

**Analisis Sentimen Menggunakan *Pseudo Nearest Neighbor* dan  
TF-IDF Text Vectorizer**

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



Oleh:

Yogi Pratama  
NIM : 09021381621133

**JURUSAN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA  
2021**

**LEMBAR PENGESAHAN TUGAS AKHIR**

**ANALISIS SENTIMEN MENGGUNAKAN PSEUDO NEAREST NEIGHBOR DAN  
TF-IDF TEXT VECTORIZER**

Oleh :

**YOGI PRATAMA**  
NIM : 09021381621133

Pembimbing I,



Dr. Abdiansah, S.Kom., M.CS.  
NIP. 198410012009121005

Palembang, Agustus 2021  
Pembimbing II,



Kanda Januar Miraswan, MT  
NIP. 199001092019031012

Mengetahui,  
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## LETTER OF PASS FINAL PROJECT EXAMINATION

On Friday, July 30 2021, the final examination was held by Department of Informatics Engineering, Faculty of Computer Science, Sriwijaya University.

Name : Yogi Pratama  
NIM : 09021381621133  
Title : Sentiment Analysis Using Pseudo Nearest Neighbor and TF-IDF  
Vectorizer

1. Supervisor I

Dr. Abdiansah, S.Kom., M.CS  
NIP. 198410012009121005



.....

2. Supervisor II

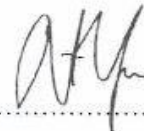
Kanda Januar Miraswan, MT  
NIP. 199001092019031012



.....

3. Examiner I

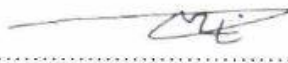
Novi Yusliani, M.T  
NIP. 198211082012122001



.....

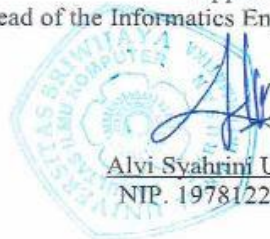
4. Examiner II

Osvari Arsalan, MT  
NIP. 198806282018031001



.....

Approved,  
Head of the Informatics Engineering Department



.....

Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Yogi Pratama  
NIM : 09021381621133  
Program Studi : Teknik Informatika  
Judul Skripsi : Analisis sentimen menggunakan *Pseudo Nearest Neighbor* dan *TF-IDF Vectorizer*

Hasil Pengecekan Software *iThenticate/Turnitin* : 20 %

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

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



Palembang, Agustus 2021



Yogi Pratama  
NIM. 09021381621133

## **MOTTO DAN PERSEMBAHAN**

“Lakukanlah hal yang ingin kau lakukan, jangan sampai kau menyesal  
saat kau mengingatnya kembali”

Kupersembahkan Karya tulis ini kepada:

- Kedua Orang tua
- Keluarga Besar
- Teman – Teman Seperjuangan
- Fakultas Ilmu Komputer
- Universitas Sriwiaya

## SENTIMENT ANALYSIS USING PSEUDO NEAREST NEIGHBOR AND TF-IDF TEXT VECTORIZER


Yogi Pratama  
09021381621133

### ABSTRACT

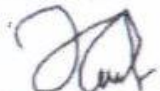
Twitter is one of the social media that is often used by researchers as an object of research to conduct sentiment analysis. Twitter is also a good indicator in influencing research, problems that often arise in research in the field of sentiment analysis are the many factors such as the use of colloquial or informal language and other factors that can affect sentiment results. To improve the results of sentiment classification, it is necessary to carry out a good information extraction process. One of the word weighting methods resulting from the information extraction process is the TF-IDF Vectorizer. This study examines the effect of the TF-IDF Vectorizer weighting results in sentiment analysis using the Pseudo Nearest Neighbor method. The results of the f-measure classification of sentiment using the TF-IDF Vectorizer at parameters k-2 = 89%, k-3 = 89%, k-4 = 71% and k-5 = 75% while without using the TF-IDF Vectorizer on the parameters k-2 = 90%, k-3 = 92%, k-4 = 84% and k-5 = 89%. From the results of the classification of sentiment analysis that does not use the TF-IDF Vectorizer, the f-measure value is slightly better than using it.

**Keywords:** Sentiment Analysis, TF-IDF Vectorizer, Pseudo Nearest Neighbor


Supervisor I,

  
Dr. Abdiansah, S.Kom., M.CS.  
NIP. 198410012009121005

Palembang, August 2021  
Supervisor II,

  
Kanda Januar Miraswan, MT  
NIP. 199001092019031012

Approved,  
Head of the Informatics Engineering Department

  
Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

# ANALISIS SENTIMEN MENGGUNAKAN *PSEUDO NEAREST NEIGHBOR* DAN *TF-IDF TEXT VECTORIZER*

Yogi Pratama  
09021381621133

## ABSTRAK

Twitter merupakan salah satu media sosia yang sering digunakan oleh peneliti sebagai objek riset untuk melakukan analisis sentimen. Twitter juga menjadi indikator yang baik dalam memberi pengaruh dalam penelitian, masalah yang sering muncul dalam penelitian di bidang analisis sentimen adalah banyaknya faktor-faktor seperti penggunaan bahasa sehari-hari atau bahasa tidak formal dan faktor lainnya yang dapat mempengaruhi hasil sentimen. Untuk meningkatkan hasil klasifikasi sentimen, perlu dilakukan proses ekstraksi informasi yang baik. Salah satu metode pembobotan kata hasil proses ekstraksi informasi adalah *TF-IDF Vectorizer*. Penelitian ini menguji pengaruh hasil pembobotan *TF-IDF Vectorizer* dalam analisis sentimen menggunakan metode *Pseudo Nearest Neighbor*. Hasil *f-measure* klasifikasi sentimen dengan menggunakan *TF-IDF Vectorizer* pada parameter  $k=2 = 89\%$ ,  $k=3 = 89\%$ ,  $k=4 = 71\%$  dan  $k=5 = 75\%$  sedangkan tanpa menggunakan *TF-IDF Vectorizer* pada parameter  $k=2 = 90\%$ ,  $k=3 = 92\%$ ,  $k=4 = 84\%$  dan  $k=5 = 89\%$ . Dari hasil klasifikasi analisis sentimen yang tidak menggunakan *TF-IDF Vectorizer* mempunyai nilai *f-measure* yang sedikit lebih baik daripada menggunakannya.

**Kata Kunci:** Analisis Sentimen, *TF-IDF Vectorizer*, *Pseudo Nearest Neighbor*

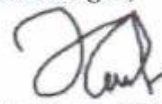
Pembimbing I,



Dr. Abdiansah, S.Kom., M.CS.  
NIP. 198410012009121005

Palembang, Agustus 2021

Pembimbing II,



Kanda Januar Miraswan, MT  
NIP. 199001092019031012

Mengetahui,

Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## **KATA PENGANTAR**

Selama penelitian dan penyusunan laporan penelitian skripsi ini, penulis tidak luput dari kendala dan hambatan. Namun demikian kendala dan hambatan tersebut dapat penulis atasi berkat bantuan, bimbingan dan dukungan dari berbagai pihak. Oleh karena itu dalam kesempatan ini, penulis ingin menyampaikan rasa terima kasih sebesar-besarnya kepada:

1. Kedua orang tua saya dan keluarga yang telah memberikan dukungan baik moril maupun materik serta doa yang tiada henti-hentinya kepada penulis.
2. Bapak Dr. Abdiansah, S.Kom., M.Cs. selaku dosen pembimbing I dan Bapak Kanda Januar Miraswan, M.T. selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi kepada penulis dalam proses pengerjaan Tugas Akhir, hingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
3. Bapak Danny Matthew Saputra, M.Sc. selaku dosen pembimbing akademik yang telah membimbing, mengarahkan dan memberikan motivasi kepada penulis dalam proses perkuliahan dan pengerjaan Tugas Akhir.
4. Ibu Novi Yusliani, M.T. selaku dosen penguji I, dan Bapak Osvari Arsalan, MT. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
5. Seluruh dosen program studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.



6. Mbak Wiwin dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
7. Luthfi dan Astero, yang telah banyak membantu, memberikan dukungan, informasi dan referensi selama proses pengerjaan Tugas Akhir hingga penulis dapat menyelesaikan Tugas Akhir dengan baik.
8. Diananata, Fadhlana, Reza, Angga, Hara serta seluruh teman jurusan Teknik Informatika yang tidak dapat disebutkan satu persatu yang telah berbagi keluh kesah, motivasi, semangat, dan hiburan selama masa perkuliahan.

Penulis menyadari dalam penyusunan Tugas Akhir ini Masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Agustus 2021

Yogi Pratama

## DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
DAFTAR ISI.....	v
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	x
BAB I PENDAHULUAN.....	I-1
1.1 Pendahuluan.....	I-1
1.2 Latar Belakang Masalah.....	I-1
1.3 Rumusan Masalah.....	I-3
1.4 Tujuan Penelitian.....	I-4
1.5 Manfaat Penelitian.....	I-4
1.6 Batasan Masalah.....	I-5
1.7 Sistematika Penulisan.....	I-5
1.7.1 Bab I. Pendahuluan.....	I-5
1.7.2 Bab II. Kajian Literatur.....	I-5
1.7.3 Bab III. Metodologi Penelitian.....	I-6
1.8 Kesimpulan.....	I-7
BAB II KAJIAN LITERATUR.....	II-1
2.1 Pendahuluan.....	II-1
2.2 Landasan Teori.....	II-1
2.2.1 <i>Text Mining</i> .....	II-1
2.2.2 Analisis Sentimen.....	II-2
2.2.3 <i>Pseudo Nearest Neighbor</i> .....	II-2
2.2.4 TF-IDF.....	II-3
2.2.5 <i>Confusion Matrix</i> .....	II-4
2.2.5.1 <i>Recall (Sensitivity)</i> .....	II-5
2.2.5.2 <i>Precision</i> .....	II-5

2.2.5.3 <i>F-Measure</i> .....	II-6
2.2.6 UML .....	II-6
2.2.7 <i>Rational Unified Process</i> .....	II-7
2.3 Penelitian Lain Yang Relevan .....	II-9
2.4 Kesimpulan.....	II-10
<b>BAB III METODE PENELITIAN.....</b>	<b>III-1</b>
3.1 Pendahuluan.....	III-1
3.2 Pengumpulan Data.....	III-1
3.2.1 Jenis dan Sumber Data.....	III-1
3.2.2 Metode Pengumpulan Data .....	III-1
3.3 Tahapan Penelitian.....	III-3
3.3.1 Menetapkan Kerangka Kerja / Framework.....	III-4
3.3.2 Menetapkan Kriteria Pengujian .....	III-5
3.3.3 Menetapkan Format Data Pengujian .....	III-5
3.3.4 Menentukan Alat yang Digunakan dalam Penelitian .....	III-6
3.3.5 Melakukan Pengujian Penelitian .....	III-6
3.3.6 Melakukan Analisis Pengujian dan Membuat Kesimpulan.....	III-7
3.4 Metode Pengembangan Perangkat Lunak .....	III-7
3.4.1 Fase Insepsi .....	III-8
3.4.2 Fase Elaborasi.....	III-8
3.4.3 Fase Kontruksi.....	III-9
3.4.4 Fase Transisi .....	III-9
<b>BAB IV PENGEMBANGAN PERANGKAT LUNAK .....</b>	<b>IV-1</b>
4.1 Pendahuluan .....	IV-1
4.2 Fase Insepsi .....	IV-1
4.2.1 Pemodelan Bisnis .....	IV-1
4.2.2 Kebutuhan Sistem.....	IV-1
4.2.3 Analisis Desain.....	IV-3
4.3 Fase Elaborasi.....	IV-24
4.3.1 Pemodelan Bisnis .....	IV-24

4.3.2 Perancangan Data .....	IV-24
4.3.3 Perancangan Antarmuka.....	IV-25
4.3.4 Kebutuhan Sistem.....	IV-25
4.3.5 Diagram Aktivitas.....	IV-26
4.3.6 Diagram Alur .....	IV-31
4.4 Fase Kontruksi .....	IV-36
4.4.1 Kebutuhan Sistem.....	IV-36
4.4.2 Diagram Kelas .....	IV-36
4.4.3 Implementasi.....	IV-37
4.5 Fase Transisi .....	IV-39
4.5.1 Pemodelan Bisnis.....	IV-39
4.5.2 Kebutuhan Sistem.....	IV-40
4.5.3 Rencana Pengujian.....	IV-40
4.5.4 Implementasi.....	IV-41
4.6 Kesimpulan.....	IV-44
<b>BAB V HASIL DAN ANALISA PENELITIAN .....</b>	<b>V-1</b>
5.1 Pendahuluan.....	V-1
5.2 Hasil Penelitian.....	V-1
5.2.1 Konfigurasi Percobaan.....	V-1
5.2.2 Data Hasil Konfigurasi .....	V-1
5.3 Analisa Hasil Penelitian.....	V-3
5.4 Kesimpulan.....	V-5
<b>BAB VI KESIMPULAN DAN SARAN .....</b>	<b>VI-1</b>
6.1 Pendahuluan.....	VI-1
6.2 Kesimpulan.....	VI-1
6.3 Saran .....	VI-1
<b>DAFTAR PUSTAKA .....</b>	<b>xi</b>
<b>LAMPIRAN.....</b>	<b>L-1</b>

## DAFTAR TABEL

	Halaman
II-1	Tabel Hasil <i>Confusion Matrix</i> .....II-4
III-1	Contoh Data yang Digunakan ..... III-1
III-2	Rancangan Tabel <i>Confusion Matrix</i> Hasil Klasifikasi..... III-5
III-3	Rancangan Tabel Hasil Analisis Klasifikasi ..... III-7
III-3	Rencana Jadwal Penelitian ..... III-9
IV-1	Tabel Kebutuhan Fungsional ..... IV-2
IV-2	Tabel Kebutuhan Non-Fungsional ..... IV-3
IV-3	Contoh Data Cuitan..... IV-5
IV-4	Hasil Penghapusan <i>Unicode</i> dari Contoh Data Cuitan..... IV-6
IV-5	Hasil Penghapusan Kata Tidak Baku Menjadi Kata Baku..... IV-7
IV-6	Hasil <i>Case Folding</i> dari Data Cuitan ..... IV-8
IV-7	Hasil Tokenizing dari Contoh Data Cuitan ..... IV-9
IV-8	Hasil <i>Stopword Removal</i> dari Contoh Data Cuitan.....IV-10
IV-9	Hasil <i>Stemming</i> dari Contoh Data Cuitan.....IV-10
IV-10	Hasil TF dan IDF dari Contoh Data Cuitan.....IV-12
IV-11	Hasil Pembobotan Kata dari Contoh Data Cuitan.....IV-13
IV-12	Contoh Data <i>Testing</i> .....IV-14
IV-13	Hasil <i>Preprocessing</i> Data <i>Testing</i> .....IV-14
IV-14	Hasil TF-IDF <i>Vectorizer</i> Data <i>Testing</i> .....IV-15
IV-15	Definisi Actor.....IV-19
IV-16	Definisi <i>Use Case</i> .....IV-19
IV-17	Skenario <i>Use Case</i> Melakukan Load Data.....IV-20
IV-18	Skenario <i>Use Case</i> Melakukan Pelatihan.....IV-21
IV-19	Skenario <i>Use Case</i> Melakukan Klasifikasi PNN.....IV-22
IV-20	Skenario <i>Use Case</i> Melakukan Klasifikasi PNN + TF-IDF.....IV-23
IV-21	Tabel Implementasi Kelas.....IV-38
IV-22	Rencana Pengujian <i>Use Case</i> Melakukan Load Data.....IV-40
IV-23	Rencana Pengujian <i>Use Case</i> Melakukan Pelatihan.....IV-41

IV-24	Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi PNN.....	IV-41
IV-25	Rencana Pengujian <i>Use Case</i> Melakukan Klasifikasi PNN + TF-IDF..	IV-41
IV-26	Pengujian <i>Use Case</i> Melakukan <i>Load Data</i> .....	IV-42
IV-27	Pengujian <i>Use Case</i> Melakukan Pelatihan.....	IV-43
IV-28	Pengujian <i>Use Case</i> Melakukan Klasifikasi PNN.....	IV-44
IV-29	Pengujian <i>Use Case</i> Melakukan Klasifikasi PNN + TF-IDF.....	IV-45
V-1	Tabel <i>Confusion Matrix</i> Hasil Klafisikasi.....	V-2
V-2	Nilai <i>Performance</i> .....	V-2
V-3	Data Teks FP Pada Nilai Yang Sama.....	V-3

## DAFTAR GAMBAR

	Halaman
II-1	Arsitektur <i>Rational Unified Process</i> .....II-8
III-1	Diagram Tahapan Penelitian .....III-3
III-2	Diagram Tahapan Proses Perangkat Lunak.....III-4
IV-1	Diagram <i>Use Case</i> .....IV-18
IV-2	Rancangan Antarmuka Perangkat Lunak .....IV-25
IV-3	Diagram Aktivitas <i>Load Data</i> .....IV-27
IV-4	Diagram Aktivitas Pelatihan .....IV-28
IV-5	Diagram Aktivitas Klasifikasi PNN.....IV-29
IV-6	Diagram Aktivitas Klasifikasi PNN + TF-IDF .....IV-30
IV-7	Diagram Alur <i>Load Data</i> .....IV-31
IV-8	Diagram Alur Pelatihan.....IV-32
IV-9	Diagram Alur Klasifikasi PNN .....IV-33
IV-10	Diagram Alur Klasifikasi PNN + TF-IDF.....IV-34
IV-11	Diagram Alur <i>Preprocessing (Subsequence)</i> .....IV-35
IV-12	Diagram Kelas Perangkat Lunak.....IV-37
IV-13	Antarmuka Perangkat Lunak.....IV-39
V-1	Perbandingan Hasil <i>Performance</i> .....V-4

## DAFTAR LAMPIRAN

	Halaman
<b>Lampiran 1</b> Data Penelitian.....	L-1
<b>Lampiran 2</b> Source Code Program .....	L-35



# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab ini membahas tentang latar belakang masalah, rumusan masalah, tujuan, manfaat penelitian serta batasan masalah. Bab ini juga memberikan penjelasan umum mengenai keseluruhan penelitian.

### **1.2 Latar Belakang Masalah**

Twitter sekarang ini menjadi perangkat komunikasi yang sangat populer di kalangan pengguna internet. Pada konferensi resmi pengembang Twitter Chirp 2010, perusahaan tersebut menyampaikan statistik mengenai situs dan pengguna Twitter. Statistik tersebut menyebutkan bahwa pada bulan April 2010, Twitter memiliki 106 juta akun dan sebanyak 180 juta pengunjung unik setiap bulannya. Jumlah pengguna Twitter disebutkan terus meningkat 300.000 user setiap harinya (Buntoro, 2017). Adapun masalah yang sering muncul dalam penelitian di bidang Analisis Sentimen adalah banyaknya faktor-faktor seperti penggunaan bahasa yang tidak baku atau bahasa sehari-hari di kalangan masyarakat dan faktor lainnya yang dapat mempengaruhi hasil klasifikasi Sentimen.

Analisis Sentimen Adalah mengacu pada penggunaan pemrosesan bahasa alami, analisis teks, linguistik komputasi, dan biometrik untuk mengidentifikasi, mengekstraksi, mengukur, dan mempelajari status afektif dan informasi subjektif

secara sistematis, dalam tugas akhir ini akan digunakan metode *Pseudo Nearest Neighbor* yang dimana akan digunakan untuk melakukan Analisis Sentimen. Pada tahun 2018 dulu pernah ramai masalah tentang LRT yang digunakan untuk ASEAN Games menyebabkan banyaknya komentar dari masyarakat mengenai fenomena tersebut. Sehingga fenomena dinilai menarik untuk dilakukan Analisis Sentimen.

Metode yang bisa digunakan untuk melakukan Analisis Sentimen seperti Supervise Learning. Supervised learning adalah sebuah pendekatan dimana sudah terdapat data yang dilatih, dan terdapat variable yang ditargetkan sehingga tujuan dari pendekatan ini adalah mengelompokkan suatu data ke data yang sudah ada, macam-macam metode yang digunakan untuk klasifikasi seperti K-NN, Naïve Bayes, SVM, dan salah satunya adalah P-NN.

*Pseudo Nearest Neighbor* adalah merupakan varian terbaru dari K-NN, yang dimana desain ini digunakan untuk mengatasi kelemahan metode K-NN yang umumnya memberikan performansi yang rendah untuk data yang mengandung derau. P-NN bekerja dengan menghitung jarak total antara pola masukan (tak berlabel) dengan sejumlah k pola terdekat pada setiap kelas dengan pembobotan yang proporsional berdasarkan jarak, lalu memutuskan kelas dengan jarak total paling minimum sebagai kelas keputusan untuk pola tersebut (Suyanto, 2017).

Pada Penelitian Pambudi dkk, (2019), yang berjudul *Multi-label classification of Indonesian news topics using Pseudo Nearest Neighbor Rule*

melakukan penelitian menggunakan PNN. Hasil penelitiannya dimana ternyata algoritma *Pseudo Nearest Neighbor* juga bisa digunakan untuk mengklasifikasikan dari berbagai macam teks berita dalam bahasa Indonesia. Nilai algoritma P-NN untuk parameter k dan tipe proximity yang digunakan mempengaruhi performansi algoritma. Dari tiga jenis kedekatan yang diuji, kedekatan *Cosine* memberikan kinerja terbaik dibandingkan dengan perkiraan *Manhattan* dan *Euclidean*.

Pada penelitian Pratiwi dkk, (2019), yang berjudul *Credit scoring Analysis using Pseudo Nearest Neighbor* melakukan penelitian menggunakan PNN. Hasil penelitiannya berupa evaluasi ketepatan klasifikasi untuk masing-masing metode yang menggunakan perhitungan APER terlihat bahwa akurasi untuk metode K-NN adalah 98,11% dengan k= 1 sedangkan untuk metode P-NN adalah 79,25% dengan k = 13. Berdasarkan data tersebut, klasifikasi data Kredit Usaha Rakyat di Bank X Kabupaten Wonogiri Jawa Tengah sebaiknya dilakukan dengan metode K-NN bukan PNN. Pada penelitian ini akan menguji penggunaan metode *Pseudo Nearest Neighbour* pada Analisis Sentimen.

### **1.3 Rumusan Masalah**

Rumusan masalah dari tugas akhir ini adalah bagaimana penggunaan metode *Pseudo Nearest Neighbor* ( P-NN ) dan TF-IDF *Vectorizer* dapat diterapkan untuk dilakukan analisis sentimen pada pendapat masyarakat terhadap transportasi baru yang ada diwilayahnya ( LRT ).

Untuk menyelesaikan masalah di atas maka disusun menjadi beberapa pertanyaan penelitian :

1. Bagaimana menerapkan algoritma *Pseudo Nearest Neighbor* dan TF-IDF *Vectorizer* dalam melakukan sentimen?
2. Bagaimana mengukur kinerja algoritma *Pseudo Nearest Neighbor* dan TF-IDF *Vectorizer* untuk klasifikasi sentimen menggunakan *recall*, *precision*, dan *f-measure* ?

#### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah penelitian di atas maka ada dua tujuan dalam penelitian ini :

1. Membuat perangkat lunak analisis sentimen menggunakan metode *Pseudo Nearest Neighbor* dan TF-IDF *Vectorizer*.
2. Mengetahui hasil kinerja dari algoritma *Pseudo Nearest Neighbor* dan TF-IDF *Vectorizer* untuk klasifikasi sentimen menggunakan *recall*, *precision*, dan *f-measure*.

#### **1.5 Manfaat Penelitian**

Adapun manfaat yang di harapkan dari usulan penelitian ini adalah sebagai berikut:

1. Bisa menjadi acuan atau referensi bagi penelitian-penelitian lain di bidang *Pseudo Nearest neighbor* dan TF-IDF *Vectorizer*.

2. Dapat digunakan dalam prapengolahan pada penelitian NLP lainnya.
3. Dapat digunakan dalam mengetahui polaritas dari sebuah sentimen.

## **1.6 Batasan Masalah**

Adapun batasan masalah penelitian ini adalah sebagai berikut:

1. Klasifikasi ini hanya terbatas pada platform twitter.
2. Data teks yang diambil berupa teks berbahasa Indonesia, karena kaidah pra-pengolahan diatur dalam tata Bahasa Indonesia.
3. Komentar yang dibahas hanya mengenai dampak dari transportasi baru (LRT) yang ada di Palembang.

## **1.7 Sistematika Penulisan**

Sistematika penulisan tugas akhir akan ini mengikuti standar penulisan tugas akhir Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain:

### **1.7.1 Bab I. Pendahuluan**

Pada bab ini menjelaskan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah/ruang lingkup, serta sistematika penulisan.

### **1.7.2 Bab II. Kajian Literatur**

Pada bab ini akan dibahas dasar - dasar teori yang akan digunakan dalam penelitian, antara lain penelitian terkait, definisi-definisi *Text Mining*, analisis sentimen, *Pseudo Nearest Neighbor*, TF-IDF, *Confusion Matrix*, UML, dan RUP.

### **1.7.3 Bab III. Metodologi Penelitian**

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

### **1.7.4 Bab IV. Pengembangan Perangkat Lunak**

Pada bab ini akan membahas tahapan-tahapan yang dilakukan dalam proses pengembangan perangkat lunak yang didasarkan pada metode *Rational Unified Process* (RUP).

### **1.7.5 Bab V. Hasil dan Analisa Penelitian**

Pada bab ini akan membahas tentang hasil pengujian berdasarkan rencana pengujian yang telah ditentukan. Analisis hasil penelitian akan diberikan sebagai dasar dari membuat kesimpulan yang di ambil dalam penelitian ini.

### **1.7.6 Bab VI. Kesimpulan dan Saran**

Pada bab ini berisi kesimpulan dan saran dari hasil penelitian ini. Kesimpulan dan saran diambil dari semua hasil pembahasan pada bab-bab sebelumnya.

### **1.8 Kesimpulan**

Penelitian ini berfokus terhadap hasil klasifikasi analisis sentimen pada *twitter* dengan menggunakan metode *Pseudo Nearest Neighbor* dan *Tf-Idf Vectorizer*.

## DAFTAR PUSTAKA

- Adiwijaya, I. (2006). Text Mining dan Knowledge Discovery. *Kolokium Bersama Komunitas Datamining Indonesia & Soft-Computing Indonesia*, 1–9. [http://web.ipb.ac.id/~ir-lab/pdf/tm \(text summarization\).pdf](http://web.ipb.ac.id/~ir-lab/pdf/tm%20(text%20summarization).pdf)
- Aprillia (2018). “Pengaruh *Parf Of Speech Tagging* dalam Analisis Sentimen menggunakan Algoritma *Multinomial Naïve Bayes*.” Skripsi Jurusan Teknik Informatika, Fak. Ilmu Komputer, UNSRI – Tidak diterbitkan.
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *Integer Journal Maret*, 1(1), 32–41. [https://www.researchgate.net/profile/Ghulam\\_Buntoro/publication/316617194\\_Analisis\\_Sentimen\\_Calon\\_Gubernur\\_DKI\\_Jakarta\\_2017\\_Di\\_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf](https://www.researchgate.net/profile/Ghulam_Buntoro/publication/316617194_Analisis_Sentimen_Calon_Gubernur_DKI_Jakarta_2017_Di_Twitter/links/5907eee44585152d2e9ff992/Analisis-Sentimen-Calon-Gubernur-DKI-Jakarta-2017-Di-Twitter.pdf)
- Kruchten, P., & Kroll, P. (2003). The Rational Unified Process Made Easy. In *Rational Unified Process Made Easy: A Practitioner’s Guide to the RUP*.
- Langgeni, D. P. 2010 . ‘Komunitas Data *mining* Indonesia & *Soft-computing* Indonesia.’, *Text Mining dan Knowledge Discovery*., 0(semnasIF), pp. 1–10. doi: 10.1007/s13398-014-0173-7.2.
- Liu, B. (2010). *Sentiment Analysis and Subjectivity. Handbook of Natural Language Processing, Second Edition*, 2, 568.
- Pambudi, R. A., Adiwijaya, & Mubarok, M. S. (2019). Multi-label classification of Indonesian news topics using Pseudo Nearest Neighbor Rule. *Journal of Physics: Conference Series*, 1192(1). <https://doi.org/10.1088/1742-6596/1192/1/012031>
- Pan, W., Narasimhan, H., Protopapas, P., Kar, P., & Ramaswamy, H. G. (2017). *Optimizing the multiclass F-measure via biconcave programming. Proceedings - IEEE International Conference on Data Mining, ICDM*, 1101–



1106. <https://doi.org/10.1109/ICDM.2016.184>.

Patterson, J. & Gibson A. 2017. *Deep Learning A Practitioner's Approach*, O'Reilly Media, Sebastopol, USA.

Pratiwi, H., Mukid, M. A., Hoyyi, A., & Widiharih, T. (2019). Credit scoring analysis using pseudo nearest neighbor. *Journal of Physics: Conference Series*, 1217(1). <https://doi.org/10.1088/1742-6596/1217/1/012100>

Ramadhan M. L. (2020). “*Email Spam Filtering Menggunakan Multilayer Perceptron dengan Metode Pelatihan Stochastic Gradient Descent ( SGD ) dan Momentum.*” Skripsi Jurusan Teknik Informatika, Fak. Ilmu Komputer, UNSRI – Tidak diterbitkan.

Suyanto (2017). *Data Mining untuk Klasifikasi dan Klusterisasi Data*. Bandung, Jawa Barat, Indonesia: Informatika Bandung