumber: Peneliti

Kesimpulan pengujian ini algoritma k-NN dapat meningkatkan nilai akurasi yang merupakan solusi yang baik dalam permasalahan pada klasifikasi sentimen review aplikasi pada android.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Klasifikasi text dengan data berupa review aplikasi android, salah satu pengklasifikasian yang dapat digunakan adalah k-Nearest Neighbors (k-NN). Hal ini dikarenakan k-NN metode yang dapat sesuai dengan klasifikasi data dan mudah dipahami. 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 sebelum menggunakan penggabungan metode pemilihan fitur mencapai 74.50% Sedangkan setelah menggunakan penggabungan metode selection fitur Particle Swarm Optimization (PSO) akurasinya meningkat hingga mencapai 89.00%. Peningkatan akurasi mencapai 14.5%. Untuk memudahkan penelitian, dibuatlah aplikasi review apliaksi pada android untuk mengklasifikasikan review positif dan negatif yang ditampilkan dalam bentuk chart menggunakan bahasa pemrograman PHP. Model yang terbentuk dapat diterapkan pada seluruh data review aplikasi android dari berbagai sumber, sehingga dapat dilihat secara langsung hasilnya dalam bentuk positif dan negatif (chart). Hal ini dapat membantu seseorang untuk menghemat waktu saat mencari aplikasi yang akan digunakan baik atau tidak.

5.2. Saran

1. Dalam pengembangan pengujian selanjutnya adalah memperbaiki proses data preparation dimana kualitas data yang akan diolah menjadi lebih baik sehingga pengolahan pada proses text mining menjadi lebih optimal.
2. Mencari model fitur selection yang dapat digunakan untuk klasifikasi data berupa text seperti Genetic Algorithhm (GA), Chi Square, dan masih banyak lagi, agar ketika di bandingkan hasilnya optimal.
3. Mengimplementasikan model k-NN berbasis PSO pada data review yang berbeda, contohnya review produk, review barang elektronik, review online shop, dan banyak lainnya.

DAFTAR REFERENSI

Alpaydin, Ethem. (2010). *Introduction to Machine Learning*. London: The MIT Press.

B. Sanz, I. Santos, C. Laorden, X. Ugarte-Pedrero, P.G. Bringas, G. Álvarez, PUMA: permission usage to detect malware in Android, in: Proceedings of the International Joint Conference CISIS'12-ICEUTE'12-SOCO'12 Special Sessions, in: Advances in Intelligent Systems and Computing, vol. 189, Springer, Berlin, Heidelberg, 2013, pp. 289–298.

Belur V. Dasarathy, "Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques", Mc Graw-Hill Computer Science Series, *IEEE Computer Society Press*, Las Alamitos, California, pp. 217-224, 1991.

Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). *Feature Selection for Text Classification with* Naïve Bayes. Expert Systems with Applications, 36(3), 5432–5435

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

Han, J. dan Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.

Hu, and Liu, "Mining and summarizing customer reviews", Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, 2005,pp. 168–177.

Khairull, Baharudin&Aurnagzep, Asraf Ullah, Khan. (2014). *Mining Opinion Components From Unstructured Review*: A. Review. King Saud University

Langgeni, D. P., Baizal, Z. K. A., & W, Y. F. A. (2010). *Clustering* Artikel Berita Berbahasa Indonesia, 2010(semnasIF), 1–10.

Liao. (2007). *Recent Advances in Data Mining of Enterprise Data*: *Algorithms and Application*. Singapore: World Scientific Publishing

M.Govindarajan, Romina M, "A Survey of Classification Methods and Applications for Sentiment Analysis", The International Journal Of Engineering And Science (IJES), ISSN(e): 2319 – 1813 ISSN(p): 2319 – 1805, 2013

N. Saguna, K. Thanushkodi, “An Improved k-Nearest Neighbors Classification Using Genetic Algorithm”, IJCSI International Journal of Computer Science Issues, Vol. 7, 2010

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.

Qingliang Miao, Qiudan Li, Ruwei Dai , “AMAZING: A sentiment mining and retrieval system”, Expert Systems with Applications 36 (2009) 7192–7198.

Rapid-I GmbH. (2010).*Rapid Miner User Manual*. Dortmund: Rapid-I GmbH

Shukla, A., Tiwari, R., & Kala, R. (2010). *Real Life Applications of Soft Computing.* United States of America on: Taylor and Francis Group, LLC

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.

Weitong Huang, Yu Zhao, Shiqiang Yang, Yuchang Lu, “Analysis of the use behavior and opinion classification based on the BBS” , Applied Mathematics and Computation 205 (2008) 668–676

Xiang Jie, XiaoHong Han, Fu Duan, Yan Qiang, XiaoYan Xiong, Yuan Lan, Haishui Chai. 2015. A novel hybrid system for feature selection based on an improvedgravitational search algorithm and k-NN method.

Y. Lihua, D. Qi, and G. Yanjun, “Study on KNN Text Categorization Algorithm”, *Micro Computer Information*, 21, pp. 269-271, 2006.

DAFTAR RIWAYAT HIDUP

Yang bertanda tangan dibawah ini :

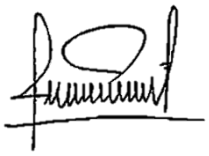
Nama	: Sucitra Sahara
Umur	: 26 Tahun
Tempat/Tgl Lahir	: Purworejo, 12 Mei 1988
Jenis Kelamin	: Wanita
Agama	: Islam
Warga Negara	: Indonesia
Alamat Kost	: Gg. Salak, Pondok Cina Depok, Jawa Barat
E-mail	: Sucitrasahara@gmail.com

Pendidikan Formal :

1. SDN Baledono III, Kab. Puworejo, Jawa Tengah - Lulus Tahun 2001.
2. SLTP Negeri 31 Purworejo, Jawa Tengah - Lulus Tahun 2004.
3. SMK Batik Perbaik Purworejo, Jawa Tengah - Lulus Tahun 2007.
4. D3 Jurusan Manajemen Informatika di Bina Sarana Informatika (BSI) Jakarta Pusat - Lulus Tahun 2010.
5. S1 Jurusan Sistem Informasi di STMIK Nusa Mandiri Jakarta - Lulus Tahun 2012.
6. S2 Jurusan Manajemen Ilmu Komputer di STMIK Nusa Mandiri mulai tahun 2012 sampai sekarang.

Demikianlah Daftar Riwayat Hidup ini saya buat dengan sebenar-benarnya. Atas perhatiannya, saya ucapkan terima kasih.

Hormat Saya,



Sucitra Sahara

STMIK NUSA MANDIRI	**LEMBAR KONSULTASI BIMBINGAN TESIS** **SEKOLAH TINGGI MANAJEMEN INFORMATIKA & KOMPUTER NUSA MANDIRI**

NIM : 14000841
Nama Lengkap : Sucitra Sahara
Dosen Pembimbing : Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd
Judul Tesis : Implementasi Particle Swarm Optimization pada *Analysis Sentiment Review* Appstore *for* Android Menggunakan K-Nearest Neighbors

No	Tanggal Bimbingan	Pokok Bahasan	Paraf dosen Pembimbing
1	14 November 2014	Pengajuan Proposal dan Judul	
2	10 Desember 2014	Bab I dan Bab II	
3	28 Januari 2015	Metode Penelitian	
4	13 Februari 2015	Pembahasan dan Rancangan Aplikasi	
5	20 Februari 2015	Pengajuan Bab IV	
6	27 Februari 2015	ACC Bab V dan Koreksi Keseluruhan	

Catatan:
Bimbingan Tesis
Dimulai pada tanggal : 14 November 2014
Diakhiri pada tanggal : 27 Februari 2015
Jumlah pertemuan bimbingan : 6 kali pertemuan

Disetujui oleh,
Dosen Pembimbing

Dr. Mochamad Wahyudi, MM, M.Kom, M.Pd