



**PERBANDINGAN PEMBOBOTAN *TF-IDF* DAN *BAG OF WORDS*  
DENGAN SELEKSI FITUR *CHI-SQUARE* PADA ANALISIS SENTIMEN  
NFT GHOZALI MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES***

**Skripsi**

**Untuk Memenuhi Persyaratan  
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh  
TRI PRABOWO  
NIM 1711016310032**

**PROGRAM STUDI S-1 ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMBUNG MANGKURAT  
BANJARBARU  
JULI 2023**

SKRIPSI

PERBANDINGAN PEMBOBOTAN *TF-IDF* DAN *BAG OF WORDS* DENGAN  
SELEKSI FITUR *CHI-SQUARE* PADA ANALISIS SENTIMEN NFT  
GHOZALI MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES*

Oleh :  
**TRI PRABOWO**  
NIM. 1711016310032


Telah dipertahankan di depan Dosen Penguji pada Tanggal 12 Juli 2023

Susunan Dosen Penguji:

Pembimbing I

Dosen Penguji I


  
Dwi Kartaji, S.Kom., M.Kom.  
NIP. 198704212012122003

  
Muliadi, S.Kom., M.Cs.  
NIP. 197804222010121002

Pembimbing II

Dosen Penguji II

  
Radityo Adi Nugroho, S.T., M.Kom.  
NIP. 198212042008011006

  
Andi Firmadi, S.Si., M.T.  
NIP. 197307252008011006



12 Juli 2023

Program Studi Ilmu Komputer

Budiman, S.T., M.Kom.  
NIP. 19770325 200812 1 001

## **PERNYATAAN**

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 12 Juli 2023



**Tri Prabowo**

**NIM 1711016310032**

## ABSTRAK

### PERBANDINGAN PEMBOBOTAN *TF-IDF* DAN *BAG OF WORDS* DENGAN SELEKSI FITUR *CHI-SQUARE* PADA ANALISIS SENTIMEN NFT GHOZALI MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES*

(Oleh: Tri Prabowo; Pembimbing; Dwi Kartini, S.Kom., M.Kom. dan Radityo Adi Nugroho, S.T., M.Kom.; 2023; halaman )

NFT atau ( Non-Fungible Token ) sebagai suatu aset digital yang mewakili objek dunia nyata yang umumnya sama dengan crypto. Sultan Gustaf Al Ghozali menjual NFT berupa foto selfie yang mendapatkan keuntungan miliaran rupiah, mendadak trending menjadi perbincangan di dunia maya dan semakin dikenal sebagian masyarakat salah satunya media sosial twitter. Analisis sentimen dilakukan bertujuan untuk memahami opini masyarakat tentang kasus NFT Ghozali Ghozalu. Pada penelitian ini melakukan perbandingan ekstraksi fitur *Bag Of Words* dan *TF-IDF* dengan menggunakan seleksi fitur *Chi-Square* kemudian dilakukan klasifikasi menggunakan metode *Naïve Bayes*. Data yang digunakan adalah data tweet dari twitter yang berjumlah 1000 data dan terdiri dari 2 kelas yaitu positif dan negatif dengan masing-masing kelas sebanyak 500 data. Hasil dari perbandingan tersebut diperoleh bahwa performa akurasi tertinggi dihasilkan dari ekstraksi fitur *TF-IDF* dengan menggunakan seleksi fitur *Chi-Square* dengan hasil akurasi 86,90% dari 2255 fitur setelah diseleksi fitur *Chi-square*, sedangkan ekstraksi fitur *Bag Of Words* dengan seleksi fitur *Chi-Square* dihasilkan 84,90% dari 2255 fitur setelah diseleksi fitur *Chi-Square*. Sehingga penelitian ini memberikan hasil kinerja terbaik pada pengujian dengan menggunakan *TF-IDF* dengan seleksi fitur *Chi-Square*.

Kata kunci : Bag Of Words, TF-IDF, Naïve Bayes, Klasifikasi, Text Mining, ChiSquare, Twitter, NFT, GhozaliGhozalu

## ***ABSTRACT***

**A COMPARISON OF TF-IDF AND BAG OF WORDS WEIGHTING WITH CHI-SQUARE FEATURE SELECTION IN SENTIMENT ANALYSIS OF NFT GHOZALI USING NAÏVE BAYES CLASSIFICATION** (By: Tri Prabowo; Advisors: Dwi Kartini, S.Kom., M.Kom. and Radityo Adi Nugroho, S.T., M.Kom.; 2023; pages)

NFT, or Non-Fungible Token, represents a digital asset that represents real-world objects and is commonly associated with cryptocurrencies. Sultan Gustaf Al Ghozali sold an NFT in the form of a selfie photo and gained billions of rupiah in profit. This sudden trend became a hot topic of discussion in the online world, especially on the Twitter social media platform. Sentiment analysis was conducted with the aim of understanding public opinion regarding the NFT Ghozali Ghozalu case. This study compared the feature extraction methods of *Bag Of Words* and *TF-IDF* using *Chi-Square* feature selection, followed by classification using the Naïve Bayes method. The data used consisted of 1000 tweets from Twitter, divided into two classes: positive and negative, with each class containing 500 tweets. The results of the comparison showed that the highest accuracy performance was achieved using *TF-IDF* feature extraction with *Chi-Square* feature selection, with an accuracy of 86,90% using 2255 selected features. On the other hand, the *Bag Of Words* feature extraction with *Chi-Square* feature selection resulted in an accuracy of 84,90% using 2255 selected features. Therefore, this research indicates that the best performance was achieved in the testing phase using *TF-IDF* with *Chi-Square* feature selection.

Keywords: *Bag Of Words*, *TF-IDF*, *Naïve Bayes*, Classification, Text Mining, *Chi-Square*, Twitter, NFT, GhozaliGhozalu

## PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “PERBANDINGAN PEMBOBOTAN *TF-IDF* DAN *BAG OF WORDS* DENGAN SELEKSI FITUR *CHI-SQUARE* PADA ANALISIS SENTIMEN NFT GHOZALI MENGGUNAKAN KLASIFIKASI *NAÏVE BAYES* ” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

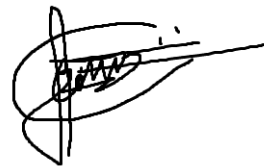
1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini, terutama untuk Bapak, Ibu, Mas Rinto, Mas Rio dan Mba Tina.
2. Ibu Dwi Kartini, S.Kom.,M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Radityo Adi Nugroho, S.T.,M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Ucapan terima kasih kepada Mita Azzahra Ramadhan, Ryan, Barok, Arif, Shega, Eza, Aldo, Ipang, Ubai, Ilham, Zamzam, Yudhit, Said, Luthfi, Aprix yang memberikan dukungan dalam proses mengerjakan skripsi.

8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 30 Juni 2023

A handwritten signature in black ink, appearing to read 'Tri Prabowo', with a long horizontal stroke extending to the right.

Tri Prabowo

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>i</b>
<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN.....</b>	<b>iii</b>
<b>ABSTRAK .....</b>	<b>iv</b>
<b>ABSTRACT .....</b>	<b>v</b>
<b>PRAKATA.....</b>	<b>vi</b>
<b>DAFTAR ISI.....</b>	<b>viii</b>
<b>DAFTAR TABEL .....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>Error! Bookmark not defined.</b>
1.1 Latar Belakang.....	<b>Error! Bookmark not defined.</b>
1.2 Rumusan Masalah.....	<b>Error! Bookmark not defined.</b>
1.3 Batasan Masalah .....	<b>Error! Bookmark not defined.</b>
1.4 Tujuan .....	<b>Error! Bookmark not defined.</b>
1.5 Manfaat Penelitian.....	<b>Error! Bookmark not defined.</b>
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>Error! Bookmark not defined.</b>
2.1 Kajian Terdahulu .....	<b>Error! Bookmark not defined.</b>
2.2 Landasan Teori .....	<b>Error! Bookmark not defined.</b>
2.2.1 NFT .....	<b>Error! Bookmark not defined.</b>
2.2.2 Analisis Sentimen .....	<b>Error! Bookmark not defined.</b>
2.2.3 Data Pre-Processing .....	<b>Error! Bookmark not defined.</b>
2.2.4 Pembobotan <i>TF-IDF</i> .....	<b>Error! Bookmark not defined.</b>
2.2.5 Pembobotan <i>Bag Of Words</i> .....	<b>Error! Bookmark not defined.</b>



2.2.6	Seleksi Fitur <i>Chi-Square</i> .....	<b>Error! Bookmark not defined.</b>
2.2.7	<i>Naïve Bayes</i> .....	<b>Error! Bookmark not defined.</b>
2.2.8	<i>K-Fold Cross Validation</i> .....	<b>Error! Bookmark not defined.</b>
2.2.9	<i>Confussion Matrix</i> .....	<b>Error! Bookmark not defined.</b>
<b>BAB III METODE PENELITIAN</b> .....		<b>Error! Bookmark not defined.</b>
3.1	Bahan Penelitian .....	<b>Error! Bookmark not defined.</b>
3.2	Alat Penelitian .....	<b>Error! Bookmark not defined.</b>
3.3	Variabel Penelitian .....	<b>Error! Bookmark not defined.</b>
3.4	Prosedur Penelitian .....	<b>Error! Bookmark not defined.</b>
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		<b>Error! Bookmark not defined.</b>
4.1	Hasil .....	<b>Error! Bookmark not defined.</b>
4.1.1	Pengumpulan Data .....	<b>Error! Bookmark not defined.</b>
4.1.2	Data <i>Pre-Processing</i> .....	<b>Error! Bookmark not defined.</b>
4.1.2.1	<i>Cleansing</i> .....	<b>Error! Bookmark not defined.</b>
4.1.2.2	<i>Case Folding</i> .....	<b>Error! Bookmark not defined.</b>
4.1.2.3	<i>Stemming</i> .....	<b>Error! Bookmark not defined.</b>
4.1.2.4	<i>StopWord</i> .....	<b>Error! Bookmark not defined.</b>
4.1.3	Ekstraksi Fitur .....	<b>Error! Bookmark not defined.</b>
4.1.3.1	Ekstraksi Fitur <i>Bag Of Words</i> .....	<b>Error! Bookmark not defined.</b>
4.1.3.2	Ekstraksi Fitur <i>TF-IDF</i> .....	<b>Error! Bookmark not defined.</b>
4.1.5	Data Mining .....	<b>Error! Bookmark not defined.</b>
4.1.6	Klasifikasi <i>Naïve Bayes</i> .....	<b>Error! Bookmark not defined.</b>
4.1.6.1	Klasifikasi <i>Bag Of Words</i> dengan seleksi fitur <i>Chi-Square</i> menggunakan <i>Naïve Bayes</i> .....	<b>Error! Bookmark not defined.</b>
4.1.6.2	Klasifikasi <i>TF-IDF</i> dengan seleksi fitur <i>Chi-Square</i> menggunakan <i>Naïve Bayes</i> .....	<b>Error! Bookmark not defined.</b>

4.1.7	Evaluasi.....	<b>Error! Bookmark not defined.</b>
4.1.7.1	Evaluasi <i>Bag Of Words</i> dengan seleksi fitur <i>Chi-Square</i> .....	<b>Error! Bookmark not defined.</b>
4.1.7.2	Evaluasi <i>TF-IDF</i> dengan seleksi fitur <i>Chi-Square</i> .....	<b>Error! Bookmark not defined.</b>
4.2	Pembahasan .....	<b>Error! Bookmark not defined.</b>
<b>BAB V PENUTUP.....</b>		<b>Error! Bookmark not defined.</b>
5.1	Kesimpulan.....	<b>Error! Bookmark not defined.</b>
5.2	Saran .....	<b>Error! Bookmark not defined.</b>
<b>DAFTAR PUSTAKA .....</b>		<b>Error! Bookmark not defined.</b>
<b>LAMPIRAN.....</b>		<b>Error! Bookmark not defined.</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
<b>Tabel 1.</b> Keaslian Penelitian.....	8
<b>Tabel 2.</b> Perancangan Penelitian.....	<b>Error! Bookmark not defined.</b>
<b>Tabel 3.</b> Contoh Kalimat Setiap Dokumen.....	<b>Error! Bookmark not defined.</b>
<b>Tabel 4.</b> <i>Bag of words</i> .....	<b>Error! Bookmark not defined.</b>
<b>Tabel 5.</b> <i>Confusion Matrix</i> .....	<b>Error! Bookmark not defined.</b>
<b>Tabel 6.</b> Data Tweet Crawling .....	<b>Error! Bookmark not defined.</b>
<b>Tabel 7.</b> Jumlah Data Perkelas .....	<b>Error! Bookmark not defined.</b> 0
<b>Tabel 8.</b> Data Tweet dengan Label.....	<b>Error! Bookmark not defined.</b> 0
<b>Tabel 9.</b> Data <i>Cleansing</i> .....	<b>Error! Bookmark not defined.</b> 1
<b>Tabel 10.</b> Data Hasil <i>Cleansing</i> .....	<b>Error! Bookmark not defined.</b> 1
<b>Tabel 11.</b> Data Hasil <i>Case Folding</i> .....	<b>Error! Bookmark not defined.</b>
<b>Tabel 12.</b> Data <i>Stemming</i> .....	<b>Error! Bookmark not defined.</b>
<b>Tabel 13.</b> Kamus <i>Stopwords</i> .....	<b>Error! Bookmark not defined.</b> 4
<b>Tabel 14.</b> Data Penghapusan <i>Stopwords</i> .....	<b>Error! Bookmark not defined.</b> 5
<b>Tabel 15.</b> Proses Perhitungan <i>Bag Of Word</i> .....	<b>Error! Bookmark not defined.</b> 6
<b>Tabel 16.</b> Pemberian Bobot <i>Bag Of Word</i> .....	<b>Error! Bookmark not defined.</b> 6
<b>Tabel 17.</b> Data Hasil <i>Bag Of Word</i> .....	<b>Error! Bookmark not defined.</b> 6
<b>Tabel 18.</b> Contoh Hasil Ekstraksi Fitur <i>TF-IDF</i> .....	<b>Error! Bookmark not defined.</b> 7
<b>Tabel 19.</b> Contoh Perhitungan <i>TF-IDF</i> .....	<b>Error! Bookmark not defined.</b> 8
<b>Tabel 20.</b> Perhitungan <i>TF-IDF</i> per-fitur.....	<b>Error! Bookmark not defined.</b> 8
<b>Tabel 21.</b> Pemangkatan hasil <i>TF-IDF</i> .....	<b>Error! Bookmark not defined.</b> 8
<b>Tabel 22.</b> Penjumlahan hasil pangkat <i>TF-IDF</i> pada setiap baris	<b>Error! Bookmark not defined.</b> 9

<b>Tabel 23.</b> Perhitungan penyebut <i>enclidean normalization</i>	<b>Error! Bookmark not defined.</b>
<b>Tabel 24.</b> Membagi nilai <i>TF-IDF</i> awal dengan nilai <i>enclidean normalization</i>	<b>Error! Bookmark not defined.</b>
<b>Tabel 25.</b> Tabel Kontingensi Seleksi Fitur <i>Chi-Square</i>	30
<b>Tabel 26.</b> Hasil perhitungan nilai <i>Chi-Square Bag Of Word</i>	30
<b>Tabel 27.</b> Contoh Perhitungan <i>Chi-Square Bag Of Words</i>	31
<b>Tabel 28.</b> Tabel <i>Chi-Square</i>	31
<b>Tabel 29.</b> Peringkat Hasil Seleksi Fitur <i>Chi-Square Bag Of Words</i>	32
<b>Tabel 30.</b> Hasil <i>TF (Term Frequence)</i> dari <i>TF-IDF</i>	32
<b>Tabel 31.</b> Perhitungan <i>Chi-Square TF-IDF</i>	33
<b>Tabel 32.</b> Contoh perhitungan <i>Chi-Square TF-IDF</i>	33
<b>Tabel 33.</b> Tabel <i>Chi-Square</i>	34
<b>Tabel 34.</b> Hasil Rangking <i>Chi-Square TF-IDF</i>	34
<b>Tabel 35.</b> 10 K-Fold Cross Validation	34
<b>Tabel 36.</b> Frekuensi term <i>Bag Of Words</i> dengan seleksi fitur <i>Chi-Square</i>	35
<b>Tabel 37.</b> Probabilitas term <i>Bag Of Words</i> dengan seleksi fitur <i>Chi-Square</i>	35
<b>Tabel 38.</b> Hasil kali probabilitas <i>Bag Of Words</i> dengan seleksi fitur <i>Chi-Square</i>	37
<b>Tabel 39.</b> Probabilitas <i>Prior</i> Kelas c pada Data Training	38
<b>Tabel 40.</b> Probabilitas Fitur	38
<b>Tabel 41.</b> Hasil Kali Probabilitas <i>TF-IDF</i> dengan seleksi fitur <i>Chi-Square</i>	39
<b>Tabel 42.</b> Hasil Akurasi setiap fold <i>Bag Of Words</i> dengan <i>Chi-Square</i>	39
<b>Tabel 43.</b> <i>Confusion Matrix Naïve Bayes + Bag Of Words + Chi-Square</i>	40
<b>Tabel 44.</b> Perhitungan <i>Accuracy Bag Of Words + Chi-Square + Naïve Bayes</i>	40
<b>Tabel 45.</b> Hasil Akurasi setiap fold <i>TF-IDF</i> dengan <i>Chi-Square</i>	41
<b>Tabel 46.</b> <i>Confusion Matrix Naïve Bayes + TF-IDF + Chi-Square</i>	42

<b>Tabel 47.</b> Perhitungan <i>Accuracy TF-IDF + Chi-Square + Naïve Bayes</i> .....	42
--	----

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
<b>Gambar 1.</b> <i>K-Fold Cross Validation (K-10)</i> .....	<b>Error! Bookmark not defined.</b>
<b>Gambar 2.</b> Alur Penelitian.....	<b>Error! Bookmark not defined.</b>
<b>Gambar 3.</b> Grafik Perbandingan Akurasi <i>Bag Of Word</i> dan <i>TF-IDF</i> dengan <i>Chi-Square</i> .....	<b>Error! Bookmark not defined.</b>

## DAFTAR LAMPIRAN

### **Lampiran**

Lampiran 1 Deskripsi Dataset

Lampiran 2 Kamus Tala

Lampiran 3 Data Set

Lampiran 4 Source Code

