



UNIVERSITAS
Dinamika

**ANALISIS SENTIMEN *REVIEW* WISATAWAN TERHADAP DESTINASI
WISATA DI INDONESIA MENGGUNAKAN ALGORITMA *NAIVE*
BAYES CLASSIFIER BERBASIS *PARTICLE SWARM OPTIMIZATION***



TUGAS AKHIR

**Program Studi
S1 Sistem Informasi**

UNIVERSITAS
Dinamika

Oleh :

Shofa Wardatul Jannah

17410100115

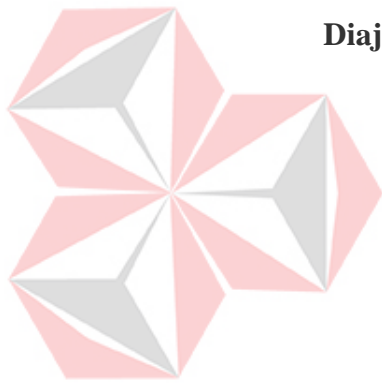
FAKULTAS TEKNOLOGI DAN INFORMATIKA

UNIVERSITAS DINAMIKA

2021

**ANALISIS SENTIMEN *REVIEW* WISATAWAN TERHADAP DESTINASI
WISATA DI INDONESIA MENGGUNAKAN ALGORITMA *NAIVE*
BAYES CLASSIFIER BERBASIS *PARTICLE SWARM OPTIMIZATION***

TUGAS AKHIR



**Diajukan sebagian salah satu syarat untuk menyelesaikan
Program Sarjana Komputer**

**UNIVERSITAS
Dinamika**

Oleh:

**Nama : Shofa Wardatul Jannah
NIM : 17410100115
Program : S1 Sistem Informasi**

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA**

2021

Tugas Akhir

ANALISIS SENTIMEN *REVIEW* WISATAWAN TERHADAP DESTINASI WISATA DI INDONESIA MENGGUNAKAN ALGORITMA *NAIVE* *BAYES CLASSIFIER* BERBASIS *PARTICLE SWARM OPTIMIZATION*

Dipersiapkan dan disusun oleh

Shofa Wardatul Jannah

NIM : 17410100115

Telah diperiksa, dibahas dan disetujui oleh Dewan Pembahas

Pada : 15 Januari 2021

Susunan Dewan Pembahas

Pembimbing:

I. Vivine Nurcahyawati, M.Kom.

NIDN: 0723018101


II. Julianto Lemantara, S.Kom., M.Eng.


NIDN: 0722108601


Pembahas:

Dr. Jusak

NIDN: 0708017101


Digitally signed by Universitas
Dinamika
Date: 2021.02.10
11:34:08 +07'00'


Digitally signed by Julianto Lemantara
DN: cn=Julianto Lemantara,
ou=Universitas Dinamika, ou=S1 Sistem
Informasi,
email=julianto@dinamika.ac.id, c=ID
Date: 2021.02.10 12:16:11 +07'00'


Digitally signed by
Universitas
Dinamika
Date: 2021.02.10
12:33:34 +07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana


Digitally signed by
Universitas Dinamika
Date: 2021.02.17
11:43:40 +07'00'

Dr. Jusak

NIDN: 0708017101

Dekan Fakultas Teknologi dan Informatika
UNIVERSITAS DINAMIKA

SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KEASLIAN KARYA ILMIAH

Sebagai mahasiswa Universitas Dinamika, saya :

Nama : Shofa Wardatul Jannah.
Nim : 17410100115
Program Studi : SI Sistem Informasi
Fakultas : Fakultas Teknologi dan Informatika
Jenis Karya : Tugas Akhir
Judul Karya : **ANALISIS SENTIMEN REVIEW WISATAWAN
TERHADAP DESTINASI WISATA DI INDONESIA
MENGUNAKAN ALGORITMA NAIVE BAYES
CLASSIFIER BERBASIS PARTICLE SWARM
OPTIMIZATION**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, saya menyetujui memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalti Free Right*) atas seluruh isi/ sebagian karya ilmiah saya tersebut di atas untuk disimpan, diahlmediakan dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta
2. Karya tersebut di atas adalah karya asli saya, bukan plagiat baik sebagian maupun keseluruhan, Kutipan karya atau pendapat orang lain yang ada dalam karya ilmiah ini adalah semata hanya rujukan yang dicantumkan dalam Daftar Pustaka saya
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiat pada karya ilmiah ini, maka saya bersedia untuk menerima pencabut terhadap gelar kerjasama yang telah diberikan kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Surabaya, 15 Januari 2021

Yang menyatakan



NIM : 17410100115

Just do it, Get it done!



UNIVERSITAS
Dinamika

Ku persembahkan, karya ini untuk orang-orang yang kucintai:

Orangtuaku

dan

Diriku sendiri.



UNIVERSITAS
Dinamika

ABSTRAK

Berdasarkan hasil survey dari CultureTrip, didapatkan 7 *The Most Beautiful Places To Visit in Indonesia* dengan 7 peringkat pertama yaitu *Dieng Plateau, Borobudur Temple, Komodo National Park, Raja Ampat, Mount Bromo, Ubud, dan Lake Toba*. Review dari para wisatawan yang telah berkunjung ke objek wisata merupakan hal yang terpenting karena ulasan tersebut dapat dijadikan informasi mengenai objek wisata dan banyak wisatawan menghabiskan waktu berjam-jam untuk mencari informasi tersebut di internet. Penelitian ini menggunakan metode *naïve bayes classifier* yaitu pencarian dengan menghitung peluang berdasarkan frekuensi pengalaman sebelumnya dan *naïve bayes – particle swarm optimization* dengan menggunakan dua pencarian yaitu pengalaman pribadinya (*local best*) dan pengalaman sekitar (*global best*). Maka penelitian ini bertujuan untuk mengetahui sentimen review wisatawan dan membandingkan akurasi *naïve bayes* dan *naïve bayes* berbasis PSO menggunakan RapidMiner. Tahap evaluasi menggunakan *confusion matrix* dan validasi menggunakan *10-fold cross-validation*. Penerapan *particle swarm optimization* pada *naïve bayes* dapat meningkatkan nilai *accuracy* sebesar 16.23%, nilai *precision* atau PPV sebesar 2.58%, nilai *sensitivity* sebesar 22.73%, dan nilai *false negative rate* sebesar 19.23%. Namun penggunaan PSO juga dapat menurunkan hasil nilai dari *specificity* sebesar 19.30%, *negative prediction value* sebesar 3.20%, dan *false negative rate* sebesar 22.64%. Waktu eksekusi dari *naïve bayes* di semua destinasi adalah 14.6 ms, sedangkan dari *naïve bayes - PSO* rata-ratanya 513.2ms.

Kata kunci: review wisatawan, analisis sentimen, *naïve bayes classifier*, *particle swarm optimization*.

KATA PENGANTAR

Segala puji bagi Tuhan yang Maha Esa, Allah SWT atas segala rahmat, karunia dan hidayah-Nya, sehingga dapat menyelesaikan Laporan Tugas Akhir dengan judul “Analisis Sentimen Pada Review Wisatawan Terhadap Destinasi Wisata Di Indonesia Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization” dengan baik dan lancar meskipun penulis sadari bahwa masih ada banyak kekurangan yang ada didalamnya.

Dalam penyelesaian Tugas Akhir ini, penulis ingin memberikan apresiasi atas bantuannya baik berupa saran, kritik, nasihat maupun dukungan kepada penulis. Oleh karena itu, penulis ingin mengucapkan banyak terima kasih kepada:

1. Tuhan yang Maha Esa, Allah SWT karena selalu mendengarkan doa yang penulis selalu panjatkan supaya diperlancar untuk pengerjaan tugas akhir ini.
2. Orang tua dan keluarga penulis yang selalu mendoakan dan memberi dukungan kepada penulis.
3. Vivine Nurcahyawati, M.Kom., OCP selaku dosen pembimbing pertama yang selalu sabar dalam memberikan bimbingan dan mendukung dalam menyelesaikan Tugas Akhir.
4. Julianto Lemantara, S.Kom., M.Eng selaku dosen pembimbing kedua yang memberi yang selalu sabar dalam memberikan bimbingan dan mendukung dalam menyelesaikan Tugas Akhir.
5. Dr. Jusak selaku dosen pembahas yang telah menyempurnakan laporan tugas akhir ini.
6. dan semua pihak yang tidak dapat disebutkan satu-persatu, terima kasih banyak.

Penulis menyadari bahwa laporan tugas akhir ini jauh dari kesempurnaan, oleh sebab itu penulis ingin meminta maaf jika terdapat banyak kesalahan dalam penulisan laporan. Penulis juga berharap akan masukan, kritik, dan saran dari para pembaca untuk menyempurnakan laporan tugas akhir ini.

Akhir kata, semoga isi laporan ini dapat bermanfaat bagi penulis pada khususnya dan para pembaca pada umumnya.

Surabaya, 15 Januari 2021



Penulis



UNIVERSITAS
Dinamika

DAFTAR ISI

Halaman

ABSTRAK	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xvi
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	4
1.4 Tujuan	5
1.5 Manfaat	5
BAB II LANDASAN TEORI	6
2.1 Penelitian Terdahulu	6
2.2 Pariwisata	7
2.3 Objek Wisata.....	7
2.4 Analisis Sentimen	7
2.5 <i>Web Crawler</i>	8
2.6 <i>Text Mining</i>	8
2.6.1 Text Preprocessing.....	8
2.7 <i>Naïve Bayes</i>	9
2.8 <i>Particle Swarm Optimization</i>	10
2.9 <i>Confusion Matrix</i>	11
2.10 RapidMiner.....	13
2.11 Visualisasi	13
BAB III METODE PENELITIAN	14
3.1 Tahapan Awal Penelitian	15
3.2 Tahap Pengembangan Penelitian	19
3.2.1 Analisis	20
3.2.2 Penambangan Data.....	21

3.2.3 Pelabelan	22
3.2.4 Data Preprocessing	22
3.2.5 <i>Training data</i> dan <i>Testing data</i>	22
3.2.6 <i>Naive Bayes Classifier</i>	22
3.2.7 <i>Particle Swarm Optimization</i>	23
3.2.8 <i>Apply Model</i>	23
3.2.9 Evaluasi dan Validasi Data	23
3.2.10 Visualisasi	24
3.3 Tahap Akhir Penelitian	25
3.3.1 Kesimpulan	25
3.3.2 Saran	25

BAB IV HASIL DAN PEMBAHASAN 26

4.1 Analisis	26
4.2 Penambahan Data	26
4.3 Pelabelan	26
4.4 <i>Data Preprocessing</i>	27
4.1.1 Frekuensi Review <i>Dieng Plateau</i>	28
4.1.2 Frekuensi Review <i>Borobudur Temple</i>	29
4.1.3 Frekuensi Review <i>Komodo National Park</i>	29
4.1.4 Frekuensi Review <i>Raja Ampat</i>	29
4.1.5 Frekuensi Review <i>Mount Bromo</i>	29
4.1.6 Frekuensi Review <i>Lake Toba</i>	30
4.1.7 Frekuensi Review <i>Ubud</i>	30
4.5 <i>Training Data</i> dan <i>Testing Data</i>	30
4.6 Klasifikasi <i>Naïve Bayes</i>	30
4.7 Evaluasi dan Validasi <i>Naïve Bayes</i>	32
4.7.1 Evaluasi dan Validasi NB <i>Dieng Plateau</i>	32
4.7.2 Evaluasi dan Validasi NB <i>Borobudur Temple</i>	32
4.7.3 Evaluasi dan Validasi NB <i>Komodo National Park</i>	32
4.7.4 Evaluasi dan Validasi NB <i>Raja Ampat</i>	33
4.7.5 Evaluasi dan Validasi NB <i>Mount Bromo</i>	33
4.7.6 Evaluasi dan Validasi NB <i>Lake Toba</i>	33
4.7.7 Evaluasi dan Validasi NB <i>Ubud</i>	34
4.8 Klasifikasi <i>Naïve Bayes</i> – PSO	36

4.9 Evaluasi dan Validasi Naïve Bayes – PSO	37
4.9.1 Evaluasi dan Validasi NB – PSO <i>Dieng Plateau</i>	37
4.9.2 Evaluasi dan Validasi NB – PSO <i>Borobudur Temple</i>	37
4.9.3 Evaluasi dan Validasi NB – PSO <i>Komodo National Park</i>	38
4.9.4 Evaluasi dan Validasi NB – PSO <i>Raja Ampat</i>	38
4.9.5 Evaluasi dan Validasi NB – PSO <i>Mount Bromo</i>	38
4.9.6 Evaluasi dan Validasi NB – PSO <i>Lake Toba</i>	38
4.9.7 Evaluasi dan Validasi NB – PSO <i>Ubud</i>	39
4.10 Hasil Akhir dari <i>Naïve Bayes</i> dan <i>Naïve Bayes - PSO</i>	41
4.11 Visualisasi	43
4.11.1 Visualisasi Data <i>Dieng Plateau</i>	43
4.11.2 Visualisasi Data <i>Borobudur Temple</i>	44
4.11.3 Visualisasi Data <i>Komodo National Park</i>	44
4.11.4 Visualisasi Data <i>Raja Ampat</i>	45
4.11.5 Visualisasi Data <i>Mount Bromo</i>	45
4.11.6 Visualisasi Data <i>Lake Toba</i>	46
4.11.7 Visualisasi Data <i>Ubud</i>	46
BAB V PENUTUP	48
5.1 Kesimpulan	48
5.2 Saran	49
DAFTAR PUSTAKA	50
LAMPIRAN	52

DAFTAR TABEL

Halaman

Tabel 4. 2 Hasil Tahap <i>Case Folding</i>	27
Tabel 4. 3 Hasil Tahap <i>Tokenizing</i>	27
Tabel 4. 4 Hasil Tahap <i>Filtering</i>	28
Tabel 4. 5 Hasil Tahap <i>Stemming</i>	28
Tabel 4. 6 Nilai Akurasi NB pada 7 Destinasi Bahasa Indonesia.....	35
Tabel 4. 7 Nilai Akurasi NB pada 7 Destinasi Bahasa Inggris	35
Tabel 4. 8 Nilai Akurasi NB - PSOpada 7 Destinasi Bahasa Indonesia	40
Tabel 4. 9 Nilai Akurasi NB - PSOpada 7 Destinasi Bahasa Inggris	40
Tabel 4. 10 Hasil Komparasi NB dan NB – PSO Bahasa Indonesia	41
Tabel 4. 11 Hasil Komparasi NB dan NB – PSO Bahasa Inggris	41
Tabel L1. 1 Pengelompokan Data.....	52
Tabel L1. 2 Kebutuhan Pengguna.....	52
Tabel L3. 1 Contoh Hasil <i>Processing Review</i>	54
Tabel L3. 2 Contoh Frekuensi Kata Uji Pada <i>Training data</i>	55
Tabel L3. 3 Contoh Hasil Perhitungan Probabilitas Kata.....	55
Tabel L3. 4 Hasil Akhir Klasifikasi	56
Tabel L4. 1 Jumlah Data	56
Tabel L7. 1 Klasifikasi Naïve Bayes	59
Tabel L8. 1 Hasil pBest dan gBest.....	62
Tabel L8. 2 Nilai Update Kecepatan.....	63
Tabel L12. 1 Probabilitas NB <i>Dieng Pleteau</i> Bahasa Indonesia	74
Tabel L12. 2 Probabilitas NB <i>Dieng Plateau</i> Bahasa Inggris	74
Tabel L12. 3 Probabilitas NB <i>Borobudur Temple</i> Bahasa Indonesia.....	75
Tabel L12. 4 Probabilitas NB <i>Borobudur Temple</i> Bahasa Inggris	75
Tabel L12. 5 Probabilitas NB <i>Komodo National Park</i> Bahasa Indonesia.....	76
Tabel L12. 6 Probabilitas NB <i>Komodo National Park</i> Bahasa Inggris	77
Tabel L12. 7 Probabilitas NB <i>Raja Ampat</i> Bahasa Indonesia	77
Tabel L12. 8 Probabilitas NB <i>Raja Ampat</i> Bahasa Inggris	78
Tabel L12. 9 Probabilitas NB <i>Mount Bromo</i> Bahasa Indonesia.....	78

Tabel L12. 10 Probabilitas NB <i>Mount Bromo</i> Bahasa Inggris	79
Tabel L12. 11 Probabilitas NB <i>Lake Toba</i> Bahasa Indonesia	79
Tabel L12. 12 Probabilitas NB <i>Lake Toba</i> Bahasa Inggris.....	80
Tabel L12. 13 Probabilitas NB <i>Ubud</i> Bahasa Indonesia	80
Tabel L12. 14 Probabilitas NB <i>Ubud</i> Bahasa Inggris.....	81
Tabel L12. 15 Probabilitas NB – PSO <i>Dieng Plateau</i> Bahasa Indonesia	82
Tabel L12. 16 Probabilitas NB – PSO <i>Dieng Plateau</i> Bahasa Inggris	82
Tabel L12. 17 Probabilitas NB – PSO <i>Borobudur Temple</i> Bahasa Indonesia.....	83
Tabel L12. 18 Probabilitas NB – PSO <i>Borobudur Temple</i> Bahasa Inggris.....	83
Tabel L12. 19 Probabilitas NB – PSO <i>Komodo National Park</i> Bahasa Indonesia	84
Tabel L12. 20 Probabilitas NB – PSO <i>Komodo National Park</i> Bahasa Inggris...	84
Tabel L12. 21 Probabilitas NB – PSO <i>Raja Ampat</i> Bahasa Indonesia	85
Tabel L12. 22 Probabilitas NB – PSO <i>Raja Ampat</i> Bahasa Inggris	85
Tabel L12. 23 Probabilitas NB – PSO <i>Mount Bromo</i> Bahasa Indonesia.....	86
Tabel L12. 24 Probabilitas NB – PSO <i>Mount Bromo</i> Bahasa Inggris.....	86
Tabel L12. 25 Probabilitas NB – PSO <i>Lake Toba</i> Bahasa Indonesia.....	87
Tabel L12. 26 Probabilitas NB – PSO <i>Lake Toba</i> Bahasa Inggris	87
Tabel L12. 27 Probabilitas NB – PSO <i>Ubud</i> Bahasa Indonesia	88
Tabel L12. 28 Probabilitas NB – PSO <i>Ubud</i> Bahasa Inggris	89

DAFTAR GAMBAR

Halaman

Gambar 3. 1 Metode Penelitian	14
Gambar 3. 2 Tahap Awal	15
Gambar 3. 3 Rating Dan Review Wisata Dari Google Maps Reviews.....	16
Gambar 3. 4 Proses Pengumpulan Dan Preprocessing Data.....	17
Gambar 3. 5 Proses Naïve Bayes dan Naïve Bayes + PSO	18
Gambar 3. 6 Tahap Pengembangan	19
Gambar 3. 7 Tahap Akhir	25
Gambar L4. 1 Penambahan Data Menggunakan WebHarvy	56
Gambar L4. 2 Hasil Penambahan Data Menggunakan Webharvy	57
Gambar L4. 3 Hasil Penambahan Data Menggunakan Pemrograman Python ...	57
Gambar L5. 1 Hasil Pelabelan Kelas	57
Gambar L6. 1 <i>Data Preprocessing</i>	58
Gambar L6. 2 Proses <i>Data Preprocessing</i> Bahasa Inggris	58
Gambar L6. 3 Proses <i>Data Preprocessing</i> Bahasa Indonesia	58
Gambar L6. 4 Hasil Tahap <i>Preprocessing</i>	59
Gambar L7. 1 Konfigurasi Model <i>Naïve Bayes</i>	60
Gambar L8. 1 Konfigurasi Model <i>Naïve Bayes + PSO</i>	64
Gambar L9. 1 Desain Pengujian Dengan <i>Naïve Bayes</i>	65
Gambar L9. 2 Desain Pengujian Dengan <i>Naïve Bayes + PSO</i>	65
Gambar L10. 1 Visualisasi Dengan <i>Wordcloud</i>	66
Gambar L10. 2 Visualisasi Dengan <i>Bar Chart</i>	66
Gambar L10. 3 Visualisasi Dengan <i>Pie Chart</i>	67
Gambar L11. 1 Frekuensi Kemunculan Kata <i>Deing Plateau</i>	67
Gambar L11. 2 Frekuensi Kemunculan Kata <i>Borobudur Temple</i>	68
Gambar L11. 3 Frekuensi Kemunculan Kata <i>Komodo National Park</i>	69
Gambar L11. 4 Frekuensi Kemunculan Kata <i>Raja Ampat</i>	70
Gambar L11. 5 Frekuensi Kemunculan Kata <i>Mount Bromo</i>	71
Gambar L11. 6 Frekuensi Kemunculan Kata <i>Lake Toba</i>	72
Gambar L11. 7 Frekuensi Kemunculan Kata <i>Ubud</i>	73

Gambar L13. 1 <i>Wordcloud</i> pada <i>Dieng Plateau</i>	89
Gambar L13. 2 Jumlah Data Review Dari 7 Wisata.....	90
Gambar L13. 3 Emosi pada <i>Dieng Plateau</i>	90
Gambar L13. 4 Persentase Sentimen pada <i>Dieng Plateau</i>	91
Gambar L13. 5 Akun pada <i>Dieng Plateau</i>	91
Gambar L13. 6 <i>Wordcloud</i> pada <i>Borobudur Temple</i>	92
Gambar L13. 7 Emosi pada <i>Borobudur Temple</i>	92
Gambar L13. 8 Persentase Sentimen pada <i>Borobudur Temple</i>	93
Gambar L13. 9 Akun pada <i>Borobudur Temple</i>	93
Gambar L13. 10 <i>Wordcloud</i> pada <i>Komodo National Park</i>	94
Gambar L13. 11 Emosi pada <i>Komodo National Park</i>	94
Gambar L13. 12 Persentase Sentimen pada <i>Komodo National Park</i>	95
Gambar L13. 13 Akun pada <i>Komodo National Park</i>	95
Gambar L13. 14 <i>Wordcloud</i> pada <i>Raja Ampat</i>	96
Gambar L13. 15 Emosi pada <i>Raja Ampat</i>	96
Gambar L13. 16 Persentase Sentimen pada <i>Raja Ampat</i>	96
Gambar L13. 17 Akun pada <i>Raja Ampat</i>	97
Gambar L13. 18 <i>Wordcloud</i> pada <i>Mount Bromo</i>	97
Gambar L13. 19 Emosi pada <i>Mount Bromo</i>	98
Gambar L13. 20 Persentase Sentimen pada <i>Mount Bromo</i>	98
Gambar L13. 21 Akun pada <i>Mount Bromo</i>	99
Gambar L13. 22 <i>Wordcloud</i> pada <i>Lake Toba</i>	99
Gambar L13. 23 Emosi pada <i>Lake Toba</i>	100
Gambar L13. 24 Persentase Sentimen pada <i>Lake Toba</i>	100
Gambar L13. 25 Akun pada <i>Lake Toba</i>	101
Gambar L13. 26 <i>Wordcloud</i> pada <i>Ubud</i>	101
Gambar L13. 27 Emosi pada <i>Ubud</i>	102
Gambar L13. 28 Persentase Sentimen pada <i>Ubud</i>	102
Gambar L13. 29 Akun pada <i>Ubud</i>	103

DAFTAR LAMPIRAN

	Halaman
Lampiran 1. Teori Analisis Sentiment	52
Lampiran 2. Text Preprocessing	53
Lampiran 3. Teori <i>Naïve Bayes</i>	54
Lampiran 4. Penambahan Data	56
Lampiran 5. Pelabelan Kelas	57
Lampiran 6. <i>Data Preprocessing</i>	58
Lampiran 7. <i>Naïve Bayes Classifier</i>	59
Lampiran 8. <i>Particle Swarm Optimization</i>	61
Lampiran 9. Evaluasi dan Validasi Data.....	64
Lampiran 10. Visualisasi	65
Lampiran 11. Hasil <i>Data Preprocessing</i>	67
Lampiran 12. Hasil Evaluasi.....	74
Lampiran 13. Hasil Visualisasi	89
Lampiran 14. Biodata Penulis.....	104



UNIVERSITAS
Dinamika

BAB I

PENDAHULUAN

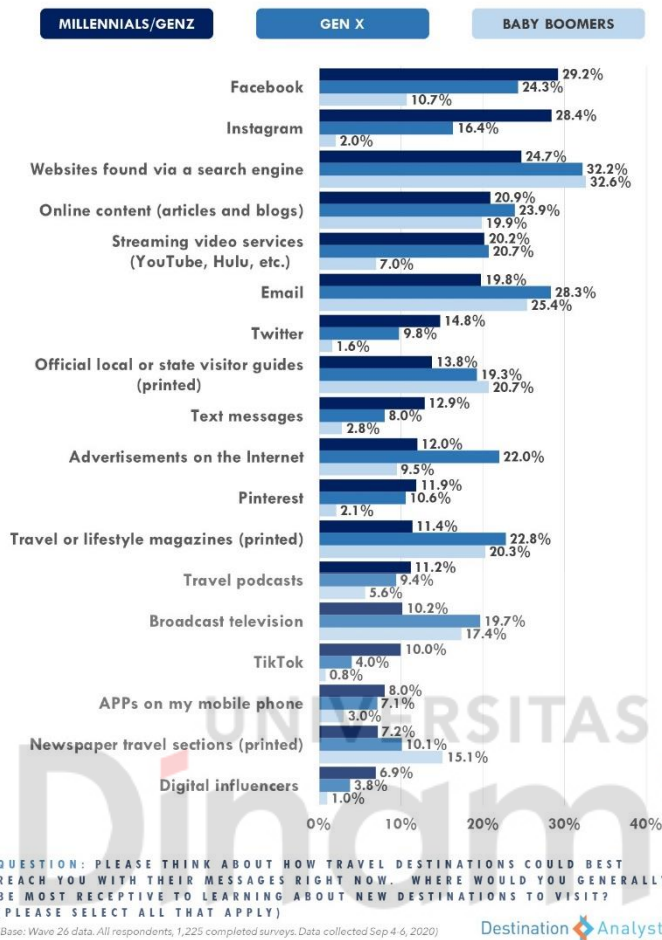
1.1 Latar Belakang

Pariwisata merupakan aktivitas orang yang berpergian atau tinggal dengan waktu yang tidak lebih dari satu tahun dengan untuk mengetahui sesuatu dan mendapatkan kesenangan, Gretzel, Zhong, and Koo (2016). Kegiatan pariwisata merupakan salah satu zone sektor ekonomi yang tumbuh paling cepat, membawa pendapatan devisa ke negara-negara dan menciptakan lapangan kerja (Ferdiansyah, Endayana, Rachmat, and Khadijah, 2020).

Dilansir dari CultureTrip Putri (2019), didapatkan informasi *The 15 Most Beautiful Places To Visit in Indonesia* dengan tujuh peringkat pertama adalah *Dieng Plateau, Borobudur Temple, Komodo National Park, Raja Ampat, Mount Bromo, Ubud, dan Lake Toba*, disusul oleh *Tanah Toraja, Tanjung Puting National Park, Derawan Islands, Bangka Belitung Islands, Wae Rebo Village, Lombok, Nusa Islands, dan Bukittinggi*. Tujuh destinasi wisata dipilih untuk dijadikan objek penelitian karena memiliki banyak ulasan pada destinasi wisata tersebut.

Menurut Rahmat, Rahayu, & Pariyati (2020), banyak tujuan wisata yang ada di Indonesia masih kurang akan informasi yang menjelaskan rentang tempat tujuan wisata tersebut yang mengakibatkan wisatawan mencari informasi tersebut melalui internet, buku dan komunitas.

MARKETING CHANNELS TRAVELERS FEEL MOST RECEPTIVE TO AS OF SEPTEMBER 6TH



Gambar 1. 1 Sumber informasi *Analysts* (2020).

Berdasarkan grafik diatas, disimpulkan bahwa wisatawan sekarang lebih memilih sumber dari Internet (blog perjalanan & *social media*) untuk menentukan tujuan wisata. Menurut Windasari & Eridani (2017), Internet menyediakan media untuk berbagi informasi dan pengalaman wisatawan melalui saluran komunikasi elektronik, seperti media sosial dan situs ulasan wisatawan. Banyak wisatawan mancanegara menggali informasi tentang wisata di Indonesia melalui internet. Hal tersebut bisa diketahui melalui *review* orang-orang yang pernah datang ke tempat wisata tersebut sebelumnya. *Review* dari para wisatawan yang sudah berkunjung ke wisata tersebut merupakan hal yang penting karena *review* tersebut dapat dijadikan informasi mengenai wisata yang ada di Indonesia dan banyak juga wisatawan yang berjam-jam menghabiskan waktu untuk mendapatkan informasi

tentang tujuan wisata di internet. Menurut Bajo (2020), salah memilih objek wisata maka akan fatal akibatnya baik dari sisi materi ataupun non materi seperti biaya dan barang yang akan dibawa untuk berwisata maupun perasaan yang akan didapat dari mengunjungi tempat wisata tersebut.

Analisis sentimen sendiri adalah jenis *naturale language* yaitu pengolahan kata untuk melacak *mood* atau opini masyarakat tentang suatu produk, jasa atau suatu topik tertentu dari *media social* atau website. Disini peneliti menggunakan *social media* TripAdvisor, Twitter, dan Google Maps Reviews karena Menurut Nicoli & Papadopoulou (2017), TripAdvisor adalah salah satu media sosial wisatawan terbesar dan terkenal di dunia dengan pengguna yang berpartisipasi mengutip ulasan deskriptif dan bermanfaat (70%), keakuratan konten (62%), dan cakupan konten wisata yang luas (62%) sebagai alasan utama mereka mengunjungi situs.

Menurut Rozi, Hamdana, & Alfahmi (2018) *twitter* merupakan social media yang populer karena pertumbuhan dari twitter terus meningkat setiap waktu yang mana hal itu di dimanfaatkan oleh wisatawan untuk menyampaikan sesuatu yang berupa saran dan kritikan yang dibutuhkan peneliti untuk memperoleh gambaran umum persepsi masyarakat akan suatu produk. Menurut Haq & Rachmat (2020), Google Maps Reviews adalah salah satu hal dari era big data saat ini di mana setiap orang dapat menaruh jejak setelah mereka mendatangi sebuah tempat dan termasuk kedalam *smart tourism* yang mana bisa menciptakan pengalaman yang pintar.

Metode klasifikasi sentimen dapat digunakan untuk menentukan sentimen dari sebuah tweet atau review seperti *naive bayes classifier*. Metode ini memiliki beberapa kelebihan, yaitu algoritma yang mudah, sederhana, cepat dalam perhitungan dan berakurasi tinggi (Muhamad, Prasajo, Sugianto, Surtiningsih, and Cholissodin, 2017). Untuk meningkatkan seleksi fitur banyak penelitian terdahulu mencoba untuk melakukan penambahan algoritma optimasi cerdas. Menurut Ernawati (2016), algoritma pada *particle swarm optimization* yaitu cepat dan sederhana untuk mendapatkan nilai optimasi. Maka pada penelitian ini menggabungkan algoritma *particle swarm optimization* dengan metode *naive*

bayes (PSO-NB) sebagai *feature select* yang diterapkan untuk klasifikasi pada review wisatawan.

Dengan kondisi yang terjadi, maka perlu adanya solusi untuk masalah tersebut. Solusi dari permasalahan tersebut adalah mengetahui tujuan wisata mana yang terbaik dengan sumber information mengenai audit wisatawan yang bisa didapat melalui *site* atau *web-based* media. Kemudian data tersebut nanti akan diklasifikasikan menjadi dua jenis klasifikasi yaitu *positive* dan *negative* menggunakan *naive bayes classifier* dengan menerapkan *particle swarm optimization* sebagai *feature select* untuk meningkatkan akurasi.

1.2 Rumusan Masalah

Berdasarkan permasalahan yang didapat maka dirumuskan bahwa masalah yang ada yaitu: bagaimana melakukan analisis sentimen *review* wisatawan terhadap destinasi wisata di Indonesia menggunakan algoritma *naive bayes classifier* berbasis *particle swarm optimization*.

1.3 Batasan Masalah

Batasan Masalah yang digunakan dalam melakukan analisis sentimen adalah:

1. Data sekunder yang digunakan adalah media sosial TripAdvisor, Twitter, dan Google Maps Reviews.
2. Analisis ini hanya membahas tujuh destinasi wisata, yakni, *Dieng Plateau*, *Borobudur Temple*, *Komodo National Park*, *Raja Ampat*, *Mount Bromo*, *Ubud*, dan *Lake Toba*.
3. Analisis sentiment meliputi pengelompokan *negative* dan *positive*.
4. Tidak membahas analisis sentimen dalam kelompok *neutral*.
5. Dalam melakukan ekstraksi data, penelitian hanya menggunakan metode *text preprocessing* pada *text mining*.
6. Proses klasifikasi sentimen dilakukan pada *review* bahasa Indonesia dan bahasa Inggris.

1.4 Tujuan

Tujuan dari penelitian ini adalah bagaimana melakukan analisis sentimen *review* wisatawan terhadap destinasi wisata di Indonesia menggunakan algoritma *naive bayes classifier* berbasis *particle swarm optimization*.

1.5 Manfaat

Berdasarkan analisis sentimen yang dilakukan, diharapkan hasil analisis sentimen dapat digunakan sebagai tolak ukur wisatawan dalam menentukan wisata mana yang akan dikunjungi



UNIVERSITAS
Dinamika

BAB II

LANDASAN TEORI

Bab kedua adalah landasan teori. Pada bab ini menjelaskan mengenai literatur apa saja yang digunakan pada penelitian ini yaitu :

2.1 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

Judul	Penulis	Hasil	Akurasi
Tourist sentiment analysis on TripAdvisor using text mining: A case study using hotels in Ubud, Bali	Nasa Zata Dina	Studi kasus : Analisis sentimen <i>review</i> hotel di Ubud menggunakan <i>Decision Tree</i> dan <i>Naive Bayes</i> . Tools : RapidMiner Testing : Confusion Matrix	<i>Decision Tree</i> Scenario pertama = 91.77%, kedua = 90.61%, dan ketiga = 90.72%. Akurasi dengan algoritma <i>Naive Bayes</i> pertama = 71.40%, kedua = 72.42% dan ketiga = 75.26%
Pemilihan Fitur Pada Analisis Sentimen Review Travel Online Menggunakan Algoritma <i>Naive Bayes</i> Dalam Penerapan Mutual Information Dan <i>Particle Swarm Optimization</i> (PSO)	Lisda Widiastuti	Studi kasus : Analisis sentimen travel <i>review</i> prediksi secara <i>online</i> menggunakan algoritma <i>Naive Bayes</i> dan <i>Particle Swarm Optimization</i> (PSO) tanpa pemilihan fitur dari sumber sitejabber.com. Tools : RapidMiner Testing : Confusion Matrix	<i>Naive Bayes</i> tanpa fitur = 81.50%, <i>Naive Bayes</i> berbasis PSO = 93.50%
Accuracy for Sentimen Analysis of Twitter Students on E-Learning in Indonesia using <i>Naive Bayes</i> Algorithm Based on <i>Particle Swarm Optimization</i>	Mohamad Kartiko dan Sfenrianto	Studi kasus : Analisis sentimen akurasi siswa tentang <i>E-Learning</i> menggunakan Bahasa Indonesia dari media sosial twitter. Algoritma yang digunakan adalah <i>Naive Bayes</i> (NB) dan untuk mengoptimalkan hasil perhitungan akurasi, menggunakan <i>Particle Swarm Optimization</i> (NB-PSO) . Tools : Rapidminer Testing : Confusion Matrix	<i>Naive Bayes</i> Scenario pertama = 71.67%, kedua = 56.67%, dan ketiga = 40.00%. Akurasi dengan algoritma <i>Naive Bayes</i> berbasis PSO pertama = 81,67%, kedua = 70.00 dan ketiga = 67.22%

2.2 Pariwisata

Sebagai antisipasi perkembangan dunia pariwisata yang telah mengglobal sifatnya, pemerintah Indonesia mengeluarkan Undang-Undang No. 10 Tahun 2009 tentang kepariwisataan yang terdiri atas tujuh belas bab dan tujuh puluh pasal yang mengandung ketentuan meliputi delapan hal, yaitu :

1. Wisata adalah kegiatan perjalanan yang dilakukan oleh seseorang atau sekelompok orang dengan mengunjungi tempat tertentu untuk tujuan rekreasi, pengembangan pribadi atau mempelajari keunikan daya tarik wisata yang dikunjungi dalam jangka waktu sementara.
2. Wisatawan adalah orang yang melakukan wisata.
3. Pariwisata adalah berbagai macam kegiatan wisata dan didukung berbagai fasilitas serta layanan yang disediakan oleh masyarakat, pengusaha, pemerintah dan pemerintah daerah.

2.3 Objek Wisata

Objek wisata adalah segala sesuatu yang ada di daerah tujuan wisata yang merupakan daya tarik agar orang-orang ingin datang berkunjung ke tempat tersebut. Objek dan daya tarik wisata menurut Undang-undang No 10 tentang kepariwisataan yaitu daya tarik wisata adalah segala sesuatu yang memiliki keunikan, keindahan, dan nilai yang berupa keanekaragaman kekayaan alam, budaya, dan hasil buatan manusia yang menjadi sasaran atau tujuan kunjungan wisatawan dan daerah tujuan pariwisata yang selanjutnya disebut destinasi pariwisata.

Kirom, Sudarmiati, and Putra (2016) objek wisata adalah segala sesuatu yang memiliki keindahan, keunikan, keanekaragaman budaya, alam dan hasil buatan manusia yang menjadi objek atau tujuan wisatawan untuk melakukan wisata.

2.4 Analisis Sentimen

Menurut Wibowo & Jumiati (2018), Analisis sentimen adalah bidang studi dengan melakukan analisis pada sentimen, opini, evaluasai, sikap, emosi dan penilaian terhadap produk, organisasi, layanan, individu, masalah, topik, dan peristiwa. Analisis sentimen memiliki banyak nama penelitian seperti ekstraksi

pendapat, penambangan sentimen, analisis subjektivitas, penambangan opini, analisis emosi, analisis pengarus dan penambangan *review*. Contoh table pengelompokan data terdapat pada Tabel L1.1 Pengelompokan Data Lampiran 1. Teori Analisis Sentimen.

2.5 Web Crawler

Menurut Shrivastava (2018), *Web crawler (spider atau robot website)* adalah program yang otomatis melintasi sejumlah halaman website dengan mengikuti acuan (*hyperlink* atau pranala), menentukan urutan bagian-bagian atau unit-unit dari kata(mengindeks), dan menyimpan sumber halaman *website* yang dilalui. Data yang tersedia pada *website* berupa data tidak terstruktur dan terstruktur. Penggunaan yang paling luas dari *crawler* yaitu untuk mendukung mesin pencari (*search engine*). *Crawler* istimewa lebih bertarget. Mereka berusaha untuk men-download halaman untuk topik tertentu (Abdillah, Taufik, & Jumadi, 2016).

2.6 Text Mining

Sedangkan, Menurut Jalal (2020), *Text mining* adalah penambangan teks yang mengacu pada penggalian informasi yang berguna dari bahasa manusia teks dengan menganalisis sejumlah besar informasi seperti dalam dokumen elektronik dan halaman web online. Sistem penambangan teks mengklasifikasikan dan mengatur data, bergantung pada pola leksikal atau linguistik. Konsep text mining berkisar pada penggalian informasi dari data tekstual itu ditulis dalam tidak terstruktur atau semi-terstruktur melalui pemrosesan bahasa alami. Salah satu langkah yang dilakukan dalam *text mining* adalah.

2.6.1 Text Preprocessing.

Menurut Jaka H (2015), *Preprocessing task* atau *preprocessing text* adalah proses *text mining* yang diharapkan mampu menghilangkan atau mengurangi kata-kata atau text yang tidak mempunyai arti atau teks yang tidak perlu dari dokumen. *Preprocessing task* pada umumnya mengubah informasi dari masing-masing sumber data asli ke dalam format kata dasar dengan menerapkan berbagai jenis metode fitur ekstraksi.

1. *Case Folding*

Menurut Sasmita (2019), *Case folding* adalah proses mengonversi semua kata dalam dokumen ke huruf kecil. Hanya huruf 'a - z' yang diterima.

2. *Tokenizing*

Menurut Indraloka & Santosa (2017), *Tokenizing* adalah memotong kalimat berdasarkan tiap kata yang menyusunnya. Di dalam tahap *tokenizing* untuk symbol dan karakter yang selain a-z akan dihilangkan, kata akan dilakukan pemotongan yang berdasarkan spasi di dalam kalimat tersebut dan juga menghilangkan karakter-karakter seperti tanda baca. Contoh dari tahap ini terletak pada Gambar L2.1 Lampiran 2. *Text Preprocessing*.

3. *Filtering*

Menurut Rahutomo & Ririd (2019), *filtering* (Stopword Removal) adalah penghilangan kata yang tidak relevan. Kata yang tidak berguna seperti kata ganti, kata preposisi, dan kata sambung. Contoh dari tahap ini terletak pada Gambar L2.2 Lampiran 2. *Text Preprocessing*.

4. *Stemming*

Menurut Indraloka & Santosa (2017), *stemming* (Lemmatizing) adalah merubah berbagai kata berimbuhan menjadi kata dasarnya. Contoh proses *stemming* seperti kata "mempresentasikan", "dipresentasikan", dan "berpresentasi" diubah menjadi kata dasar "presentasi". Contoh dari tahap ini terletak pada Gambar L2.3 Lampiran 2. *Text Preprocessing*.

2.7 *Naïve Bayes*

Menurut Saleh (2015), *Naive Bayes* adalah salah satu algoritma yang sering digunakan untuk pengklasifikasian teks, juga salah satu metode untuk pembelajaran mesin menggunakan perhitungan probabilitas. *Naïve bayes* dikemukakan oleh Thomas Bayes yaitu ilmuwan Inggris, yaitu untuk memprediksi peluang di masa depan berdasarkan pengalaman masa lalu dengan menggunakan statistik untuk menghitung peluang suatu hipotesis. Menurut Rizqiyani et al (2017), *Naive Bayes* adalah algoritma yang baik dalam hal waktu komputasi atau waktu yang dibutuhkan untuk membangun suatu model. Waktu komputasi dari

naïve bayes tidak perlu diragukan karena merupakan yang tercepat. Kecepatan komputasi adalah salah satu kelebihan yang baik untuk diterapkan dalam sebuah sistem klasifikasi buku. Berikut adalah rumus dari *Naive Bayes*:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)}$$

Dimana:

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

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

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

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

$P(X)$: Probabilitas X

Dimana *flowchart* dari metode *Naive Bayes* ini terletak pada Gambar L2.4 *Flowchart* algoritma *Naive Bayes* dan contoh dari penerapan *naïve bayes* terdapat pada Lampiran 2. Teori *Naïve Bayes*.

2.8 Particle Swarm Optimization

Menurut Que, Iriane, & Purnomo (2020), PSO adalah suatu metode *optimasi* yang sederhana untuk memodifikasi beberapa *boundary*. PSO didasarkan pada perilaku berkelompok dan kerja sama sosial dari burung dan dengan penekanan khusus pada pengembangan, penerapan, dan peningkatan dari yang paling dasar serta beberapa implementasi mutakhir. Optimasi PSO dapat dilakukan dengan cara *attribute select* dan *feature select*, dan juga meningkatkan *attribute weight* pada semua variabel atau *attribute* yang digunakan. Kelebihan PSO adalah mudah diterapkan, sederhana dan cepat.

Setiap individu atau partikel menyampaikan informasi posisi terbaiknya kepada partikel yang lain dan menyesuaikan posisi dan kecepatan masing-masing berdasarkan informasi yang diterima mengenai posisi tersebut. Setiap individu diperlakukan seperti sebuah titik pada suatu dimensi ruang tertentu.

Kemudian terdapat dua faktor yang memberikan karakter terhadap status partikel pada ruang pencarian yaitu posisi partikel dan kecepatan partikel

Pencarian pada PSO menggunakan dua pencarian yaitu pencarian local dan global. Pencarian dilakukan dengan cara setiap partikel melakukan penyesuaian terhadap posisi terbaik dari partikel tersebut (*pbest*) dan penyesuaian terhadap posisi partikel terbaik dari seluruh kelompok (*gbest*) selama melintasi ruang pencarian. Jadi, penyebaran pengalaman atau informasi terjadi di dalam partikel itu sendiri dan antara suatu partikel dengan partikel terbaik dari seluruh kawanan selama proses pencarian solusi. Dalam PSO terdapat beberapa proses sebagai berikut:

2.8.1 Inisialisasi

a. Inisialisasi kecepatan awal

Untuk iterasi ke 0, dapat dipastikan untuk nilai kecepatan awal disemua partikel adalah 0

b. Inisialisasi posisi awal partikel

$$x = x_{min} + rand[0,1] \times (x_{max} - x_{min}) \quad (5)$$

c. Inisialisasi *pbest* dan *gbest*

Untuk iterasi ke 0, *pbest* dipastikan sama dengan nilai posisi awal partikel. Sedangkan *gbest* didapat dari hasil *pbest* dengan *fitness* tertinggi.

2.8.2 Update Kecepatan

$$v_{i,j}^{t+1} = w \cdot v_{i,j}^t + c_1 \cdot r_1 (pBest_{i,j}^t - x_{i,j}^t) + c_2 \cdot r_2 (gBest_{g,j}^t - x_{i,j}^t)$$

2.8.3 Update posisi dan hitung fitness

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1}$$

2.8.4 Update *pBest* dan *gBest*

Tahap ini melakukan perbandingan antara *pbest* yang sebelumnya dengan hasil dari update posisi. Nilai *pbest* diketahui dari *fitness* tertinggi. Nilai *gbest* diketahui dari nilai *pbest* dengan *fitness* tertinggi.

2.9 Confusion Matrix

Menurut Nugroho (2019), *Confusion matrix* atau *error matrix* berisi tentang informasi klasifikasi aktual dan prediksi yang dilakukan oleh klasifikasi sistem.

Performa sistem seperti itu biasanya dilakukan evaluasi menggunakan data dalam matriks. Tabel berikut menunjukkan *confusion matriks* untuk pengklasifikasi dua kelas.

TP (True Positive)	FP (False Positive) Type I Error
FN (False Negative) Type II Error	TN (True Negative)

Tabel 2. 2 *Confusion Matrix*

2.9.1 Accuracy

Akurasi digunakan untuk mengetahui nilai akurasi dari hasil klasifikasi pada penelitian. Berikut adalah rumus dari akurasi *confusion matrix*:

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

2.9.2 Precision

Precision atau *positive predictive value* adalah jumlah data yang ditemukan dan dianggap relevan untuk kebutuhan pencari informasi dan itu merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Berikut adalah rumus dari *precision*:

$$\text{Prec} = \frac{TP}{TP + FP}$$

2.9.3 Sensitivity

Sensitivity, *recall*, *true positive rate* atau *hit rate* adalah jumlah data yang ditemukan oleh sebuah proses pencarian dalam system dan itu merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif. Berikut adalah rumus dari *recall*:

$$\text{SENS} = \frac{TP}{TP + FN}$$

2.9.4 Specificity

Specificity, *selectivity*, atau *true negative rate* merupakan kebenaran memprediksi negatif dibandingkan dengan keseluruhan data negatif. Berikut

adalah rumus dari *specificity*:

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

2.9.5 Negative Predictive Value

Negative predictive value merupakan rasio prediksi benar negatif dibandingkan dengan keseluruhan hasil yang diprediksi negatif . Berikut adalah rumus dari NPV:

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

2.9.6 False Negative Rate

False negative rate merupakan rasio positif yang diketahui hasilnya negatif. FNR ini kadang-kadang disebut miss rate. Berikut adalah rumus dari FNR:

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

2.9.7 False Positive Rates

False positive rates merupakan rasio negatif yang diketahui hasilnya positif. Berikut adalah rumus dari FNP:

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

2.10 RapidMiner

Menurut Silalahi (2018), *RapidMiner* merupakan software untuk pengolahan informasi menggunakan prinsip dan algoritma data mining. Rapidminer melakukan ekstrak pola dari suatu dokumen yang besar dengan menggabungkan metode statistic, *artificial intelligence* dan *database*.

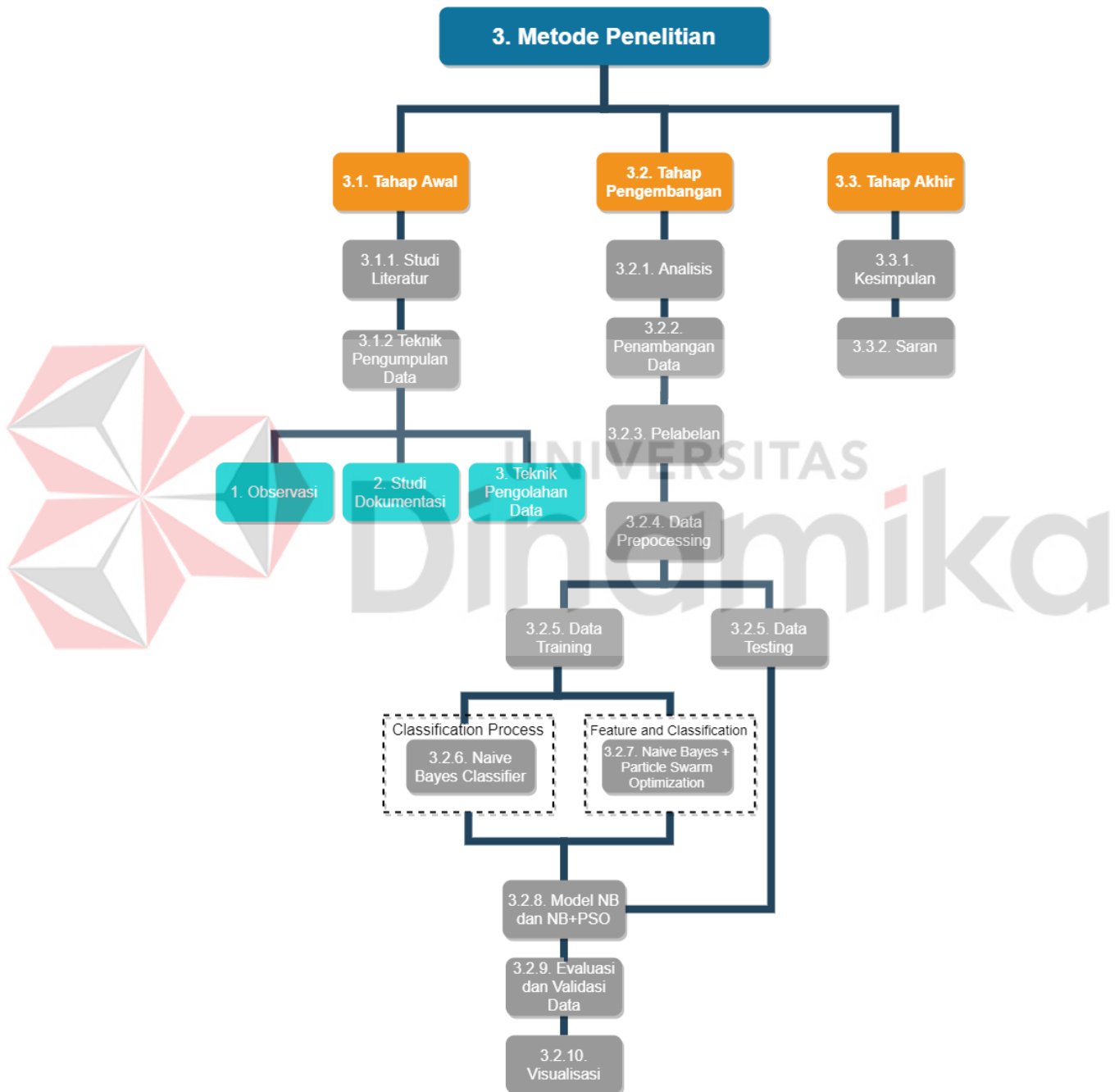
2.11 Visualisasi

Menurut Prasetyo (2019), visualisasi adalah alat komunikasi untuk menginformasikan secara jelas dan efisien tentang informasi melalui grafik atau tabel kepada pengguna. Visualisasi membantu penggunaanya dalam melakukan Analisa dan penalaran tentang informasi dan bukti sehingga informasi yang kompleks menjadi lebih mudah untuk dipahami.

BAB III

METODE PENELITIAN

Pada bab metode penelitian ini, didapatkan alur untuk menjelaskan tahapan metode dari penelitian yang akan dilakukan 3 tahapan. Gambaran dari metodologi penelitian terdapat pada gambar 3.1. Metode Penelitian.



Gambar 3. 1 Metode Penelitian

3.1 Tahapan Awal Penelitian

Hal yang dilakukan pada penelitian ini:



Gambar 3. 2 Tahap Awal

Gambar diatas merupakan tahapan penelitian yang peneliti lakukan dalam melakukan penelitian ini adapun rincian jelasnya adalah sebagai berikut:

3.1.1 Studi Literatur

Pada tahap awal metode penelitian, hal yang pertama kali dilakukan adalah melakukan review pada penelitian-penelitian terdahulu dalam bentuk jurnal dan tugas akhir (TA) tentang analisis sentimen *review* menggunakan metode *naïve bayes* dan PSO dengan menggunakan RapidMiner. Setelah mereview penelitian terdahulu, dapat ditentukan studi masalah dan tahapan-tahapan penyelesaian yang dipergunakan untuk melakukan penelitian.

3.1.2 Teknik Pengumpulan Data

Pada tahap pengumpulan data ini dilakukan *crawling* dari media social Tripadvisor, Twitter dan Google Maps Reviews yang memuat komentar mengenai tujuh destinasi wisata di Indonesia, yaitu , *Dieng Plateau*, *Borobudur Temple*, *Komodo National Park*, *Raja Ampat*, *Mount Bromo*, *Ubud*, dan *Lake Toba*.

Berikut adalah tahapan dari pengumpulan data yang dilakukan dalam penelitian tugas akhir ini, diantaranya:

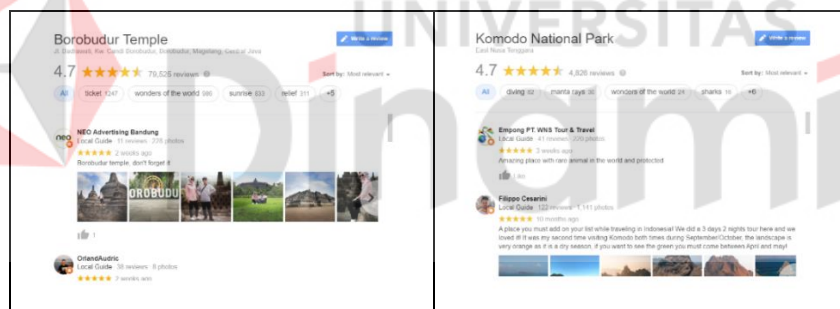
1. Observasi

Observasi ini dilakukan untuk mengamati objek yang akan diteliti. Pada penelitian ini dilakukan dengan mengamati review wisatawan dari social media Twitter, website TripAdvisor dan Google Maps Reviews. Adanya banyak macam tweet dan review yang diberikan oleh pengguna pada destinasi wisata yang membuat peneliti melakukan penelitian dalam presentase klasifikasi review wisatawan terhadap wisata di Indonesia pada social media Twitter, website TripAdvisor dan Google Maps Reviews.

2. Studi Dokumentasi

Dalam studi dokumentasi ini didapatkan data tentang rating dan review yang diberikan oleh wisatawan terhadap wisata *Dieng Plateau*, *Borobudur Temple*, *Komodo*, *Raja Ampat*, *Mount Bromo*, *Lake Toba* dan *Ubud*.

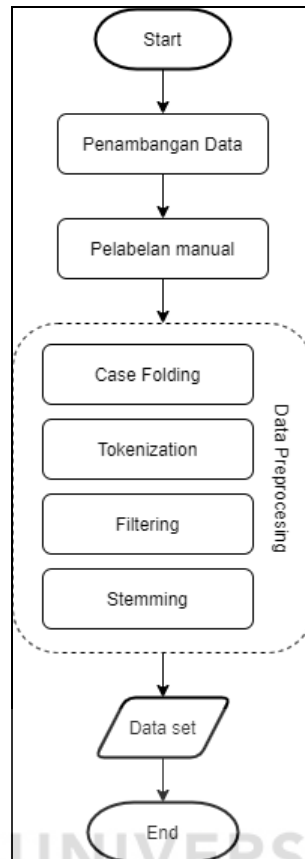
Rating Dieng Plateau, Borobudur Temple, dan Komodo National Park dari website Google Maps Reviews



Gambar 3. 3 Rating Dan Review Wisata Dari Google Maps Reviews

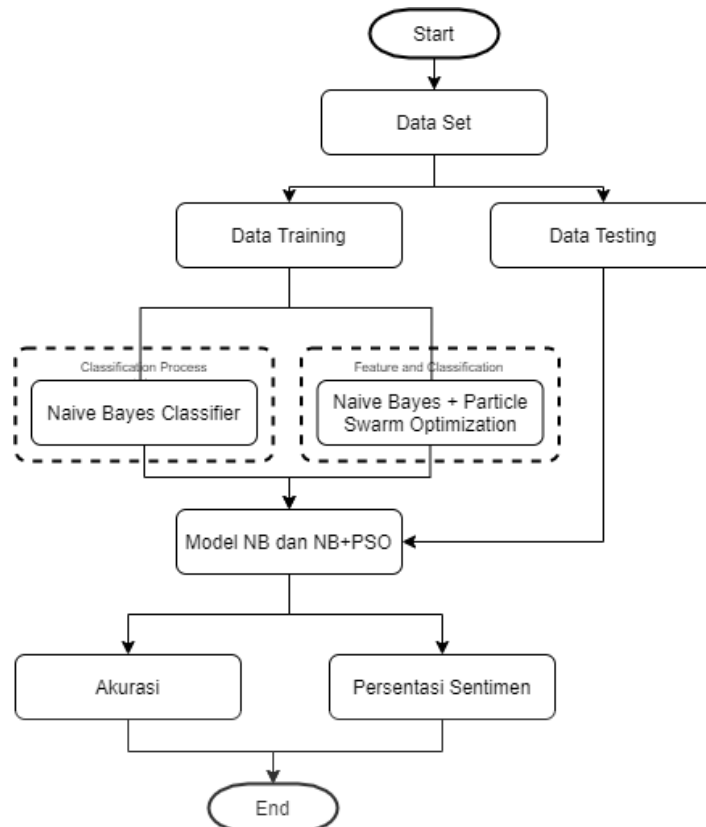
3. Teknik pengolahan data

Tahap ini dilakukan dengan membagi data menjadi dua set data yaitu *training data* dan *testing data*. Data tersebut akan melalui beberapa proses untuk menghasilkan data yang siap untuk klasifikasi hasil. Gambar 3.4 di bawah ini menunjukkan proses pengumpulan dan pra-pemrosesan data hingga tersedia untuk segera digunakan.



Gambar 3. 4 Proses Pengumpulan Dan Preprocessing Data

Pada Gambar 3.4, proses pertama yang dijelaskan adalah penambahan data. Penambahan data dilakukan dengan mencrawl data dari TripAdvisor dan Google Maps Reviews dan API Twitter yang akan dijadikan sebagai objek. Selanjutnya, untuk mengambil data dari TripAdvisor menggunakan *tools* Webharvy, untuk mengambil data dari Twitter dan Google Maps Reviews menggunakan bahasa pemrograman Python dan data yang di ambil *review* dalam bahasa Indonesia dan bahasa Inggris. Data mentah akan dilakukan pelabelan yang diberikan sesuai dengan classnya yaitu *positive* dan *negative* dengan bantuan Setelah dilakukan pelabelan, data akan menjalani proses preprocessing data. Pada preprocessing data dilewatkan 4 sub. Setelah proses *preprocessing* data, akan dihasilkan kumpulan data yang siap menggunakan *Naive Bayes Classifier* untuk klasifikasi. Proses klasifikasi dari *Naive Bayes classifier* akan ditunjukkan pada Gambar 3.5 di bawah ini.

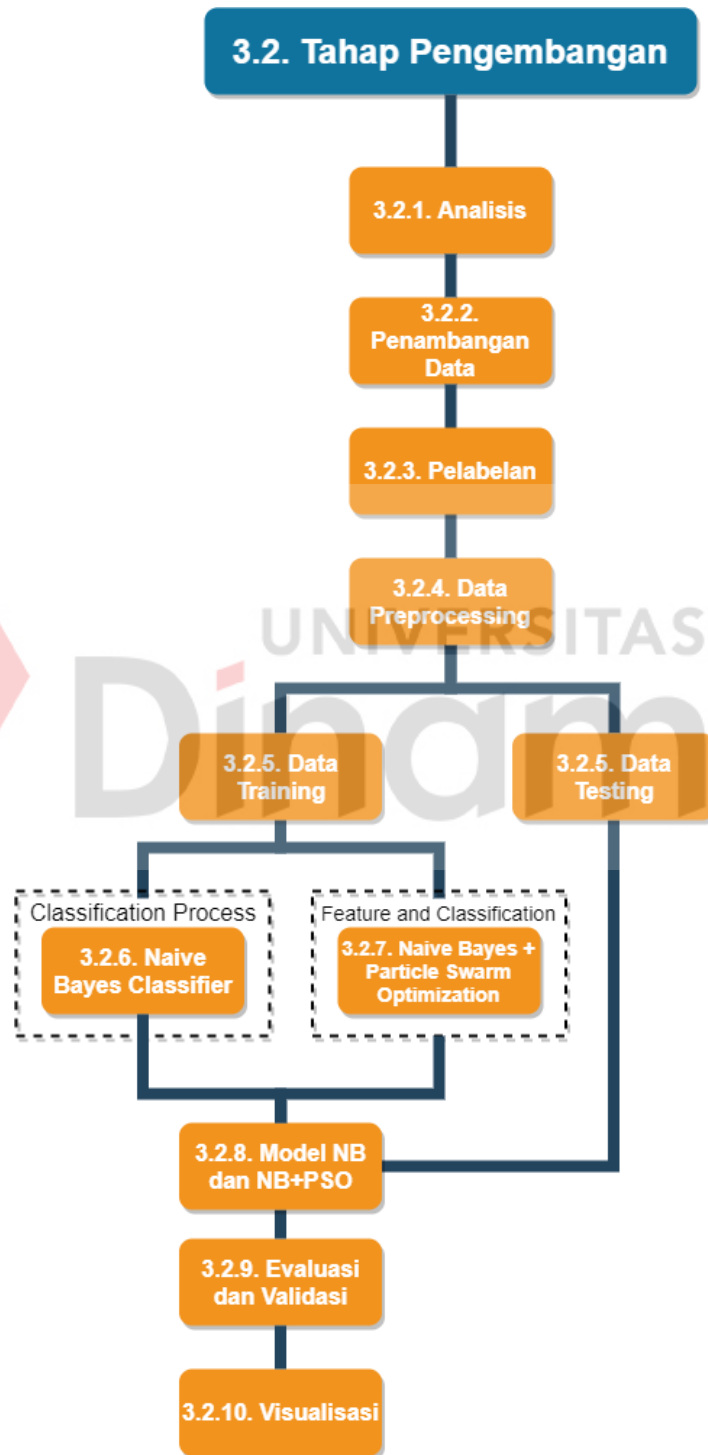


Gambar 3. 5 Proses *Naive Bayes* dan *Naive Bayes + PSO*

Pada Gambar 3.5, dapat dijelaskan bahwa data yang siap diolah akan dibagi menjadi 2 data yaitu data training dan data testing. Data yang sudah dilabeli diambil sebesar 90% sebagai data training dari dataset. Training data digunakan acuan untuk menentukan klasifikasi sentimen dari data *tweet* yang belum ditandai dengan *Naive Bayes Classifier* dan *Naive Bayes + Particle Swarm Optimization* yaitu data testing. Setelah semua data memiliki label, selanjutnya akan diukur tingkat akurasi metode *Naive Bayes Classifier* dan *Naive Bayes + Particle Swarm Optimization* menggunakan *Confusion Matrix* yang akan diukur dalam aplikasi RapidMiner.

3.2 Tahap Pengembangan Penelitian

Tahap ini menjelaskan tentang pengembangan untuk menyelesaikan tugas akhir ini.



Gambar 3. 6 Tahap Pengembangan

3.2.1 Analisis

Tahapan selanjutnya yang dilakukan pada Tahap Pengembangan yaitu Analisis. Tahap ini untuk mengetahui kebutuhan yang dibutuhkan pada tugas akhir ini.

1. Identifikasi Masalah

Identifikasi masalah dilakukan dengan mempelajari dari jurnal dan tugas akhir, melakukan pengamatan pada *source* yang diangkat yaitu media social sehingga mendapatkan informasi yang berkaitan dengan masalah yang diangkat yaitu dibutuhkan analisis membantu untuk mengetahui *review* atau pendapat wisatawan terhadap objek wisata di Indonesia.

2. Analisis Kebutuhan Pengguna

Analisis kebutuhan pengguna dilakukan dengan tujuan untuk mengetahui kebutuhan apa saja yang pengguna butuhkan. Kebutuhan pengguna dapat dilihat pada Tabel L1.2.

3. Analisis Kebutuhan Fungsional

Analisis kebutuhan fungsional dilakukan dengan tujuan untuk mengetahui kebutuhan fungsional berupa fungsi penambangan data untuk mengumpulkan *information*, fungsi untuk melakukan ekstraksi data, fungsi pengelompokan menurut kelompok *positive* dan *negative*, dan fungsi visualisasi hasil *review* dari wisatawan.

4. Analisis Kebutuhan Non Fungsional

Analisis kebutuhan non fungsional dilakukan dengan tujuan untuk mengetahui kebutuhan yang dibutuhkan seperti kebutuhan akses dan kebutuhan dalam menyajikan *information* dalam bentuk visualisasi dengan waktu yang efektif.

5. Analisis Kebutuhan Sistem

a. Kebutuhan Perangkat Lunak

Kebutuhan *software requirement* adalah komponen perangkat lunak yang digunakan dalam penelitian ini. Adapun *software* yang digunakan sebagai berikut:

- 1) Bahasa pemrograman yang digunakan pada penelitian ini yaitu Python yang digunakan untuk memproses *web scraping* dari Twitter dan Google Maps Reviews

- 2) Alat pengkodean yang digunakan pada penelitian ini adalah Anaconda promp (Miniconda3) karena ini bersifat *Open Source* dan tidak dibebankan biaya.
- 3) Tools yang digunakan dalam melakukan *Web Scrapping* untuk website TripAdvisor adalah WebHarvy. Webharvy sangat mudah digunakan dan tidak memerlukan kode untuk penggunaannya.

b. Kebutuhan Perangkat Keras

Berdasarkan analisis kebutuhan *software*, maka dapat diketahui kebutuhan *hardware* yang mendukung komputer dalam menjalankan fungsinya. Kebutuhan yang dibutuhkan adalah:

- 1) Pada penelitian ini menggunakan Processor Generasi ke 8 Intel® Core™ i5-8250U.
- 2) Pada penelitian ini menggunakan Random Access Memory (RAM) sebesar 8GB.
- 3) Pada penelitian ini menggunakan Monitor dengan resolusi 1920x1080 pixel
- 4) Pada penelitian ini menggunakan Graphic Processing Unit (GPU) NVIDIA® GeForce 940 MX

3.2.2 Penambangan Data

Tahap kedua dari Tahap Pengembangan adalah melakukan penambangan data yang digunakan untuk mengambil data yang dibutuhkan pada penelitian ini. Untuk penambangan data menggunakan *software* WebHarvy untuk website TripAdvisor seperti yang terdapat pada Gambar L4.1 Penambangan data menggunakan WebHarvy.

Jumlah data yang diambil dari Tripadvisor, Twitter dan Google Maps Reviews mengenai tujuh destinasi wisata di Indonesia dapat dilihat di Tabel L4.1. Contoh hasil data yang sudah di crawl dengan tools Webharvy dapat dilihat pada Gambar L4.2. Hasil penambangan data meggunakan Webharvy dan contoh hasil data yang sudah di crawl dengan pemrograman python dapat dilihat pada Gambar L4.3 Hasil penambangan data menggunakan pemrograman Python. Hasil dari penambangan menggunakan webharvy maupun menggunakan pemrograman python berupa file excel .

3.2.3 Pelabelan

Pelabelan atau penentuan kelas atribut dibagikan menjadi dua kelas yaitu sentimen *positive* untuk review wisatawan yang bersifat *positive* dan sentimen *negative* untuk review wisatawan yang bersifat *negative*. Untuk pelabelan review dilakukan dengan melibatkan bantuan dari sukarelawan1 melakukan pelabelan pada destinasi *Borobudur Temple*, sukarelawan2 melakukan pelabelan pada destinasi *Bromo*, sukarelawan3 melakukan pelabelan pada destinasi *Komodo National Park*, sukarelawan4 melakukan pelabelan pada destinasi *Raja Ampat*, sukarelawan5 melakukan pelabelan pada destinasi *Dieng Plateau*, sukarelawan6 melakukan pelabelan pada destinasi *Toba Lake*, dan sukarelawan7 melakukan pelabelan pada destinasi *Ubud*. Untuk hasil pelabelan dapat dilihat pada Gambar L5.1 Hasil Pelabelan Kelas.

3.2.4 Data Preprocessing

Tahapan selanjutnya yang dilakukan pada Tahap Pengembangan yaitu *Data Preprocessing*. Tahap *Data Preprocessing* yang dilakukan dalam tahap ini adalah melakukan ekstraksi data. Ekstraksi data bertujuan untuk mengolah hasil penambangan data yang formatnya tidak terstruktur hingga menjadi format kata dasar. Ekstraksi data yang dilakukan pada penelitian ini terdiri dari 4 (empat) tahap, yaitu *case folding*, *tokenizing*, *filtering*, dan *stemming*. Untuk tahap *Data Preprocessing* dapat dilihat pada Lampiran 6. *Data Preprocessing*.

3.2.5 Training data dan Testing data

Dataset yang telah dikumpulkan pertama akan dipecah (*split*) menjadi *training data* dan *testing data* dengan perbandingan 90:10. *Training data* digunakan untuk melatih algoritma klasifikasi, data yang akan digunakan sebagai *training data* harus dilabeli terlebih dahulu dan juga harus melalui *preprocessing text*. Kemudian *testing data* akan digunakan untuk menguji algoritma atau model yang telah dilatih. Sistem akan memberikan hasil akhir dari hasil klasifikasi sistem yaitu *negative* atau *positive* sesuai dengan model data.

3.2.6 Naive Bayes Classifier

Tahapan selanjutnya yang dilakukan dalam tahap pengembangan adalah melakukan klasifikasi informasi menggunakan *Naive Bayes Classifier*. *Naive*

Bayes mengklasifikasikan information dengan menggunakan information menjadi *training data* dan *testing data*. Saat melakukan klasifikasi, *Naive Bayes* akan menggunakan rumus untuk mencari nilai probabilitas tertinggi dari pengalaman sebelumnya atau pengalaman pribadinya. Klasifikasi data pada penelitian ini memiliki fungsi untuk menentukan class pada information yaitu *Positive* dan *Negative*. Contoh perhitungan dari *Borobudur Temple* dan klasifikasi dari *naïve bayes* dapat dilihat pada Lampiran 7. *Naïve Bayes Classifier*.

3.2.7 Particle Swarm Optimization

Tahapan selanjutnya yang dilakukan dalam tahap pengembangan adalah peningkatan *attribute weighting* yaitu *Particle Swarm Optimization* (PSO). Penerapan dari metode *naïve bayes* – PSO ada beberapa atribut yaitu jumlah iterasi, bobot inersia, konstanta kecepatan dari masing-masing atribut untuk melakukan pencarian nilai optimal. Jumlah partikel digunakan untuk menentukan banyaknya *population size* pada PSO. Contoh perhitungan dari *Borobudur Temple* dan klasifikasi dari *naïve bayes + particle swarm optimization* dapat dilihat pada Lampiran 7. *Particle Swarm Optimization*.

3.2.8 Apply Model

Tahap ini untuk melakukan penerapan dari model yang sudah dibuat. Dengan menggunakan *training data* sebagai *testing data* dan dilakukan perhitungan untuk mengetahui hasil dari kesuksesan klasifikasi pada *training data*. Tahap *training* dan *testing* dilakukan validasi menggunakan *10-fold cross validation*.

3.2.9 Evaluasi dan Validasi Data

Tahap evaluasi untuk mengetahui hasil akurasi, waktu komputasi dan kinerja dari PSO terhadap algoritma *naïve bayes*. Validasi bertujuan untuk mengetahui hasil dari akurasi dari kedua metode. Teknik validasi yang digunakan adalah *10-folds cross-validation* dengan membagi *training data* menjadi beberapa bagian, akurasi akan diukur menggunakan *confusion matrix*. Hasil dari perhitungan *confusion matrix* untuk *true positive* didapat dari hasil pelabelan dan hasil prediksi yang hasilnya sama *positive*, untuk *false positive* didapat dari hasil prediksi *positive* dan hasil pelabelan *negative*, untuk *false negative* didapat dari

hasil prediksi *negative* dan hasil pelabelan *positive*, sedangkan untuk *true negative* didapat dari hasil prediksi dan pelabelan sama-sama *negative*. Gambaran dari proses pengujian *naïve bayes* dapat dilihat pada Lampiran 9. Evaluasi dan Validasi Data.

Pengujian *naïve bayes* berbasis *Particle Swarm Optimization* juga menggunakan *10-Fold Cross Validation*. Pada pengujian di gunakan nilai *population size* yang telah di tentukan sebelumnya. Nilai pada parameter *population size* dan *maximum number of generation* bernilai default yaitu 5 dan 30. Pengujian ini dapat dilihat pada Gambar L9.2 Desain pengujian dengan Naïve Bayes + PSO Lampiran 9. Evaluasi dan Validasi Data

3.2.10 Visualisasi

Tahap visualisasi data telah dilakukan dengan memvisualkan data dengan *Worldcloud*, *bar chart*, *pie chart* dan *table*.

1. Visualisasi data dengan wordcloud menghasilkan kata yang sering muncul dalam teks yang telah dianalisis sebelumnya, penggunaan font-size huruf yang berbeda menggambarkan frekuensi kemunculan kata pada data yang dianalisis. Visualisasi data dengan wordcloud dapat dilihat Gambar L10.1 pada Lampiran 10. Visualisasi.
2. Visualisasi data dengan *bar chart* digunakan untuk mengetahui jumlah data komentar dari 7 wisata yang nantinya bisa menentukan wisata mana yang paling banyak diulas dan emosi dari *review* wisatawan. Visualisasi dengan *bar chart* dapat dilihat Gambar L10.2 pada Lampiran 10. Visualisasi.
3. Visualisasi data dengan *pie chart* digunakan untuk hasil presentase *Naive Bayes*, dan akun yang sering memberi *review* dan *tweet*. Visualisasi data dengan *pie chart* dapat dilihat Gambar L10.3 pada Lampiran 10. Visualisasi.

3.3 Tahap Akhir Penelitian

Bagian ini menjelaskan tentang Tahap Akhir untuk menyelesaikan penelitian ini.



Gambar 3. 7 Tahap Akhir

3.3.1 Kesimpulan

Tahapan selanjutnya yaitu kesimpulan. Tahap ini dilakukan dengan tujuan untuk mengetahui hasil perbandingan analisis sentiment, evaluasi akurasi dan waktu komputasi dari metode *naïve bayes* dan *naïve bayes* – PSO pada tujuh destinasi yakni *Dieng Plateau*, *Borobudur Temple*, *Komodo National Park*, *Raja Ampat*, *Mount Bromo*, *Lake Toba* dan *Ubud*.

3.3.2 Saran

Tahapan selanjutnya yang dilakukan pada tahap akhir yaitu saran. Tahap ini dilakukan dengan tujuan untuk memberi masukan bagaimana meningkatkan destinasi *Dieng Plateau*, *Borobudur Temple*, *Komodo National Park*, *Raja Ampat*, *Mount Bromo*, *Lake Toba* dan *Ubud*.

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas hasil analisis berdasarkan pengolahan data yang telah dilakukan pada social media TripAdvisor, Twitter, dan Google Maps Reviews yang memuat komentar mengenai tujuh destinasi wisata di Indonesia yaitu , *Dieng Plateau, Borobudur Temple, Komodo National Park, Raja Ampat, Mount Bromo, Ubud, dan Lake Toba*.

4.1 Analisis

Analisis dimulai dengan tahap penambangan data, melakukan sentimen manual atau pelabelan manual, melakukan *data preprocessing*, membagi data menjadi *training data* dan *testing data*, pengklasifikasian dengan metode *naïve bayes*, evaluasi dan validasi *naïve bayes*, pengklasifikasian dengan metode *naïve bayes* - PSO, evaluasi dan validasi *naïve bayes* - PSO, dan visualisasi..

Analisis tugas akhir ini dikerjakan dengan spesifikasi perangkat keras dengan Processor Generasi ke 8 Intel® Core™ i5-8250U dan Memori 8192 MB RAM dan perangkat lunak dengan Sistem Operasi Windows 10 Home, Perangkat Pembantu seperti Microsoft Excel 2016 dan Microsoft Word 2016.

4.2 Penambangan Data

Setelah tahap analisis, dilanjut untuk melakukan penambangan data. Penambangan data dipergunakan untuk *scraping* data dari *source* yang akan digunakan pada penelitian ini. Untuk penambangan data menggunakan *software* WebHarvy dan pemrograman Python yang mana data hasil dari penambangan data ini diunduh dalam bentuk file excel (.xls). Keseluruhan jumlah data yang di crawl dapat dilihat pada Tabel L4.1 Jumlah Data.

4.3 Pelabelan

Selanjutnya adalah pelabelan. Untuk tahap pelabelan ini diambil dengan data dari Tripadvisor 50 review, Twitter 50 review dan Google Maps Reviews 50 review dengan total jumlah masing-masing destinasi adalah 150 review. Pada tahapan ini dilakukan pemberian *name* kelas pada data yang sudah di *crawl*. Kelas yang diberikan adalah *positive* dan *negative*. Gambar L5.1 adalah beberapa contoh

data yang telah diberi label kelas dengan bantuan sukarelawan..

4.4 Data Preprocessing

Setelah dilakukan pelabelan, selanjutnya adalah tahap *data preprocessing*. Pada tahap ini dilakukan pengolahan terhadap hasil penambangan yang formatnya tidak terstruktur hingga menjadi kata dasar. Hal pertama yang dilakukan adalah melakukan *case folding*, *tokenizing*, *filtering*, dan *stemming*. *Case folding* bertujuan untuk mengubah semua review menjadi huruf kecil (lowercases), karena pada text review banyak memiliki huruf yang tidak sama. Contoh data sebelum dan sesudah melalui tahap *case folding* ditampilkan pada Tabel 4.1

Tabel 4. 1 Hasil Tahap *Case Folding*

Sebelum <i>case folding</i>	Sesudah <i>case folding</i>
Sunrise in <i>borobudur temple</i> . Mystical place full of history, beautiful but always crowded both at sunrise and sunset. If you are in Indonesia it is a must .. to visit without doubts	sunrise in <i>borobudur temple</i> . mystical place full of history, beautiful but always crowded both at sunrise and sunset. if you are in Indonesia it is a must .. to visit without doubts
Tempatnya sangat luas, asri. Di masa pandemi ini tetap buka dengan protokol kesehatan pastinya, tetapi tidak bisa naik ke candinya hanya boleh di bawah candi. Tapi masih banyak tempat wisata didalamnya seperti museum dll.	tempatnya sangat luas, asri. di masa pandemi ini tetap buka dengan protokol kesehatan pastinya, tetapi tidak bisa naik ke candinya hanya boleh di bawah candi. tapi masih banyak tempat wisata didalamnya seperti museum dll.

Setelah melakukan *case folding*, langkah selanjutnya adalah tahap *tokenizing* yang dilakukan pemotongan kalimat menjadi beberapa token atau bagian dan penghilangan symbol, tanda baca dan yang bukan huruf. Contoh information sebelum dan sesudah melalui tahap *tokenizing* ditampilkan pada Tabel 4.2..

Tabel 4. 2 Hasil Tahap *Tokenizing*

Sebelum <i>tokenizing</i>	Sesudah <i>tokenizing</i>
sunrise in <i>borobudur temple</i> . mystical place full of history, beautiful but always crowded both at sunrise and sunset. if you are in Indonesia it is a must .. to visit without doubts	sunrise in <i>borobudur temple</i> mystical place full of history beautiful but always crowded both at sunrise and sunset if you are in Indonesia it is a must to visit without doubts
tempatnya sangat luas, asri. di masa pandemi ini tetap buka dengan protokol kesehatan pastinya, tetapi tidak bisa naik ke candinya hanya boleh di bawah candi. tapi masih banyak tempat wisata didalamnya seperti museum dll.	tempatnya sangat luas asri di masa pandemi ini tetap buka dengan protokol kesehatan pastinya tetapi tidak bisa naik ke candinya hanya boleh di bawah candi tapi masih banyak tempat wisata didalamnya seperti museum dll

Setelah melakukan *tokenizing*, langkah selanjutnya adalah tahap *filtering* untuk penghilangan kata yang tidak relevan, penghilangan kata yang tidak ada arti seperti kata ganti, kata sambung dan kata ganti. Contoh data sebelum dan sesudah melalui tahap *filtering* ditampilkan pada Tabel 4.3.

Tabel 4. 3 Hasil Tahap *Filtering*

Sebelum <i>filtering</i>	Sesudah <i>filtering</i>
sunrise in <i>borobudur temple</i> mystical place full of history beautiful but always crowded both at sunrise and sunset if you are in Indonesia it is a must to visit without doubts	sunrise <i>borobudur temple</i> mystical place full history beautiful crowded sunrise sunset Indonesia visit doubts
tempatnnya sangat luas asri di masa pendemi ini tetap buka dengan protokol kesehatan pastinya tetapi tidak bisa naik ke candinya hanya boleh di bawah candi tapi masih banyak tempat wisata didalamnya seperti museum dll	tempatnnya luas asri masa pendemi tetap buka protokol kesehatan pastinya tidak bisa naik candinya boleh bawah candi masih banyak tempat wisata didalamnya museum

Setelah melakukan proses *filtering* pada data review, langkah selanjutnya adalah proses *stemming* dilakukan untuk proses pencarian kata dasar dengan menghilangkan imbuhan. Contoh data sebelum dan sesudah melalui tahap *stemming* ditampilkan pada Tabel 4.4.

Tabel 4. 4 Hasil Tahap *Stemming*

Sebelum <i>stemming</i>	Sesudah <i>stemming</i>
sunrise <i>borobudur temple</i> mystical place full history beautiful crowded sunrise sunset Indonesia visit doubts	sunrise <i>borobudur temple</i> mystic place full history beautiful crowd sunrise sunset Indonesia visit doubt
tempatnnya luas asri masa pendemi tetap buka protokol kesehatan pastinya tidak bisa naik candinya boleh bawah candi masih banyak tempat wisata didalamnya museum	tempat luas asri masa pendemi tetap buka protokol kesehatan pasti tidak naik candi boleh bawah candi masih banyak tempat wisata dalam museum

Berdasarkan tahap dari *data preprocessing*, dapat diketahui *word frequency* tiap atribut atau kata. *Word frequency* menunjukkan kata yang paling populer dari data yang telah dilakukan *data preprocessing*. Frekuensi dari kata yang sering muncul sangat penting karena digunakan untuk memprediksi kelas sentimen.

4.1.1 Frekuensi Review *Dieng Plateau*

Sepuluh kata yang sering muncul dari hasil penambangan data pada *Dieng Plateau* dapat dilihat pada Gambar L11.1. Gambar L11.1 (a) menunjukkan bahwa

kata-kata dengan sentimen *negative* lebih sering ditulis oleh wisatawan pada destinasi *Dieng Plateau* review bahasa Indonesia. Gambar L11.1 (b) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Dieng Plateau* review bahasa Inggris.

4.1.2 Frekuensi Review *Borobudur Temple*

Sepuluh kata yang sering muncul dari hasil penambangan data pada *Dieng Plateau* dapat dilihat pada Gambar L11.2. Gambar L11.2 (a) menunjukkan bahwa kata-kata dengan sentimen *negative* lebih sering ditulis oleh wisatawan pada destinasi *Borobudur Temple* review bahasa Indonesia. Gambar L11.2 (b) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Borobudur Temple* review bahasa Inggris.

4.1.3 Frekuensi Review *Komodo National Park*

Sepuluh kata yang sering muncul dari hasil penambangan data pada *Dieng Plateau* dapat dilihat pada Gambar L11.3. Gambar L11.3 (a) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Komodo National Park* review bahasa Indonesia. Gambar L11.3 (b) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Komodo National Park* review bahasa Inggris.

4.1.4 Frekuensi Review *Raja Ampat*

Sepuluh kata yang sering muncul dari hasil penambangan data pada *Dieng Plateau* dapat dilihat pada Gambar L11.4. Gambar L11.4 (a) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Raja Ampat* review bahasa Indonesia. Gambar L11.4 (b) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Raja Ampat* review bahasa Inggris.

4.1.5 Frekuensi Review *Mount Bromo*

Sepuluh kata yang sering muncul dari hasil penambangan data pada *Dieng Plateau* dapat dilihat pada Gambar L11.5. Gambar L11.5 (a) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Mount Bromo* review bahasa Indonesia. Gambar L11.5 (b) menunjukkan

bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Mount Bromo* review bahasa Inggris.

4.1.6 Frekuensi Review *Lake Toba*

Sepuluh kata yang sering muncul dari hasil penambangan data pada *Dieng Plateau* dapat dilihat pada Gambar L11.6. Gambar L11.6 (a) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Lake Toba* review bahasa Indonesia. Gambar L11.6 (b) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Lake Toba* review bahasa Inggris.

4.1.7 Frekuensi Review *Ubud*

Sepuluh kata yang sering muncul dari hasil penambangan data pada *Dieng Plateau* dapat dilihat pada Gambar L11.7. Gambar L11.7 (a) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Ubud* review bahasa Indonesia. Gambar L11.7 (b) menunjukkan bahwa kata-kata dengan sentimen *positive* lebih sering ditulis oleh wisatawan pada destinasi *Ubud* review bahasa Inggris.

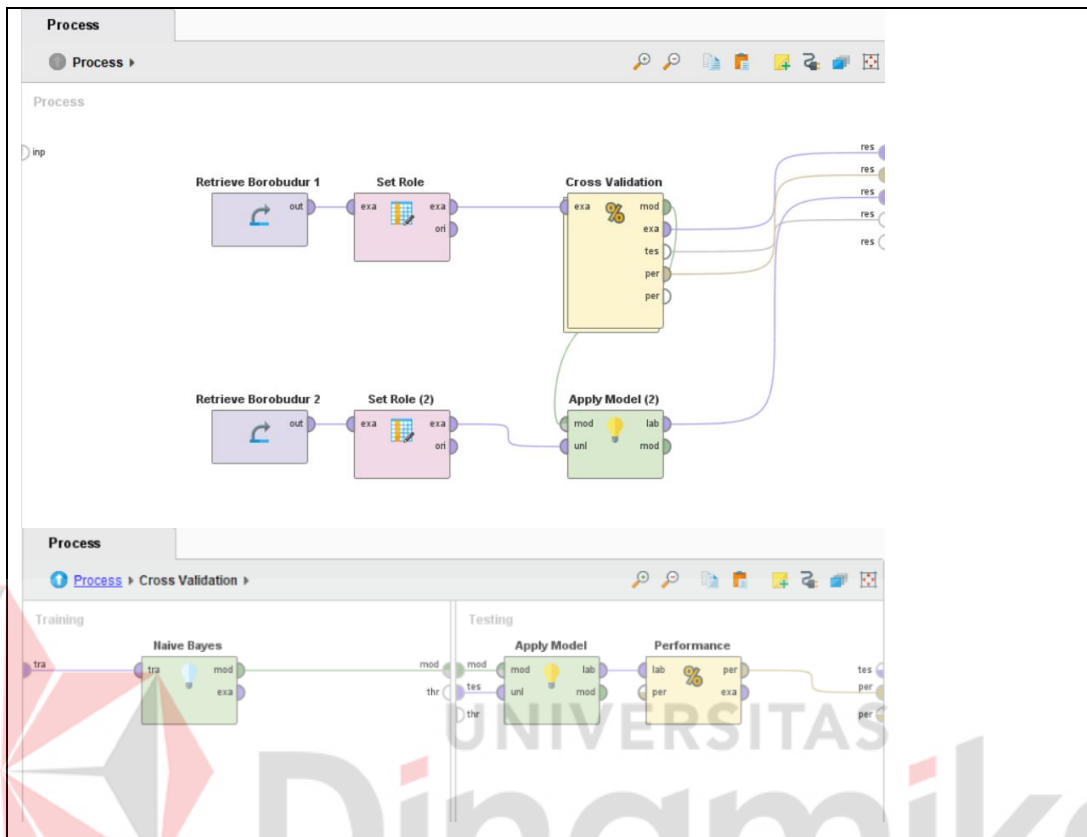
4.5 Training Data dan Testing Data

Data review akan dipecah (*split*) menjadi *training data* dan *testing data* dengan perbandingan 90:10. *Training data* untuk menyesuaikan model dengan kelas *positive* dan *negative*, serta jumlah dari *training data* lebih banyak yaitu sebesar 0.9 atau 90% dari masing-masing destinasi. *Testing data* untuk memberikan evaluasi tentang kesesuaian model pada set data pelatihan dan jumlah dari *testing data* sebesar 10% atau 0.1.

4.6 Klasifikasi *Naïve Bayes*

Naïve Bayes Classification adalah metode untuk memprediksi peluang suatu kelas. Hal pertama yang dilakukan dalam pengklasifikasian adalah menguji model menggunakan *training data*. Kemudian model yang sudah diuji dengan *training data* digunakan untuk mengklasifikasikan *testing data* ke *negative* dan *positive*. Untuk penerapan validasi *10-fold cross validation* digunakan untuk

menghasilkan akurasi *naïve bayes*. Desain model *Naïve Bayes* yang terdapat pada RapidMiner seperti Gambar 4.6.



Gambar 4. 1 Desain Model Validation *Naïve Bayes Classifier*

Pada gambar diatas, proses diawali dengan *retrieve* yang berfungsi untuk menginputkan data *review* ke dalam RapidMiner. Setelah itu, dilakukan *set Role* yaitu untuk membedakan baris penamaan atribut dan prediksi posisi yang akan di masukan kedalam kategori *label* untuk kolom sentimen, agar pada saat pengkategorian data *label* tidak ikut serta terhitung dan merubah hasil. Langkah selanjutnya adalah *split data* yang digunakan untuk membagi data *retrieve* menjadi dua, yaitu *training data* sebesar 0.9 atau 90% dan *testing data* sebesar 0.1 atau 10%. Selanjutnya dilakukan *validation* menggunakan *10-fold cross validation*. Didalam *validation* terdapat dua proses, *training* dan *testing*. Didalam proses *training* terdapat algoritma kalsifikasi yang diterapkan yaitu *Naïve Bayes*, sedangkan di dalam proses *testing* terdapat *Apply Model* untuk menjalankan algoritma atau model *Naïve Bayes* dan *Performance* untuk mengukur performa dari model *Naïve Bayes* tersebut.

4.7 Evaluasi dan Validasi *Naïve Bayes*

Ditahap ini adalah hasil pemodelan yang telah diproses oleh *Tools RapidMiner*. Setelah dilakukan pemodelan *Naive Bayes*, dapat diketahui probabilitas klasifikasi dengan metode *naïve bayes* pada 7 destinasi yang dapat dilihat pada subbab 1. *Naïve Bayes* Lampiran 12. Hasil Evaluasi. Jika probabilitas sentimen *positive* lebih tinggi daripada probabilitas sentimen *negative*, maka review tersebut akan menghasilkan hasil keputusan *positive* begitupun sebaliknya.

4.7.1 Evaluasi dan Validasi NB *Dieng Plateau*

Pada Table L12.1, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.754052 dan confidence *negativenya* 0.245948. Dan pada Table L12.2, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.017905 dan confidence *negativenya* 0.982095.

4.7.2 Evaluasi dan Validasi NB *Borobudur Temple*

Pada Table L12.3, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.058985 dan confidence *negativenya* 0.941015. Dan pada Table L12.4, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.90711 dan confidence *negativenya* 0.09289.

4.7.3 Evaluasi dan Validasi NB *Komodo National Park*

Pada Table L12.5, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.556139 dan confidence *negativenya* 0.443861. Dan pada Table L12.6, data

komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.154101 dan confidence *negativenya* 0.845899.

4.7.4 Evaluasi dan Validasi NB *Raja Ampat*

Pada Table L12.7, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.660983 dan confidence *negativenya* 0.339017. Dan pada Table L12.8, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.01096 dan confidence *negativenya* 0.98904.

4.7.5 Evaluasi dan Validasi NB *Mount Bromo*

Pada Table L12.9, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.831934 dan confidence *negativenya* 0.168066. Dan pada Table L12.10, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.411685461 dan confidence *negativenya* 0.588314539.

4.7.6 Evaluasi dan Validasi NB *Lake Toba*

Pada Table L12.11, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.666722 dan confidence *negativenya* 0.333278. Dan pada Table L12.12, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive*

yang mana itu karena confidence *positivenya* sebesar 0.064595 dan confidence *negativenya* 0.935405.

4.7.7 Evaluasi dan Validasi NB Ubud

Pada Table L12.13, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *negative* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.104291 dan confidence *negativenya* 0.895709. Dan pada Table L12.14, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.999748 dan confidence *negativenya* 2.52e-04.

Langkah selanjutnya adalah validitas data untuk mengetahui tingkat keakuratan data yang dihasilkan dengan menggunakan Rapidminer. Pada pengujian validasi data menggunakan *10-Fold Cross Validation*, proses pengujian dapat dilihat pada Gambar 4.6 dan data dibagi menjadi *training testing* dengan perbandingan 90:10. Hasil dari uji validitas dan waktu eksekusi data menggunakan *Naïve Bayes* pada 7 destinasi dalam review bahasa Indonesia dapat dilihat pada Tabel 4.5 dan 7 destinasi dalam review bahasa Inggris dapat dilihat pada Tabel 4.6.

Tabel 4. 5 Nilai Akurasi NB pada 7 Destinasi Bahasa Indonesia

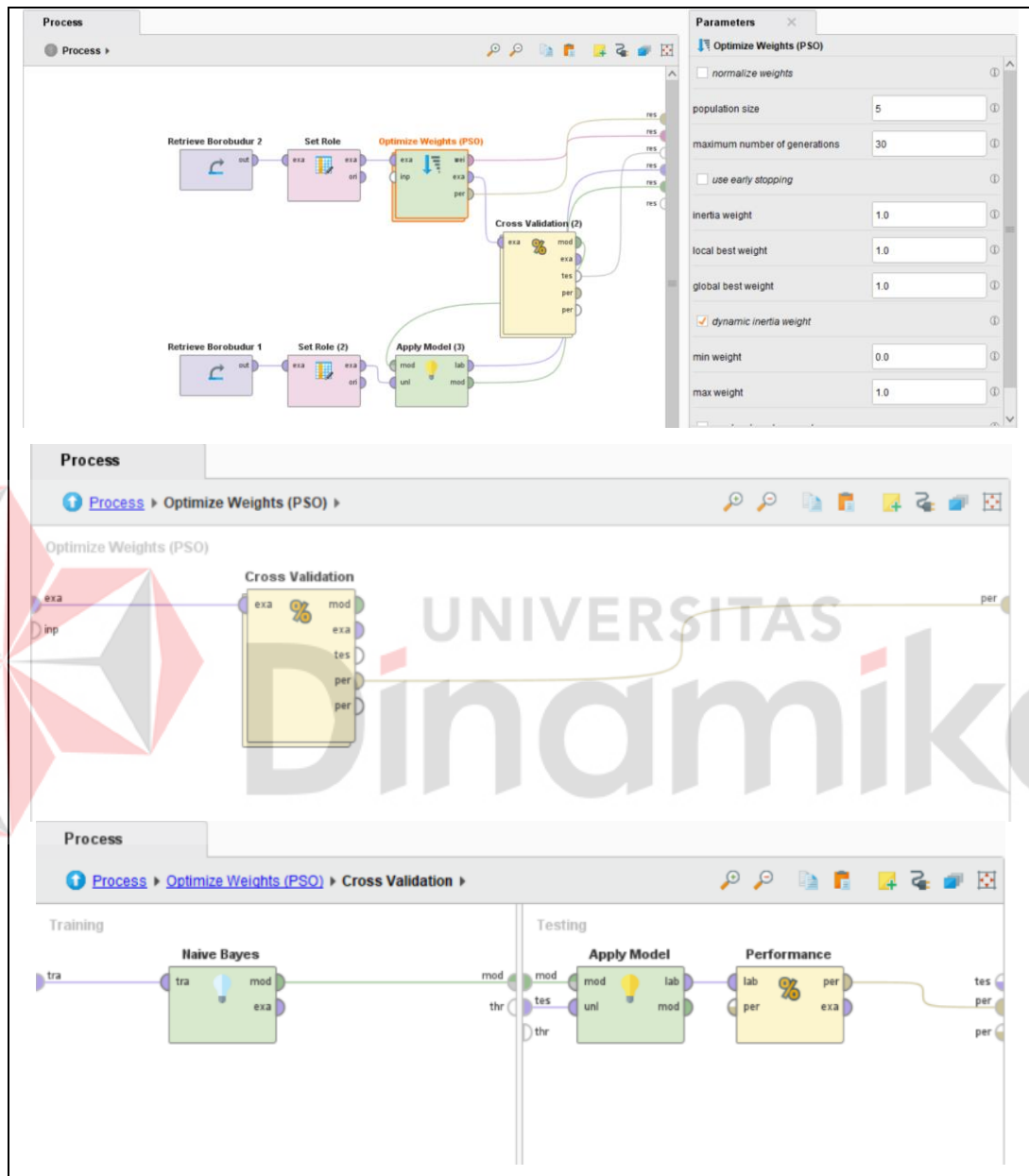
Fold ke	<i>Dieng Plateau</i>	Waktu	<i>Borobudur Temple</i>	Waktu	<i>Komodo National</i>	Waktu	<i>Raja Ampat</i>	Waktu	<i>Mount Bromo</i>	Waktu	<i>Lake Toba</i>	Waktu	<i>Ubud</i>	Waktu
2	33.04%	8ms	33.04%	7ms	26.79%	5ms	40.18%	9ms	40.18%	6ms	66.07%	7ms	46.43%	9ms
3	60.00%	8ms	60.00%	7ms	26.79%	6ms	46.67%	9ms	40.00%	9ms	60.00%	6ms	33.33%	9ms
4	60.42%	8ms	47.92%	9ms	26.79%	9ms	43.75%	11ms	39.58%	10ms	66.67%	9ms	45.83%	11ms
5	60.00%	12ms	53.33%	9ms	33.33%	9ms	46.67%	14ms	40.00%	12ms	60.00%	7ms	46.67%	11ms
6	61.11%	12ms	50.00%	10ms	25.00%	8ms	50.00%	17ms	41.67%	16ms	69.44%	14ms	47.22%	8ms
7	59.52%	13ms	61.90%	10ms	45.24%	10ms	52.38%	17ms	40.48%	14ms	64.29%	11ms	47.62%	6ms
8	62.50%	12ms	43.75%	11ms	25.00%	9ms	56.25%	13ms	37.50%	20ms	68.75%	14ms	50.00%	9ms
9	55.56%	9ms	61.11%	11ms	22.22%	12ms	50.00%	15ms	38.89%	19ms	72.22%	17ms	50.00%	11ms
10	60.00%	10ms	60.00%	11ms	20.00%	11ms	50.00%	17ms	35.00%	17ms	70.00%	15ms	50.00%	9ms
AVG	56.91%	11.2ms	52.34%	9.4ms	27.91%	8.8ms	48.43%	13.6ms	39.26%	13.7ms	66.38%	11.1ms	46.34%	9.2ms

Tabel 4. 6 Nilai Akurasi NB pada 7 Destinasi Bahasa Inggris

Fold ke	<i>Dieng Plateau</i>	Waktu	<i>Borobudur Temple</i>	Waktu	<i>Komodo National</i>	Waktu	<i>Raja Ampat</i>	Waktu	<i>Mount Bromo</i>	Waktu	<i>Lake Toba</i>	Waktu	<i>Ubud</i>	Waktu
2	47.32%	12ms	79.46%	64ms	73.21%	9ms	60.71%	10ms	53.57%	14ms	86.61%	5ms	46.43%	9ms
3	60.00%	13ms	66.67%	24ms	86.67%	13ms	73.33%	13ms	66.67%	13ms	73.33%	7ms	60.00%	11ms
4	52.08%	10ms	93.75%	33ms	75.00%	14ms	62.50%	16ms	52.08%	14ms	75.00%	10ms	72.92%	12ms
5	53.33%	10ms	86.67%	21ms	80.00%	15ms	73.33%	15ms	46.67%	17ms	73.33%	11ms	73.33%	15ms
6	63.89%	17ms	88.89%	34ms	86.11%	11ms	63.89%	19ms	58.33%	18ms	66.67%	11ms	75.00%	14ms
7	52.38%	15ms	85.71%	40ms	78.57%	16ms	73.81%	21ms	52.38%	18ms	66.67%	15ms	76.19%	16ms
8	56.25%	20ms	87.50%	34ms	81.25%	16ms	62.50%	24ms	43.75%	31ms	68.75%	16ms	75.00%	18ms
9	61.11%	17ms	77.78%	40ms	83.33%	18ms	66.67%	12ms	50.00%	30ms	77.78%	15ms	72.22%	13ms
10	60.00%	16ms	80.00%	41ms	80.00%	17ms	60.00%	18ms	40.00%	30ms	75.00%	15ms	75.00%	13ms
AVG	56.26%	14.4ms	82.94%	36.8ms	80.46%	14.3ms	66.30%	16.4ms	51.49%	20.6ms	73.68%	11.7ms	69.57%	13.4ms

4.8 Klasifikasi Naïve Bayes – PSO

Desain Model *Naïve Bayes* berbasis *PSO* pada pembobotan atribut menggunakan Rapidminer dapat dilihat pada gambar 4.7.



Gambar 4. 2 Desain Model Validation Naïve Bayes berbasis PSO

Optimize Weights digunakan untuk penerapan PSO pada pembobotan atribut. Untuk validasi menggunakan *10-fold cross-validation* yang terdapat dua proses yaitu *training* dan testing. Didalam proses *training* terdapat algoritma *naïve bayes* dan di proses *testing* terdapat *apply model* untuk mengeksekusi hasil

dari model dan *performance* untuk menghasilkan performa dari model *naïve bayes - pso*.

4.9 Evaluasi dan Validasi Naïve Bayes – PSO

Pada pengujian ini dilakukan evaluasi dan validasi menggunakan Algoritma *Naïve Bayes* dengan *PSO*. Tujuan untuk menghasilkan akurasi dari Algoritma *Naïve Bayes* dengan *PSO* akan di bandingkan dengan pengujian sebelumnya. Pengujian validitas data dari metode *Naïve Bayes - PSO* pada 7 destinasi. Pengujian ini menggunakan *10-Fold Cross Validation* dan data dibagi menjadi *training testing* dengan perbandingan 90:10. Nilai pada parameter *population size* dan *maximum number of generation* bernilai default yaitu 5 dan 30.

4.9.1 Evaluasi dan Validasi NB – PSO Dieng Plateau

Pada Table L12.15, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.754052 dan confidence *negativenya* 0.245948. Dan pada Table L12.16, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.017905 dan confidence *negativenya* 0.982095.

4.9.2 Evaluasi dan Validasi NB – PSO Borobudur Temple

Pada Table L12.17, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.058985 dan confidence *negativenya* 0.941015. Dan pada Table L12.18, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.90711 dan confidence *negativenya* 0.09289.

4.9.3 Evaluasi dan Validasi NB – PSO Komodo National Park

Pada Table L12.19, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.556139 dan confidence *negativenya* 0.443861. Dan pada Table L12.20, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.154101 dan confidence *negativenya* 0.845899.

4.9.4 Evaluasi dan Validasi NB – PSO Raja Ampat

Pada Table L12.21, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.660983 dan confidence *negativenya* 0.339017. Dan pada Table L12.22, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.01096 dan confidence *negativenya* 0.98904.

4.9.5 Evaluasi dan Validasi NB – PSO Mount Bromo

Pada Table L12.23, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.831934 dan confidence *negativenya* 0.168066. Dan pada Table L12.24, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.411685461 dan confidence *negativenya* 0.588314539.

4.9.6 Evaluasi dan Validasi NB – PSO Lake Toba

Pada Table L12.25, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen

menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.666722 dan confidence *negativenya* 0.333278. Dan pada Table L12.26, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.064595 dan confidence *negativenya* 0.93540.

4.9.7 Evaluasi dan Validasi NB – PSO Ubud

Pada Table L12.27, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *negative* dan hasil prediksi analisis sentimen menunjukkan hasil *negative* yang mana itu karena confidence *positivenya* sebesar 0.104291 dan confidence *negativenya* 0.895709. Dan pada Table L12.28, data komentar pertama review yang sudah dilabeli manual dikategorikan sebagai review *positive* dan hasil prediksi analisis sentimen menunjukkan hasil *positive* yang mana itu karena confidence *positivenya* sebesar 0.999748 dan confidence *negativenya* 2.52e-04.

Langkah selanjutnya adalah validitas data untuk mengetahui tingkat keakuratan data yang dihasilkan menggunakan *tools* Rapidminer. Pada pengujian validasi data menggunakan *10-Fold Cross Validation*, proses pengujian dapat dilihat pada Gambar 4.7 dan data dibagi menjadi *training testing* dengan perbandingan 90:10. Nilai pada parameter *population size* dan *maximum number of generation* bernilai default yaitu 5 dan 30. Hasil dari uji validitas dan waktu eksekusi data menggunakan *Naïve Bayes* pada 7 destinasi dalam review bahasa Indonesia dapat dilihat pada Tabel 4.5 dan 7 destinasi dalam review bahasa Inggris dapat dilihat pada Tabel 4.6.

Tabel 4. 7 Nilai Akurasi NB – PSO pada 7 Destinasi Bahasa Indonesia

Fold ke	<i>Dieng Plateau</i>	Waktu	<i>Borobudur Temple</i>	Waktu	<i>Komodo National</i>	Waktu	<i>Raja Ampat</i>	Waktu	<i>Mount Bromo</i>	Waktu	<i>Lake Toba</i>	Waktu	<i>Ubud</i>	Waktu
2	66.96%	276ms	60.71%	124ms	58.93%	115ms	59.82%	367ms	73.21%	145ms	73.21%	120ms	93.75%	110ms
3	66.67%	163ms	86.67%	159ms	53.33%	197ms	60.00%	435ms	73.33%	147ms	73.33%	330ms	93.33%	146ms
4	66.67%	371ms	87.50%	219ms	70.83%	192ms	60.00%	551ms	72.92%	246ms	68.75%	416ms	93.75%	185ms
5	66.67%	521ms	86.67%	196ms	60.00%	219ms	60.00%	602ms	73.33%	210ms	73.33%	506ms	93.33%	263ms
6	66.67%	622ms	86.11%	253ms	61.11%	252ms	58.33%	692ms	72.22%	246ms	77.78%	252ms	94.44%	612ms
7	66.67%	744ms	85.71%	294ms	69.05%	264ms	59.52%	643ms	73.81%	231ms	76.19%	299ms	92.86%	799ms
8	62.50%	746ms	87.50%	665ms	68.75%	345ms	56.25%	373ms	75.00%	315ms	68.75%	355ms	81.25%	830ms
9	66.67%	768ms	83.33%	848ms	61.11%	818ms	61.11%	358ms	50.00%	460ms	77.78%	377ms	94.44%	796ms
10	70.00%	791ms	85.00%	919ms	45.00%	925ms	60.00%	334ms	75.00%	413ms	80.00%	507ms	80.00%	1034ms
AVG	66.61%	556ms	83.24%	409ms	60.90%	370ms	59.45%	484ms	70.98%	268ms	74.35%	351ms	90.79%	531ms

Tabel 4. 8 Nilai Akurasi NB – PSO pada 7 Destinasi Bahasa Inggris

Fold ke	<i>Dieng Plateau</i>	Waktu	<i>Borobudur Temple</i>	Waktu	<i>Komodo National</i>	Waktu	<i>Raja Ampat</i>	Waktu	<i>Mount Bromo</i>	Waktu	<i>Lake Toba</i>	Waktu	<i>Ubud</i>	Waktu
2	66.96%	267ms	93.75%	167ms	93.75%	153ms	93.75%	230ms	66.96%	245ms	86.67%	118ms	73.21%	143ms
3	80.00%	361ms	93.33%	183ms	93.33%	148ms	93.33%	289ms	73.33%	340ms	87.50%	151ms	73.33%	194ms
4	75.00%	405ms	93.75%	510ms	93.75%	185ms	93.75%	371ms	75.00%	191ms	87.50%	294ms	81.25%	272ms
5	80.00%	664ms	93.33%	483ms	93.33%	525ms	93.33%	233ms	73.33%	216ms	86.67%	482ms	73.33%	308ms
6	83.33%	701ms	94.44%	595ms	94.44%	654ms	94.44%	296ms	72.22%	249ms	88.89%	602ms	77.78%	598ms
7	61.90%	850ms	95.24%	680ms	95.24%	692ms	95.24%	263ms	73.81%	281ms	88.10%	720ms	76.19%	882ms
8	68.75%	674ms	93.75%	354ms	93.75%	918ms	93.75%	325ms	75.00%	354ms	87.50%	792ms	81.25%	777ms
9	77.78%	809ms	94.44%	353ms	94.44%	983ms	94.44%	244ms	77.78%	389ms	88.89%	410ms	83.33%	805ms
10	85.00%	380ms	95.00%	416ms	95.00%	971ms	90.00%	375ms	80.00%	439ms	90.00%	384ms	85.00%	907ms
AVG	75.41%	568ms	94.11%	416ms	94.11%	581ms	93.56%	292ms	74.16%	300ms	87.97%	439ms	78.30%	543ms

4.10 Hasil Akhir dari Naïve Bayes dan Naïve Bayes - PSO

Setelah dilakukan pemodelan, kemudian dilakukan perbandingan hasil dari masing-masing metode. Maka diperoleh sebagai berikut:

Tabel 4. 9 Hasil Komparasi NB dan NB – PSO Bahasa Indonesia

	<i>Dieng Plateau</i>		<i>Borobudur</i>		<i>Komodo</i>		<i>Raja Ampat</i>		<i>Mount Bromo</i>		<i>Lake Toba</i>		<i>Ubud</i>	
	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO
<i>Accuracy</i>	62.50%	70.00%	61.90%	87.50%	45.24%	70.83%	56.25%	61.11%	41.67%	75.00%	72.22%	80.00%	50.00%	94.44%
<i>Prec</i>	100.00%	100.00%	61.90%	86.66%	50.00%	75.00%	100.00%	61.11%	100.00%	75.00%	83.33%	69.23%	27.27%	100.00%
<i>Sens</i>	40.00%	40.00%	100.00%	100.00%	7.14%	75.00%	22.22%	100.00%	25.00%	100.00%	55.56%	100.00%	100.00%	66.67%
<i>Spec</i>	100.90%	0.00%	100.00%	0.00%	12.50%	75.00%	100.00%	0.00%	100.00%	0.00%	55.56%	100.00%	100.00%	66.67%
NPV	45.45%	0.00%	25.00%	0.00%	50.00%	75.00%	46.15%	0.00%	30.77%	0.00%	83.83%	69.23%	27.27%	100.00%
FNR	60.00%	0.00%	46.15%	0.00%	14.28%	28.57%	77.77%	0.00%	81.81%	0.00%	16.67%	66.66%	66.68%	0.00%
FPR	0.00%	100.00%	0.00%	100.00%	87.50%	25.00%	0.00%	100.00%	0.00%	100.00%	44.44%	0.00%	0.00%	33.33%
Waktu	12ms	791ms	10ms	665ms	10ms	192ms	13ms	358ms	16ms	315ms	17ms	507ms	9ms	612ms

Tabel 4. 10 Hasil Komparasi NB dan NB – PSO Bahasa Inggris

	<i>Dieng Plateau</i>		<i>Borobudur</i>		<i>Komodo</i>		<i>Raja Ampat</i>		<i>Mount Bromo</i>		<i>Lake Toba</i>		<i>Ubud</i>	
	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO
<i>Accuracy</i>	63.89%	85.00%	93.75%	95.24%	86.67%	95.24%	73.81%	95.24%	66.67%	80.00%	86.61%	90.00%	76.19%	85.00%
<i>Prec</i>	55.56%	75.00%	91.67%	83.33%	100.00%	100.00%	88.89%	95.24%	83.33%	63.64%	83.33%	83.33%	76.19%	70.00%
<i>Sens</i>	83.33%	85.71%	100.00%	100.00%	83.33%	92.86%	64.29%	100.00%	61.11%	100.00%	100.00%	100.00%	100.00%	100.00%
<i>Spec</i>	71.42%	85.71%	100.00%	100.00%	100.00%	100.00%	75.00%	75.00%	57.14%	100.00%	60.00%	60.00%	100.00%	100.00%
NPV	55.55%	75.00%	83.33%	83.33%	66.66%	80.00%	50.00%	10.0%	66.66%	63.63%	100.00%	100.00%	63.63%	70.00%

	<i>Dieng Plateau</i>		<i>Borobudur</i>		<i>Komodo</i>		<i>Raja Ampat</i>		<i>Mount Bromo</i>		<i>Lake Toba</i>		<i>Ubud</i>	
	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO	NB	NB-PSO
FNR	50.00%	25.00%	10.00%	10.00%	18.18%	9.09%	27.27%	0.00%	25.00%	50.00%	0.00%	0.00%	50.00%	37.50%
FPR	28.57%	14.28%	0.00%	0.00%	0.00%	0.00%	25.00%	25.00%	42.85%	0.00%	40.00%	40.00%	0.00%	0.00%
Waktu	17ms	380ms	33ms	680ms	13ms	692ms	21ms	263ms	13ms	439ms	5ms	384ms	16ms	907ms

Hasil dari table diatas dapat disimpulkan bahwa :

1. Hasil akurasi dari model *naïve bayes* mempunyai rata-rata 66.96% dan *naïve bayes* - PSO memiliki akurasi yang lebih baik yang nilai akurasi rata-ratanya 83.19%. Pengujian dengan menggunakan PSO dapat meningkatkan akurasi hingga 16.23%. Maka artinya penggunaan PSO memberikan kinerja lebih baik atau keakuratan dari metode *naïve bayes*.
2. Hasil *precision or positive predictive value* (PPV) dari model *naïve bayes* mempunyai rata-rata 78.68% dan *naïve bayes* - PSO memiliki nilai yang lebih baik yang nilai rata-ratanya 81.25%. Pengujian dengan menggunakan PSO dapat meningkatkan nilai *precision* hingga 2.58%. Maka artinya penggunaan PSO meningkatkan keefektifan dari metode *naïve bayes*.
3. Hasil *sensitivity, recall, hit rate, or true positive rate* (TPR) dari model *naïve bayes* mempunyai rata-rata 67.28% dan *naïve bayes* - PSO memiliki nilai yang lebih baik yang nilai rata-ratanya 90.02%. Pengujian dengan menggunakan PSO dapat meningkatkan nilai *recall* hingga 22.73%. Maka artinya penggunaan PSO meningkatkan hasil analisis yang benar-benar *positive*.
4. Hasil *specificity, selectivity or true negative rate* (TNR) dari model *naïve bayes* mempunyai rata-rata 80.89% dan *naïve bayes* - PSO memiliki nilai yang lebih baik yang nilai rata-ratanya 61.60%. Pengujian dengan menggunakan PSO menurunkan hasil nilai *specificity* hingga 19.30%. Maka artinya penggunaan PSO menurunkan hasil analisis yang benar-benar *negative*.
5. Hasil *negative predictive value* (NPV) dari model *naïve bayes* memiliki nilai akurasi yang lebih baik dengan nilai akurasi rata-rata 56.74% dan *naïve bayes* - PSO memiliki nilai akurasi rata-ratanya 55.09%. Pengujian dengan menggunakan PSO menurunkan

hasil nilai NPV sebesar 1.64%. Maka artinya penggunaan PSO menurunkan hasil analisis yang benar-benar *negative* tapi hasilnya *positive*.

6. Hasil *false negative rate* (FNR) dari model *naïve bayes* mempunyai rata-rata 38.84% dan *naïve bayes* - PSO memiliki nilai yang lebih baik yang nilai rata-ratanya 16.20%. Pengujian dengan menggunakan PSO menurunkan hasil nilai NPV sebesar 22.64%. Maka artinya penggunaan PSO menurunkan rasio *positive* yang hasilnya *negative*.
7. Hasil *false positive rate* (FPR) dari model *naïve bayes* mempunyai rata-rata 19.17% dan *naïve bayes* - PSO memiliki nilai yang lebih baik yang nilai rata-ratanya 38.40%. Pengujian dengan menggunakan PSO dapat meningkatkan nilai *precision* hingga 19.23%. Maka artinya penggunaan PSO menurunkan rasio *negative* yang hasilnya *positive*.
8. Waktu eksekusi dari *naïve bayes* semua destinasi rata-rata 14.6 ms.
9. Waktu eksekusi dari *naïve bayes* + PSO semua destinasi rata-rata 513.2 ms.

Maka dari hasil perhitungan *accuracy*, *precision*, *sensitivity*, dan FPR terbukti bahwa PSO dapat meningkatkan hasil dari *naïve bayes*. Sedangkan penerapan PSO juga dapat menurunkan hasil dari *specificity*, NPV, dan FNR.

4.11 Visualisasi

4.11.1 Visualisasi Data *Dieng Plateau*

Berdasarkan Gambar L13.1 (a) diketahui bahwa kata *negative* yang sering muncul adalah “dingin” dan “tinggi”, dan dari Gambar L13.1 (b) diketahui bahwa kata *positive* yang sering muncul adalah *beautiful* dan *enjoy*. Jumlah data review dapat dilihat pada Gambar L13.2. Pada Gambar L13.3, menunjukkan bahwa emosi yang didapat dari mengunjungi *Dieng Plateau* adalah *excited* dan *happy*.

Pada Gambar L13.4, menunjukkan persentase review yang sudah dilakukan klasifikasi. Presentase keseluruhan pada review bahasa Indonesia sebesar 76.3% untuk *negative* dan 23.7% untuk *positive*, sedangkan pada review bahasa Inggris sebesar 47.4% untuk *negative* dan 52.6% untuk *positive*.

Pada Gambar L13.5, menunjukkan akun mana yang sering mereview *Dieng Plateau* dari 3 *source*. Di review bahasa Indonesia, akun abby mereview sebanyak 9 kali. Di review bahasa Inggris, akun Soutoureu? mereview sebanyak 4 kali.

4.11.2 Visualisasi Data *Borobudur Temple*

Berdasarkan Gambar L13.6 (a) diketahui bahwa kata *negative* lah yang sering muncul yaitu kata “panas” dan “salah”, dan dari Gambar L13.6 (b) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata *amazing* dan *worth*.

Pada Gambar L13.7, menunjukkan bahwa emosi yang didapat dari mengunjungi *Borobudur Temple* adalah *excited* dan *happy*. Pada Gambar L13.8, menunjukkan persentase review yang sudah dilakukan klasifikasi. Presentase keseluruhan pada review bahasa Indonesia sebesar 78,5% untuk *negative* dan 21.5% untuk *positive*, sedangkan pada review bahasa Inggris sebesar 20% untuk *negative* dan 80% untuk *positive*.

Pada Gambar L13.9, menunjukkan akun mana yang sering mereview *Borobudur Temple* dari 3 *source*. Di review bahasa Indonesia, akun bento ger dan JALAN TENGAH ? mereview sebanyak 3 kali. Di review bahasa Inggris, akun RAPISU mereview sebanyak 5 kali.

4.11.3 Visualisasi Data *Komodo National Park*

Gambar L13.10 (a) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata “indah” dan “bagus”, dan dari Gambar L13.10 (b) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata *great* dan *good*.

Pada Gambar L13.11, menunjukkan bahwa emosi yang didapat dari mengunjungi *Komodo National Park* adalah *excited* dan *happy*. Pada Gambar

L13.12, menunjukkan persentase review yang sudah dilakukan klasifikasi. Presentase keseluruhan pada review bahasa Indonesia sebesar 9.6% untuk *negative* dan 90.4% untuk *positive*, sedangkan pada review bahasa Inggris sebesar 14.8% untuk *negative* dan 85.2% untuk *positive*.

Pada Gambar L13.13, menunjukkan akun mana yang sering mereview *Komodo National Park* dari 3 *source*. Di review bahasa Indonesia, akun Sadboi mereview sebanyak 11 kali. Di review bahasa Inggris, akun Mark Komodo ??? B mereview sebanyak 77 kali

4.11.4 Visualisasi Data *Raja Ampat*

Berdasarkan Gambar L13.14 (a) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata “indah” dan “keindahan”, dan dari Gambar L13.14 (b) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata *top* dan *beautiful*

Pada Gambar L13.15, menunjukkan bahwa emosi yang didapat dari mengunjungi *Raja Ampat* adalah *excited* dan *happy*. Pada Gambar L13.16, menunjukkan persentase review yang sudah dilakukan klasifikasi. Presentase keseluruhan pada review bahasa Indonesia sebesar 72.6% untuk *negative* dan 24.4% untuk *positive*, sedangkan pada review bahasa Inggris sebesar 63.7% untuk *negative* dan 36.3% untuk *positive*.

Pada Gambar L13.5, menunjukkan akun mana yang sering mereview *Raja Ampat* dari 3 *source*. Di review bahasa Indonesia, akun Tour Hemat Bromo mereview sebanyak 4 kali. Di review bahasa Inggris, akun Alex dan Agys mereview sebanyak 2 kali.

4.11.5 Visualisasi Data *Mount Bromo*

Berdasarkan Gambar L13.18 (a) diketahui bahwa kata yang sering muncul yaitu kata “indah”, dan dari Gambar L13.18 (b) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata *top* dan *amazing*.

Pada Gambar L13.19, menunjukkan bahwa emosi yang didapat dari mengunjungi *Mount Bromo* adalah *excited* dan *happy*. Pada Gambar L13.20, menunjukkan persentase review yang sudah dilakukan klasifikasi. Presentase

keseluruhan pada review bahasa Indonesia sebesar 83.7% untuk *negative* dan 16.3% untuk *positive*, sedangkan pada review bahasa Inggris sebesar 23% untuk *negative* dan 77% untuk *positive*.

Pada Gambar L13.21, menunjukkan akun mana yang sering mereview *Mount Bromo* dari 3 *source*. Di review bahasa Indonesia, akun Tour Hemat Bromo mereview sebanyak 14 kali. Di review bahasa Inggris, akun ACC Corporation mereview sebanyak 8 kali.

4.11.6 Visualisasi Data *Lake Toba*

Berdasarkan Gambar L13.22 (a) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata “indah” dan “sejuk”, dan dari Gambar L13.22 (b) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata *beauti* dan *good*.

Pada Gambar L13.23, menunjukkan bahwa emosi yang didapat dari mengunjungi *Lake Toba* adalah *excited* dan *happy*. Pada Gambar L13.24, menunjukkan persentase review yang sudah dilakukan klasifikasi. Presentase keseluruhan pada review bahasa Indonesia sebesar 6.7% untuk *negative* dan 93.3% untuk *positive*, sedangkan pada review bahasa Inggris sebesar 56.3% untuk *negative* dan 43.7% untuk *positive*.

Pada Gambar L13.25, menunjukkan akun mana yang sering mereview *Lake Toba* dari 3 *source*. Di review bahasa Indonesia, akun Badan Pelaksana mereview sebanyak 5 kali. Di review bahasa Inggris, akun TM mereview sebanyak 78 kali.

4.11.7 Visualisasi Data *Ubud*

Berdasarkan Gambar L13.26 (a) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata “bagus” dan “sejuk”, dan dari Gambar L13.26 (b) diketahui bahwa kata *positive* lah yang sering muncul yaitu kata *great* dan *love*.

Pada Gambar L13.27, menunjukkan bahwa emosi yang didapat dari mengunjungi *Ubud* adalah *fear* dan *happy*. Pada Gambar L13.28, menunjukkan persentase review yang sudah dilakukan klasifikasi. Presentase keseluruhan pada review bahasa Indonesia sebesar 67.4% untuk *negative* dan 32.6% untuk *positive*,

sedangkan pada review bahasa Inggris sebesar 71.9% untuk *negative* dan 28.1% untuk *positive*.

Pada Gambar L13.29, menunjukkan akun mana yang sering mereview *Ubud* dari 3 *source*. Di review bahasa Indonesia dan bahasa Inggris, akun butter cup bali mereview sebanyak 38 kali dan 59 kali.

Dari penjelasan hasil masing-masing visualisasi, dapat disimpulkan bahwa:

1. Wisata *Komodo National Park* mendapatkan hasil sentiment *positive* terbanyak baik dari review bahasa Indonesia dan bahasa Inggris. Dilanjut *Toba* mendapatkan hasil peringkat kedua sentimen *positive*. Peringkat ketiga yaitu *Borobudur Temple*. Peringkat keempat yaitu *Mount Bromo*. Peringkat keempat yaitu *Dieng Plateau*. Peringkat keempat yaitu *Raja Ampat*. Peringkat keempat yaitu *Ubud*
2. Wisata *Toba Lake* mendapatkan emosi baik yang paling banyak dari 7 destinasi. Dilanjut oleh *Raja Ampat* mendapatkan peringkat kedua dalam emosi terbaik. Peringkat ketiga didapatkan oleh *Komodo National park*. Peringkat keempat didapatkan oleh *Mount Bromo*. Peringkat kelima didapatkan oleh *Dieng Plateau*. Peringkat keenam didapatkan oleh *Borobudur Temple* dan yang terakhir peringkat ketujuh didapatkan oleh *Ubud* dengan perolehan emosi yang kurang baik dari wisatawan.
3. Hal *negative* dari *Dieng Plateau* yaitu dingin dan tinggi, *Borobudur* adalah panas, *Komodo National Park* yaitu jauh dan panas, *Raja Ampat* yaitu jauh, *Mount Bromo* yaitu dingin, *Toba Lake* yaitu jauh dan *Ubud* adalah harganya yang mahal dan ramai.

BAB V

PENUTUP

Bab ini berisi hasil dari penelitian terkait pengerjaan tugas akhir ini. Pada kesimpulan, berisi kesimpulan dari hasil pengejaan tugas akhir. Sedangkan pada bagian saran, berisi saran yang berguna untuk mengembangkan penelitian tugas akhir ini.

5.1 Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan, dapat disimpulkan bahwa:

1. Data yang diolah menggunakan Algoritma *Naïve Bayes* mempunyai *accuracy* dengan rata-rata sebesar 66.96%, *precision* atau PPV dengan rata-rata sebesar 78.68%, *sensitivity* dengan rata-rata sebesar 67.28%, *specificity* dengan rata-rata sebesar 80.89%, *negative prediction value* dengan rata-rata sebesar 56.74%, *false negative rate* dengan rata-rata sebesar 38.84%, dan *false positive rate* dengan rata-rata sebesar 19.17%.
2. Data yang diolah menggunakan *Naïve Bayes* berbasis *Particle Swarm Optimization* mempunyai *accuracy* dengan rata-rata sebesar 83.19%, *precision* atau PPV dengan rata-rata sebesar 81.25%, *sensitivity* dengan rata-rata sebesar 90.02%, *specificity* dengan rata-rata sebesar 61.60%, *negative prediction value* dengan rata-rata sebesar 55.09%, *false negative rate* dengan rata-rata sebesar 16.20%, dan *false positive rate* dengan rata-rata sebesar 38.40%.
3. Penerapan *Particle Swarm Optimization* pada *Naïve Bayes* dapat meningkatkan nilai *accuracy* sebesar 16.23%, nilai *precision* atau PPV sebesar 2.58%, nilai *sensitivity* sebesar 22.73%, dan nilai *false negative rate* sebesar 19.23%. Namun penggunaan PSO dapat menurunkan hasil nilai dari *specificity* sebesar 19.30%, *negative prediction value* sebesar 1.64%, dan *false negative rate* sebesar 22.64%.
4. Waktu eksekusi dari *naïve bayes* di semua destinasi adalah 14.6 ms, sedangkan dari *naïve bayes* - PSO rata-ratanya 513.2 ms.
5. Hasil visualisasi dari persentase sentimen didapatkan *Komodo National*

Park mendapatkan hasil sentiment *positive* terbanyak. Peringkat kedua *Toba Lake*. Peringkat ketiga yaitu *Borobudur Temple*. Peringkat keempat yaitu Mount Bromo. Peringkat kelima yaitu Dieng Plateau. Peringkat keenam yaitu Raja Ampat dan terakhir yaitu Ubud.

6. Hasil visualisasi dari emosi didapatkan *Toba Lake* mendapatkan emosi baik yang paling banyak dari 7 destinasi. Dilanjut oleh Raja Ampat, kemudian Komodo National park, Mount Bromo, Dieng. Borobudur dan terakhir Ubud
7. Hal *negative* dari *Dieng Plateau* yaitu dingin dan tinggi, *Borobudur* adalah panas, *Komodo National Park* yaitu jauh dan panas, *Raja Ampat* yaitu jauh, *Mount Bromo* yaitu dingin, *Toba Lake* yaitu jauh dan *Ubud* adalah harganya yang mahal dan ramai.

5.2 Saran

Dalam pengerjaan tugas akhir ini masih terdapat kekurangan yang perlu diperbaiki. Berikut adalah beberapa saran yang bisa digunakan dalam melakukan pengembangan penelitian ke depan yang lebih baik.

1. Tahap *stemming* yang dilakukan menggunakan *stem porter* kurang tepat, karena banyak terjadi perubahan dalam perubahan kata dasar, seperti kata *amazing* menjadi *amaz*, dan *fabulous* menjadi *fabul*.
2. Mengganti pembagian proporsi penggunaan *training data* dan *testing data*.
3. Melakukan kamus tentang bahasa gaul, karena pada social media twitter review bahasa Indonesia terlalu banyak bahasa yang kurang baku.
4. Meningkatkan nilai *population size* serta menggunakan indikator yang berbeda untuk meningkatkan kinerja dari *feature selection Particle Swarm Optimization*.

DAFTAR PUSTAKA

- Abdillah, F., Taufik, I., & Jumadi. (2016). News Web Crawler Dengan Metode Xpath. *News Web Crawler Dengan Metode Xpath*, 2.
- Analysts, D. (2020, September 6). *insights-updates*. Retrieved from [www.destinationanalysts.com: https://www.destinationanalysts.com/insights-updates/](https://www.destinationanalysts.com/insights-updates/)
- Bajo, A. F. (2020, Mar 6). *tips/memilih-tempat-wisata*. Retrieved from [firstlabuanbajo.com: https://firstlabuanbajo.com/tips/memilih-tempat-wisata](https://firstlabuanbajo.com/tips/memilih-tempat-wisata)
- Gretzel, U., Zhong, L., & Koo, C. (2016). Application of Smart Tourism to Cities – Introduction to the Special Issue. *National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A3A2043345)*.
- Haq, F. U., & Rachmat, H. (2020). PENGGUNAAN GOOGLE REVIEW SEBAGAI PENILAIAN KEPUASAN PENGUNJUNG DALAM PARIWISATA. *Tornare - Journal of Sustainable Tourism Research Vol. 2, No. 1, Januari 2020: 10 - 12, 12*.
- Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *JURNAL SAINS DAN SENI ITS Vol. 6, No. 2 (2017) 2337-3520 (2301-928X Print), A52*.
- Jalal, A. A. (2020). Text Mining: Design of Interactive Search Engine Based Regular Expressions of Online Automobile Advertisements. *iJEP – Vol. 10, No. 3, 2020, 35 - 36*.
- Kirom, N. R., Sudarmiatin, & Putra, I. J. (2016). FAKTOR-FAKTOR PENENTU DAYA TARIK WISATA BUDAYA DAN PENGARUHNYA TERHADAP KEPUASAN WISATAWAN. *EISSN: 2502-471X, Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan Volume: 1 Nomor: 3 Bulan Maret Tahun 2016 Halaman: 536—546, 537*.
- Muhamad, H., Prasajo, C. A., Sugianto, N. A., Surtiningsih, L., & Cholissodin, I. (2017). OPTIMASI NAÏVE BAYES CLASSIFIER DENGAN MENGGUNAKAN PARTICLE SWARM OPTIMIZATION PADA DATA IRIS. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK) p-ISSN: 2355-7699 Vol. 4, No. 3, September 2017, hlm. 180-184 e-ISSN: 2528-6579, 180*.
- Nicoli, N., & Papadopoulou, E. (2017). TripAdvisor and reputation: a case study of the hotel industry in Cyprus. *EuroMed Journal of Business Vol. 12 No. 3, 2017 pp. 316-334 © Emerald Publishing Limited 1450-2194 DOI 10.1108/EMJB-11-2016-0031, 317*.
- Nugroho, K. S. (2019, Nov 13). Retrieved from Confusion Matrix untuk Evaluasi Model pada Supervised Learning: <https://medium.com/@ksnugroho/confusion-matrix-untuk-evaluasi-model-pada-unsupervised-machine-learning-bc4b1ae9ae3f>
- Prasetio, N. (2019, September 19). *Data Visualisasi : Definisi (Part 1)*. Retrieved from [medium.com: https://medium.com/data-folks-Indonesia/data-visualisasi-definisi-part-1-63d68599be29](https://medium.com/data-folks-Indonesia/data-visualisasi-definisi-part-1-63d68599be29)

- Putri, E. (2019, Juni 12). *The 15 Most Beautiful Places To Visit in Indonesia*. Retrieved from theculturetrip.com: <https://theculturetrip.com/asia/Indonesia/articles/the-10-most-beautiful-places-to-visit-in-Indonesia/>
- Que, V. K., Iriane, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis *Particle Swarm Optimization*. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* / Vol. 9, No. 2, Mei 2020, 165.
- Rahmat, A., Rahayu, S., & Pariyati, Y. (2020). Rancang Bangun Perangkat Lunak Aplikasi Katalog Elektronik Wisata Edukasi Berbasis Android. *Jurnal Algoritma* Vol. 17; No. 1; 2020; Hal 75-81, 80.
- Rahutomo, F., & Ririd, A. R. (2019). EVALUASI DAFTAR STOPWORD BAHASA INDONESIA. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* DOI: 10.25126/jtiik.201861226 Vol. 6, No. 1, Februari 2019, hlm. 41-48 p-ISSN: 2355-7699 Akreditasi KEMENRISTEKDIKTI, No. 30/E/KPT/2018 e-ISSN: 2528-6579, 42.
- Rozi, I. F., Hamdana, E. N., & Alfahmi, M. B. (2018). *PENGEMBANGAN APLIKASI ANALISIS SENTIMEN TWITTER MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER*, 149.
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes. *Citec Journal*, Vol. 2, No. 3, Mei 2015 – Juli 2015, 1-12.
- Sasmita, R. A. (2019). *PEMANFAATAN ALGORITMA TF/IDF UNTUK SISTEM INFORMASI e-COMPLAINT HANDLING*, 3.
- Shrivastava, V. (2018). A Methodical Study of Web Crawler. *ISSN : 2248-9622* Vol. 8, Issue 11 (Part -I), 1-8.
- Silalahi, M. (2018). ANALISIS CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS TERHADAP PENJUALAN PRODUK PADAPT BATAMAS NIAGA JAYA. *CBIS JOURNAL - VOL. 06 NO. 02 (2018) : SEPTEMBER*, 25-27.
- Wibowo, A. P., & Jumiati, E. (2018). Sentiment Analysis Masyarakat Pekalongan Terhadap Pembangunan Jalan Tol Pemalang-Batang Di Media Sosial. *IC-Tech Volume XIII No. 1 April 2018* <http://ejournal.stmik-wp.ac.id>, 43.
- Windasari, I. P., & Eridani, D. (2017). Sentiment Analysis on Travel Destination in Indonesia. *Proc. of 2017 4th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE)*, Oct 18-19, 2017, Semarang, Indonesia, 276.