

# Implementasi Ensemble Learning Adaboost Pada Algoritma Klasifikasi Decision Tree dan SVM Untuk Klasifikasi SMS Berbahasa Indonesia

M. Ibnu Umar Rosyidi<sup>1</sup>, Naim Rochmawati<sup>2</sup>

<sup>1,2</sup>Jurusan Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

<sup>1</sup>[m.ibnu.19048@mhs.unesa.ac.id](mailto:m.ibnu.19048@mhs.unesa.ac.id)

<sup>2</sup>[naimrochmawati@unesa.ac.id](mailto:naimrochmawati@unesa.ac.id)

**Abstrak**— Perkembangan teknologi di berbagai sektor yang sangat cepat selama satu dekade ini, membuat kita semakin dimudahkan dalam melakukan aktivitas sehari-hari. Kebutuhan dalam berkomunikasi terus mengalami kemajuan dari yang awalnya menggunakan surat dengan jangkauan terbatas dan lama, hingga sekarang menggunakan layanan provider SMS (*Short Message Service*). Layanan SMS sangat digemari karena harganya yang murah dan dapat memilih untuk menjawab nanti jika tidak ada waktu pada saat tersebut. Bertumbuhnya pengguna SMS dibersamai dengan orang-orang yang tidak bertanggung jawab memanfaatkan situasi untuk keuntungan sendiri dengan melakukan tindak penipuan melalui SMS. Untuk mencegah tersebut, diperlukan sistem yang dapat memilah SMS agar tidak semua masuk ke pengguna. Salah satu cara yang dapat digunakan adalah membuat model *machine learning* yang dapat mengenali dan mengelompokkan SMS apakah SMS yang masuk tersebut adalah SPAM atau tidak. Algoritma SVM dan *decision tree* adalah dua algoritma klasifikasi yang mudah digunakan untuk memecahkan masalah klasifikasi, pada penelitian ini juga digunakan teknik ensemble *adaboost* yang dapat meningkatkan nilai akurasi dari *base* algoritma yang digunakan. Data yang digunakan adalah dataset SMS dari penelitian Rahmi dan Wibisono[1]. Algoritma SVM memiliki nilai akurasi tertinggi dengan nilai 0.96 atau 96%, ditemukan juga bahwa algoritma SVM + *Adaboost* hanya membuahkan hasil yang baik saat diterapkan data jenis unigram dan bigram tanpa *tf-idf* dan unigram tanpa *tf-idf*. Algoritma SVM + *Adaboost* tidak cocok menggunakan *tf-idf* karena dapat menyebabkan penurunan nilai akurasi.

**Kata Kunci**— *Adaboost*, Klasifikasi, SMS, *Decision Tree*, SVM, TF-IDF, N-Gram

## I. PENDAHULUAN

Dalam beberapa dekade terakhir, kemajuan dalam teknologi komunikasi telah mengubah cara orang berinteraksi dan berkomunikasi satu sama lain. Pesan singkat, atau SMS, menjadi salah satu bentuk komunikasi yang paling populer di seluruh dunia, terutama sejak awal tahun 2000-an. Meskipun SMS dirancang sebagai alat komunikasi yang efisien dan efektif, sayangnya, perkembangan teknologi ini juga membawa dampak negatif, salah satunya adalah maraknya penipuan melalui SMS.

Pada era modern, penipuan melalui SMS telah menjadi masalah yang meresahkan bagi masyarakat. Penipu menggunakan SMS sebagai alat untuk menerapkan berbagai strategi penipuan, termasuk penawaran palsu, hadiah palsu, dan peringatan palsu yang mengancam orang. Dalam beberapa

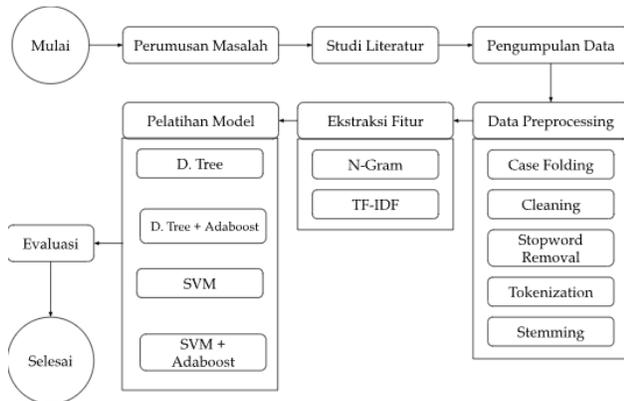
kasus, penipuan melalui SMS bahkan dapat menyebabkan kerugian finansial yang signifikan.

Penelitian ini akan mencoba mengatasi permasalahan tersebut dengan menggunakan teknik klasifikasi untuk membantu mengenali apakah sms yang diterima adalah sms *fraud* atau penipuan, algoritma klasifikasi yang akan digunakan adalah *decision tree* dan SVM, untuk melihat bagaimana performa kedua algoritma tersebut dalam mengelompokkan SMS berdasarkan kelasnya. Algoritma juga akan diterapkan teknik *adaboost* untuk melihat bagaimana pengaruh teknik *ensemble* ini pada algoritma yang akan dijadikan sebagai *base model* yaitu SVM dan *decision tree*.

Beberapa penelitian sudah dilakukan untuk mengatasi masalah klasifikasi SMS yaitu penelitian oleh Devi[5], menggunakan 4 algoritma yang akan diuji mengklasifikasikan SMS, algoritma yang digunakan diantaranya Support Vector Machine (SVM), Naive Bayes, Random Forest dan Bagging Classifier. Dari penelitian yang dilakukan oleh Edi[6], performa dari algoritma tersebut masih dapat ditingkatkan menggunakan ensemble learning. Ensemble learning adalah gabungan dari beberapa pembelajaran mesin untuk meningkatkan performa sistem secara keseluruhan[7]. Studi empiris telah menunjukkan bahwa masalah klasifikasi dan regresi menggunakan ensemble learning seringkali menghasilkan performa model yang akurat daripada model *base individual*[8, 9, 10]. Mengapa penelitian ini menggunakan data SMS karena *dataset* mudah didapat dan tersedia. Perbedaan penelitian ini dengan yang sebelumnya adalah digunakannya 2 parameter *base model* yang berbeda yaitu *decision tree* dan SVM untuk diterapkan teknik *adaboost* sehingga 4 model akan dihasilkan dari penelitian ini.

## II. METODE PENELITIAN

Dalam melakukan penelitian, diperlukan pedoman sebagai acuan agar penelitian sesuai dengan tujuan yang diharapkan sehingga dapat menjawab rumusan masalah yang sudah ditentukan.



Gbr. 1 Gambaran Umum Penelitian

Berdasarkan Gambar 1. Gambaran umum penelitian, dapat diketahui langkah langkah yang akan diterapkan dalam penelitian ini, yaitu:

### A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini adalah dataset sekunder yang didapat dari penelitian Rahmi dan Wibisono[1]. Dataset berjumlah 1143 pesan SMS dengan 3 label kelas dengan rincian label terdapat pada table berikut.

TABEL I  
 KETERANGAN LABEL DATASET

Label	Keterangan
0	SMS Normal
1	SMS <i>Fraud</i> atau Penipuan
2	SMS Promo

### B. Pra Pemrosesan Data

Sebelum data dibagi menjadi data latih dan uji, data akan melalui tahap preprocessing data yang bertujuan untuk memudahkan dalam pelatihan model. Teknik yang digunakan dalam penelitian ini, yaitu:

#### 1) Case Folding

Tahap merubah semua huruf pada kalimat menjadi huruf kecil atau *lowercase*.

#### 2) Cleaning

Tahap membersihkan data dari simbol, tanda baca dan angka yang tidak diperlukan dalam klasifikasi.

#### 3) Remove Stopwords

Tahap menghapus stopwords atau kata yang kurang memiliki makna saat data digunakan sebagai data latih pada machine learning.

#### 4) Tokenization

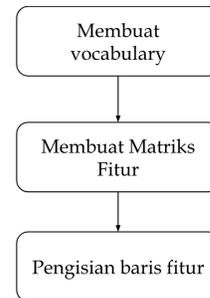
Tahap memisahkan teks kalimat menjadi token perkata sebagai langkah mempermudah saat menggunakan teknik stemming.

### 5) Stemming

Tahap mengembalikan kata menjadi kata dasar, manfaat lain dari teknik stemming adalah mengurangi dimensi cospus atau fitur data.

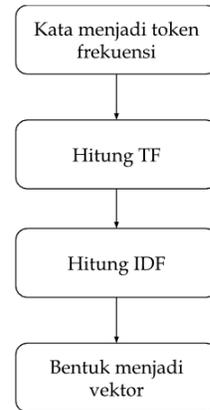
### C. Ekstraksi Fitur

Penelitian ini akan menggunakan dua teknik n-gram untuk ekstraksi fitur: unigram dan bigram. Teknik n-gram menggunakan kata sejumlah n. Berikut adalah bagaimana n-gram direpresentasikan pada gambar dibawah.



Gbr. 2 Proses n-gram

Langkah pertama adalah mengumpulkan kosa kata unik dari data yang digunakan; langkah kedua adalah memasukkan setiap kata ke dalam kolom yang berbeda untuk membuat matriks fitur; dan langkah terakhir adalah mengisi nilai 1 untuk kata yang muncul dan 0 untuk kata yang tidak muncul. Teknik ekstraksi fitur kedua adalah tf-idf, berikut adalah bagaimana tf-idf direpresentasikan pada gambar dibawah.

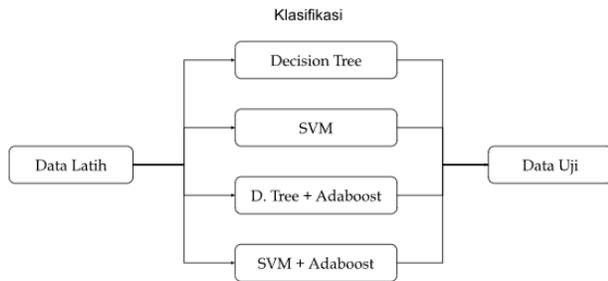


Gbr. 3 Proses tf-idf

Langkah pertama dalam implementasi tf-idf adalah membuat kata menjadi token frekuensi, artinya tiap kata unik akan dihitung berapa kali kata tersebut muncul pada dokumen, langkah kedua menghitung nilai tf dengan membag jumlah kata pada dokumen, langkah ketiga adalah menghitung idf. Terakhir data akan direpresentasikan pada bentuk table dan nilai tf-idf didapat dari perkalian tf dan idf.

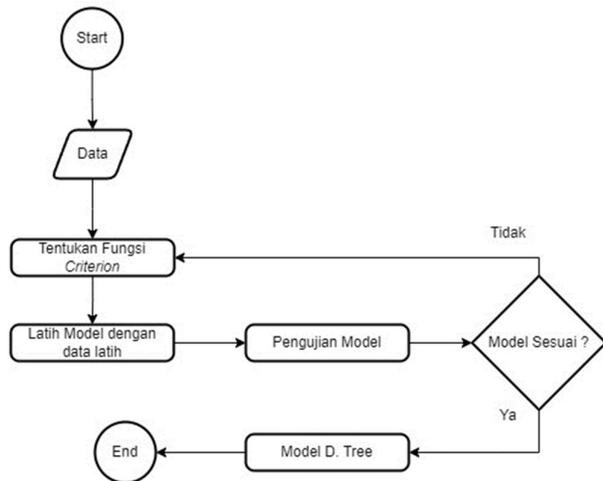
D. Pelatihan Model

Setelah melalui tahap pra pemrosesan data dan ekstraksi fitur, data siap digunakan sebagai data latih dan uji dari model machine learning. Penelitian ini akan menggunakan teknik adaboost dengan dua jenis *base model* algoritma decision tree dan SVM.



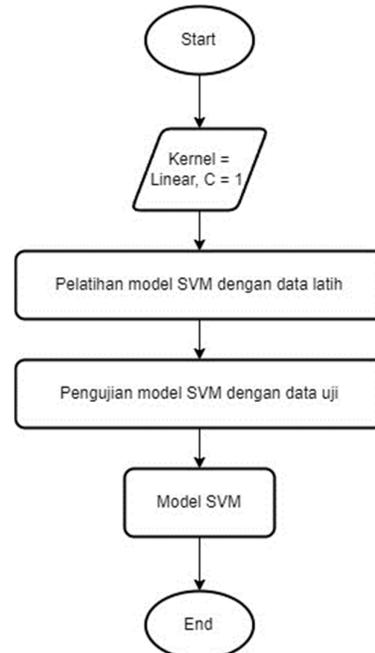
Gbr. 4 Proses Pelatihan Model

Metode pertama yang akan diuji adalah algoritma decision tree, gambar dibawah adalah diagram bagaimana melakukan metode decision tree pada penelitian ini.



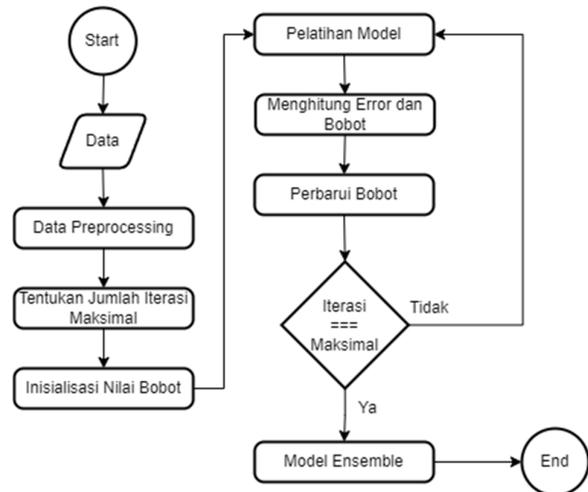
Gbr. 5 Decision tree

Decision tree perlu untuk membagi data menjadi data latih dan uji. Setelah dibagi, tentukan fungsi *criterion* dan jika model sudah sesuai maka proses selesai. Metode kedua adalah SVM, bagaimana implementasi dari algoritma SVM pada penelitian ini akan direpresentasikan pada diagram dibawah.



Gbr. 6 SVM

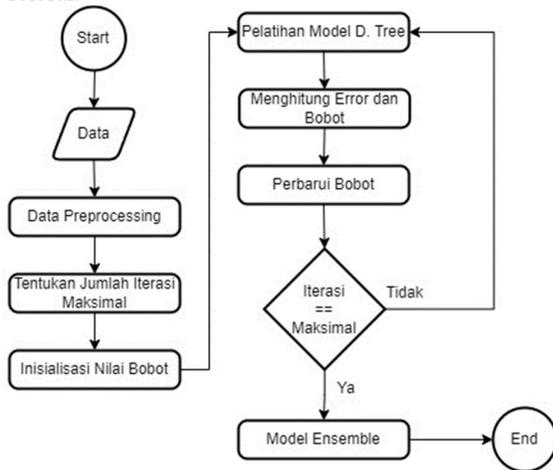
Parameter algoritma SVM yang digunakan dalam penelitian ini adalah kernel linear dan nilai  $C = 1$ , setelah model melalui proses pelatihan dengan data latih, model akan digunakan untuk pengujian. Metode ketiga atau terakhir yang digunakan pada penelitian ini adalah teknik adaboost, teknik ensemble yang sering digunakan untuk meningkatkan kinerja dari algoritma *machine learning*. Bagaimana adaboost diimplementasikan pada penelitian ini akan direpresentasikan pada diagram dibawah.



Gbr. 7 Adaboost

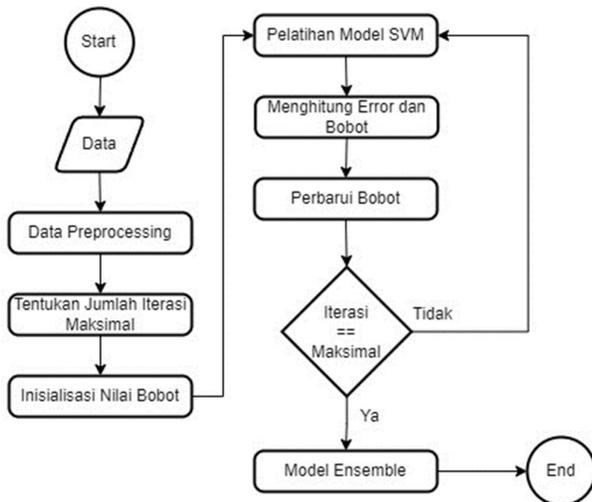
Gambar 7 diatas adalah bagaimana metode adaboost dilakukan, langkah pertama adalah membuat data menjadi

siap digunakan sebagai data latih, inialisasi nilai bobot, menuju ke pelatihan model, pada tahap ini model akan melakukan update bobot untuk digunakan pada iterasi berikutnya hingga iterasi maksimal maka iterasi telah selesai.



Gbr. 8 Adaboost + D. Tree

Pada gambar 8, adalah posisi dari algoritma decision tree saat diterapkan teknik *adaboost*. Tidak jauh berbeda dengan SVM pada gambar 9, posisi algoritma SVM saat diterapkan teknik *adaboost* juga sama, hanya saja base model atau bergantung pada basis dari teknik *adaboost* yang digunakan.



Gbr. 9 Adaboost + SVM

E. Evaluasi

Proses evaluasi dilakukan penghitungan nilai akurasi dari skenario pembagian jumlah dataset latih dan uji 70:30, 80:20 dan 90:10. Dalam penelitian ini juga akan

menggunakan k-fold cross validation dalam evaluasi dan validasi model. Nilai K yang digunakan adalah 5 dan 10.

III. HASIL DAN PEMBAHASAN

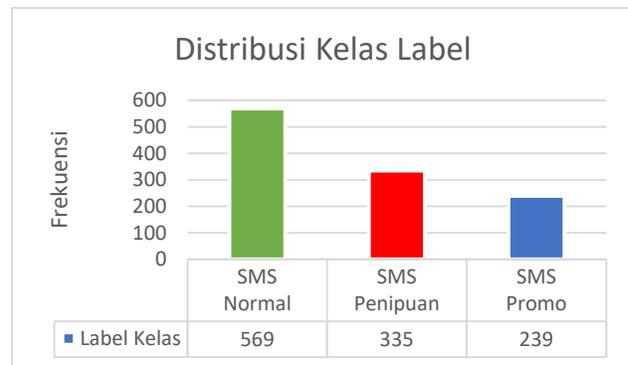
A. Analisis Eksploratif Data

Dataset yang telah didapat dari penelitian Rahmi dan Wibisono[1] ditampilkan pada panel data pandas untuk mengetahui bagaimana karakteristik dari data.

TABEL II  
 PANEL DATA DATASET

Label	Teks
0	Ada semua di panduan pla. dishare di grup ini dan fb
1	De inih bapa pake no temen, tolong kirimin pulsa ke no inih soalnya penring banget, nanti di rumah digantiin
2	4.5GB/30 hari hanya Rp 55 Ribu Spesial buat anda yang terpilih. Aktifkan sekarang juga di *550*907# Buruan.! SKB

Untuk lebih mengenal tentang karakteristik dataset, juga dilakukan analisa distribusi kelas label pada dataset. Berikut adalah diagram batang distribusi kelas label pada dataset.



Gbr. 10 Distribusi Kelas Label

Pada gambar 8 adalah bagaimana distribusi kelas label dataset, jumlah kelas label normal sebanyak 569, kelas penipuan 335 dan kelas promo 239.

B. Pra Pemrosesan Data

Pada analisis eksploratif data, terlihat data masih belum dapat digunakan sebagai data latih, jika dipaksakan menjadi data latih maka akan mempengaruhi hasil dari model pembelajaran mesin. Dalam dunia *machine learning* ada istilah "garbage in garbage out" yaitu jika sesuatu yang dimasukkan kurang baik maka hasil keluaran juga akan mengikuti. Teknik prapemrosesan data pertama adalah *case folding*, pada tabel III dibawah adalah hasil data SMS setelah melalui pra pemrosesan data.

TABEL III  
 HASIL PRA PEMROSESAN DATA

Sebelum	Sesudah
<b>Case Folding</b>	
4.5GB/30 hari hanya Rp 55 Ribu Spesial buat anda yang terpilih. Aktifkan sekarang juga di *550*907# Buruan..! SKB	4.5gb/30 hari hanya rp 55 ribu spesial buat anda yang terpilih. aktifkan sekarang juga di *550*907# buruan..! skb
<b>Cleaning</b>	
4.5GB/30 hari hanya Rp 55 Ribu Spesial buat anda yang terpilih. Aktifkan sekarang juga di *550*907# Buruan..! SKB	hari hanya rp ribu spesial buat anda yang terpilih aktifkan sekarang juga di buruan skb
<b>Remove Stopwords</b>	
hari hanya rp ribu spesial buat anda yang terpilih aktifkan sekarang juga di buruan skb	hari hanya rp ribu special buat terpilih aktifkan sekarang juga buruan skb
<b>Tokenization</b>	
hari hanya rp ribu special buat terpilih aktifkan sekarang juga buruan skb	{“hari”, “hanya”, “rp”, “ribu”, “special”, “buat”, “terpilih”, “aktifkan”, “sekarang”, “juga”, “buruan”, “skb”}
<b>Stemming</b>	
{“hari”, “hanya”, “rp”, “ribu”, “special”, “buat”, “terpilih”, “aktifkan”, “sekarang”, “juga”, “buruan”, “skb”}	hari hanya rp ribu spesial buat pilih aktif sekarang juga buru skb

C. Ekstraksi Fitur

Setelah melakukan pelatihan model data yang sudah melalui pra pemrosesan, perlu dilakukan ekstraksi fitur untuk mengenali pola dari tiap kelas SMS. Ekstraksi fitur pertama yang digunakan adalah N-Gram.

TABEL IV  
 TABEL FITUR UNIGRAM DAN BIGRAM

Fitur Unigram						
advice	aja	ayo	...	ball	trust	yth
<b>Fitur Bigram</b>						
advice text	aja data	aktif laku	...	trust in	wow aktif	yth rekening

Pada tabel IV diatas adalah fitur unigram dan bigram yang didapat dari dataset dengan menyeleksi kata unik menjadi pasangan kata jika ingin menggunakan lebih dari unigram atau 2 pasangan kata bahkan lebih.

TABEL V  
 PENGISIAN TABEL FITUR UNIGRAM DAN BIGRAM

Fitur Unigram						
advice	aja	ayo	...	ball	trust	yth
0	0	0	...	1	0	0
0	1	0	...	0	0	0
1	0	0	...	0	1	0
...	...	...	...	...	...	...
0	0	1	...	0	0	1
0	0	0	...	0	0	0
0	0	0	...	0	0	0

advice text	aja data	aktif laku	...	trust in	wow aktif	yth rekening
0	0	0	...	0	0	0
0	1	0	...	0	0	0
1	0	0	...	1	0	0
...	...	...	...	...	...	...
0	0	1	...	0	0	1
0	0	0	...	0	1	0
0	0	0	...	0	0	0

Pada table V diatas adalah hasil dari pengisian baris fitur dari unigram dan bigram, jika kalimat mengandung kata dari salah satu fitur maka akan diberi nilai 1 dan jika tidak maka diberi nilai 0. Proses ekstraksi fitur kedua yang akan digunakan pada penelitian ini adalah tf-idf, setelah diterapkan teknik n-gram data melalui pembobotan tf-idf.

	diskon	kacamata	maaf	promo	bayar	gratis
0.217670	0.556839	0.000000	0.157912	0.000000	0.000000	0.000000
0.000000	0.000000	0.271116	0.000000	0.31051	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.295538	0.000000	0.000000	0.000000	0.000000
0.228875	0.000000	0.000000	0.166041	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.252293	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.41398	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.188515	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

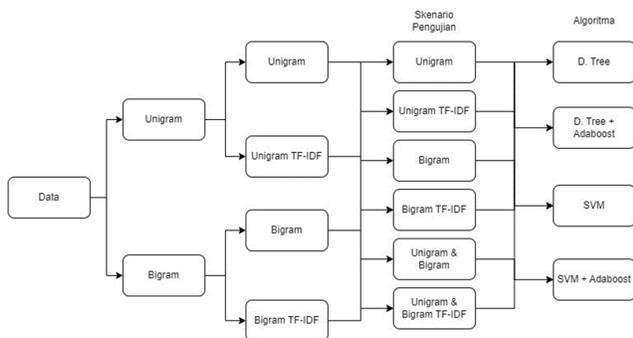
Gbr. 11 Hasil Implementasi Unigram tf-idf

abadi big	advice text	aja data	aktif laku	aman klik	appreciation trust	ayo coba	ball pool	banget loh	banget main
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.27735	0.27735	0.27735
0.000000	0.000000	0.57735	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.218218	0.000000	0.000000	0.000000	0.218218	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.305386	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.000000	0.000000	0.000000	0.000000	0.353553	0.000000	0.353553	0.000000	0.000000	0.000000
0.26062	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Gbr. 12 Hasil Implementasi Bigram tf-idf

Pada gambar 9 dan 10 adalah hasil dari teknik n-gram setelah dilakukan pembobotan tf-idf. Berbeda dengan n-gram saja yang menggunakan notasi 0 dan 1, pada tf-idf menggunakan persamaan matematika sehingga didapat bobot yang berbeda antar kalimatnya.

D. Pelatihan Model

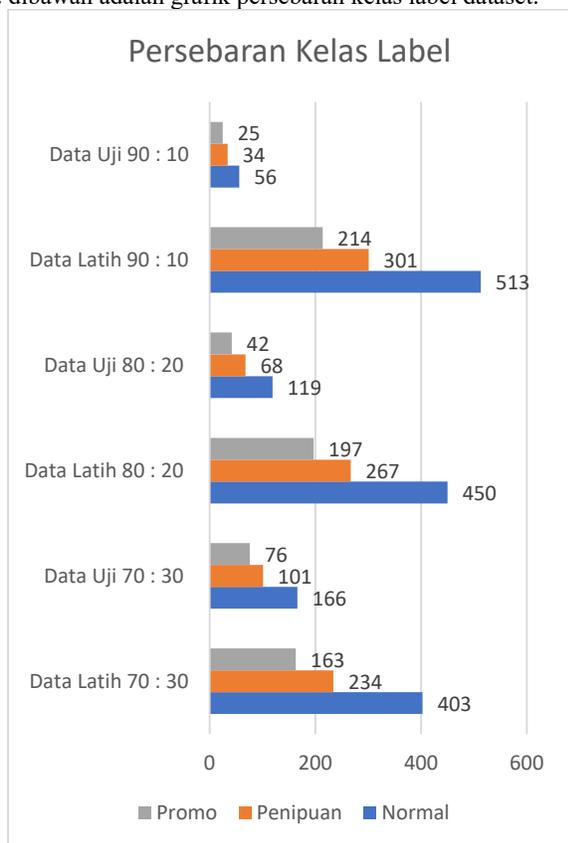


Gbr. 13 Skenario Pelatihan

Pada gambar 11 menunjukkan skenario pelatihan dimana data yang utuh akan dibagi menjadi unigram dan bigram. Setelah itu dipecah lagi menjadi 2 jenis lagi yaitu menggunakan pembobotan tf-idf dan tidak sehingga data siap digunakan pada model pembelajaran mesin.

E. Evaluasi

Tahap terakhir dalam pembelajaran mesin adalah melakukan evaluasi pada model yang telah di latih dan di uji, pada gambar 12 dibawah adalah grafik persebaran kelas label dataset.



Gbr. 14 Persebaran Kelas Label

Hasil nilai akurasi pengujian pada tiap skenario ditunjukkan pada tabel VI dibawah, pada skenario pertama adalah perbandingan data latih dan uji sebanyak 70:30, diikuti 80:20 dan terakhir 90:10.

TABEL VI  
HASIL NILAI AKURASI BERDASARKAN SKENARIONYA

Algoritma	Perbandingan Data Latih dan Uji 70:30					
	Unigram	Unigram tf-idf	Bigram	Bigram tf-idf	Unigram & Bigram	Unigram & Bigram tf-idf
D. Tree	0.85	0.85	0.8	0.77	0.86	0.84
D. Tree + Adaboost	0.9	0.9	0.8	0.79	0.88	0.91
SVM	0.92	0.93	0.75	0.88	0.91	0.94
SVM + Adaboost	0.85	0.48	0.48	0.48	0.58	0.48
Algoritma	Perbandingan Data Latih dan Uji 80:20					
	Unigram	Unigram tf-idf	Bigram	Bigram tf-idf	Unigram & Bigram	Unigram & Bigram tf-idf
D. Tree	0.9	0.89	0.79	0.78	0.9	0.86
D. Tree + Adaboost	0.92	0.93	0.8	0.8	0.91	0.91
SVM	0.93	0.94	0.76	0.88	0.92	0.96
SVM + Adaboost	0.7	0.52	0.52	0.52	0.83	0.52
Algoritma	Perbandingan Data Latih dan Uji 90:10					
	Unigram	Unigram tf-idf	Bigram	Bigram tf-idf	Unigram & Bigram	Unigram & Bigram tf-idf
D. Tree	0.89	0.82	0.81	0.8	0.89	0.85
D. Tree + Adaboost	0.9	0.9	0.83	0.8	0.91	0.9
SVM	0.91	0.91	0.76	0.84	0.9	0.91
SVM + Adaboost	0.83	0.49	0.49	0.49	0.61	0.49

Pada tabel VI menunjukkan nilai akurasi berdasarkan skenario pengujian pada perbandingan data latih dan uji, pada tiap skenario nilai tertinggi ada pada SVM yang memiliki nilai 96% saat menggunakan data unigram & bigram tf-idf.

TABEL VII  
HASIL NILAI AKURASI TERTINGGI TIAP ALGORITMA

Perbandingan Data Latih 70:30		
Algoritma	Jenis Data Latih	Nilai Akurasi
D. Tree	Unigram dan Bigram	0.86
D. Tree + Adaboost	Unigram & Bigram tf-idf	0.91
SVM	Unigram & Bigram tf-idf	0.94
SVM + Adaboost	Unigram	0.85
Perbandingan Data Latih 80:20		
Algoritma	Jenis Data Latih	Nilai Akurasi
D. Tree	Unigram dan Unigram & Bigram	0.9
D. Tree + Adaboost	Unigram tf-idf	0.93
SVM	Unigram & Bigram tf-idf	0.96
SVM + Adaboost	Unigram & Bigram	0.83
Perbandingan Data Latih 90:10		
Algoritma	Jenis Data Latih	Nilai Akurasi
D. Tree	Unigram dan Unigram & Bigram	0.89
D. Tree + Adaboost	Unigram & Bigram	0.91
SVM	Unigram, Unigram tf-idf dan unigram & bigram tf-idf	0.91
SVM + Adaboost	Unigram	0.83

Pada tabel VII adalah hasil nilai akurasi tertinggi tiap algoritma, pattern atau pola yang dapat diketahui dari data tersebut adalah jenis data unigram selalu menjadi penyebab dari nilai akurasi tertinggi.

TABEL VIII  
 HASIL NILAI K-FOLD CROSS VALIDATION

Nilai k = 5						
Algoritma	Data					
	Unigram	Unigram tf-idf	Bigram	Bigram tf-idf	Unigram dan Bigram	Unigram dan Bigram tf-idf
D. Tree	0.86	0.87	0.79	0.79	0.87	0.85
D. Tree + Adaboost	0.89	0.89	0.81	0.8	0.9	0.9
SVM	0.93	0.93	0.76	0.76	0.91	0.92
SVM + Adaboost	0.85	0.5	0.5	0.5	0.67	0.5
Nilai k = 10						
Algoritma	Data					
	Unigram	Unigram tf-idf	Bigram	Bigram tf-idf	Unigram dan Bigram	Unigram dan Bigram tf-idf
D. Tree	0.88	0.88	0.78	0.79	0.88	0.87
D. Tree + Adaboost	0.9	0.9	0.82	0.81	0.9	0.9
SVM	0.93	0.93	0.77	0.76	0.93	0.92
SVM + Adaboost	0.76	0.5	0.5	0.5	0.67	0.5

Pengujian k-fold menggunakan k = 5, Peningkatan signifikan pada algoritma D. Tree setelah diterapkan teknik adaboost terjadi saat menggunakan data unigram dan bigram tf-idf dengan kenaikan akurasi dari 85% menjadi 90%.

#### IV. KESIMPULAN

Penerapan teknik *adaboost* berpengaruh signifikan pada peningkatan akurasi algoritma decision tree. Algoritma SVM +

Adaboost hanya membuahkan hasil yang baik saat diterapkan data jenis unigram & bigram tanpa tf-idf dan unigram tanpa tf-idf. Algoritma SVM + Adaboost tidak cocok menggunakan tf-idf karena dapat menyebabkan penurunan nilai akurasi. Data bigram tidak memberikan hasil yang maksimal pada nilai akurasi model. D. Tree + Adaboost memiliki nilai akurasi k-fold yang cukup tinggi dibandingkan algoritma lain saat diterapkan tipe dataset bigram dan bigram tf-idf. Algoritma SVM memiliki nilai akurasi tertinggi dengan nilai 0.96 atau 96%.

#### REFERENSI

- [1] Rahmi F. & Wibisono Y., "Aplikasi SMS Spam Filtering pada Android menggunakan Naïve Bayes", 2021.
- [2] Sunardi, H. Murti dan H. Listiyono, "Aplikasi SMS Gateway", vol. XIV, no. 1, 2009.
- [3] Novanema Duha, "Short Message Services (SMS) Fraud Against Mobile Telephone Provider Consumer Review From Law Number 8 Of 1999 Concerning Consumer Protection", vol. III, no. 1, 2021.
- [4] Nabila Affah Mumtazahh, "Perbandingan Hasil Metode Support Vector Machine (SVM) dengan Ensemble SMOTE Bagging dan SMOTE Boosting pada Data Kelulusan Mahasiswa UNIMUS", 2021.
- [5] Devi Irawan, Eza Budi Perkasa dkk, "Perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naïve Bayes Classifier, Random Forest dan Bagging Classifier", vol. X, 2021.
- [6] Edi Zuviyanto, Teguh Bharata Adji, dkk, "Perbandingan Algoritme-Algoritme Pembelajaran Mesin Pada Klasifikasi SMS Spam", 2018.
- [7] Azmiardhy Zulkifli F., "Deteksi Surel SPAM dan Non-Spam Bahasa Indonesia Menggunakan Metode Naïve Bayes", 2021.
- [8] Panji Bimo Nugroho, Dkk., "Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5", 64-71, 2020.
- [9] Yogo Aryo Jatmiko, Dkk., " Analisis Perbandingan Kinerja CART Konvensional, Bagging dan Random Forest Pada Klasifikasi Objek: Hasil dari Dua Simulasi", 2019.
- [10] Agung Nugroho, Yoga Religia, "Analisis Optimasi Algoritma Klasifikasi Naïve Bayes menggunakan Genetic Algorithm dan Bagging", 2021.