

Peningkatan Akurasi Algoritma C4.5 pada Analisis Sentimen menggunakan *Particle Swarm Optimization*

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Azhary Arliansyah

09021181419007

Jurusan Teknik Informatika

FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA

2018

LEMBAR PENGESAHAN TUGAS AKHIR

**PENINGKATAN AKURASI ALGORITMA C4.5 PADA ANALISIS
SENTIMEN MENGGUNAKAN *PARTICLE SWARM
OPTIMIZATION***

Oleh :

AZHARY ARLIANSYAH
NIM : 09021181419007

Indralaya, November 2018

Pembimbing I,

Pembimbing II,



Rifkie Primartha, M.T.
NIP. 197706012009121004



Kanda Januar Miraswan, M.T.
NIPUS. 1671080901900006

Mengetahui,

Ketua Jurusan Teknik Informatika,



Rifkie Primartha, M.T.
NIP. 197706012009121004


TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Kamis tanggal 01 November 2018 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Azhary Arliansyah
N I M : 09021181419007
Judul : Peningkatan Akurasi Algoritma C4.5 pada Analisis Sentimen Menggunakan *Particle Swarm Optimization*


1. Pembimbing I

Rifkie Primartha, M.T.
NIP. 197706012009121004


.....

2. Pembimbing II

Kanda Januar Miraswan, M.T.
NIPUS. 1671080901900006


.....

3. Penguji I

Dian Palupi Rini, M.Kom., Ph.D.
NIP. 197802232006042002


.....

4. Penguji II

Danny Matthew Saputra, M.Sc.
NIP. 198505102015041002


.....

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha S.T. M.T.
NIP 197706012009121004

HALAMAN PERNYATAAN BEBAS PLAGIAT

Yang bertanda tangan di bawah ini :

Nama : Azhary Arliansyah
NIM : 09021181419007
Program Studi : Teknik Informatika
Judul Skripsi : Peningkatan Akurasi Algoritma C4.5 pada Analisis Sentimen Menggunakan *Particle Swarm Optimization*
Hasil Pengecekan Software *iThenticate/Turnitin* : 14 %

Menyatakan bahwa Laporan Tugas Akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Palembang, 30 September 2018



Azhary Arliansyah
NIM. 09021181419007

OPTIMIZING ACCURACY OF C4.5 ALGORITHM FOR SENTIMENT
ANALYSIS WITH PARTICLE SWARM OPTIMIZATION

by:
Azahry Arliansyah
09021181419007


ABSTRACT

Sentiment analysis can be considered as a classification task in natural language processing because it uses classification algorithm to determine the class of the text data. In classification, feature extraction is a process to extract the features from the data so that it can be used as the inputs for the classification algorithm. However, not all features are relevant for classification. Irrelevant features tend to decrease accuracy of the classification algorithm. In order to avoid such problem, the features of the data should go through a feature selection mechanism to select only relevant features. In this research, C4.5 algorithm is used to classify class sentiment of public review of online transportation in Indonesia, and PSO is used as the feature selection mechanism to improve C4.5 accuracy. The result shows that C4.5 achieves 1,69% improvement in accuracy by selecting features using PSO.

Keyword: Sentiment Analysis, Natural Language Processing, C4.5 Algorithm, PSO, Classification, Feature Selection, Online Transportation

Pembimbing I,

Indralaya, November 2018
Pembimbing II,


Rifkie Primartha, M.T.
NIP. 197706012009121004


Kanda Januar Miraswan, M.T.
NIPUS. 1671080901900006

Mengetahui,

Ketua Jurusan Teknik Informatika,


Rifkie Primartha, M.T.
NIP 197706012009121004

PENINGKATAN AKURASI ALGORITMA C4.5 PADA ANALISIS
SENTIMEN MENGGUNAKAN *PARTICLE SWARM OPTIMIZATION*

oleh:
Azhary Arliansyah
09021181419007

ABSTRAK


Analisis sentimen dapat dianggap sebagai suatu proses klasifikasi pada bidang pemrosesan bahasa alami karena pada prosesnya menggunakan algoritma klasifikasi untuk menentukan suatu kelas pada data yang berupa teks. Pada klasifikasi, terdapat suatu proses ekstraksi fitur atau atribut pada data yang keluarannya berupa fitur-fitur data yang akan digunakan sebagai masukan pada algoritma klasifikasi. Namun, tidak semua fitur merupakan fitur yang relevan digunakan untuk klasifikasi. Fitur yang tidak relevan cenderung dapat mengurangi akurasi dari algoritma klasifikasi. Untuk menghindari permasalahan tersebut, fitur-fitur tersebut harus melalui suatu mekanisme seleksi fitur untuk memilih fitur-fitur yang relevan. Pada penelitian ini, algoritma C4.5 digunakan untuk melakukan klasifikasi kelas sentimen pada ulasan masyarakat mengenai jasa transportasi *online* di Indonesia, dan PSO digunakan sebagai mekanisme seleksi fitur untuk meningkatkan akurasi C4.5. Hasil yang didapat menunjukkan peningkatan akurasi C4.5 sebesar 1,69% dengan seleksi fitur menggunakan PSO.

Kata Kunci: Analisis Sentimen, Pemrosesan Bahasa Alami, Algoritma C4.5, PSO, Klasifikasi, Seleksi Fitur, Transportasi *Online*

Pembimbing I,

Indralaya, November 2018
Pembimbing II,


Rifkie Primartha, M.T.
NIP. 197706012009121004


Kanda Januar Miraswan, M.T.
NIPUS. 1671080901900006

Mengetahui,

Ketua Jurusan Teknik Informatika,


Rifkie Primartha, M.T.
NIP 197706012009121004

KATA PENGANTAR



Puji syukur kepada Allah SWT atas berkah dan rahmat-Nya penulis dapat menyelesaikan tugas akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tak terhingga kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi selama proses penelitian ini dilaksanakan. Secara khusus penulis ingin menyampaikan rasa terima kasih kepada:

1. Orang tuaku, Dahlia Yustina dan Syamsudin, saudara kandungku Kak Diaz, Kak Azmi, dan Ayuk Ipar Hesti, serta seluruh keluarga besarku yang selalu mendoakan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Rifkie Primartha, S.T., M.T. selaku Ketua Jurusan Teknik Informatika dan dosen pembimbing I yang telah memberikan kemudahan, arahan, bimbingan, dan saran selama proses pengerjaan tugas akhir ini .
4. Kak Kanda Januar Miraswan, M.T selaku dosen pembimbing 2 yang telah memberikan arahan, bimbingan, bantuan dan mau mendengarkan lalu mengkoreksi segala pendapat penulis selama proses pengerjaan tugas akhir ini.
5. Ibu Dian Palupi Rini, M.Kom Ph,D dan Pak Danny Matthew Saputra, M.Sc selaku dosen penguji I dan II yang telah memberikan masukan, saran dan

motivasi yang tinggi agar penulis dapat memahami dan menyelesaikan tugas akhir ini dengan baik.

6. Bapak Megah Mulya, M. T selaku pembimbing akademik yang telah memberikan kemudahan dalam hal perkuliahan dan pengajuan judul tugas akhir ini.
7. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya yang telah memberikan ilmunya kepada penulis selama proses perkuliahan.
8. Seluruh staf administrasi dan pegawai yang telah membantu dan mendukung Penulis dalam hal administrasi perkuliahan.
9. Ayu Lestari, PudingLab, dan seluruh teman-teman mahasiswa pejuang skripsi yang telah bersama-sama berbagi keluh kesah, nasihat, motivasi dan hiburan di tengah proses pengerjaan tugas akhir ini.
10. Semua pihak yang telah membantu dalam penyelesaian tugas akhir ini dan tidak dapat disebutkan satu persatu.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Indralaya, November 2018

A handwritten signature in black ink, consisting of a large, stylized letter 'A' followed by a horizontal line and a small flourish.

Azhary Arliansyah

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN BEBAS PLAGIAT	iv
ABSTRACT	v
ABSTRAK	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB I PENDAHULUAN	
1.1 Pendahuluan	I-1
1.2 Latar Belakang Masalah	I-1
1.3 Rumusan Masalah	I-4
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR	
2.1 Pendahuluan	II-1
2.2 Analisis Sentimen.....	II-1
2.3 Praproses	II-3
2.3.1 Casefolding	II-3
2.3.2 Tokenizing	II-4
2.3.3 Stopword Removal	II-4
2.3.4 Stemming	II-5
2.4 Term Frequency - Inverse Document Frequency	II-5
2.4.1 Term Frequency	II-6
2.4.2 Inverse Document Frequency	II-6
2.5 Pohon Keputusan.....	II-7
2.6 Algoritma C4.5	II-8
2.7 Particle Swarm Optimization	II-10
2.8 Algoritma C4.5 Berbasis Particle Swarm Optimization (C4.5 - PSO) ...	II-11
2.9 Cross Validation	II-14
2.10 Confusion Matrix	II-15
2.11 Rational Unified Process	II-17
2.12 Penelitian Lain Yang Relevan	II-18

2.12.1 Walaa Medhat, Ahmed Hassan, Hoda Korashy (2014): Sentiment Analysis Algorithms and Applications: A Survey, School of Electronic Engineering, Ain Shams University, Cairo, Egypt.....	II-19
2.12.2 Meng-Chang Tsai, Kun-Huang Chen, Chao-Ton Su, Hung-Chun Lin (2012): An Application of PSO Algorithm and Decision Tree for Medical Problem, Division of Physics, Institute of Nuclear Energy Research, Taoyuan, Taiwan.	II-20
2.12.3 Harvinder Chauhan and Anu Chauhan (2013): Implementation of Decision Tree Algorithm C4.5, Department of Computer Science, Kamla Nehru College For Women, Punjab Phagwara, India.....	II-21
2.12.4 Phu Vo Ngoc, Chau Vo Thi Ngoc, Tran Vo Thi Ngoc, Dat Nguyen Duy (2017): A C4.5 Algorithm for English Emotional Classification, Institute of Research and Development, Duy Tan University, Danang, Vietnam ...	II-22
2.13 Kesimpulan.....	II-22

BAB III METODOLOGI PENELITIAN

3.1 Pendahuluan	III-1
3.2 Unit Penelitian.....	III-1
3.3 Data	III-1
3.3.1 Jenis dan Sumber Data.....	III-1
3.3.2 Metode Pengumpulan Data.....	III-2
3.4 Tahapan Penelitian	III-2
3.4.1 Menetapkan Kerangka Kerja / <i>Framework</i>	III-2
3.4.2 Menetapkan Kriteria Pengujian	III-7
3.4.3 Menetapkan Format Data Pengujian.....	III-8
3.4.4 Menentukan Alat yang Digunakan dalam Pelaksanaan Penelitian..	III-10
3.4.5 Melakukan Pengujian Penelitian	III-12
3.4.6 Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-13
3.5 Metode Pengembangan Perangkat Lunak	III-13
3.5.1 Fase Insepsi.....	III-13
3.5.2 Fase Elaborasi	III-14
3.5.3 Fase Konstruksi.....	III-14
3.5.4 Fase Transisi	III-15
3.6 Manajemen Proyek Penelitian.....	III-15

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1 Pendahuluan	IV-1
4.2 Fase Insepsi	IV-1
4.2.1 Permodelan Bisnis	IV-2
4.2.2 Kebutuhan Sistem	IV-3
4.2.3 Analisis dan Desain	IV-5
4.2.3.2 Desain Perangkat Lunak	IV-24
4.3 Fase Elaborasi.....	IV-36
4.3.1 Permodelan Bisnis	IV-36
4.3.2 Kebutuhan Sistem.....	IV-39
4.3.3 Sequence Diagram	IV-39
4.4 Fase Konstruksi	IV-48
4.4.1 Kebutuhan Sistem.....	IV-48

4.4.2 Diagram Kelas	IV-48
4.4.3 Implementasi	IV-51
4.5 Fase Transisi.....	IV-55
4.5.1 Permodelan Bisnis	IV-56
4.5.2 Kebutuhan Sistem	IV-56
4.5.3 Rencana Pengujian.....	IV-57
4.5.4 Implementasi.....	IV-60
4.6 Kesimpulan.....	IV-69

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Data Hasil Percobaan	V-1
5.2.1 Konfigurasi Percobaan.....	V-1
5.2.2 Data Hasil Konfigurasi I.....	V-3
5.2.3 Data Hasil Konfigurasi II.....	V-6
5.2.4 Data Hasil Konfigurasi III	V-9
5.3 Analisis Hasil Penelitian	V-12
5.3.1 Analisis Peningkatan Akurasi	V-12
5.3.2 Analisis Outliers	V-16
5.3.3 Analisis Performa Klasifikasi	V-18
5.4 Kesimpulan.....	V-21

BAB VI KESIMPULAN DAN SARAN

6.1 Pendahuluan	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran.....	VI-2

DAFTAR PUSTAKA	xv
LAMPIRAN.....	xvii

DAFTAR TABEL

	Halaman
III-1. Rancangan Tabel <i>Confusion Matrix</i> Untuk Setiap Hasil Pengujian ..	III-9
III-2. Rancangan Tabel Hasil Pengujian Klasifikasi Sentimen Dengan Dan Tanpa Proses Seleksi Atribut	III-9
III-3. Rancangan Tabel Hasil Rata-rata Pengujian Perbandingan Klasifikasi Dengan Dan Tanpa Proses Seleksi Atribut	III-10
III-4. Tabel Penjadwalan Penelitian dalam Bentuk <i>Work Breakdown Structure</i> (WBS).....	III-16
IV-1. Kebutuhan Fungsional.....	IV-5
IV-2. Kebutuhan Non-Fungsional	IV-5
IV-3. Hasil <i>Tokenizing</i> Contoh Ulasan	IV-9
IV-4. Hasil <i>Stopword Removal</i> Contoh Ulasan	IV-10
IV-5. Hasil <i>Stemming</i> Contoh Ulasan	IV-10
IV-6. Hasil Pembobotan Kata Contoh Ulasan	IV-11
IV-7. Contoh Data Ulasan Beserta Kelas Sentimennya.....	IV-12
IV-8. Hasil Perhitungan <i>Entropy-based Discretization</i>	IV-15
IV-9. Hasil Perhitungan <i>Entropy-based Discretization</i>	IV-17
IV-10. Daftar Atribut yang Terbentuk dan Notasinya	IV-20
IV-11. Parameter PSO	IV-22
IV-12. Nilai <i>Fitness</i> Tiap Partikel.....	IV-23
IV-13. Definisi Aktor.....	IV-25
IV-14. Definisi <i>Use Case</i>	IV-26
IV-15. Skenario <i>Use Case</i> Klasifikasi dengan C4.5	IV-27
IV-16. Skenario <i>Use Case</i> Memuat Data.....	IV-29
IV-17. Skenario <i>Use Case</i> Praproses Data	IV-30
IV-18. Skenario <i>Use Case</i> Melihat Data.....	IV-31
IV-19. Skenario <i>Use Case</i> Optimasi C4.5 dengan PSO	IV-32
IV-20. Implementasi Kelas	IV-51
IV-21. Pengujian <i>Use Case</i> Klasifikasi dengan C4.5	IV-61
IV-22. Pengujian <i>Use Case</i> Memuat Data.....	IV-63
IV-23. Pengujian <i>Use Case</i> Praproses Data.....	IV-65
IV-24. Pengujian <i>Use Case</i> Melihat Data.....	IV-66
IV-25. Pengujian <i>Use Case</i> Seleksi Atribut Menggunakan <i>Particle Swarm Optimization</i>	IV-67
V-1. Konfigurasi Percobaan I.....	V-2
V-2. Konfigurasi Percobaan II	V-2
V-3. Konfigurasi Percobaan III	V-3
V-4. Hasil Akurasi Percobaan Konfigurasi I.....	V-3
V-5. <i>Precision, Recall, dan F-Measure</i> Percobaan Konfigurasi I.....	V-4
V-6. Hasil Akurasi Percobaan Konfigurasi II	V-6
V-7. <i>Precision, Recall, dan F-Measure</i> Percobaan Konfigurasi II	V-7
V-8. Hasil Akurasi Percobaan Konfigurasi III	V-9
V-9. <i>Precision, Recall, dan F-Measure</i> Percobaan Konfigurasi III.....	V-10

DAFTAR GAMBAR

	Halaman
II-1. Arsitektur RUP	II-15
III-1. Tahapan Pengujian Penelitian	III-12
III-2. Penjadwalan untuk Tahap Menentukan Ruang Lingkup dan Unit Penelitian	III-20
III-3. Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian	III-20
III-4. Penjadwalan untuk Tahap Menentukan Kriteria Pengujian	III-21
III-5. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Insepsi	III-21
III-6. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Elaborasi	III-22
III-7. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Konstruksi	III-22
III-8. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian pada Fase Transisi	III-23
III-9. Penjadwalan untuk Tahap Melakukan Pengujian Penelitian	III-23
III-10. Penjadwalan untuk Tahap Melakukan Analisa Hasil Pengujian dan Membuat Kesimpulan	III-24
IV-1. <i>Root Node</i> dan Percabangan yang Terbentuk	IV-16
IV-2. Pohon Keputusan yang Terbentuk dari C4.5	IV-18
IV-3. Diagram <i>Use Case</i> Perangkat Lunak	IV-25
IV-4. Diagram Aktivitas <i>Use Case</i> Klasifikasi dengan C4.5	IV-33
IV-5. Diagram Aktivitas <i>Use Case</i> Memuat Data	IV-34
IV-6. Diagram Aktivitas <i>Use Case</i> Praproses Data	IV-34
IV-7. Diagram Aktivitas <i>Use Case</i> Melihat Data	IV-35
IV-8. Diagram Aktivitas <i>Use Case</i> Optimasi C4.5 dengan PSO	IV-35
IV-9. Rancangan Antarmuka Menu Praproses	IV-37
IV-10. Rancangan Antarmuka Menu Pelatihan - Pengujian	IV-38
IV-11. Diagram Sekuen Klasifikasi dengan C4.5	IV-41
IV-12. Diagram Sekuen Memuat Data	IV-42
IV-13. Diagram Sekuen Praproses Data	IV-43
IV-14. Diagram Sekuen Melatih Algoritma C4.5	IV-44
IV-15. Diagram Sekuen Pruning C4.5	IV-45
IV-16. Diagram Sekuen Melihat Data	IV-46
IV-17. Diagram Sekuen Optimasi C4.5 dengan PSO	IV-47
IV-18. Diagram Kelas Perangkat Lunak	IV-50
IV-19. Implementasi Antarmuka Menu Praproses	IV-54
IV-20. Implementasi Antarmuka Menu Pelatihan - Pengujian	IV-55
IV-21. Rencana Pengujian <i>Use Case</i> Klasifikasi dengan C4.5	IV-57
IV-22. Rencana Pengujian <i>Use Case</i> Memuat Data	IV-58
IV-23. Rencana Pengujian <i>Use Case</i> Praproses Data	IV-58
IV-24. Rencana Pengujian <i>Use Case</i> Melihat Data	IV-59
IV-25. Rencana Pengujian <i>Use Case</i> Optimasi C4.5 dengan PSO	IV-59

V-1.	Perbandingan Akurasi C4.5 dan PSO - C4.5 dengan Konfigurasi I...	V-13
V-2.	Perbandingan Akurasi C4.5 dan PSO - C4.5 dengan Konfigurasi II.	V-13
V-3.	Perbandingan Akurasi C4.5 dan PSO - C4.5 dengan Konfigurasi III	V-14
V-4.	Perbandingan Akurasi, <i>Precision</i> , <i>Recall</i> , dan <i>F-Measure</i> C4.5 dan PSO - C4.5 dengan Konfigurasi I.....	V-15
V-5.	Perbandingan Akurasi, <i>Precision</i> , <i>Recall</i> , dan <i>F-Measure</i> C4.5 dan PSO - C4.5 dengan Konfigurasi II.....	V-15
V-6.	Perbandingan Akurasi, <i>Precision</i> , <i>Recall</i> , dan <i>F-Measure</i> C4.5 dan PSO - C4.5 dengan Konfigurasi III	V-16
V-7.	Visualisasi Sebaran Data Dua Dimensi.....	V-17
V-8.	Visualisasi Sebaran Data Tiga Dimensi	V-17
V-9.	<i>F-Measure</i> PSO-C4.5 Tiap Label pada Konfigurasi I	V-19
V-10.	<i>F-Measure</i> PSO-C4.5 Tiap Label pada Konfigurasi II	V-20
V-11.	<i>F-Measure</i> PSO-C4.5 Tiap Label pada Konfigurasi III.....	V-20

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, serta batasan masalah. Bab ini akan memberikan penjelasan umum mengenai keseluruhan penelitian.

Pendahuluan dimulai dengan penjelasan mengenai bisnis angkutan *online* beserta kaitannya terhadap sentimen masyarakat mengenai bisnis tersebut. Serta penelitian yang berkaitan dengan peningkatan akurasi dari proses klasifikasi sentimen yang menjadi latar belakang penelitian ini.

1.2 Latar Belakang Masalah

Persaingan bisnis antara angkutan *online* dan angkutan konvensional menjadi suatu topik yang pro dan kontra. Angkutan *online* menjadi solusi bagi sebagian masyarakat karena memberikan kemudahan dalam pencarian jasa angkutan melalui aplikasi dan juga dapat memastikan tarif penggunaannya melalui aplikasi tersebut. Namun disisi lain angkutan *online* dapat menurunkan pendapatan angkutan konvensional. Dalam Undang-undang Nomor 22 Tahun 2009 tentang Lalu Lintas dan Angkutan Jalan tertera, yang termasuk angkutan

umum adalah kendaraan yang beroda tiga atau lebih. Sedangkan untuk sepeda motor merupakan kendaraan beroda dua, sehingga tidak termasuk dalam kategori angkutan umum berdasarkan regulasi yang berlaku di Indonesia. Berdasarkan topik permasalahan tersebut, maka perlu dilakukan analisis apakah sebagian besar sentimen masyarakat mendukung adanya jasa angkutan *online* tersebut.

Klasifikasi sentimen merupakan bidang komputasi yang mempelajari opini, sikap, atau emosi dari suatu entitas. Suatu entitas bisa saja berupa individu, topik, maupun kejadian. Topik tersebut biasanya akan menjadi suatu ulasan atau *review* (Medhat et. al., 2014). Dalam analisis sentimen biasanya menggunakan metode klasifikasi pada kumpulan entitas yang dibagi menjadi kelas-kelas sentimen tertentu, misalnya positif, negatif, atau netral. Salah satu metode klasifikasi yang dapat digunakan pada analisis sentimen adalah pohon keputusan (*decision tree*) yang dibentuk menggunakan algoritma C4.5.

Pohon keputusan merupakan salah satu metode klasifikasi yang menggunakan struktur pohon dengan membagi-bagi data menjadi cabang pada pohon berdasarkan atribut pada *dataset* sehingga diperoleh label atau kelas (Quinlan, J. R., 1986). Pohon keputusan tersusun atas *node* yang pada tiap *node*-nya merepresentasikan suatu pertanyaan atau label berdasarkan percabangan dari nilai atribut. Proses pembentukan pohon keputusan yang efektif dapat menggunakan algoritma C4.5.

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan berdasarkan nilai *information gain* dan *entropy* pada *dataset*.

Nilai-nilai tersebut dapat digunakan untuk menentukan *root node* dan percabangan suatu *node* pada pohon keputusan (Han et. al., 2012). Algoritma C4.5 juga dapat menghindari *overfitting* pada pohon keputusan, yaitu suatu kondisi dimana pohon keputusan tersebut sulit untuk menggeneralisasi data. *Overfitting* juga dapat diartikan suatu kondisi dimana suatu metode klasifikasi hanya “mengingat” data dari pada “mempelajari” data.

Permasalahan pada pohon keputusan akan timbul ketika digunakan pada *dataset* dengan dimensi yang besar. Tidak seluruh atribut pada data merupakan atribut yang relevan pada klasifikasi. Atribut yang tidak relevan tersebut akan mengurangi akurasi dari pohon keputusan sehingga perlu proses seleksi atribut (*attribute selection*) pada data (Jensen, R., 2005). Analisis sentimen melibatkan proses klasifikasi pada data berupa teks dimana data yang berupa teks tersebut akan membuat dimensi atribut yang sangat besar. Untuk itu diperlukan proses seleksi atribut menggunakan metode optimasi heuristik, salah satunya adalah *Particle Swarm Optimization*(PSO).

Particle Swarm Optimization adalah metode optimasi heuristik global yang diperkenalkan oleh Eberhart dan Kennedy (1995) yang terinspirasi dari perilaku kawanan burung dan ikan. PSO merupakan suatu metode komputasi yang mengoptimasi suatu masalah dengan meningkatkan kandidat solusi yang dilakukan secara iteratif. Pada penelitian ini, PSO akan digunakan untuk melakukan seleksi atribut untuk meningkatkan akurasi algoritma C4.5 pada analisis sentimen.

Berdasarkan uraian di atas, pada penelitian ini akan dilakukan analisis sentimen mengenai angkutan *online* menggunakan algoritma C4.5 yang diharapkan dapat digunakan untuk melakukan klasifikasi kelas-kelas sentimen, dan menggunakan PSO yang diharapkan dapat digunakan untuk meningkatkan akurasi dari algoritma C4.5 pada analisis sentimen tersebut.

1.3 Rumusan Masalah

Masalah yang dirumuskan pada penelitian ini adalah apakah *Particle Swarm Optimization* dapat meningkatkan akurasi algoritma C4.5 pada analisis sentimen mengenai angkutan *online*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut

1. Menerapkan *Particle Swarm Optimization* dengan algoritma C4.5 pada analisis sentimen mengenai angkutan *online*
2. Mengukur akurasi dari algoritma C4.5 sebelum dan sesudah dioptimasi dengan *Particle Swarm Optimization*

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut :

1. Peneliti dapat menggunakan algoritma C4.5 dengan akurasi yang lebih tinggi yang dioptimasi menggunakan *Particle Swarm Optimization*
2. Dapat mengetahui sentimen masyarakat dengan lebih akurat mengenai jasa angkutan *online*
3. Dapat digunakan sebagai landasan bagi pembuat kebijakan mengenai jasa angkutan *online*

1.6 Batasan Masalah

Batasan masalah penelitian ini adalah sebagai berikut :

1. Data yang digunakan berupa teks
2. Data yang digunakan adalah teks dalam Bahasa Indonesia
3. Data diambil dari platform aplikasi *Google Play* dan survei secara *online*
4. Data yang dianalisis adalah opini mengenai angkutan *online*

1.7 Sistematika Penulisan

Sistematika penulisan pada penelitian ini adalah sebagai berikut :

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang masalah, perumusan masalah, tujuan, dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi-definisi analisis sentimen, praproses, TF-IDF, pohon keputusan, algoritma C4.5, *particle swarm optimization*, *cross validation*, *confusion matrix*, dan *rational unified process*. Pada akhir bab akan disertakan penelitian-penelitian lain yang relevan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini diuraikan mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Pada akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai perancangan dan implementasi perangkat lunak dengan metode pemrograman berorientasi objek berdasarkan panduan *Rational Unified Process* yang di dalamnya terdapat fase inepsi, elaborasi, konstruksi, dan transisi.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan dibahas mengenai hasil klasifikasi algoritma C4.5 dan hasil optimasinya menggunakan *Particle*

Swarm Optimization. Pada akhir bab ini berisi analisis dari hasil yang telah didapatkan.

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini akan menjelaskan kesimpulan dan saran berdasarkan hasil analisis dalam meningkatkan akurasi algoritma C4.5 dengan *Particle Swarm Optimization* pada analisis sentimen mengenai angkutan *online*.

1.8 Kesimpulan

Berdasarkan uraian di atas, pada penelitian ini akan dilakukan peningkatan akurasi algoritma C4.5 pada analisis sentimen mengenai angkutan *online* menggunakan *particle swarm optimization* dengan batasan masalah yang telah ditentukan.

DAFTAR PUSTAKA

- Aggarwal, C. C., and Zhai, C. (2012). "Mining Text Data". Springer Science & Business Media.
- Beer, J., Langer, S., and Gipp, B. (2017). "TF-IDuF: A Novel Term-Weighting Scheme for User Modeling based on Users' Personal Document Collections". 12th iConference.
- Ben-Gal, I. (2005). "Outlier Detection". Kluwer Academic Publishers
- Chauhan, H., and Chauhan, A. (2013). "Implementation of Decision Tree Algorithm C4.5". International Journal of Scientific and Research Publications.
- Eberhart, R., and Kennedy, J., (1995). "A New Optimizer Using Particle Swarm Theory". The Micro Machine and Human Science.
- Grinstein, G., Trutschi, M., and Cvek, U. (2001). "High-Dimensional Visualizations". Institute for Visualization and Perception Research.
- Han, J., Kamber, M., and J. Pei. (2012). "Data Mining Concepts and Techniques (3rd Edition).".
- Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment". Computing in Science & Engineering: 90-95.
- Kennedy, J., and Eberhart, R. (1997). "A Discrete Binary Version of the Particle Swarm Algorithm". IEEE International Conference on Systems, Man, and Cybernetics.
- Kohavi, R., and Provost, F. (1998). "On Applied Research in Machine Learning". Columbia University.
- Kohavi, R., and Sahami, M. (1996). "Error-based and Entropy-based Discretization of Continuous Features". Second International Conference on Knowledge Discovery and Data Mining: 114-119.
- Kotsiantis, S., Kanellopoulos, D., and Pintelas, P. (2006). "Data Preprocessing for Supervised Learning". International Journal of Computer Science: 111-117.
- Kruchten, P. (2002). "Tutorial: Introduction to the Rational Unified Process/sup /spl reg/". 24rd International Conference on Software Engineering (ICSE).
- Kusrini, dan Luthfi, T. E. (2009). "Algoritma Data Mining". Yogyakarta: Andi Publisher.
- Medhat, W., Hassan, A., and Korashy, H. (2014). "Sentiment Analysis Algorithms and Applications: A Survey". Ain Shams Engineering Journal: 1093-1113.

- Ngoc, P. V., Ngoc, C. V. T, Ngoc, T. V. T, and Duy, D. N. A. (2017). “A C4.5 Algorithm for English Emotional Classification”. Institute of Research and Development Duy Tan University.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research* 12: 2825-2830
- Quinlan, J. R. (1986). “Induction of Decision Trees”. *Machine Learning*: 81-106.
- Quinlan, J. R. (1993). “C4.5: Programs for Machine Learning”. San Francisco: Morgan Kaufmann Publisher.
- Tsai, M., -C., Chen, K., Su, C., and Lin, H. (2012). “An Application of PSO Algorithm and Decision Tree for Medical Problem”. *Second International Conference on Intelligent Computational System (ICS)*.
- Wong, T. -T., Yang, N. -Y. (2017). “Dependency Analysis of Accuracy Estimates in k-Fold Cross Validation”. *IEEE Transaction on Knowledge and Data Engineering*: 2417-2427.
- Xue, B., Zhang, M., and Browne, W. N. (2013). “Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach”. *IEEE Transactions on Cybernetics*: 1656-1671.