

Perbandingan Analisis Sentimen menggunakan  
Metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*  
dengan Seleksi Fitur *Information Gain*

Diajukan Sebagai Syarat untuk Menyelesaikan  
Pendidikan Program Strata-1 pada  
Jurusan Teknik Informatika



Oleh :

Adityo Aji Pambudi  
NIM : 09021281722049

**Jurusan Teknik Informatika**  
**FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA**  
**2022**

**LEMBAR PENGESAHAN SKRIPSI**

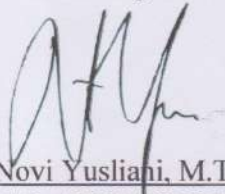
Perbandingan Analisis Sentimen menggunakan  
Metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*  
dengan Seleksi Fitur *Information Gain*

Oleh :

Adityo Aji Pambudi  
NIM : 09021281722049

Palembang, 26 Januari 2022

Pembimbing I



Novi Yusliani, M.T.  
NIP. 198211082012122001

Pembimbing II



Kanda Januar Miraswan, M.T.  
NIP. 199001092019031012

Mengetahui,  
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

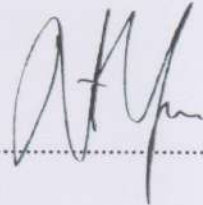
## TANDA LULUS UJIAN SIDANG SKRIPSI

Pada hari **Senin** tanggal **10 Januari 2022** telah dilaksanakan ujian sidang skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Adityo Aji Pambudi  
NIM : 09021281722049  
Judul : Perbandingan Analisis Sentimen menggunakan Metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* dengan Seleksi Fitur *Information Gain*

1. Pembimbing I

Novi Yusliani, M.T.  
NIP. 198211082012122001



.....

2. Pembimbing II

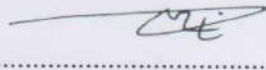
Kanda Januar Miraswan, M.T.  
NIP. 199001092019031012



.....

3. Penguji I

Osvari Arsalan, M.T.  
NIP. 198806282018031001



.....

4. Penguji II

Hadipurnawan Satria, Ph.D.  
NIP. 198004182020121001



.....

Mengetahui,  
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.  
NIP. 197812222006042003

## HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Adityo Aji Pambudi  
NIM : 09021281722049  
Program Studi : Teknik Informatika Bilingual  
Judul Skripsi : Perbandingan Analisis Sentimen  
menggunakan Metode *K-Nearest  
Neighbor* dan *Modified K-Nearest  
Neighbor* dengan Seleksi Fitur  
*Information Gain*  
Hasi Pengecekan Software  
iThenticate/Turnitin : 8%

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

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



Palembang, Januari 2022



Adityo Aji Pambudi

NIM. 09021281722049

## MOTTO DAN PERSEMBAHAN

MOTTO:

*“You cannot stop the rain, no way*

*Holdin' an umbrella when the grey clouds come over again”*

- Ed Sheeran

*“The first step is to establish that something is possible,*

*then probability will occur”*

- Elon Musk

*“Jangan kebanyakan pikiran ya sayang! Jalanin aja.”*

- Penulis

Kupersembahkan karya tulis ini kepada :

- *Allah SWT & Nabi Muhammad SAW*
- *Ibu & ayahku tercinta*
- *Keluarga besarku*
- *Dosen pembimbing & dosen penguji*
- *Teman seperjuanganku*
- *Almamater*

Comparison of Sentiment Analysis using  
*K-Nearest Neighbor* and *Modified K-Nearest Neighbor* Methods  
with *Information Gain* Feature Selection

by:  
Adityo Aji Pambudi  
09021281722049

**ABSTRACT**

Netizens can express opinions on social media in written form to express their feelings or emotions. The number can be very large and the language is not always standard. Therefore, sentiment analysis is needed as a system that can analyze these opinions. In sentiment analysis, the problem with high dimensions of data attributes can have an effect on accuracy so that system performance is not satisfactory. This requires feature selection. In this study, sentiment analysis was carried out using the *K-Nearest Neighbor* (KNN) and *Modified K-Nearest Neighbor* (MK-NN) classification methods without feature selection and with *Information Gain* feature selection. In the PILKADA 1 dataset, the best accuracy results were obtained in the MK-NN test with the *Information Gain* feature selection by threshold 60% which is 59.8%. In the PILKADA 2 dataset, the best accuracy results were obtained in the MK-NN test with the *Information Gain* feature selection by threshold 40%, which is 78.2%. From the test results, it is found that MK-NN is not always better than K-NN and vice versa. *Information Gain* can increase accuracy but depends on the selected threshold. MK-NN and *Information Gain* can produce better accuracy but depends on the choice of threshold. Decrease and increase in accuracy of less than 7%.

Keywords: Sentiment analysis, K-Nearest Neighbor, Modified K-Nearest Neighbor, Feature Selection, Information Gain

Perbandingan Analisis Sentimen menggunakan  
Metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor*  
dengan Seleksi Fitur *Information Gain*

oleh:  
Adityo Aji Pambudi  
09021281722049

**ABSTRAK**

Opini-opini di media sosial bisa dikeluarkan oleh warganet dalam bentuk tulisan untuk mengekspresikan perasaan atau emosinya. Jumlahnya bisa sangat banyak dan bahasanya tidak selalu baku. Maka dari itu analisis sentimen diperlukan sebagai sistem yang dapat menganalisis opini-opini tersebut. Pada analisis sentimen masalah tingginya dimensi atribut data dapat berefek pada akurasi sehingga kinerja sistem tidak memuaskan. Untuk itu diperlukan seleksi fitur. Pada penelitian ini, analisis sentimen dilakukan menggunakan metode klasifikasi *K-Nearest Neighbor (KNN)* dan *Modified K-Nearest Neighbor (MK-NN)* tanpa seleksi fitur dan dengan seleksi fitur *Information Gain*. Pada dataset PILKADA 1, hasil akurasi terbaik diperoleh pada pengujian MK-NN dengan seleksi fitur *Information Gain threshold* 60% yaitu 59,8%. Pada dataset PILKADA 2 hasil akurasi terbaik diperoleh pada pengujian MK-NN dengan seleksi fitur *Information Gain threshold* 40% yaitu 78.2%. Dari hasil pengujian didapatkan MK-NN tidak selalu lebih baik daripada K-NN begitupun sebaliknya. *Information Gain* dapat meningkatkan akurasi namun bergantung pada *threshold* yang dipilih. MK-NN dan *Information Gain* dapat menghasilkan akurasi yang lebih baik namun bergantung pada pilihan *threshold*. Penurunan dan peningkatan akurasi kurang dari 7%.

Kata kunci: Analisis sentimen, K-Nearest Neighbor, Modified K-Nearest Neighbor, Seleksi Fitur, Information Gain

## KATA PENGANTAR

Assalamualaikum warahmutallahi wabarakatuh. Puji syukur kepada Allah SWT atas berkah dan rahmat-Nya kepada penulis, skripsi yang berjudul “Perbandingan Analisis Sentimen menggunakan Metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* dengan Seleksi Fitur *Information Gain*” ini dapat diselesaikan. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Untuk selanjutnya penulis mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian skripsi ini, yaitu:

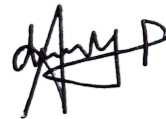
1. Bapak Jaidan Jauhari, S. Pd. M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
2. Ibu Alvi Syahrini Utami, M.Kom. selaku Kepala Jurusan Teknik Informatika Universitas Sriwijaya.
3. Ibu Novi Yusliani, S.Kom., M.T selaku dosen pembimbing skripsi 1 yang telah membimbing penulis dalam menyelesaikan skripsi ini.
4. Bapak Kanda Januar, M.T selaku dosen pembimbing skripsi 2 yang telah membimbing penulis dalam menyelesaikan skripsi ini.
5. Bapak Dr. Abdiansah, M.Cs selaku dosen pembimbing akademik yang telah membimbing penulis selama masa pendidikan di Fakultas Ilmu Komputer Universitas Sriwijaya
6. Bapak Osvari Arsalan, M.T selaku dosen penguji 1 yang telah memberikan sarannya untuk memperbaiki kekurangan pada skripsi ini
7. Bapak Hadipurnawan Satria, Ph.D selaku dosen penguji 2 yang telah memberikan sarannya untuk memperbaiki kekurangan yang pada skripsi ini
8. Seluruh staff pengajar Fakultas Ilmu Komputer yang telah memberikan ilmu pengetahuan yang tak ternilai selama penulis menempuh pendidikan di Fakultas Ilmu Komputer Universitas Sriwijaya.
9. Kedua orang tua penulis, Kahar Zaeni dan Taswen yang selalu memberikan kasih sayang, doa, fasilitas, dan dukungan di setiap langkah penulis
10. Saudara-saudaraku yang selalu menyayangi penulis
11. Teman-teman kuliah yang telah memberikan kesan indah selama masa perkuliahan, yang selalu berbagi suka dan duka dan juga berbagi ilmu.
12. Teman-teman di luar kuliah yang telah memberikan penyegaran mental dan dukungannya sehingga dalam menempuh pendidikan dan menyelesaikan skripsi ini penulis tidak merasa terlalu stress



13. Semua pihak yang tidak dapat disebutkan satu persatu, yang telah memberikan doa dengan ikhlas serta dukungan sehingga dapat menyelesaikan skripsi ini.

Penulis sangat menyadari dalam penyusunan skripsi ini masih banyak kekurangan yang disebabkan kurangnya pengalaman dan pengetahuan. Penulis mohon maaf atas itu. Karena itulah kritik dan saran pembaca sangat diterima untuk kemajuan penelitian berikutnya. Akhir kata semoga skripsi ini dapat berguna dan bermanfaat bagi kita semua. Wassalam.

Palembang, Januari 2022



Penulis,

Adityo Aji Pambudi

## DAFTAR ISI

|  | Halaman |
|--|---------|
| HALAMAN JUDUL.....                     | i       |
| LEMBAR PENGESAHAN SKRIPSI .....        | ii      |
| TANDA LULUS UJIAN SIDANG SKRIPSI ..... | iii     |
| HALAMAN PERNYATAAN .....               | iv      |
| MOTTO DAN PERSEMBAHAN .....            | v       |
| ABSTRACT.....                          | vi      |
| ABSTRAK.....                           | vii     |
| KATA PENGANTAR .....                   | viii    |
| DAFTAR ISI.....                        | x       |
| DAFTAR TABEL.....                      | xiv     |
| DAFTAR GAMBAR .....                    | xvii    |
| DAFTAR LAMPIRAN.....                   | xix     |
| <br>                                   |         |
| BAB I PENDAHULUAN.....                 | I-1     |
| 1.1    Pendahuluan .....               | I-1     |
| 1.2    Latar Belakang Masalah .....    | I-1     |
| 1.3    Rumusan Masalah .....           | I-4     |
| 1.4    Tujuan Penelitian.....          | I-4     |
| 1.5    Manfaat Penelitian.....         | I-5     |
| 1.6    Batasan Masalah.....            | I-5     |
| 1.7    Sistematika Penulisan.....      | I-5     |
| 1.8    Kesimpulan.....                 | I-7     |
| <br>                                   |         |
| BAB II KAJIAN LITERATUR .....          | II-1    |
| 2.1    Pendahuluan .....               | II-1    |
| 2.2    Landasan Teori .....            | II-1    |
| 2.2.1    Analisis Sentimen .....       | II-1    |
| 2.2.2    Preprocessing .....           | II-3    |

|  |        |
|--|--------|
| 2.2.3 Metode Pembobotan TF-IDF.....                                  | II-7   |
| 2.2.4 <i>Information Gain</i> .....                                  | II-9   |
| 2.2.5 <i>K-Nearest Neighbor</i> (K-NN) .....                         | II-10  |
| 2.2.6 <i>Modified K-Nearest Neighbor</i> (MK-NN) .....               | II-11  |
| 2.2.7 <i>Cross Validation</i> .....                                  | II-13  |
| 2.2.8 <i>Confusion Matrix</i> .....                                  | II-14  |
| 2.2.9 <i>Rational Unified Process (RUP)</i> .....                    | II-15  |
| 2.3 Penelitian Lain yang Relevan .....                               | II-17  |
| 2.4 Kesimpulan.....  | II-21  |
| <br>   |        |
| BAB III METODOLOGI PENELITIAN.....                                   | III-1  |
| 3.1 Pendahuluan .....  | III-1  |
| 3.2 Pengumpulan Data .....   | III-1  |
| 3.2.1 Jenis Data.....  | III-1  |
| 3.2.2 Sumber Data .....  | III-1  |
| 3.2.3 Metode Pengumpulan Data.....                                   | III-2  |
| 3.3 Tahapan Penelitian .....   | III-2  |
| 3.3.1 Menetapkan Kerangka Kerja/Framework.....                       | III-3  |
| 3.3.2 Menetapkan Kriteria Pengujian .....                            | III-16 |
| 3.3.3 Menetapkan Format Data Pengujian.....                          | III-17 |
| 3.3.4 Menentukan Alat yang Digunakan dalam Penelitian .....          | III-18 |
| 3.3.5 Melakukan Pengujian Penelitian .....                           | III-19 |
| 3.3.6 Melakukan Analisis Hasil Pengujian dan Kesimpulan Penelitian.. | III-20 |
| 3.4 Metode Pengembangan Perangkat Lunak .....                        | III-20 |
| 3.4.1 Fase Insepsi.....  | III-20 |
| 3.4.2 Fase Elaborasi .....   | III-21 |
| 3.4.3 Fase Konstruksi.....   | III-21 |
| 3.4.4 Fase Transisi .....  | III-21 |
| 3.5 Manajemen Proyek Penelitian.....                                 | III-22 |
| 3.6 Kesimpulan.....  | III-31 |
| <br>   |        |
| BAB IV PENGEMBANGAN PERANGKAT LUNAK .....                            | IV-1   |

|  |   |       |
|--|---|-------|
| 4.1                                      | Pendahuluan .....                             | IV-1  |
| 4.2                                      | Fase Insepsi .....                            | IV-1  |
| 4.2.1                                    | Permodelan Bisnis .....                       | IV-1  |
| 4.2.2                                    | Analisis dan Desain .....                     | IV-7  |
| 4.3                                      | Fase Elaborasi.....                           | IV-22 |
| 4.3.1                                    | Permodelan Bisnis .....                       | IV-23 |
| 4.3.2                                    | Persiapan Data .....                          | IV-23 |
| 4.3.3                                    | Perancangan Antarmuka .....                   | IV-24 |
| 4.3.4                                    | Kebutuhan Sistem .....                        | IV-27 |
| 4.3.6                                    | Diagram Alur ( <i>Sequence Diagram</i> )..... | IV-35 |
| 4.4                                      | Fase Konstruksi .....                         | IV-41 |
| 4.4.1                                    | Kebutuhan Sistem .....                        | IV-41 |
| 4.4.2                                    | Diagram Kelas .....                           | IV-41 |
| 4.4.3                                    | Implementasi.....                             | IV-43 |
| 4.5                                      | Fase Transisi.....                            | IV-52 |
| 4.5.1                                    | Permodelan Bisnis .....                       | IV-53 |
| 4.5.2                                    | Kebutuhan Sistem .....                        | IV-53 |
| 4.5.3                                    | Rencana Pengujian.....                        | IV-53 |
| 4.5.4                                    | Implementasi.....                             | IV-58 |
| 4.6                                      | Kesimpulan.....                               | IV-69 |
| BAB V HASIL DAN ANALISIS PENELITIAN..... |   | V-1   |
| 5.1                                      | Pendahuluan .....                             | V-1   |
| 5.2                                      | Data Hasil Penelitian .....                   | V-1   |
| 5.2.1                                    | Konfigurasi Percobaan.....                    | V-1   |
| 5.2.2                                    | Hasil Pengujian Pada Data PILKADA 1 .....     | V-2   |
| 5.2.3                                    | Hasil Pengujian Pada Data PILKADA 2 .....     | V-21  |
| 5.3                                      | Analisis Hasil Pengujian Keseluruhan .....    | V-41  |
| 5.4                                      | Kesimpulan.....                               | V-45  |
| BAB VI KESIMPULAN DAN SARAN .....        |   | VI-1  |
| 6.1                                      | Pendahuluan .....                             | VI-1  |

|                      |      |
|----------------------|------|
| 6.2 Kesimpulan.....  | VI-1 |
| 6.3 Saran.....       | VI-3 |
| DAFTAR PUSTAKA ..... | xxi  |
| LAMPIRAN.....        | L-1  |

## DAFTAR TABEL

Halaman

|  |        |
|--|--------|
| Tabel II- 1. Kombinasi awalan dan akhiran yang tidak diperbolehkan .....   | II-6   |
| Tabel II- 2. Contoh tahapan <i>preprocessing</i> .....   | II-6   |
| Tabel II- 3. <i>Confusion Matrix</i> .....   | II-14  |
|  |        |
| Tabel III- 1. Data Latih dan Data Uji Tetangga Terdekat .....  | III-10 |
| Tabel III- 2. Hasil Perhitungan Validitas Setiap Data Latih .....  | III-11 |
| Tabel III- 3. Data Latih dengan Jarak Terdekat dari Data Uji ( $k=5$ ) .....   | III-12 |
| Tabel III- 4. Contoh <i>Confusion Matrix</i> .....   | III-14 |
| Tabel III- 5. Contoh Hasil <i>Cross Validation</i> Klasifikasi .....   | III-15 |
| Tabel III- 6. Rancangan Tabel <i>Confusion Matrix</i> Hasil Klasifikasi dengan <i>k-cross validation</i> .....   | III-17 |
| Tabel III- 7. Rancangan Tabel Hasil Analisis Klasifikasi .....   | III-20 |
|  |        |
| Tabel IV- 1. Kebutuhan Fungsional .....  | IV-6   |
| Tabel IV- 2. Kebutuhan Non-Fungsional .....  | IV-6   |
| Tabel IV- 3. Definisi Aktor .....  | IV-11  |
| Tabel IV- 4. Definisi <i>Use Case</i> .....  | IV-12  |
| Tabel IV- 5. Skenario Memilih Dataset .....  | IV-13  |
| Tabel IV- 6. Skenario Membagi Dataset .....  | IV-14  |
| Tabel IV- 7. Skenario Menguji Sistem Analisis Sentimen dengan Metode K-NN Tanpa Seleksi Fitur .....  | IV-15  |
| Tabel IV- 8. Skenario Menguji Sistem Analisis Sentimen dengan Metode MK-NN Tanpa Seleksi Fitur .....   | IV-17  |
| Tabel IV- 9. Skenario Menguji Sistem Analisis Sentimen dengan Metode K-NN dan Seleksi Fitur <i>Information Gain</i> .....                              | IV-19  |
| Tabel IV- 10. Skenario Menguji Sistem Analisis Sentimen dengan Metode MK-NN dan Seleksi Fitur <i>Information Gain</i> .....                            | IV-21  |
| Tabel IV- 11. Rancangan Data .....   | IV-23  |
| Tabel IV- 12. Tabel Implementasi Kelas .....   | IV-43  |
| Tabel IV- 13. Rencana Pengujian <i>Use Case</i> Memilih Dataset .....  | IV-54  |
| Tabel IV- 14. Rencana Pengujian <i>Use Case</i> Membagi Dataset .....  | IV-54  |
| Tabel IV- 15. Rencana Pengujian <i>Use Case</i> pada Sistem Analisis Sentimen menggunakan <i>K-Nearest Neighbor</i> tanpa Seleksi Fitur .....          | IV-55  |
| Tabel IV- 16. Rencana Pengujian <i>Use Case</i> pada Sistem Analisis Sentimen menggunakan <i>Modified K-Nearest Neighbor</i> tanpa Seleksi Fitur ..... | IV-56  |

|  |       |
|--|-------|
| Tabel IV- 17. Rencana Pengujian <i>Use Case</i> pada Sistem Analisis Sentimen menggunakan <i>K-Nearest Neighbor</i> dengan Seleksi Fitur <i>Information Gain</i> ...           | IV-57 |
| Tabel IV- 18. Rencana Pengujian <i>Use Case</i> pada Sistem Analisis Sentimen menggunakan <i>Modified K-Nearest Neighbor</i> dengan Seleksi Fitur .....                        | IV-58 |
| Tabel IV- 19. Pengujian <i>Use Case</i> Memilih Dataset .....  | IV-59 |
| Tabel IV- 20. Pengujian <i>Use Case</i> Membagi Dataset.....   | IV-61 |
| Tabel IV- 21. Pengujian <i>Use Case</i> Sistem pada Sistem Analisis Sentimen menggunakan <i>K-Nearest Neighbor</i> tanpa Seleksi Fitur.....                                    | IV-63 |
| Tabel IV- 22. Pengujian <i>Use Case</i> Sistem pada Sistem Analisis Sentimen menggunakan menggunakan <i>Modified K-Nearest Neighbor</i> tanpa Seleksi Fitur .....              | IV-64 |
| Tabel IV- 23. Pengujian <i>Use Case</i> Sistem pada Sistem Analisis Sentimen menggunakan <i>K-Nearest Neighbor</i> dengan Seleksi Fitur <i>Information Gain</i> ...            | IV-66 |
| Tabel IV- 24. Pengujian <i>Use Case</i> Sistem pada Sistem Analisis Sentimen menggunakan <i>Modified K-Nearest Neighbor</i> dengan Seleksi Fitur <i>Information Gain</i> ..... | IV-68 |
|  |       |
| Tabel V- 1. Hasil Evaluasi K-NN dengan Berbagai k pada Data PILKADA 1. .   | V-2   |
| Tabel V- 2. Hasil Evaluasi MK-NN dengan Berbagai k pada Data PILKADA 1..   | V-2   |
| Tabel V- 3. Hasil Evaluasi K-NN k=3 untuk Data PILKADA 1 .....   | V-4   |
| Tabel V- 4. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 40% untuk Data PILKADA 1 .....  | V-5   |
| Tabel V- 5. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 50% pada Data PILKADA 1 .....   | V-6   |
| Tabel V- 6. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 60% pada Data PILKADA 1 .....   | V-7   |
| Tabel V- 7. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 75% pada Data PILKADA 1 .....   | V-8   |
| Tabel V- 8. Perbandingan Rata-rata Hasil Evaluasi K-NN dan K-NN dengan <i>Information Gain</i> Berbagai <i>Threshold</i> pada Data PILKADA 1 .....                             | V-9   |
| Tabel V- 9. Hasil Evaluasi MK-NN k=3 pada Data PILKADA 1 .....   | V-11  |
| Tabel V- 10. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 40% pada Data PILKADA 1 .....   | V-12  |
| Tabel V- 11. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 50% pada Data PILKADA 1 .....   | V-13  |
| Tabel V- 12. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 60% pada Data PILKADA 1 .....   | V-14  |
| Tabel V- 13. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 75% pada Data PILKADA 1 .....   | V-15  |
| Tabel V- 14. Perbandingan Rata-rata Hasil Evaluasi MK-NN dan MK-NN dengan <i>Information Gain</i> Berbagai <i>Threshold</i> pada Data PILKADA 1 .....                          | V-17  |

|   |      |
|---|------|
| Tabel V- 15. Hasil Evaluasi K-NN dengan Berbagai k pada Data PILKADA 2 .....  | V-22 |
| Tabel V- 16. Hasil Evaluasi MK-NN dengan Berbagai k pada Data PILKADA 2 .....   | V-22 |
| Tabel V- 17. Hasil Evaluasi K-NN k=3 untuk Data PILKADA 2.....  | V-23 |
| Tabel V- 18. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 40% untuk Data PILKADA 2.....   | V-24 |
| Tabel V- 19. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 50% pada Data PILKADA 2 .....   | V-25 |
| Tabel V- 20. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 60% pada Data PILKADA 2 .....   | V-26 |
| Tabel V- 21. Hasil Evaluasi K-NN k=3 dengan <i>Information Gain threshold</i> 75% pada Data PILKADA 2 .....   | V-27 |
| Tabel V- 22. Perbandingan Rata-rata Hasil Evaluasi K-NN dan K-NN dengan <i>Information Gain</i> Berbagai <i>Threshold</i> pada Data PILKADA 2 .....   | V-29 |
| Tabel V- 23. Hasil Evaluasi MK-NN k=3 pada Data PILKADA 2.....  | V-31 |
| Tabel V- 24. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 40% pada Data PILKADA 2.....   | V-32 |
| Tabel V- 25. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 50% pada Data PILKADA 2.....   | V-33 |
| Tabel V- 26. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 60% pada Data PILKADA 2.....   | V-34 |
| Tabel V- 27. Hasil Evaluasi MK-NN k=3 dengan <i>Information Gain threshold</i> 75% pada Data PILKADA 2.....   | V-35 |
| Tabel V- 28. Perbandingan Rata-rata Hasil Evaluasi MK-NN dan MK-NN dengan <i>Information Gain</i> Berbagai <i>Threshold</i> pada Data PILKADA 2 ..... | V-36 |



## DAFTAR GAMBAR

Halaman

|  |        |
|--|--------|
| Gambar II- 1. Proses analisis sentimen (Sihwi, Jati & Anggrainingsih, 2018) ..   | II-2   |
| Gambar II- 2. Fase RUP (Krutchen, 1998) .....  | II-17  |
| Gambar III- 1. Diagram Kerangka Kerja Perangkat Lunak.....   | III-5  |
| Gambar III- 2. Diagram Tahapan <i>Preprocessing</i> Text .....   | III-6  |
| Gambar III- 3 Contoh Perhitungan <i>Information Gain</i> .....   | III-8  |
| Gambar III- 4. Contoh Perhitungan Jarak dengan <i>Euclidean Distance</i> .....   | III-9  |
| Gambar III- 5. Pehitungan Bobot Kelas dan Penentuan Kelas .....  | III-10 |
| Gambar III- 6. Contoh Perhitungan <i>Validity</i> .....  | III-11 |
| Gambar III- 7 Contoh Pehitungan Bobot Tiap Data Uji Terdekat.....  | III-13 |
| Gambar III- 8. Pehitungan Penentuan Kelas pada MK-NN.....  | III-13 |
| Gambar III- 9. Perhitungan Akurasi, <i>Precision</i> , <i>Recall</i> dan <i>F-Measure</i> .....                                    | III-15 |
| Gambar III- 10. Perhitungan Rata-Rata Skor Kinerja Sistem.....   | III-16 |
| Gambar III- 11. Diagram Tahapan Pengujian Penelitian.....  | III-19 |
| Gambar III- 12. Penjadwalan untuk Tahap Menentukan Ruang Lingkup dan Unit Penelitian.....  | III-23 |
| Gambar III- 13. Penjadwalan untuk Tahap Menentukan Dasar Teori yang Berkaitan dengan Penelitian.....                               | III-24 |
| Gambar III- 14. Penjadwalan untuk Tahap Menentukan Kriteria Pengujian... III-25  |        |
| Gambar III- 15. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian fase <i>Inception</i> .....    | III-26 |
| Gambar III- 16. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian fase <i>Elaboration</i> .....  | III-27 |
| Gambar III- 17. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian fase <i>Construction</i> ..... | III-28 |
| Gambar III- 18. Penjadwalan untuk Tahap Menentukan Alat yang Digunakan untuk Pelaksanaan Penelitian fase <i>Transition</i> .....   | III-29 |
| Gambar III- 19. Penjadwalan untuk Tahap Melakukan Pengujian Penelitian dan Melakukan Analisa Hasil Pengujian .....                 | III-30 |
| Gambar IV- 1. Diagram <i>Use case</i> .....  | IV-11  |
| Gambar IV- 2. Rancangan Antarmuka Pilih Dataset .....  | IV-24  |
| Gambar IV- 3. Rancangan Antarmuka Halaman Homepage.....  | IV-25  |
| Gambar IV- 4. Rancangan Antarmuka Halaman Hasil Evaluasi.....  | IV-26  |
| Gambar IV- 5. Rancangan Antarmuka Waiting Dialog.....  | IV-26  |
| Gambar IV- 6. Rancangan Antarmuka Warning Dialog.....  | IV-27  |

|   |       |
|---|-------|
| Gambar IV- 7. Diagram Aktivitas Memilih Dataset.....  | IV-29 |
| Gambar IV- 8. Diagram Aktivitas Membagi Dataset .....   | IV-30 |
| Gambar IV- 9. Diagram Aktivitas Menguji Sistem Analisis Sentimen dengan Metode K-NN Tanpa Seleksi Fitur .....   | IV-31 |
| Gambar IV- 10. Diagram Aktivitas Menguji Sistem Analisis Sentimen dengan Metode MK-NN Tanpa Seleksi Fitur.....  | IV-32 |
| Gambar IV- 11. Diagram Aktivitas Menguji Sistem Analisis Sentimen dengan Metode K-NN dengan Seleksi Fitur <i>Information Gain</i> .....                                     | IV-33 |
| Gambar IV- 12. Diagram Aktivitas Menguji Sistem Analisis Sentimen dengan Metode MK-NN dengan Seleksi Fitur <i>Information Gain</i> .....                                    | IV-34 |
| Gambar IV- 13. Diagram Alur Memilih Dataset .....   | IV-35 |
| Gambar IV- 14. Diagram Alur Membagi Dataset.....  | IV-36 |
| Gambar IV- 15. Diagram Alur Menguji Sistem Analisis Sentimen dengan Metode K-NN tanpa Seleksi Fitur.....  | IV-37 |
| Gambar IV- 16. Diagram Alur Menguji Sistem Analisis Sentimen dengan Metode MK-NN tanpa Seleksi Fitur .....  | IV-38 |
| Gambar IV- 17. Diagram Alur Menguji Sistem Analisis Sentimen dengan Metode K-NN dengan Seleksi Fitur <i>Information Gain</i> .....  | IV-39 |
| Gambar IV- 18. Diagram Alur Menguji Sistem Analisis Sentimen dengan Metode MK-NN dengan Seleksi Fitur <i>Information Gain</i> .....   | IV-40 |
| Gambar IV- 19. Diagram Kelas .....  | IV-42 |
| Gambar IV- 20 Implementasi Antarmuka Pilih Dataset .....  | IV-48 |
| Gambar IV- 21. Implementasi Antarmuka Halaman Homepage.....   | IV-49 |
| Gambar IV- 22. Implementasi Antarmuka Halaman Hasil Evaluasi.....   | IV-50 |
| Gambar IV- 23. Implementasi Antarmuka Waiting Dialog.....   | IV-51 |
| Gambar IV- 24. Implementasi Antarmuka Warning Dialog.....   | IV-52 |
| <br>  |       |
| Gambar V- 1. Grafik Perbandingan Kinerja K-NN vs MK-NN dengan k=3 pada Data PILKADA 1.....  | V-19  |
| Gambar V- 2 Grafik Perbandingan K-NN + <i>Information Gain</i> dan MK-NN + <i>Information Gain</i> dengan k=3 pada Data PILKADA 1 .....                                     | V-20  |
| Gambar V- 3. Grafik Perbandingan Kinerja K-NN vs MK-NN dengan k=3 pada Data PILKADA 2.....  | V-38  |
| Gambar V- 4. Grafik Perbandingan K-NN + <i>Information Gain</i> dan MK-NN + <i>Information Gain</i> dengan k=3 pada Data PILKADA 2 .....                                    | V-39  |
| Gambar V- 5. Grafik Perbandingan hasil K-NN, MK-NN, K-NN dan <i>Information Gain</i> , MK-NN dan <i>Information Gain</i> dengan Akurasi Tertinggi pada Data PILKADA 1 ..... | V-41  |
| Gambar V- 6. Grafik Perbandingan hasil K-NN, MK-NN, K-NN dan <i>Information Gain</i> , MK-NN dan <i>Information Gain</i> dengan Akurasi Tertinggi pada Data PILKADA 2 ..... | V-41  |

## DAFTAR LAMPIRAN

Halaman

|  |      |
|--|------|
| <b>Lampiran 1.</b> Contoh Data Kicauan.....  | L-1  |
| <b>Lampiran 2.</b> Hasil <i>Casefolding</i> pada Contoh Data Kicauan .....   | L-5  |
| <b>Lampiran 3.</b> Hasil Tahapan <i>Cleaning</i> pada Contoh Data Kicauan .....  | L-8  |
| <b>Lampiran 4.</b> Hasil Tahapan <i>Tokenizing</i> pada Contoh Data Kicauan .....  | L-11 |
| <b>Lampiran 5.</b> Hasil Tahapan Normalisasi Kata Gaul pada Contoh Data Kicauan .....  | L-15 |
| <b>Lampiran 6.</b> Hasil Tahapan <i>Stopwords Removal</i> pada Contoh Data Kicauan .....   | L-19 |
| <b>Lampiran 7.</b> Hasil Tahapan <i>Stemming</i> pada Contoh Data Kicauan .....  | L-21 |
| <b>Lampiran 8.</b> Hasil Perhitungan <i>Information Gain</i> pada Contoh Data Kicauan.....                                       | L-23 |
| <b>Lampiran 9.</b> Tabel Tahapan Perhitungan Kata pada Contoh Data Kicauan ....  | L-29 |
| <b>Lampiran 10.</b> Tabel Hasil Tahapan Pembobotan dengan TF-IDF pada Contoh Data Kicauan.....                                   | L-35 |
| <b>Lampiran 11.</b> Jarak Antara Data Uji dan Data Latih pada Contoh Data Kicauan .....  | L-43 |
| <b>Lampiran 12.</b> Perhitungan Jarak Masing-masing Data Latih ke Data Latih Lainnya pada Contoh Data Kicauan .....              | L-44 |
| <b>Lampiran 13.</b> Penjadwalan Penelitian dalam Bentuk WBS .....  | L-46 |
| <b>Lampiran 14.</b> Hasil <i>Confusion Matrix</i> KNN k=3 pada Data PILKADA 1 .....  | L-57 |
| <b>Lampiran 15.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain threshold</i> 40% pada Data PILKADA 1.....  | L-58 |
| <b>Lampiran 16.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain threshold</i> 50% pada Data PILKADA 1.....  | L-59 |
| <b>Lampiran 17.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain threshold</i> 60% pada Data PILKADA 1.....  | L-60 |
| <b>Lampiran 18.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain threshold</i> 75% pada Data PILKADA 1.....  | L-61 |
| <b>Lampiran 19.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 pada Data PILKADA 1 ..  | L-62 |
| <b>Lampiran 20.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain threshold</i> 40% pada Data PILKADA 1..... | L-63 |
| <b>Lampiran 21.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain threshold</i> 50% pada Data PILKADA 1..... | L-64 |
| <b>Lampiran 22.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain threshold</i> 60% pada Data PILKADA 1..... | L-65 |
| <b>Lampiran 23.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain threshold</i> 75% pada Data PILKADA 1..... | L-66 |
| <b>Lampiran 24.</b> Hasil <i>Confusion Matrix</i> KNN k=3 pada Data PILKADA 2 .....  | L-67 |

|   |      |
|---|------|
| <b>Lampiran 25.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 40% pada Data PILKADA 2.....  | L-68 |
| <b>Lampiran 26.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 50% pada Data PILKADA 2.....  | L-69 |
| <b>Lampiran 27.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 60% pada Data PILKADA 2.....  | L-70 |
| <b>Lampiran 28.</b> Hasil <i>Confusion Matrix</i> KNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 75% pada Data PILKADA 2.....  | L-71 |
| <b>Lampiran 29.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 pada Data PILKADA 2 ..   | L-72 |
| <b>Lampiran 30.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 40% pada Data PILKADA 2..... | L-73 |
| <b>Lampiran 31.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 50% pada Data PILKADA 2..... | L-74 |
| <b>Lampiran 32.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 60% pada Data PILKADA 2..... | L-75 |
| <b>Lampiran 33.</b> Hasil <i>Confusion Matrix</i> MKNN k=3 dengan <i>Information Gain</i> <i>threshold</i> 75% pada Data PILKADA 2..... | L-76 |
| <b>Lampiran 34.</b> 10 Kata Teratas Hasil Seleksi Fitur <i>Information Gain</i> pada Data PILKADA 1 .....                               | L-77 |
| <b>Lampiran 35.</b> 10 Kata Teratas Hasil Seleksi Fitur <i>Information Gain</i> pada Data PILKADA 2 .....                               | L-78 |
| <b>Lampiran 36.</b> Dokumentasi <i>Source Code</i> .....  | L-79 |

# **BAB I**

## **PENDAHULUAN**

### **1.1 Pendahuluan**

Bab ini berisi pokok - pokok pikiran yang melandasi pembuatan skripsi ini. Pokok - pokok pikiran tersebut yaitu latar belakang, masalah penelitian, pertanyaan penelitian, tujuan penelitian, manfaat penelitian, dan batasan masalah.

### **1.2 Latar Belakang Masalah**

Internet telah menjadi bagian dari kehidupan manusia. Seiring berkembangnya teknologi, pengaksesan internet semakin mudah. Di Indonesia menurut APJII pada kuartal II/2020 pengguna internet telah mencapai 196,7 Juta atau 73,7% dari populasi Indonesia. Salah satu yang banyak diakses oleh pengguna internet ialah media sosial (*social media*). Menurut survey yang dilakukan APJII, penggunaan internet yang paling tinggi adalah media sosial dengan persentase 51,5 %.<sup>1)</sup>

Di media sosial masyarakat bisa mengekspresikan perasaan atau emosinya tentang apapun. Tentu ini menjadi peluang bagi perusahaan, pemerintah, ataupun pihak lainnya untuk mendapatkan opini-opini tersebut. Opini masyarakat sangat dibutuhkan dalam mengambil keputusan yang krusial. Namun, opini-opini di media

---

<sup>1)</sup> Kompas.com, Pengguna Internet Indonesia hingga Kuartal II 2020 Capai 196,7 Juta Orang, <https://money.kompas.com/read/2020/11/09/213534626/pengguna-internet-indonesia-hingga-kuartal-ii-2020-capai-1967-juta-orang> , diakses 9 November 2020

sosial tidak selalu terstruktur. Opini tersebut bisa dikeluarkan oleh warganet dengan bahasa yang tidak selalu baku dan tersebar di platform manapun. Jumlahnya bisa sangat banyak. Karena itulah diperlukan suatu sistem yang berguna untuk menganalisis sentimen tersebut yaitu analisis sentimen (*sentiment analysis*).

Analisis sentimen adalah suatu sistem yang berguna untuk mengetahui polaritas dari sentimen atau pendapat yang diberikan oleh pengguna atas subjek, area, atau produk tertentu di media daring. Sentimen tersebut dapat dikelompokkan menjadi dua jenis: 1) positif atau 2) negatif yang menentukan sikap umum orang terhadap suatu topik tertentu (Huq et al., 2017). Untuk melakukan analisis sentimen diperlukan metode pembelajaran. Salah satu metode yang sering digunakan adalah *K-Nearest Neighbor*. Metode ini memiliki kelebihan yaitu waktu *training* yang cepat dan relatif mudah untuk diimplementasikan (Hemmatian & Sohrabi, 2017).

Terdapat juga pengembangan dari metode K-NN yaitu *Modified K-Nearest Neighbor* (MK-NN). Metode ini mirip dengan K-NN namun terdapat peningkatan perhitungan untuk validitas dan *voting* bobot (Okfalisa et al., 2017). Pada K-NN jarak ketetanggaan hanya dihitung berdasarkan data uji dan data latih kemudian ketetanggaan terdekat dihitung berdasarkan k. Namun pada M-KNN terdapat proses perhitungan validitas pada data latih. Setelah itu, dihitung jarak antara data latih dan data uji. Selanjutnya bobot tertinggi pemungutan suara dari tetangga terdekat berdasarkan jumlah k atau tetangga akan dihitung.

Pada analisis sentimen masalah tingginya dimensi atribut data dapat berkontribusi pada akurasi sehingga kinerja sistem tidak memuaskan. Pengurangan sejumlah atribut dapat efektif untuk meningkatkan kinerja (Kurniawati & Pardede,

2018). Untuk itulah diperlukan metode seleksi fitur (*feature selection*) agar dapat memilih fitur-fitur yang penting saja. Salah satu teknik pemilihan fitur yang dapat digunakan adalah *Information Gain*. Metode ini menggunakan entropi untuk menghitung gain value. Berdasarkan *gain value*, fitur dipilih dengan *threshold* atau batas tertentu. Fitur yang terpilih itulah yang menjadi fitur terbaik dan dapat dilanjutkan ke tahap klasifikasi (Sihwi, Jati & Anggrainingsih 2018). Kelebihan dari metode ini yaitu sederhana untuk diimplementasikan dan kinerjanya cepat (Ahmad, Bakar, & Yaakub 2019).

Penelitian mengenai analisis sentimen dengan *K-Nearest Neighbor* pernah dilakukan oleh (Huq, Ali & Rahman, 2017). Pada penelitian tersebut K-NN dibandingkan dengan *Support Vector Machine (SVM)* menggunakan empat fitur dan lima fitur. Dari proses klasifikasi didapatkan K-NN memiliki akurasi tertinggi sekitar 80% dibandingkan SVM yang akurasi tertingginya sekitar 70% pada klasifikasi menggunakan lima fitur.

Selanjutnya penelitian tentang perbandingan *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor (M-KNN)* pernah dilakukan oleh (Okfalisa *et al.*, 2017) pada data Conditional Cash Transfer Implementation Unit (Unit Pelaksana Program Keluarga Harapan). Hasilnya K-NN mendapatkan akurasi tertinggi 94,95% dengan rata-rata akurasi 93.94% dan MK-NN mendapatkan akurasi tertinggi 99.51% dengan dengan rata-rata akurasi 99,20%.

Selanjutnya ada penelitian “Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor” oleh (Daeli & Adiwijaya, 2020). Pada penelitian ini digunakan polarity dataset v2.0 dari Cornell movie review. Hasilnya

K-NN mendapatkan akurasi tertinggi yaitu 60% pada K=3. Setelah menggunakan *Information Gain*, akurasi meningkat menjadi 96.8% pada K=3. Bahkan akurasi tertinggi dibanding metode lainnya yaitu *Naive Bayes*, *Support Vector Machine*, dan *Random Forest*.

Pada penelitian kali ini penulis akan membandingkan kinerja *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* pada analisis sentimen Bahasa Indonesia menggunakan seleksi fitur *Information Gain*. Penelitian ini dilakukan karena metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* dengan *Information Gain* sudah pernah digunakan untuk berbagai macam klasifikasi namun kinerjanya pada suatu data belum pernah dibandingkan secara langsung untuk analisis sentimen Bahasa Indonesia.

### **1.3 Rumusan Masalah**

Berdasarkan latar belakang permasalahan, timbul pertanyaan yang menjadi dasar penelitian ini yaitu:

1. Bagaimana kinerja *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* pada analisis sentimen?
2. Bagaimana pengaruh seleksi fitur *Information Gain* terhadap kinerja klasifikasi *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* pada analisis sentimen?

### **1.4 Tujuan Penelitian**

Tujuan yang akan dicapai dari penelitian ini adalah :



1. Membandingkan kinerja *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* pada analisis sentimen.
2. Mengetahui pengaruh seleksi fitur *Information Gain* terhadap kinerja klasifikasi *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* pada analisis sentimen.

### **1.5 Manfaat Penelitian**

Manfaat dari penelitian ini adalah:

1. Sistem dapat digunakan untuk mengetahui polaritas dari sebuah sentimen
2. Dapat menjadi referensi untuk penelitian - penelitian selanjutnya pada analisis sentimen.

### **1.6 Batasan Masalah**

Masalah pada penelitian ini terbatas pada berikut:

1. Data yang digunakan adalah data sentimen dalam Bahasa Indonesia yang telah digunakan pada penelitian sebelumnya
2. Data menggunakan format .csv
3. Klasifikasi terbatas pada dua kelas yaitu positif dan negatif.

### **1.7 Sistematika Penulisan**

Sistematika penulisan penelitian ini mengikuti standar sebagai berikut:

## **BAB I. PENDAHULUAN**

Pada bab ini dijelaskan latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan

## **BAB II. KAJIAN LITERATUR**

Pada bab ini dijelaskan mengenai dasar teori yang digunakan pada penelitian seperti definisi algoritma yang digunakan dan kajian literatur yang relevan pada penelitian ini

## **BAB III. METODOLOGI PENELITIAN**

Pada bab ini dijelaskan mengenai semua tahapan yang dilakukan dalam penelitian ini. Setiap tahapan akan dideskripsikan secara detail dan berdasarkan pada sebuah kerangka kerja.

## **BAB IV. PENGEMBANGAN PERANGKAT LUNAK**

Pada bab ini dijelaskan mengenai tahapan pengembangan perangkat lunak analisis sentimen menggunakan metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* tanpa seleksi fitur dan dengan seleksi fitur *Information Gain*

## **BAB V. HASIL DAN ANALISIS PENELITIAN**

Pada bab ini dijelaskan tentang hasil pengujian dengan menggunakan perangkat lunak yang telah dibangun dan analisis dari hasil pengujian tersebut.

## **BAB VI. KESIMPULAN DAN SARAN**

Pada bab ini diberikan kesimpulan dari hasil penelitian yang didapatkan dan saran oleh penulis untuk penelitian-penelitian selanjutnya

## 1.8 Kesimpulan

Pada bab ini telah dibahas mengenai latar belakang dilakukannya penelitian ini yaitu untuk mengembangkan perangkat lunak analisis sentimen dengan metode *K-Nearest Neighbor* dan *Modified K-Nearest Neighbor* menggunakan seleksi fitur *Information Gain*. Sentimen diklasifikasikan ke dalam kelas positif atau negatif. Akurasi klasifikasi kemudian akan dibandingkan. Dari penelitian ini akan diketahui metode mana yang mampu memberikan akurasi terbaik sehingga kesalahan pada analisis sentimen dapat diminimalisir.

## DAFTAR PUSTAKA

- Ahmad, B. and Yaakub. 2019. A Review of Feature Selection Techniques in Sentiment Analysis. *Intelligent Data Analysis*. *Intelligent Data Analysis* 23(1): 159-189
- Anawula, N.M.P. 2020. Pengaruh Synthetic Minority Oversampling Technique (Smote) Pada Sentimen Analisis Menggunakan Algoritma Naive Bayes. Skripsi Program Teknik Informatika FAKULTAS ILMU KOMPUTER Palembang (tidak dipublikasikan).
- Brandao, J. and W.P.Calixto. 2019. N-Gram and TF-IDF for Feature Extraction on Opinion Mining of Tweets with SVM Classifier. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) 2019: 1-5.
- Chandrashekar, G, S. Ferat. 2014. A Survey on Feature Selection Methods. *New York: Computers and Electrical Engineering* 40(1):6–28
- Daeli,N.O.F., and Adiwijaya. 2020. Sentiment Analysis on Movie Reviews Using Information Gain and K-Nearest Neighbor. *Journal Of Data Science And Its Applications* 3(1): 1-7
- Dhaneswara, C., Y. Azhar, dan N. Hayatin. 2020. Deteksi Berita Hoax Pada Dokumen Berbahasa Indonesia Menggunakan Metode Modified K-Nearest Neighbor. *Seminar Nasional Teknologi dan Rekayasa (SENTRA) 2020*: 2527-6050
- Han, Jiawei, M.Kamber and J.Pei. 2012. *Data Mining Concepts and Techniques Third Edition*. Elsevier Inc, Waltham, US.
- Hemmatian, F. and M.K. Sohrabi. 2017. A Survey on Classification Techniques for Opinion Mining and Sentiment Analysis: *Artif Intell Rev* 2019(52): 1495–1545
- Huq, M.R, A. Ali, and A. Rahman. 2017. Sentiment Analysis on Twitter Data using KNN and SVM. *International Journal of Advanced Computer Science and Applications (IJACSA)* 8(6): 19–25.
- Kurniawati, I. and H.F.Pardede. 2018. Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis. 2018 International Conference on Information Technology Systems and Innovation (ICITSI), 1-5.
- Liu, B. 2012. *Sentiment Analysis and Opinion Mining*.Morgan & Claypool.
- Muller, A. C. and S.Guido. 2017. *Introduction to Machine Learning with Python*. O'Reilly Media Inc, Sebastopol, US.
- Okfalisa, et al. 2017. Comparative Analysis of K-Nearest Neighbor and Modified K-Nearest Neighbor Algorithm for Data Classification. *International*

- Conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE) 2017: 294-298
- Ozdemir, S. 2016. Principles of Data Science. Packt Publishing, Brimingham, UK.
- Paramitha,A.A., Indrianti dan Y.A.Sari. 2020. Analisis Sentimen Terhadap Ulasan Pengguna MRT Jakarta Menggunakan Information Gain dan Modified K-Nearest Neighbor. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 4(4): 1125-1132
- Parvin, H., H. Alizadeh, and B.M.Bidgoli. 2008. MKNN: Modified K-Nearest Neighbor. Proceedings of the World Congress on Engineering and Computer Science WCECS. 2008(1): 1-4
- Sihwi,S.W., I.P.Jati and R.Anggrainingsih. 2018. Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier. International Seminar on Application for Technology of Information and Communication 2018: 190-195
- Tripathy,A., A. Agrawal,S.K. Rath. 2016. Classification of sentiment reviews using n-gram machine learning approach. Expert Systems With Applications 2019(57): 117–126
- Wibawa, D.W., M. Nasrun, C. Setianingsih. 2018. Sentiment Analysis on User Satisfaction Level of Cellular Data Service Using the K-Nearest Neighbor (K-Nn) Algorithm. International Conference on Control, Electronics, Renewable Energy and Communications (ICCEREC) 2018: 234-240