

**KOMPARASI METODE NAIVE BAYES  
DAN K-NEAREST NEIGHBOR PADA ANALISIS  
SENTIMEN RENCANA KENAIKAN UKT**



**TESIS**  
**oleh**  
**I KETUT DENI GUNAWAN**  
**NIM 2229101049**

**PROGRAM STUDI ILMU KOMPUTER**  
**PROGRAM PASCASARJANA**  
**UNIVERSITAS PENDIDIKAN GANESHA**  
**2025**

**KOMPARASI METODE NAIVE BAYES  
DAN K-NEAREST NEIGHBOR PADA ANALISIS  
SENTIMEN RENCANA KENAIKAN UKT**

**TESIS**

Diajukan kepada  
Universitas Pendidikan Ganesha  
untuk Memenuhi Sebagian Persyaratan  
Memperoleh Gelar Magister Pendidikan  
Program Studi Ilmu Komputer

**oleh**

**I KETUT DENI GUNAWAN**

**NIM 2229101049**

**PROGRAM STUDI ILMU KOMPUTER  
PROGRAM PASCASARJANA  
UNIVERSITAS PENDIDIKAN GANESHA**

**2025**

## LEMBAR PERSETUJUAN PEMBIMBING

Tesis oleh I Ketut Deni Gunawan ini telah diperiksa dan disetujui oleh Tim Pembimbing.

Singaraja, 14 Agustus 2025

Pembimbing I



Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs  
NIP. 198307252008011008

Pembimbing II



Dr. I Gede Aris Gunadi, S.Si. M.Kom.  
NIP. 197703182008121004

## LEMBAR PERSETUJUAN TIM PENGUJI

Tesis oleh I Ketut Deni Gunawan ini telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Disetujui pada tanggal: 14 Agustus 2025

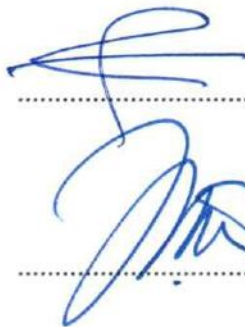
Oleh  
Tim Penguji



....., Ketua (Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs.)  
NIP. 198307252008011008



....., Anggota (Dr. I Gede Aris Gunadi, S.Si. M.Kom.)  
NIP. 197703182008121004



....., Anggota (Prof. Drs. Sariyasa, M.Sc., Ph.D.)  
NIP. 196406151989021001



....., Anggota (Dr. I Made Agus Wirawan, S.Kom., M.Cs.)  
NIP. 198408272008121001

Mengetahui Direktur  
Pascasarjana Undiksha,  
  
Prof. Dr. I Nyoman Jampel, M.Pd.  
NIP. 195910101986031003

## LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Komputer dari Program Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya saya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila di kemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagianbagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 14 Agustus 2025

Yang memberi pernyataan,

A 10,000 Rupiah Indonesian revenue stamp (Meterai Tempel) is placed over the signature. The stamp features the Garuda emblem, the text "10000", "METERAI TEMPEL", and the serial number "9BAMX416812938".

I Ketut Deni Gunawan

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadapan Tuhan Yang Maha Esa atas anugerah-Nya, sehingga tesis yang berjudul “Komparasi Metode Naive Bayes dan K-Nearest Neighbor Pada Analisis Sentimen Rencana Kenaikan UKT” dapat terselesaikan sesuai dengan yang direncanakan.

Selama penyusunan tesis ini, penulis menemui beberapa kendala dan kesulitan, namun berkat dukungan dan bantuan dari berbagai pihak, penulis dapat mengatasi hambatan dan kesulitan tersebut, sehingga pada kesempatan ini penulis ingin menyampaikan rasa terimakasih yang sebesar-besarnya kepada:

1. Bapak Rektor Universitas Pendidikan Ganesha, yang telah memberikan dukungan secara moril dan memfasilitasi berbagai kepentingan akademik, selama Penulis menempuh studi di Program Pascasarjana Universitas Pendidikan Ganesha.
2. Bapak Direktur Program Pascasarjana Universitas Pendidikan Ganesha, yang telah banyak memberikan bimbingan dan arahan serta petunjuk hingga terselesaikannya tesis ini.
3. Ketua Program Studi Ilmu Komputer Universitas Pendidikan Ganesha yang telah banyak membantu dan menyemangati penulis selama menjalani pendidikan dan penyusunan tesis ini.
4. Bapak Dr. Ir. I Made Gede Sunarya, S.Kom., M.Cs. sebagai Dosen Pembimbing I yang telah memberikan bimbingan, arahan dan petunjuk yang bermanfaat serta memacu semangat Penulis untuk menyelesaikan tesis ini sesuai yang diharapkan.
5. Bapak Dr. I Gede Aris Gunadi, S.Si. M.Kom. selaku Dosen Pembimbing II yang dengan sabar memberikan bimbingan, arahan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu mengatasi berbagai kendala dalam perjalanan studi dan penyelesaian tesis ini.
6. Bapak dan Ibu dosen dan Staf, yang telah memberikan pelajaran kepada Penulis selama mengikuti studi di Program Studi Ilmu Komputer Universitas Pendidikan Ganesha.
7. Segenap keluarga tercinta, Bapak, Ibu, dan Saudara yang telah banyak membantu memberikan semangat dan motivasi serta dukungan moril dan materiil dalam penulisan tesis ini.
8. Semua rekan, sahabat dan pihak-pihak yang tidak dapat Penulis sebutkan satu-per satu atas bantuan secara langsung maupun tidak langsung dalam penulisan tesis ini.

Akhir kata Penulis mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penulisan tesis ini dan Penulis berharap semoga tesis ini dapat bermanfaat bagi kita semua dan dunia pendidikan di masa yang akan datang.

Singaraja,     Agustus 2025

Penulis

## DAFTAR ISI

|                                               |      |
|-----------------------------------------------|------|
| KATA PENGANTAR .....                          | v    |
| ABSTRAK.....                                  | vi   |
| DAFTAR ISI.....                               | viii |
| DAFTAR TABEL.....                             | x    |
| DAFTAR GAMBAR .....                           | xi   |
| DAFTAR LAMPIRAN.....                          | xii  |
| BAB I PENDAHULUAN .....                       | 1    |
| 1.1 Latar Belakang .....                      | 1    |
| 1.2 Identifikasi Masalah .....                | 5    |
| 1.3 Batasan Penelitian .....                  | 6    |
| 1.4 Rumusan Masalah .....                     | 6    |
| 1.5 Tujuan Penelitian .....                   | 6    |
| 1.6 Manfaat Penelitian .....                  | 7    |
| BAB II LANDASAN TEORI .....                   | 8    |
| 2.1 Kajian Pustaka.....                       | 8    |
| 2.1.1 Sentimen Analisis.....                  | 8    |
| 2.1.2 Media sosial X.....                     | 9    |
| 2.1.3 Naive Bayes .....                       | 11   |
| 2.1.4 K-Nearest Neighbors.....                | 12   |
| 2.2 Kajian Penelitian yang Relevan.....       | 15   |
| 2.3 Kerangka Berpikir.....                    | 33   |
| BAB III METODE PENELITIAN.....                | 34   |
| 3.1 Metode Penelitian.....                    | 34   |
| 3.1.1 Studi Pustaka.....                      | 35   |
| 3.1.2 Pengumpulan Data .....                  | 35   |
| 3.1.3 <i>Preprocessing</i> Data .....         | 36   |
| 3.2 Proses Pembobotan Kata (Term TF-IDF)..... | 57   |
| 3.3 Analisis Data Klasifikasi .....           | 57   |
| 3.4 Pengujian.....                            | 70   |
| BAB IV HASIL DAN PEMBAHASAN.....              | 73   |
| 4.1 Deskripsi Data.....                       | 73   |
| 4.2 Data Preprocessing.....                   | 74   |
| 4.3 Pembobotan TF-IDF .....                   | 76   |
| 4.4 Skenario Uji Coba.....                    | 77   |
| 4.5 Klasifikasi .....                         | 78   |

|                      |                                                                               |    |
|----------------------|-------------------------------------------------------------------------------|----|
| 4.6                  | Perbandingan Hasil Metode Naïve Bayes dan K-Nearest Neighbor .....            | 82 |
| 4.6.1                | <i>Accuracy, Precision, Recall, dan F1-Score</i> .....                        | 83 |
| 4.6.2                | <i>Receiver Operating Characteristic – Area Under the Curve (ROC-AUC)</i> ... | 85 |
| BAB V PENUTUP.....   |                                                                               | 89 |
| 5.1                  | Rangkuman .....                                                               | 89 |
| 5.2                  | Kesimpulan .....                                                              | 91 |
| 5.2                  | Saran .....                                                                   | 93 |
| DAFTAR PUSTAKA ..... |                                                                               | 96 |

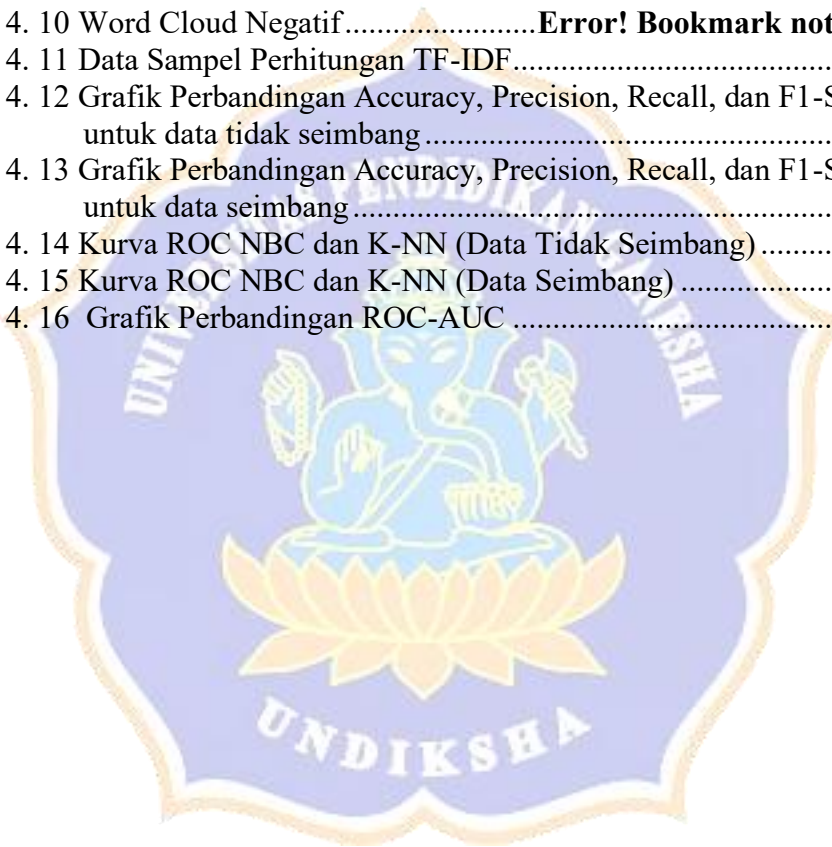


## DAFTAR TABEL

|                                                                                              |    |
|----------------------------------------------------------------------------------------------|----|
| Tabel 2. 1 Penelitian Terkait .....                                                          | 15 |
| Tabel 3. 1 Data Labeling.....                                                                | 36 |
| Tabel 3. 2 Case Folding .....                                                                | 39 |
| Tabel 3. 3 Normalisasi .....                                                                 | 44 |
| Tabel 3. 4 Tokenizing .....                                                                  | 47 |
| Tabel 3. 5 Contoh Daftar Stopword dalam Bahasa Indonesia .....                               | 51 |
| Tabel 3. 6 Contoh Stopword .....                                                             | 51 |
| Tabel 3. 7 Contoh Stemming .....                                                             | 55 |
| Tabel 3. 8 Contoh Data Latih.....                                                            | 58 |
| Tabel 3. 9 Frekuensi Kemunculan Data.....                                                    | 59 |
| Tabel 3. 10 Probabilitas Sentimen Positif dan Negatif .....                                  | 60 |
| Tabel 3. 11 Data Uji.....                                                                    | 61 |
| Tabel 3. 12 Menunjukkan Hasil Pengujian Data yang Sudah Melalui Tahap<br>Preprocessing ..... | 62 |
| Tabel 3. 13 Probabilitas Sentimen Positif dan Negatif .....                                  | 63 |
| Tabel 3. 14 contoh data latih yang digunakan dalam perhitungan KNN .....                     | 64 |
| Tabel 3. 15 Hasil Pembobotan Dokumen TF-IDF .....                                            | 65 |
| Tabel 3. 16 Perhitungan Bobot Setiap Dokumen.....                                            | 66 |
| Tabel 3. 17 Perhitungan Cosine Similiarity .....                                             | 67 |
| Tabel 3. 18 Panjang Vektor.....                                                              | 68 |
| Tabel 3. 19 Hasil Perhitungan Rumus Cosine Similarity .....                                  | 69 |
| Tabel 3. 20 Hasil Perhitungan Kemiripan Dokumen .....                                        | 69 |
| Tabel 3. 21 Menampilkan Hasil Kemiripan Tertinggi dari Dokumen Uji.....                      | 69 |
| Tabel 3. 22. Rumus akurasi, presisi, recall, dan F1-Score .....                              | 70 |
| Tabel 4. 1 Sub-iterasi Pengujian pada Metode KNN .....                                       | 77 |
| Tabel 4. 2 Hasil Pengujian Klasifikasi Naïve Bayes.....                                      | 79 |
| Tabel 4. 3 Hasil Pengujian Klasifikasi dengan K-NN .....                                     | 81 |
| Tabel 4. 4 Hasil Akurasi Metode Naive Bayes dan KNN.....                                     | 83 |
| Tabel 4. 5 Hasil Precision Metode Naive Bayes dan KNN .....                                  | 83 |
| Tabel 4. 6 Hasil Recall Metode Naive Bayes dan KNN .....                                     | 84 |
| Tabel 4. 7 Hasil F1-Score Metode Naive Bayes dan KNN.....                                    | 84 |
| Tabel 4. 8 Hasil ROC-AUC Metode Naive Bayes dan KNN .....                                    | 85 |

## DAFTAR GAMBAR

|                                                                                                              |                                     |
|--------------------------------------------------------------------------------------------------------------|-------------------------------------|
| Gambar 3. 1 Alur Penelitian.....                                                                             | 34                                  |
| Gambar 4. 1 Jumlah Tweet Positif dan Negatif.....                                                            | 74                                  |
| Gambar 4. 2 Contoh Tweet Positif Dan Negatif.....                                                            | 74                                  |
| Gambar 4. 3 Hasil Pengumpulan Data Sentimen Positif Dan Negatif.....                                         | 74                                  |
| Gambar 4. 4 Hasil Proses Normalisasi.....                                                                    | 75                                  |
| Gambar 4. 5 Hasil Proses Case Folding.....                                                                   | 75                                  |
| Gambar 4. 6 Hasil Proses Stopword Removal atau Filtering.....                                                | 75                                  |
| Gambar 4. 7 Hasil Proses Tokenizing.....                                                                     | 76                                  |
| Gambar 4. 8 Hasil Proses Stemming.....                                                                       | 76                                  |
| Gambar 4. 9 Word Cloud Positif.....                                                                          | <b>Error! Bookmark not defined.</b> |
| Gambar 4. 10 Word Cloud Negatif.....                                                                         | <b>Error! Bookmark not defined.</b> |
| Gambar 4. 11 Data Sampel Perhitungan TF-IDF.....                                                             | 77                                  |
| Gambar 4. 12 Grafik Perbandingan Accuracy, Precision, Recall, dan F1-Score<br>untuk data tidak seimbang..... | 84                                  |
| Gambar 4. 13 Grafik Perbandingan Accuracy, Precision, Recall, dan F1-Score<br>untuk data seimbang.....       | 85                                  |
| Gambar 4. 14 Kurva ROC NBC dan K-NN (Data Tidak Seimbang).....                                               | 86                                  |
| Gambar 4. 15 Kurva ROC NBC dan K-NN (Data Seimbang).....                                                     | 86                                  |
| Gambar 4. 16 Grafik Perbandingan ROC-AUC .....                                                               | 87                                  |



## DAFTAR LAMPIRAN

|                                             |    |
|---------------------------------------------|----|
| Lampiran 1. Data set setelah labeling ..... | 99 |
|---------------------------------------------|----|

