

DETEKSI SMS SPAM BERBAHASA INDONESIA MENGGUNAKAN TF-IDF DAN STOCHASTIC GRADIENT DESCENT CLASSIFIER

(*Indonesian SMS Spam Detection using TF-IDF and Stochastic Gradient Descent Classifier*)

Ramaditia Dwiyanaputra*, Gibran Satya Nugraha, Fitri Bimantoro, Arik Aranta
Dept Informatics Engineering, Mataram University
Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA
Email: rama@unram.ac.id, [gibransn, bimo, arikaranta]@unram.ac.id

*Penulis Korespondensi

Abstract

Short Message Service (SMS) has evolved in the last few decades. The simplicity of SMS makes this short message service attractive to use as a direct communication service on mobile devices. As the popularity of this service increases, it also harms attacks on mobile devices such as SMS spam. Spam SMS are short messages that the recipient doesn't want, such as advertisements and scams. Spam SMS can overwhelm your inbox and make your mobile device experience less good. One way to overcome this problem is to implement a machine learning model to automatically recognize and filter Spam SMS. This research aims to build a machine learning model that provides higher accuracy for detecting SMS spam in Indonesian using the TF-IDF method and the Stochastic Gradient Descent Classifier. Based on the test results, the model built can detect SMS spam and not spam with an accuracy of 97%.

Keywords: Klasifikasi Teks, Sms Spam, TF-IDF, Stochastic Gradient Descent

*Penulis korespondensi

1. PENDAHULUAN

Short Message Service (SMS) atau layanan pesan singkat merupakan salah satu media komunikasi jarak jauh yang masih banyak digunakan pada era sekarang ini untuk mengirim pesan singkat. Seiring dengan perkembangan layanan pesan singkat ini, muncul dampak negatif berupa serangan pada perangkat seluler seperti sms spam. SMS spam merupakan pesan singkat yang tidak dikehendaki oleh penerima, contohnya seperti iklan dan penipuan (*fraud*). SMS spam. Selain dapat menyebabkan kerugian akibat dari penipuan, sms spam ini juga dapat membanjiri kotak masuk dan membuat pengalaman penggunaan perangkat seluler menjadi kurang baik.

Beberapa upaya pencegahan yang dilakukan oleh pemerintah melalui Kementerian Komunikasi dan Informatika (Kominfo) dan Badan Regulasi Telekomunikasi Indonesia (BRTI) untuk mengatasi permasalahan sms spam ini antara lain yakni meminta sejumlah operator seluler untuk memberikan ID

khusus bagi penjual pulsa di seluruh Indonesia pada tahun 2015 [1]. Kemudian, pada tahun yang sama juga dikeluarkan ketentuan pelaksanaan registrasi pelanggan pra bayar melalui Surat Ketua BRTI Nomor: 326/BRTI/IX/2015 [2]. Namun, upaya-upaya yang telah dilakukan tersebut belum cukup untuk dapat mengatasi permasalahan sms spam ini. SMS spam masih terus bermunculan dan banyak dikeluhkan oleh masyarakat [3]–[6].

Salah satu cara untuk menanggulangi permasalahan sms spam ini adalah dengan menggunakan teknik klasifikasi pembelajaran mesin (*machine learning*) untuk menyaring sms spam tersebut secara otomatis. Klasifikasi merupakan suatu proses untuk menemukan fungsi atau model yang dapat digunakan untuk mengenali ciri-ciri dari suatu kelas atau label (dalam hal ini adalah sms spam dan bukan spam) dengan tujuan untuk memprediksi kelas dari objek yang belum diketahui kelasnya. Jika model pembelajaran mesin yang dihasilkan dapat mengenali ciri sms spam dan bukan spam dengan baik, tentu cara

ini akan sangat efektif digunakan untuk mengatasi permasalahan sms spam.

Penelitian tentang klasifikasi sms spam berbahasa Indonesia telah cukup banyak dilakukan sebelumnya. Dalam penelitian-penelitian tersebut, metode-metode pembelajaran mesin seperti Support Vector Machine (SVM) [7]–[9] dan Naïve Bayes Classifier (NBC) [7], [10], [11] menghasilkan performa yang baik, namun belum mencapai performa ideal. Performa klasifikasi sms spam berbahasa Indonesia masih dapat ditingkatkan salah satunya dengan melakukan perbaikan dengan menerapkan metode yang dapat bekerja lebih baik dalam klasifikasi teks pada umumnya.

Salah satu metode yang memiliki performa yang baik dalam klasifikasi teks adalah metode Stochastic Gradient Descent (SGD) Classifier. SGD merupakan metode optimasi sederhana namun efisien untuk mencari nilai koefisien untuk meminimalkan *loss function* pada skala besar contohnya seperti data teks. Sedangkan, SGD Classifier adalah metode yang menggunakan SGD dalam proses pembelajarannya. Beberapa metode yang menggunakan SGD dalam proses pembelajarannya di antaranya adalah seperti metode Regresi Logistik, Smoothed hinge loss classifier dan linear Support Vector Regression.

Performa metode SGD Classifier lebih baik dibandingkan dengan SVM dalam penelitian sistem deteksi berita hoax berbahasa Indonesia [12] dan lebih juga baik dari metode SVM dan NBC dalam penelitian klasifikasi dokumen teks Bangla [13]. SGD Classifier juga memiliki performa yang baik saat dipilih sebagai pembanding metode-metode pembelajaran terbimbing untuk teks klasifikasi [14]. Metode SGD Classifier dinilai cocok untuk klasifikasi data teks yang memiliki multidimensi matriks (memiliki banyak fitur) dan cocok digabungkan dengan metode pembobotan fitur yang umum digunakan dalam klasifikasi teks yakni metode TF-IDF [12].

Berdasarkan uraian di atas, pada penelitian ini akan digunakan metode Stochastic Gradient Descent Classifier untuk deteksi sms spam berbahasa Indonesia. Metode SGD Classifier akan digabungkan dengan metode TF-IDF sebagai pembobotan fitur. Model yang diusulkan kemudian akan dibandingkan performanya dengan metode-metode yang memiliki performa terbaik dalam penelitian klasifikasi sms spam berbahasa Indonesia sebelumnya yakni metode Support Vector Machine dan Naïve Bayes Classifier.

2. TINJAUAN PUSTAKA

Penelitian tentang deteksi sms spam berbahasa Indonesia telah cukup banyak dilakukan sebelumnya.

Penelitian-penelitian tersebut menggunakan metode-metode *machine learning* yang cukup beragam dan memiliki performa yang cukup baik. Dari sisi pembobotan fitur, metode TF-IDF yang paling banyak digunakan.

Pada penelitian [7] dilakukan perbandingan metode klasifikasi C4.5, KNN, Naïve Bayes dan SVM untuk klasifikasi sms spam berbahasa Indonesia. Metode pembobotan fitur yang digunakan adalah pembobotan fitur TF-IDF. Jumlah dataset yang digunakan adalah sebanyak 100 sms berbahasa Indonesia yang terdiri dari masing-masing 50 sms spam dan bukan spam (ham). Teks pra-proses seperti case folding, normalisasi kalimat, stopword removal dan stemming (menggunakan librari Sastrawi) dilakukan untuk membuat data sms menjadi terstruktur. Kemudian dilakukan pengujian dengan 30 pengulangan 10-fold cross validation dan menghasilkan metode SVM yang memiliki tingkat akurasi yang paling tinggi yakni 94,06% kemudian diikuti oleh metode Naïve bayes, C4.5 dan KNN dengan akurasi berturut-turut yakni 86,1%, 85,86% dan 77,5%.

Pada penelitian [8] dilakukan perbandingan antara metode SVM dan NBC dalam melakukan klasifikasi sms spam berbahasa Indonesia. Metode pembobotan fitur yang digunakan adalah metode TF-IDF. Data sms yang digunakan sejumlah 1143 data sms yang dibagi menjadi data pelatihan sejumlah 765 data dan data uji sejumlah 378. Hasilnya didapatkan bahwa metode NBC memiliki performa dalam recall dan presisi yang lebih baik dibandingkan dengan metode SVM yakni sebesar 94% dan 95% untuk masing-masing recall dan presisi untuk metode NBC. Sedangkan untuk metode SVM dihasilkan nilai recall sebesar 92.06% dan presisi sebesar 92.2%.

Pada penelitian [10] digunakan metode NBC dan pembobotan fitur TF-IDF untuk klasifikasi pesan sms spam berbahasa Inggris. Metode algoritme Genetik juga digunakan untuk seleksi fitur. Untuk dataset, pada penelitian ini digunakan data sms sejumlah 5572 sms berbahasa Inggris dari UCI Machine Learning Repository. Dari penelitian ini dihasilkan bahwa metode seleksi fitur menggunakan algoritme Genetika dapat meningkatkan akurasi klasifikasi yakni yang awalnya tanpa seleksi fitur dihasilkan akurasi sebesar 89.39%, kemudian setelah menggunakan seleksi fitur akurasi yang dihasilkan meningkat menjadi 89.73%.

Dari uraian beberapa penelitian sebelumnya di atas, didapatkan bahwa metode SVM dan NBC memiliki performa yang baik dalam melakukan klasifikasi sms spam berbahasa Indonesia. Namun, performa tersebut belum mencapai performa ideal

dan masih terus dapat ditingkatkan dengan melakukan perbaikan dari sisi pra-proses data sms, pembobotan fitur ataupun menerapkan metode yang dapat bekerja dengan lebih baik dalam klasifikasi teks pada umumnya.

Salah satu metode yang memiliki performa yang baik dalam klasifikasi teks adalah metode Stochastic Gradient Descent (SGD) Classifier. SGD merupakan metode optimasi sederhana namun efisien untuk mencari nilai koefisien untuk meminimalkan *loss function* pada skala besar contohnya seperti data teks. Sedangkan, SGD Classifier adalah metode yang menggunakan SGD dalam proses pembelajarannya. Beberapa metode yang menggunakan SGD dalam proses pembelajarannya di antaranya adalah seperti metode Regresi Logistik, Smoothed hinge loss classifier dan linear Support Vector Regression.

Metode SGD Classifier digunakan dalam penelitian [12] untuk mendeteksi berita hoax berbahasa Indonesia. Hasilnya didapatkan bahwa metode SGD Classifier memiliki performa akurasi yang lebih baik dari metode SVM. Metode SGD Classifier cocok untuk digabungkan dengan metode pembobotan fitur TF-IDF. Pada penelitian tersebut, metode SGD Classifier dengan smoothed hinge loss menghasilkan peningkatan akurasi dan presisi sebesar 4% dan 20% masing-masing.

Dalam penelitian lainnya yakni penelitian [13] metode SGD Classifier memiliki performa F1-score yang lebih baik dibandingkan dengan metode SVM dan NBC dalam klasifikasi dokumen teks Bangla. Metode pembobotan TF-IDF digunakan dengan fitur bigram sehingga didapatkan sejumlah 350000 fitur dari 9 kategori artikel. Metode SGD Classifier juga memiliki performa yang baik saat dipilih sebagai pembanding metode-metode pembelajaran terbimbing untuk teks klasifikasi yang dilakukan pada penelitian [14]. Dalam penelitian tersebut tidak digunakan metode pembobotan TF-IDF, melainkan hanya digunakan probabilitas kemunculan dari term. Beberapa korpus dataset yang digunakan untuk pengujian adalah Reuters Corpus, Browns Corpus dan Movie-review Corpus.

Berdasarkan beberapa penelitian sebelumnya mengenai metode SGD Classifier di mana metode tersebut menghasilkan performa yang baik untuk klasifikasi teks, maka pada penelitian ini akan menggunakan metode SGD Classifier dan menggabungkannya dengan metode TF-IDF untuk klasifikasi sms spam berbahasa Indonesia. Kemudian model yang diusulkan tersebut akan dibandingkan performanya dengan metode-metode yang memiliki performa terbaik dalam penelitian klasifikasi sms spam

berbahasa Indonesia sebelumnya yakni metode SVM dan NBC.

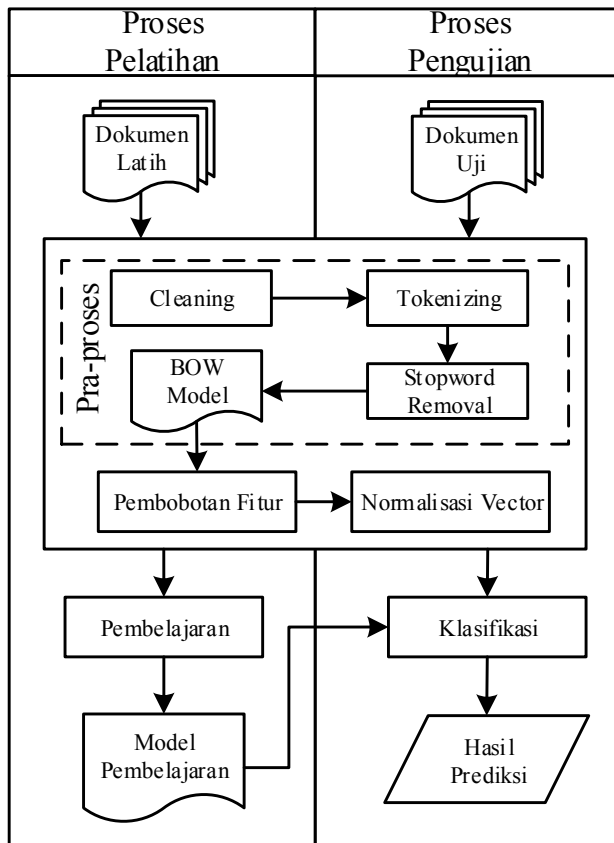
3. METODE PENELITIAN

3.1. Dataset

Dataset yang akan digunakan pada penelitian ini berjumlah 1143 data sms berbahasa Indonesia yang didapatkan dari penelitian [15]. Data sms terdiri dari 2 kategori yakni kategori sms normal (sms bukan spam) dengan jumlah 569 data sms dan kategori spam dengan jumlah 574 data sms.

3.2. Alur Klasifikasi Sms Spam

Pada Gambar 1 menunjukkan diagram proses klasifikasi sms spam yang akan dilakukan pada penelitian ini. Proses klasifikasi terbagi menjadi 2 proses yaitu proses pelatihan dan proses pengujian. Pada proses awal yaitu pelatihan, konten teks sms digunakan sebagai input sistem akan diolah agar menjadi teks terstruktur dan dapat diproses dengan baik oleh metode pembelajaran mesin. Proses pelatihan ini akan menghasilkan sebuah model pembelajaran yang akan digunakan untuk proses klasifikasi sms spam. Proses selanjutnya yakni proses pengujian bertujuan untuk mengevaluasi model yang dihasilkan pada proses pelatihan.



Gambar 1. Diagram Proses Klasifikasi Sms Spam

3.2.1. Pra-proses Data Teks

Pra-proses merupakan tahapan awal pada klasifikasi sms spam untuk mempersiapkan teks yang awalnya tidak terstruktur ke dalam bentuk token-token sehingga dapat diproses dengan baik oleh mesin klasifikasi. Token-token atau yang sering disebut juga fitur merupakan suatu ciri yang dapat diambil dari sebuah objek yang sebisa mungkin dapat merepresentasikan objek tersebut. Dalam dokumen teks, fitur yang dimaksud adalah konten teks yang terkandung dalam sms yang diambil dalam bentuk kata atau gabungan kata.

Terdapat 3 (tiga) tahapan yang dilakukan dalam proses ini yaitu tahapan *Cleaning*, *Tokenizing*, dan *Stopword Removal*.

- a. *Cleaning* merupakan suatu proses untuk menghilangkan karakter-karakter yang merupakan noise dari data teks sms. Pada penelitian ini akan fokus menggunakan konten teks sms sebagai fitur klasifikasi, maka karakter seperti *emoticon* akan dihilangkan pada proses ini.
- b. *Tokenizing* merupakan proses pemisahan deretan kata di dalam kalimat atau paragraf yang ada pada konten teks menjadi kata tunggal atau beberapa kata. Metode yang umum digunakan untuk

pemisahan kata ini adalah metode unigram yakni pemisahan menjadi kata tunggal. Sebelum dilakukan proses tokenizing, biasanya akan dilakukan proses penyeragaman kata menjadi huruf kecil atau huruf besar semua sehingga dapat menghindari perbedaan pengenalan kata "KAMU" dengan "kamu" oleh sistem

- c. *Stopword Removal* merupakan proses untuk menghilangkan kata-kata umum yang tidak terlalu berpengaruh pada teks. Contoh kata-kata tersebut dalam Bahasa Indonesia adalah kata "dan", "kamu", "ini", "itu" dan sebagainya. Pada penelitian ini digunakan stopwords yang didapatkan dari publik repositori [16] yang umum digunakan dalam klasifikasi teks secara umum.

Keluaran tahapan pra-proses ini adalah dokumen teks yang terstruktur yang disimpan dalam model representasi BOW atau "Bag-of-Word". Representasi ini merupakan model yang banyak digunakan dalam klasifikasi teks karena kemudahannya untuk tujuan klasifikasi [17]. Pada Gambar 2 dapat dilihat visualisasi model representasi BOW. D_i merupakan dokumen teks (dalam penelitian ini adalah sms), sedang T_i merupakan token hasil dari pra-proses.

$$\begin{pmatrix}
 & T_1 & T_2 & \dots & T_t \\
 D_1 & w_{11} & w_{21} & \dots & w_{t1} \\
 D_2 & w_{12} & w_{22} & \dots & w_{t2} \\
 \vdots & \vdots & \vdots & & \vdots \\
 \vdots & \vdots & \vdots & & \vdots \\
 D_n & w_{1n} & w_{2n} & \dots & w_{tn}
 \end{pmatrix}$$

Gambar 2. Representasi dengan model *bag-of-word*

3.2.2. Metode Pembobotan Fitur

Token atau kata yang dihasilkan pada tahapan pra-proses perlu diberikan bobot yang dapat merepresentasikan seberapa besar pengaruh kata tersebut terhadap suatu dokumen. Pemberian bobot yang sesuai akan memberikan pengaruh yang signifikan terhadap performa mesin klasifikasi teks.

Metode pembobotan fitur yang umum digunakan dalam penelitian sebelumnya adalah TF-IDF atau *Term Frequency - Inverse Document Frequency*. Metode TF-IDF menggunakan 2 parameter pembobotan yaitu pembobotan lokal dengan menggunakan $tf_{i,j}$ dan pembobotan global dengan menggunakan idf_i . $tf_{i,j}$ merupakan bobot yang didapat dari frekuensi

kemunculan kata i dalam dokumen j . Sedangkan idf_i merupakan bobot yang didapatkan dengan mempertimbangkan jumlah kemunculan kata i (DF_i) pada keseluruhan dokumen N . Selanjutnya, bobot total didapatkan dengan melakukan perkalian nilai bobot lokal dengan nilai bobot global. Persamaan (1) menunjukkan persamaan matematis metode TF-IDF yang digunakan pada.

$$idf_i = \log\left(\frac{N}{DF_i}\right) \quad (1)$$

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{DF_i}\right) \quad (2)$$

3.2.3. Normalisasi

Setiap dokumen teks memiliki panjang dokumen yang berbeda-beda. Variasi panjang dokumen teks ini dapat menyebabkan permasalahan sebagai berikut:

1. Besarnya frekuensi *term* atau kata
Pada dokumen yang panjang, kata yang sama cenderung muncul berulang kali sehingga menyebabkan *term frequency* cenderung besar. Besarnya *term frequency* mengakibatkan rata-rata bobot kata menjadi tinggi dan meningkatkan nilai relevansi kata pada dokumen.
2. Banyaknya *term* atau kata
Dalam dokumen yang panjang, sering ditemukan sejumlah term atau kata yang berbeda. Hal ini mengakibatkan meningkatnya sejumlah relevansi kata pada dokumen.

Untuk mengurangi pengaruh perbedaan panjang dokumen ini, pada pembobotan kata digunakan satu faktor lagi yang disebut sebagai normalisasi panjang dokumen. Normalisasi yang sering digunakan adalah normalisasi dengan cara membagi setiap bobot kata (komponen vektor) dengan panjang dokumen atau vektor (biasanya digunakan norm L2) sehingga memiliki persamaan seperti yang ditunjukkan pada Persamaan (3) [18].

$$w^n(k, j) = \frac{w^t(t_k, d_j)}{\sqrt{\sum_{s=1}^n (w^t(t_k, d_j))^2}} \quad (3)$$

3.2.4. Pembelajaran dan Proses Klasifikasi

Tujuan proses Pembelajaran ini adalah untuk membangun model pembelajaran yang akan digunakan dalam proses Klasifikasi. Algoritme pembelajaran yang digunakan pada penelitian ini adalah Stochastic Gradient Descent (SGD) Classifier. SGD merupakan metode optimasi sederhana berbasis statistik yang efisien digunakan mencari nilai koefisien

untuk meminimalkan *loss (error) function* pada skala besar. Saat diberikan suatu *function* (fungsi) yang memiliki sekumpulan parameter, SGD memulai dengan bobot nilai parameter awal, kemudian pada setiap iterasi akan bergerak menuju nilai parameter dengan bobot yang baru yang memiliki titik minimum untuk fungsi tersebut. Proses pencarian titik minimum menggunakan metode penurunan dalam kalkulus untuk mencari garis lurus yang mendekati nilai minimum.

Proses algoritme SGD adalah dengan menemukan nilai θ yang dapat meminimalkan fungsi $J(\theta)$. Untuk menentukan nilai awal θ digunakan algoritme pencarian, kemudian pada setiap iterasi nilai θ agar terus diperbaharui sampai menemukan titik minimum atau nilai J yang paling minimum. Proses pembaharuan nilai θ pada setiap iterasi menggunakan Persamaan (4). Pembaharuan dilakukan secara bersamaan untuk semua nilai $j = 0, \dots, n$. Variable α merupakan learning rate yang mengatur seberapa besar pembaharuan nilai. Persamaan nilai $J(\theta)$ dapat dilihat pada Persamaan (5), di mana L merupakan loss function yang digunakan pada data pelatihan $(x_1, y_1), \dots, (x_n, y_n)$, dan R merupakan regularisasi atau penalty terhadap kompleksitas model.

$$\theta_j = \theta_j - \alpha \frac{\partial y}{\partial x} J(\theta) \quad (4)$$

$$J(\theta) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \alpha R(w) \quad (5)$$

4. HASIL DAN PEMBAHASAN

Pada bagian ini dilaporkan hasil penelitian untuk mengukur performa model yang diusulkan yakni model untuk mendeteksi sms spam berbahasa Indonesia menggunakan TF-IDF dan SGD Classifier dengan 3 jenis loss function berbeda yakni smoothed hinge loss (modified huber) dan logistic regression (log). Performa model yang diusulkan akan dibandingkan dengan metode terbaik yang digunakan pada penelitian sebelumnya yaitu metode SVM dan NBC. Metode-metode tersebut diimplementasikan ulang menggunakan bahasa pemrograman Python dengan bantuan pustaka Scikit-learn [19]. Untuk penjelasan mengenai implementasi metode SVM dan NBC untuk melakukan klasifikasi sms spam berbahasa Indonesia dapat dilihat pada penelitian [7], [8], [10]. Selanjutnya, akan dibandingkan juga terkait pengaruh penggunaan stopword dalam pra-proses data teks dalam klasifikasi sms spam berbahasa Indonesia.

Untuk evaluasi performa digunakan metode K-fold Cross Validation dengan nilai $k=10$ dan dilakukan

pengulangan sebanyak 30 kali (*30-Stage 10-Cross Fold Validation*). Metrik pengukuran yang digunakan untuk pengukuran performa ini adalah akurasi, presisi dan recall seperti yang terlihat pada Persamaan (6), (7) dan (8). Metrik tersebut dihitung berdasarkan nilai dari true positive (TP), false positive (FP), true negative (TN) dan false negative (FN) pada tabel confusion matrix seperti yang terlihat pada Tabel I.

TABEL I. Confusion Matrix.

		Kategori Prediksi	
		Spam	Non Spam
Kategori Sebenarnya	Spam	True Positive (TP)	False Negative (FN)
	Non Spam	False Positive (FP)	True Negative (TN)

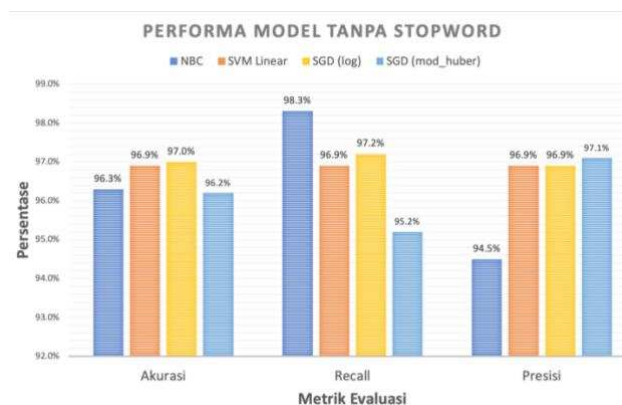
$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$Presisi = \frac{TP}{TP+FP} \tag{7}$$

$$Recall = \frac{TP}{TP+FN} \tag{8}$$

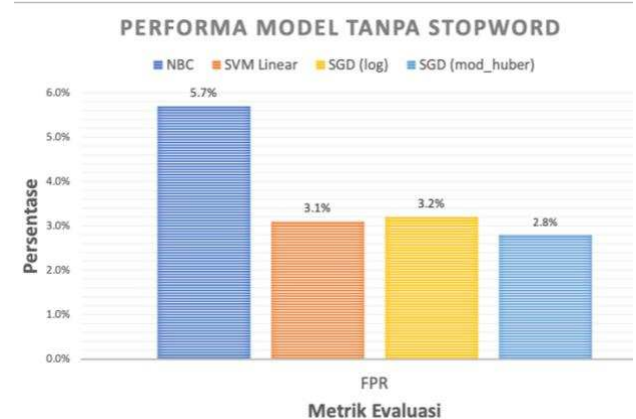
4.1. Pengujian Tanpa Stopword

Gambar 2 merupakan hasil pengujian perbandingan metode SGD Classifier dengan metode SVM dan NBC. Dari hasil tersebut dapat dilihat bahwa performa beberapa metode yang diujikan memiliki performa yang tidak jauh berbeda. Namun, secara rata-rata metode SGD Classifier dengan loss function logistik (log) memiliki performa terbaik dalam hal akurasi yakni 97%. Untuk performa dalam hal recall dan presisi secara rata-rata metode SGD Classifier dengan loss function logistik juga memiliki performa terbaik dengan rata-rata 97.05%.



Gambar 3. Perbandingan Metode SGD Classifier dengan Metode SVM dan NBC tanpa menggunakan Stopword

Dari hasil pada Gambar 3, dapat terlihat juga bahwa metode NBC memiliki nilai recall yang paling tinggi namun sebaliknya memiliki nilai presisi yang paling rendah di antara metode yang diujikan. Hal ini dapat menunjukkan bahwa metode NBC memiliki nilai False Positif Rate (FPR) yang paling tinggi di antara metode-metode yang diujikan seperti yang terlihat pada Gambar 4. Nilai FPR menunjukkan bahwa metode NBC memiliki tingkat kesalahan yang paling tinggi dalam mengidentifikasi kelas negatif (sms normal atau bukan spam) sebagai kelas positif (sms spam). Nilai FPR yang tinggi ini dapat menyebabkan *over blocking* yakni sistem akan cukup banyak mendeteksi sms normal (sms bukan spam) menjadi sms spam. Hal ini krusial karena akan merugikan pengguna jika sms normal yang dikirimkan akan difilter karena dideteksi sebagai sms spam. Pada Gambar 4, metode NBC memiliki nilai FPR sebesar 5,7% sedangkan untuk metode nilai FPR cenderung tidak jauh berbeda, namun metode SGD dengan loss function mod_huber memiliki performa yang paling baik yakni 2,8%.



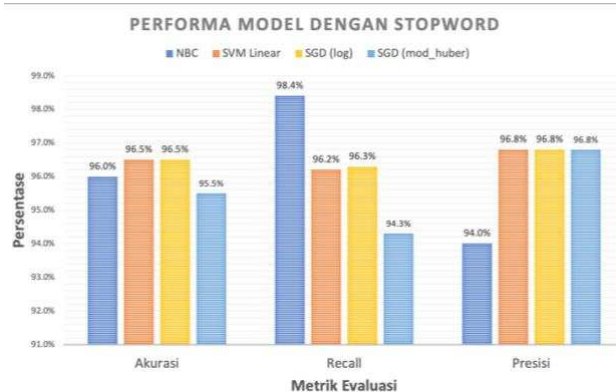
Gambar 4. Perbandingan nilai False Positif Rate (FPR)

4.2. Pengujian dengan Stopword

Hasil pengujian performa model klasifikasi sms spam dengan menggunakan stopwords yang bertujuan untuk mengurangi jumlah fitur umum memiliki hasil yang lebih rendah dibandingkan dengan tanpa penggunaan stopwords. Seperti yang terlihat pada Gambar 5, secara rata-rata model yang diujikan mengalami penurunan performa sebesar 0,5% untuk akurasi, 0,6% untuk recall dan 0,3% untuk presisi.

Penurunan performa model klasifikasi ini dapat disebabkan karena penggunaan data stopwords yang diperoleh dari publik repositori dataset yang kurang

merepresentasikan fitur-fitur umum dalam klasifikasi sms spam karena merupakan stopwords untuk klasifikasi teks berbahasa Indonesia secara umum. Konten teks sms spam cukup pendek, jadi penggunaan stopwords yang cukup banyak memungkinkan terhapusnya fitur kata penting yang dapat digunakan untuk deteksi sms spam atau sms normal dengan cukup baik.



Gambar 4. Perbandingan Metode SGD Classifier dengan Metode SVM dan NBC menggunakan Stopword

5. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dalam penelitian ini digunakan metode TF-IDF dan SGD Classifier dengan 2 metode loss function yakni logistic (log) dan smoothed hinge lost untuk mendeteksi sms spam berbahasa Indonesia. Berdasarkan hasil penelitian, metode TF-IDF dan Stochastic Gradient Descent Classifier menghasilkan performa yang cukup baik yakni yang paling tinggi dalam hal akurasi sebesar 97% dengan nilai Recall sebesar 97,2% dan Presisi sebesar 96,9% untuk mengenali sms spam dibandingkan dengan metode yang digunakan pada penelitian sebelumnya yakni SVM dan NBC.

Selanjutnya, penggunaan stopwords pada pra-proses teks yang bertujuan untuk mengurangi jumlah fitur umum menurunkan performa model klasifikasi sms spam pada penelitian ini. Hal ini dapat disebabkan karena penggunaan data stopwords yang diperoleh dari repositori untuk klasifikasi teks secara umum, sehingga perlu untuk dibentuk data stopwords yang lebih sesuai untuk klasifikasi sms spam agar tidak banyak menghapus fitur penting untuk proses klasifikasi mengingat konten teks sms yang terbatas.

5.2 Saran

Untuk penelitian selanjutnya perlu digunakan daftar stopwords yang lebih sesuai dengan data sms spam. Selanjutnya, dalam hal pembobotan fitur, dapat dicoba metode pembobotan lain selain TF-IDF yang cukup umum digunakan seperti metode pembobotan fitur terbimbing atau metode pembobotan fitur yang memanfaatkan informasi kelas dari dataset sms spam.

DAFTAR PUSTAKA

- [1] I. Rahmayani, "Atasi SMS Spam, Ini Langkah Operator Seluler," 2015. [Online]. Available: https://kominfo.go.id/content/detail/6042/ata-si-sms-spam-ini-langkah-operator-seluler/0/sorotan_media. [Accessed: 26-Feb-2021].
- [2] I. Cawidu, "Penertiban Registrasi Pelanggan Jasa Telekomunikasi Pra Bayar," *Kominfo*. [Online]. Available: https://kominfo.go.id/content/detail/6516/siaran-pers-no94pihkominfo122015-tentang-penertiban-registrasi-pelanggan-jasa-telekomunikasi-pra-bayar/0/siaran_pers. [Accessed: 25-Feb-2021].
- [3] Tim Detiknet, "Ampun! Bombardir SMS Spam Masih Terjadi," *DetikNet*, 2020. [Online]. Available: <https://inet.detik.com/telecommunication/d-5183170/ampun-bombardir-sms-spam-masih-terjadi>. [Accessed: 25-Feb-2021].
- [4] Kumparan, "Registrasi SIM Card Tak Bisa Bendung SMS Spam: Pelanggan Cuma Bisa Pasrah," *Kumparan*, 2020. [Online]. Available: <https://kumparan.com/kumparantech/registrasi-si-sim-card-tak-bisa-bendung-sms-spam-pelanggan-cuma-bisa-pasrah-1uWEDONo65h/full>. [Accessed: 25-Feb-2021].
- [5] N. F. Shalihah, "Sering Di-SMS dari Nomor Tak Dikenal? Ini Penjelasan Telkomsel," *Kompas*, 2020. [Online]. Available: <https://www.kompas.com/tren/read/2020/07/11/190500965/sering-di-sms-dari-nomor-tak-dikenal-ini-penjelasan-telkomsel>. [Accessed: 25-Feb-2021].
- [6] L. Hasibuan, "SMS Spam Marak, BRTI Minta Ini ke Telkomsel, XL, Indosat Cs," *CNBC Indonesia*, 2020. [Online]. Available: <https://www.cnbcindonesia.com/tech/20201002103826-37-191138/sms-spam-marak-brti->

- minta-ini-ke-telkomsel-xl-indosat-cs.
[Accessed: 25-Feb-2021].
- [7] E. Zuviyanto, T. B. Adji, and N. A. Setiawan, "Perbandingan Algoritme-Algoritme Pembelajaran," *Semin. Nas. Inov. dan Apl. Teknol. di Ind. 2018*, vol. 4, no. Green Technology and Sustainable Innovation, pp. 20–26, 2018.
- [8] Widyawati and Sutanto, "Perbandingan Algoritma Naive Bayes Dan Support Vector Machine (Svm)," vol. 3, no. 2, pp. 178–194, 2019.
- [9] R. I. Dona, "Identifikasi SMS SPAM Berbahasa Indonesia Menggunakan Algoritma Support Vector Machine," 2018.
- [10] I. Munitasri, S. Santosa, and C. Supriyanto, "Klasifikasi Pesan Sms Menggunakan Algoritma Naive Bayes Dengan Seleksi Fitur Genetic Algorithm," *J. Teknol. Inf.*, vol. 14, no. 1, 2018.
- [11] J. Ilmu and F. Ugm, "Aplikasi untuk Identifikasi Short Message Service (SMS) Spam Berbasis Android," *Bimipa*, vol. 24, no. 3, pp. 257–262, 2017.
- [12] A. B. Prasetijo, R. R. Isnanto, D. Eridani, Y. A. A. Soetrisno, M. Arfan, and A. Sofwan, "Hoax detection system on Indonesian news sites based on text classification using SVM and SGD," *Proc. - 2017 4th Int. Conf. Inf. Technol. Comput. Electr. Eng. ICITACEE 2017*, vol. 2018-Janua, pp. 45–49, 2017, doi: 10.1109/ICITACEE.2017.8257673.
- [13] F. Kabir, S. Siddique, M. R. A. Kotwal, and M. N. Huda, "Bangla text document categorization using Stochastic Gradient Descent (SGD) classifier," *Proc. - 2015 Int. Conf. Cogn. Comput. Inf. Process. CCIP 2015*, pp. 0–3, 2015, doi: 10.1109/CCIP.2015.7100687.
- [14] S. Z. Mishu and S. M. Rafiuddin, "Performance analysis of supervised machine learning algorithms for text classification," *19th Int. Conf. Comput. Inf. Technol. ICCIT 2016*, pp. 409–413, 2017, doi: 10.1109/ICCITECHN.2016.7860233.
- [15] F. Rahmi and Y. Wibisono, "Aplikasi SMS Spam Filtering pada Android menggunakan Naive Bayes," *Unpubl. Manuscr.*, 2016.
- [16] O. R. Hartono, "Indonesian Stoplist, Most common words (stop words) in Bahasa Indonesia," 2016. [Online]. Available: <https://www.kaggle.com/oswinrh/indonesian-stoplist>. [Accessed: 22-Jun-2020].
- [17] C. C. Aggarwal and C. Zhai, *Mining Text Data*. Springer, 2012.
- [18] Q.-R. Zhang, L. Zhang, S.-B. Dong, and J.-H. Tan, "Document indexing in text categorization," in *Proceedings of 2005 International Conference on Machine Learning and Cybernetics, 2005*, vol. 6, pp. 3792–3796 Vol. 6, doi: 10.1109/ICMLC.2005.1527600.
- [19] F. Pedregosa and G. Varoquaux, "Scikit-learn: Machine Learning in Python," *J. Mach. ...*, vol. 12, pp. 2825–2830, 2011.