

KLASIFIKASI SMS SPAM MENGGUNAKAN NAIVE BAYES DAN SVM

Jeremy Reinhart

Program Studi Teknik Informatika, Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
e-mail: jeremy.535200083@stu.untar.ac.id

ABSTRAK

Klasifikasi ini berpusat pada pengembangan kerangka kerja untuk membedakan pesan spam dalam SMS menggunakan dua strategi karakterisasi utama, khususnya Naive Bayes dan *Support Vector Machine (SVM)*. Kebutuhan akan metode efektif untuk memfilter pesan yang tidak diinginkan, khususnya pesan *spam* yang dapat mengganggu pengalaman pengguna, semakin meningkat karena pesatnya pertumbuhan komunikasi SMS. Disini saya mengumpulkan dan memproses beragam kumpulan data SMS untuk penelitian ini, termasuk pesan spam dan non-spam. Fitur-fitur yang diperoleh dari analisis teks pesan kemudian digunakan untuk melatih model dengan bantuan teknik klasifikasi Naive Bayes dan SVM. Setelah itu, kami menggunakan metrik standar seperti akurasi, presisi, perolehan, dan skor F1 untuk menilai seberapa baik kinerja kedua model.

Kata kunci: Naive Bayes, Spam, Klasifikasi, *Support Vector Machine*

ABSTRACT

This research focuses on developing a framework for distinguishing spam messages in SMS using two main characterization strategies, namely Credulous Bayes and Backing Vector Machine (SVM). The need for effective methods to filter unwanted messages, especially spam messages that can disrupt the user experience, is increasing with the rapid growth of SMS communications. We collect and process various sets of SMS data for this research, including spam and non-spam messages. The features obtained from message text analysis are then used to train the model with the help of Naive Bayes and SVM classification techniques. After that, we use standard metrics such as accuracy, precision, recall, and F1 score to assess how well the two models perform.

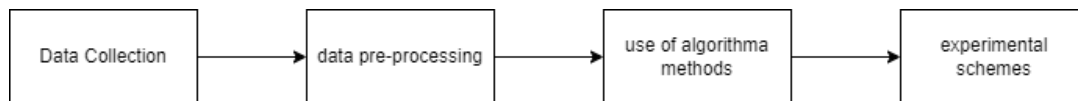
Keywords: Naive Bayes, Spam, Klasifikasi, *Support Vector Machine*

1 PENDAHULUAN

Pesatnya perkembangan inovasi persuratan, khususnya melalui *Short Message Administration (SMS)*, memberikan dampak positif bagi masyarakat dalam hal persuratan. Namun, seiring dengan kejadian tersebut [1], masalah spam SMS juga semakin meningkat, mengganggu kenyamanan pengguna dan memakan banyak waktu. Oleh karena itu, deteksi dan klasifikasi *spam* SMS merupakan komponen penting dalam mengurangi dampak negatif pesan spam [2]. Tujuan utama dari penelitian ini adalah untuk mengembangkan sistem klasifikasi pesan SMS menggunakan dua metode *Support Vector Machine* dan Naive Bayes [3]. Naive Bayes dikenal karena kecepatan dan kecukupannya dalam mengelompokkan informasi yang menerima otonomi komponen, sedangkan SVM menawarkan manfaat dalam mengelola informasi yang tinggi [4]. Dataset yang telah dikumpulkan didapatkan dari *kaggle* dengan jumlah data 5573. Selain menawarkan solusi yang dapat diandalkan dan mudah beradaptasi terhadap variasi pola spam yang terus berubah, strategi ini diharapkan dapat meningkatkan akurasi dalam membedakan pesan SMS yang sah dan tidak sah. Penerapan metode Naive Bayes dan SVM secara bersamaan diharapkan dapat mengatasi kekurangan masing-masing metode secara terpisah, sehingga menghasilkan sistem klasifikasi yang lebih kuat dan dapat dipercaya [5].

2 METODE PENELITIAN

Metode Penelitian ini dapat diilustrasikan pada gambar di bawah ini. Alur kerjanya terdiri dari *Data Collection, Data Pre-Processing, Use of Algorithm Methods, Experimental schmes*



2.1 Data Collection

Dataset yang saya gunakan pada tugas ini adalah dataset “sms” dimana data tersebut didapatkan dari <https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset/data>. Koleksi SMS Spam adalah sekumpulan pesan yang diberi tag SMS yang telah dikumpulkan untuk penelitian SMS Spam. Ini berisi satu set pesan SMS dalam bahasa Inggris dari 5.574 pesan, diberi tag menurut kategori ham (sah) atau spam.

2.2 Data Pre-Processing

Pre-Processing data mencakup dua pekerjaan utama: pemilihan fitur dan pemilihan label data untuk di olah. Setelah mengumpulkan data, dilanjutkan memilih data yang tepat untuk menentukan label dan juga fitur [6]. Pada tahap ini juga dilakukan sebuah kamus pemetaan data untuk merubah data yang tadi nya numerik menjadi data teks sesuai kamus pemetaan. Kumpulan data ini akan dilakukan klasifikasi dengan algoritma Naive Bayes, dan SVM. Tujuannya untuk mengklasifikasi email dari data yang sudah didapatkan[7].

2.3 Metode/Algoritma

- Naive Bayes

Naive Bayes adalah algoritma klasifikasi sederhana dan efektif berdasarkan teorema Bayes [8]. Disebut "naif" karena membuat asumsi bahwa fitur yang digunakan untuk mendeskripsikan observasi tidak bergantung satu sama lain, yang mungkin tidak selalu benar dalam skenario dunia nyata. Meskipun asumsinya disederhanakan, Naive Bayes sering kali berkinerja baik dalam praktiknya, terutama dalam klasifikasi teks dan pemfilteran spam. Karena kesederhanaan dan efektivitasnya, Naive Bayes sangat populer dalam tugas pemrosesan bahasa alami seperti klasifikasi teks (deteksi spam, analisis sentimen) [9]. Bernoulli Naive Bayes, Gaussian Naive Bayes, dan Multinomial Naive Bayes hanyalah beberapa dari sekian banyak variasi Naive Bayes yang ada. Penting untuk dicatat bahwa anggapan "tanpa rasa bersalah" tidak serta merta berlaku dalam kerangka berpikir tersebut, terkadang model yang lebih modern mungkin sangat penting. Meskipun demikian, Gullible Bayes tetap menjadi titik awal yang baik untuk sebagian besar tugas karakterisasi [10].

- SVM

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin terawasi yang kuat yang digunakan untuk tugas klasifikasi dan regresi [11]. Ini sangat efektif dalam ruang berdimensi tinggi dan sangat cocok untuk tugas-tugas di mana ada batas pemisah yang jelas antar kelas. Berikut gambaran umum klasifikasi SVM [12]. SVM bertujuan untuk menemukan *hyperplane* yang paling baik memisahkan data ke dalam kelas-kelas yang berbeda. *Hyperplane* yang "terbaik" adalah *hyperplane* yang memaksimalkan margin, yaitu jarak antara hyperplane dan titik data terdekat dari setiap kelas [13]. Titik-titik terdekat ini disebut vektor pendukung.

2.4 Metode Evaluasi

1. *Confusion Matrix*
2. *Accuracy*
3. *Precision*
4. *Recall*
5. *F1-Score*

2.5 Skema Eksperimen

1. Pengumpulan Data:
Mencari data sms di internet yang memiliki 2 kelas untuk di klasifikasi seperti *spam* dan ham
2. Pemrosesan Teks:
Melakukan pengolahan terhadap data yang akan diuji nanti nya menggunakan algoritma

3. Implementasi Algoritma Naive Bayes:
Latih data yang sudah di olah menggunakan algoritma naive bayes Evaluasi Model metrik akurasi *confusion matrix*, *pressi*, *recall*, dan *F-1 Score* pada data yang di uji
4. Implementasi Algoritma SVM:
Latih data yang sudah di olah menggunakan algoritma SVM
Evaluasi model menggunakan metrik akurasi, *confusion matrix*, *presisi*, *recall*, dan *F-1 Score* pada data yang di uji
5. Perbandingan Hasil
Bandingkan hasil dan kinerja ketiga model tersebut untuk menentukan algoritma mana yang paling efektif dalam mengklasifikasikan *spam* email dalam konteks tertentu

3 HASIL DAN PEMBAHASAN

3.1 Import Libraries yang diperlukan

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from collections import Counter
from sklearn.model_selection import train_test_split
import re
import nltk
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.svm import SVC
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
from sklearn.metrics import accuracy_score, confusion_matrix
```

3.2 drop unavailable attributes

```
data = pd.read_csv("spam.csv", encoding = 'latin1')
data.head()
```

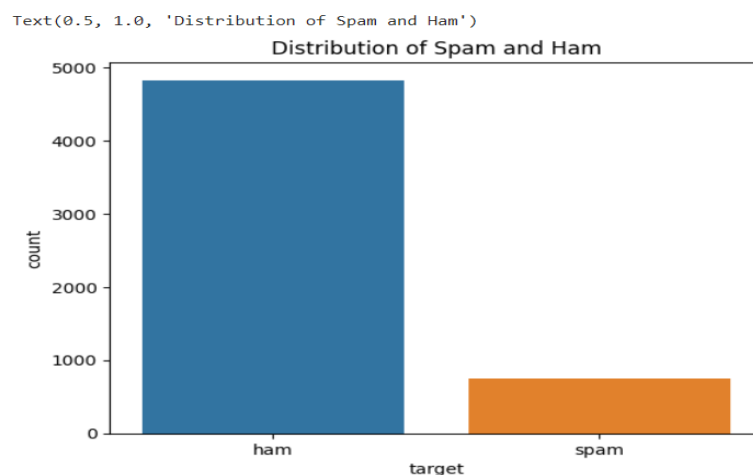
	v1	v2	Unnamed: 2	Unnamed: 3	Unnamed: 4
0	ham	Go until jurong point, crazy.. Available only ...	NaN	NaN	NaN
1	ham	Ok lar... Joking wif u oni...	NaN	NaN	NaN
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...	NaN	NaN	NaN
3	ham	U dun say so early hor... U c already then say...	NaN	NaN	NaN
4	ham	Nah I don't think he goes to usf, he lives aro...	NaN	NaN	NaN

```
# drop unavailable attributes
data = data.drop(['Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], axis = 1)

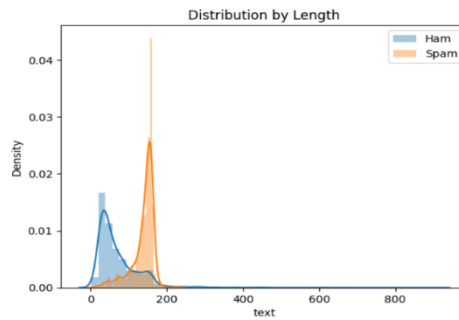
data = data.rename(columns = ({'v1': 'target', 'v2': 'text'}))
data.head()
```

	target	text
0	ham	Go until jurong point, crazy.. Available only ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...

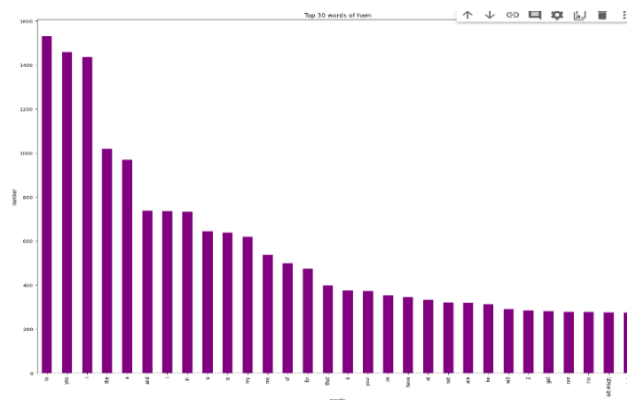
3.3 Merencanakan Grafik Untuk Distribusi



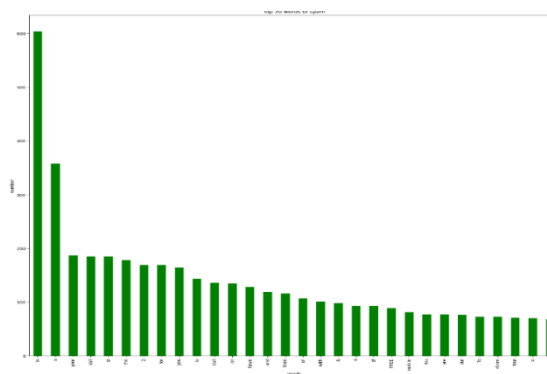
3.4 Memplot grafik berdasarkan panjang nya



3.5 Menampilkan kata 30 teratas tentang ham



3.6 Menampilkan 30 kata teratas tentang spam



3.7 Melakukan Klasifikasi dengan Algoritma Naive Bayes

```
#Naive Bayes Classification

mnb = MultinomialNB(alpha = 0.5)
mnb.fit(X_train_cv,y_train)

y_mnb = mnb.predict(X_test_cv)

print('Naive Bayes Accuracy: ', accuracy_score( y_mnb , y_test))
print('Naive Bayes confusion_matrix: ', confusion_matrix(y_mnb, y_test))

Naive Bayes Accuracy: 0.9826555023923444
Naive Bayes confusion_matrix: [[1425 15]
 [ 14 218]]

from sklearn.metrics import classification_report

classification_rep = classification_report(y_mnb , y_test)
print("Classification Report:")
print(classification_rep)
```

	precision	recall	f1-score	support
ham	0.99	0.99	0.99	1440
spam	0.94	0.94	0.94	232
accuracy			0.98	1672
macro avg	0.96	0.96	0.96	1672
weighted avg	0.98	0.98	0.98	1672

3.8 Melakukan Klasifikasi dengan Algoritma SVM

```
#SVM Classification

svc = SVC(kernel='sigmoid', gamma=1.0)
svc.fit(X_train_cv,y_train)
y_svc = svc.predict(X_test_cv)

print('SVM Accuracy: ', accuracy_score( y_svc , y_test))
print('SVM confusion_matrix: ', confusion_matrix(y_svc, y_test))

SVM Accuracy:  0.8343301435406698
SVM confusion_matrix:  [[1303  141]
 [ 136   92]]

from sklearn.metrics import classification_report

classification_rep = classification_report(y_svc , y_test)
print("Classification Report:")
print(classification_rep)

Classification Report:
              precision    recall  f1-score   support

     ham           0.91         0.90         0.90         1444
    spam           0.39         0.40         0.40          228

   accuracy                   0.83         1672
  macro avg                   0.65         1672
 weighted avg                   0.84         1672
```

4 KESIMPULAN

Berdasarkan hasil klasifikasi sms spam menggunakan metode Naive Bayes dan SVM didapatkan hasil bahwa Naive Bayes memiliki jumlah *accuracy* yang lebih baik dibandingkan dengan SVM dengan perbedaan 0.15, dan juga Naive Bayes juga memiliki jumlah pengujian yang lebih cepat dibandingkan dengan SVM. Dengan ini Naive Bayes memiliki performa yang baik dalam melakukan klasifikasi pada sms *spam* dibandingkan dengan SVM.

UCAPAN TERIMA KASIH

Terima Kasih kepada ibu Teny Handhayani yang telah membantu dan memberi arahan dalam pembuatan makalah paper ini. Dan juga berterima kasih kepada SISTEMASI yang telah menyediakan landasan bentuk format makalah ini.

DAFTAR PUSTAKA

- [1] Rahmani, Anastasya, et al. *Bunga Rampai Isu-isu Komunikasi Kontemporer 2023*. PT Rekacipta Proxy Media, 2023.
- [2] Dwiyanaputra, Ramaditia, et al. "Deteksi SMS Spam Berbahasa Indonesia menggunakan TF-IDF dan Stochastic Gradient Descent Classifier." *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTika)* 3.2 (2021): 200-207.
- [3] Irawan, Devi, et al. "Perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest dan Bagging Classifier." *Jurnal Sisfokom (Sistem Informasi Dan Komputer)* 10.3 (2021): 432-437.
- [4] Kurniawan, Candra. "A Survey on Big Data Analytics Model." *ITEJ (Information Technology Engineering Journals)* 4.1 (2019): 1-13.
- [5] Apriyani, Hilda, and Kurniati Kurniati. "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus." *Journal of Information Technology Ampera* 1.3 (2020): 133-143.
- [6] B. Gunawan, H. Sastypratiwi, dan E. E. Pratama, "Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 4, no. 2, hal. 113-118, 2018.
- [7] M. Azhari, M. Mulkan, Z. Situmorang, dan R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, hal. 640-651, 2021.

- [8] Adi, Sumarni. "Implementasi Algoritma Naive Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA Di Universitas Amikom Yogyakarta." *Jurnal Mantik Penusa* 22.1 (2018).
- [9] Anggraeni, Merry. "INTERACTIVE QUESTION-ANSWERING MENGGUNAKAN NLP DAN ALGORITMA NAÏVE BAYES UNTUK LITERASI TENTANG GANGGUAN DENGAR (I-CHAT BOT)."
- [10] Karnovi, Riki, Roni Habibi, and Mohammad Nurkamal Fauzan. *Tutorial membuat aplikasi sistem monitoring progres pekerjaan dan evaluasi pekerjaan pada job desk operational human capital menggunakan metode naive bayes*. Kreatif, 2020.
- [11] Pudjiarti, Eni. "Prediksi Spam Email Menggunakan Metode Support Vector Machine dan Particle Swarm Optimization." *Jurnal Pilar Nusa Mandiri* 12.2 (2016): 171-181.
- [12] Hadiyanoor, Amrul. "KLASIFIKASI PROPOSAL TUGAS AKHIR BERDASARKAN BIDANG KEAHLIAN DAN PEMINATAN DOSEN PEMBIMBING MENGGUNAKAN SUPPORT VECTOR MACHINE PADA JURUSAN TEKNIK INFORMATIKA DI STMIK INDONESIA BANJARMASIN." *Jurnal Ilmiah Teknologi Informasi dan Komputer Pranala* 13.2 (2018): 67-76.
- [13] fauzi Abdulloh, Ferian, and Iqbal Rilo Pambudi. "Analisis sentimen pengguna youtube terhadap program vaksin covid-19." *Csrid (computer science research and its development journal)* 13.3 (2021): 141-148.
- [14] Setiyono, Agus, and Hilman F. Pardede. "Klasifikasi Sms Spam Menggunakan Support Vector Machine." *Jurnal Pilar Nusa Mandiri* 15.2 (2019): 275-280.
- [15] Setifani, Niken Ayu, Devinta Nurul Fitriana, and Ahmad Yusuf. "Perbandingan algoritma naïve bayes, svm, dan decision tree untuk klasifikasi sms spam." *JUSIM (Jurnal Sistem Informasi Musirawas)* 5.2 (2020): 167-174.
- [16] Indiarito, Budi. "Klasifikasi Sms Spam Dengan Metode Naive Bayes Classifier Untuk Menyaring Pesan Melalui Selular." *Telematika MKOM* 8.2 (2017): 167-172.
- [17] Herwanto, Herwanto, Nuke L. Chusna, and Muhammad Syamsul Arif. "Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes." *JURNAL MEDIA INFORMATIKA BUDIDARMA* 5.4 (2021): 1316-1325.