KOMPARASI ALGORITMA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR (K-NN)* UNTUK KLASIFIKASI *FAKE JOB ADVERTISEMENT* DI *PLATFORM* MEDIA SOSIAL

SKRIPSI

Untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Matematika

Oleh:
SITI HARDIYANTI
NIM. 2211011120009

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
2026

HALAMAN PENGESAHAN

SKRIPSI

KOMPARASI ALGORITMA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR (K-NN)* UNTUK KLASIFIKASI *FAKE JOB ADVERTISEMENT* DI PLATFORM MEDIA SOSIAL

Oleh:

Siti Hardiyanti

NIM.2211011120009

telah dipertahankan di depan Dosen Penguji pada tanggal 10 Juli 2026.

Susunan Dosen Penguji:

Pembimbing I

Oni Soesanto, S.Si., M.Si.

NIP. 197301262005011003

Pembimbing II

Thresye, S.Si., M.Si.

NIP. 197205042000122002

Dosen Penguji

1. **Saman Abdurrahman, S.Si., M.Sc.** (V2)
2. **Dr. Muhammad Ahsar Karim, S.Si., M.Sc.**

Banjarbaru, 23 Juli 2026

Ketua Jurusan Matematika

FMIPA ULM

Dr. Na'imah Hijriati, S.Si., M.Si.

NIP. 197911222008012013

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, 23 Juli 2026

Siti Hardiyanti

2211011120009

ABSTRAK

KOMPARASI ALGORITMA *NAÏVE BAYES* DAN *K-NEAREST NEIGHBOR (K-NN)* UNTUK KLASIFIKASI *FAKE JOB ADVERTISEMENT* DI *PLATFOR* MEDIA SOSIAL (oleh : Siti Hardiyanti ; Oni Soesanto, Theresye,100 halaman)

Maraknya penyebaran iklan lowongan pekerjaan melalui media sosial meningkatkan kebutuhan akan metode yang mampu membedakan iklan asli dan *Fake Job Advertisement* secara otomatis. Penelitian ini bertujuan merepresentasikan teks *Fake Job Advertisement* ke dalam bentuk numerik melalui tahapan *text pre-processing* dan pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF) sebagai dasar ekstraksi fitur, serta membandingkan efektivitas algoritma *Bernoulli Naïve Bayes, Multinomial Naïve Bayes, dan K-Nearest Neighbor (K-NN)* dalam mengklasifikasikan *Fake Job Advertisement* pada platform media sosial. Dataset yang digunakan terdiri atas 246 data yang diperoleh dari berbagai platform media sosial dan sumber sekunder, kemudian dilabeli ke dalam dua kelas, yaitu penipuan dan asli. Tahapan *text pre-processing* meliputi *data cleaning, case folding, tokenization, normalization, stopword removal,* dan *stemming*. Selanjutnya, data diekstraksi menggunakan TF-IDF dan dibagi ke dalam tiga skenario pelatihan dan pengujian, yaitu 90%:10%, 80%:20%, dan 70%:30%. Evaluasi dilakukan menggunakan *confusion matrix* berdasarkan metrik *accuracy, precision, recall,* dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *K-Nearest Neighbor (K-NN*) memberikan performa terbaik dibandingkan *Bernoulli Naïve Bayes dan Multinomial Naïve Bayes.* Performa terbaik diperoleh pada skenario 90% data latih dan 10% data uji dengan *accuracy* sebesar 92,00%, serta *precision, recall,* dan *F1-score* masing-masing sebesar 92,31%, sehingga *K-NN* merupakan algoritma yang paling efektif dalam mengklasifikasikan *Fake Job Advertisement* pada platform media sosial

Kata kunci: *Fake Job Advertisement*, *Text Pre-processing*, *TF-IDF*, *Naïve Bayes*, *K-Nearest Neighbor (K-NN).*

ABSTRACT

COMPARISON OF NAÏVE BAYES AND K-NEAREST NEIGHBOR (K-NN) ALGORITHMS FOR FAKE JOB ADVERTISEMENT CLASSIFICATION ON SOCIAL MEDIA PLATFORMS (by: Siti Hardiyanti; Oni Soesanto, Theresye, 100 pages)

The widespread dissemination of job vacancy advertisements through social media has increased the need for methods capable of automatically distinguishing legitimate job advertisements from Fake Job Advertisements. This study aims to represent Fake Job Advertisement texts into numerical representations through text pre-processing and Term Frequency–Inverse Document Frequency (TF-IDF) weighting as the basis for feature extraction, as well as to compare the effectiveness of the Bernoulli Naïve Bayes, Multinomial Naïve Bayes, and K-Nearest Neighbor (K-NN) algorithms in classifying Fake Job Advertisements on social media platforms. The dataset consisted of 246 job advertisement records collected from various social media platforms and secondary sources, which were manually labeled into two classes: fraudulent and legitimate. The text pre-processing stages included data cleaning, case folding, tokenization, normalization, stopword removal, and stemming. The data were then transformed using TF-IDF and divided into three training and testing scenarios: 90%:10%, 80%:20%, and 70%:30%. Model performance was evaluated using a confusion matrix based on accuracy, precision, recall, and F1-score. The results showed that the K-Nearest Neighbor (K-NN) algorithm outperformed the Bernoulli Naïve Bayes and Multinomial Naïve Bayes algorithms across all data-splitting scenarios. The best performance was achieved under the 90% training and 10% testing scenario, with an accuracy of 92.00% and precision, recall, and F1-score of 92.31%, indicating that K-NN is the most effective algorithm for classifying Fake Job Advertisements on social media platforms.

Keywords: *Fake Job Advertisement*, *Text Pre-processing*, TF-IDF, Naïve Bayes, K-Nearest Neighbor (K-NN).

PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah Subḥānahu Wa Ta'ālā atas limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **"Komparasi Algoritma *Naïve Bayes dan K-Nearest Neighbor (K-NN)* untuk Klasifikasi *Fake Job Advertisement* di Platform Media Sosial.**

Shalawat serta salam semoga senantiasa tercurah kepada Nabi Muhammad Ṣallallāhu 'Alaihi Wa Sallam, beserta keluarga, sahabat, dan para pengikut beliau hingga akhir zaman.

Skripsi ini disusun untuk memenuhi salah satu persyaratan dalam menyelesaikan Program Sarjana (S-1) pada Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Penulis menyadari bahwa penyusunan skripsi ini dapat terselesaikan berkat bantuan, arahan, serta dukungan dari berbagai pihak. Oleh karena itu, penulis menyampaikan rasa hormat dan terima kasih kepada:

1. Bapak Prof. Ir. Sunardi, S.Si., M.Sc., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat.
2. Ibu Dr. Na'imah Hijriati, S.Si., M.Si., selaku Koordinator Program Studi Matematika beserta seluruh dosen dan staf Program Studi Matematika.
3. Bapak Nurul Huda S.Si.,M.Si selaku Dosen Penasehat Akademik yang telah memberikan arahan dan bimbingan selama menjalani perkuliahan.
4. Bapak Oni Soesanto, S.Si., M.Si**.** selaku Dosen Pembimbing I dan Ibu Thresye ,S.Si.,M.S.i selaku Dosen Pembimbing II, yang telah memberikan bimbingan, nasihat, motivasi, dan petunjuk dalam pelaksanaan penelitian serta penyusunan skripsi ini.
5. Bapak Saman Abdurrahman, S.Si.,M.Sc selaku Dosen Penguji I dan Bapak Dr. Muhammad Ahsar Karim, S.Si, M.Sc Dosen Penguji II yang telah memberikan kritik dan saran demi kesempurnaan skripsi ini.

6. Kedua orang tua serta keluarga besar yang selalu memberikan doa, dukungan, dan motivasi selama proses penyusunan skripsi.
7. Teman-teman dan seluruh pihak yang tidak dapat penulis sebutkan satu per satu, yang telah memberikan bantuan dan dukungan dalam proses penelitian dan penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih memiliki kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang . Akhir kata, penulis berharap skripsi ini dapat memberikan manfaat bagi para pembaca dan semua pihak yang berkepentingan.

Banjarbaru, 23 Juli 2026

Siti Hardiyanti

2211011120009

DAFTAR ISI

Halaman

DAFTAR TABEL

Halaman

DAFTAR GAMBAR

Halaman

DAFTAR LAMPIRAN

ARTI LAMBANG DAN SINGKATAN

c	Variabel target atau kelas
d	Data atau dokumen dengan kelas yang belum diketahui
$P(c\|d)$	Peluang kelas c dengan d diketahui (*posterior Probability)*
$P(d\|c)$	Peluang data d dengan kelas c diketahui, (*likelihood)*
$P(c)$	Peluang awal kelas c (*prior probability)*
$P(d)$	Peluang data d (*evidence)*
t	kata *(Term)*
t_k	*Term* ke-k
V	Himpunan seluruh kata *(vocabulary)* dalam data latih
n_d	Jumlah total kata dalam data d
e_i	Fitur ke-i
M	Jumlah total fitur
$P(t_k\|c)$	Peluang kemunculan (*term*) t_k pada data dengan kelas c
N_c	Jumlah data latih dalam kelas c
N	Jumlah total data latih
T_{ct}	Jumlah kemunculan kata (*term*) t pada data latih dengan kelas c
B	Jumlah kata (*term*) unik pada seluruh kelas
N_{ct}	Jumlah data latih dalam kelas $c$yang memuat *term* t
$W_{t,d}$	Bobot *term* (t) terhadap data (d)
$tf_{t,d}$	Frekuensi kemunculan *term* $t$dalam data d
D	Jumlah total dokumen
df_t	Jumlah dokumen yang mengandung *term* t
$\propto$	Proporsional
KNN	*K-Nearest Neighbor*
TP	*True Positive,* Data positif yang diprediksi benar
TN	*True Negative,* Data negatif yang diprediksi benar
FP	*False Positive,* Data negatif yang diprediksi sebagai positif
FN	*False Negative,* Data positif yang diprediksi sebagai negati