Klasifikasi Spam Pada Email Berbahasa Indonesia Menggunakan FastText dan Bernoulli Naïve Bayes

***Abstrak*—Indonesia menempati peringkat keenam di dunia dalam jumlah pengirim spam. Berbagai penelitian telah dilakukan terkait deteksi dan penyaringan spam, dengan algoritma Bayesian menjadi salah satu pendekatan yang paling banyak digunakan. Penelitian ini bertujuan untuk mengklasifikasikan pesan email berbahasa Indonesia ke dalam kategori spam dan non-spam. Dataset sekunder yang digunakan berjumlah 2.604 pesan, terdiri dari 1.362 pesan spam dan 1.242 pesan non-spam. Representasi kata dilakukan menggunakan FastText dengan pendekatan n-gram untuk menangkap informasi pada level sub-kata, sementara klasifikasi dilakukan menggunakan algoritma Bernoulli Naïve Bayes berbasis nilai biner. Eksperimen membandingkan kinerja algoritma Bernoulli Naïve Bayes dengan dan tanpa penggunaan FastText. Evaluasi dilakukan menggunakan metrik akurasi, confusion matrix, dan classification report, dengan pembagian data 70:30. Hasil penelitian menunjukkan bahwa kedua model, baik dengan maupun tanpa FastText, memberikan kinerja yang lebih seimbang pada tiap kelas dan memiliki recall yang lebih tinggi dalam mendeteksi spam. Sebaliknya, model tanpa FastText mencapai presisi dan recall sempurna untuk spam tetapi menunjukkan penurunan performa pada non-spam. Oleh karena itu, penggunaan FastText berkontribusi dalam meningkatkan sensitivitas dan keseimbangan klasifikasi email spam berbahasa Indonesia.**

Kata Kunci—Bernoulli Naïve Bayes, FastText, Bahasa Indonesia, Email Spam, Klasifikasi Teks

I. PENDAHULUAN

Email spam mengacu pada pesan massal yang tidak diminta, sering kali berisi iklan, upaya penipuan, atau konten berbahaya seperti tautan atau lampiran berbahaya. Sejak awal 1990-an, spam telah menjadi tantangan yang terus berkembang dalam komunikasi digital. Menurut statistik Av-Test, Indonesia menempati peringkat keenam di dunia dalam jumlah pengirim spam [1].

Penelitian tentang klasifikasi spam telah berkembang pesat, dengan algoritma Bayesian menjadi salah satu pendekatan yang paling banyak digunakan karena kesederhanaan dan efisiensi komputasinya [2]. Di antara variannya, Bernoulli Naïve Bayes dianggap efektif untuk tugas klasifikasi teks biner, khususnya ketika berhadapan dengan dataset yang terbatas [3].

Untuk meningkatkan akurasi klasifikasi, representasi kata menggunakan *word embedding* menjadi sangat penting [4]. Salah satu model yang menonjol adalah FastText, yang dapat menangani kata-kata di luar kosakata dengan memanfaatkan informasi sub-kata melalui karakter n-gram [5]. FastText terbukti efektif untuk bahasa yang kaya morfologi, termasuk bahasa Indonesia [6][7].

Penelitian ini bertujuan untuk menerapkan FastText sebagai model representasi kata dan Bernoulli Naïve Bayes sebagai algoritma klasifikasi untuk mendeteksi spam pada email berbahasa Indonesia. Kombinasi metode ini diharapkan mampu memberikan kinerja klasifikasi spam yang akurat dan efisien.

II. KAJIAN LITERATUR

A. *Email Spam*

Email spam adalah pesan elektronik yang dikirim secara massal tanpa diminta, biasanya berisi iklan, phishing, malware, atau konten berbahaya lainnya. Spam dapat mengganggu pengguna, membebani jaringan, dan meningkatkan risiko keamanan siber [8]. Menurut Cormack [9], meskipun berbagai teknik penyaringan spam telah dikembangkan, tantangan utama tetap ada dalam menjaga akurasi dan efisiensi sistem deteksi spam. Data latih yang diperbarui secara berkala menjadi sangat penting karena taktik spammer yang terus berkembang.

B. *Klasifikasi Teks dan Pra-pemrosesan*

Klasifikasi teks adalah proses memberikan label tertentu pada data teks berdasarkan fitur tertentu [10]. Untuk meningkatkan performa, data teks biasanya melalui tahapan pra-pemrosesan seperti:

- *Cleaning*, menghapus karakter khusus, angka, dan simbol yang tidak relevan.
- *Case Folding*, mengubah semua teks menjadi huruf kecil.
- *Tokenizing*, memecah teks menjadi kata-kata atau token.
- *Stopword Removal*, menghapus kata-kata umum yang tidak memiliki nilai semantik besar.
- *Stemming*, mengubah kata menjadi bentuk dasarnya untuk mengurangi variasi kata.

Tahapan ini penting untuk mengurangi kompleksitas data dan meningkatkan kualitas fitur sebelum klasifikasi.

C. *FastText*

FastText, dikembangkan oleh Facebook AI Research, merupakan model embedding kata dan klasifikasi teks yang mengembangkan Word2Vec dengan memasukkan informasi sub-kata (n-gram). Hal ini membuat model lebih baik dalam menggeneralisasi, terutama untuk bahasa yang kaya morfologi atau kata-kata langka [6].

$$v_w = \frac{1}{|G_w|}\sum_{g \in G_w} z_g \quad (1)$$

Dimana:

- v_w adalah vektor representasi untuk kata w.
- G_w adalah himpunan karakter n-gram dari kata w.
- $|G_w|$ adalah total jumlah n-gram dalam G_w.

- z_g adalah representasi vektor dari karakter n-gram g (dihasilkan oleh FastText selama pelatihan).

Representasi kata dihitung berdasarkan rata-rata vektor n-gram karakter yang membentuk kata, sehingga memungkinkan model menghasilkan embedding yang bermakna bahkan untuk kata di luar kosakata [7][11].

D. Bernoulli Naïve Bayes

Bernoulli Naïve Bayes adalah pengklasifikasi probabilistik yang diturunkan dari Teorema Bayes. Model ini bekerja dengan fitur biner yang merepresentasikan ada/tidak adanya kata dalam dokumen. Probabilitas sebuah dokumen termasuk ke dalam kelas tertentu dihitung berdasarkan fitur biner tersebut [12]. Model ini sangat sesuai untuk tugas klasifikasi teks dengan data terbatas, dan penelitian sebelumnya menunjukkan kinerjanya yang unggul dibandingkan varian Naïve Bayes lainnya.

$$P(x_i|C_k) = \prod_{i=1}^{n}(p_{k_i}{}^{x_i}(1 - p_{k_i})^{(1-x_i)}) \qquad (2)$$

W:

- p_{k_i} adalah probabilitas bahwa fitur x_i muncul dalam kelas C_k.
- x_i adalah nilai biner (0/1) yang menunjukkan ada atau tidak adanya fitur tersebut dalam dokumen.

III. METODOLOGI

Penelitian ini menggunakan dataset sekunder dari Kaggle dengan judul *Indonesian Email Spam Dataset*. Dataset berisi 2.604 pesan email, terdiri dari 1.362 spam dan 1.242 non-spam. Setiap record berisi isi email lengkap dan labelnya (spam atau non-spam).

Fig. 1. Kerangka Kerja

A. Pra-pemrosesan Data

Pra-pemrosesan sangat penting untuk mengubah teks yang tidak terstruktur menjadi format terstruktur yang sesuai untuk machine learning. Tahapan berikut diterapkan:

- *Cleaning*, menghapus karakter khusus, tanda baca, angka, dan simbol yang tidak relevan.
- *Cleaning*, mengubah semua teks menjadi huruf kecil.
- *Tokenizing*, memecah teks menjadi token/kata.
- *Stopword Removal*, menghapus kata-kata umum seperti “dan”, “di”, “ke”.
- *Stemming*, mengubah kata menjadi bentuk dasarnya.

B. Word Representation using FastText

Setelah pra-pemrosesan, teks bersih dikonversi ke dalam vektor menggunakan model FastText unsupervised. FastText memetakan setiap kata ke dalam ruang vektor padat berdasarkan n-gram karakter, sehingga mampu menangani kata yang tidak dikenal.

C. Classification using Bernoulli Naïve Bayes

Setelah data dipra-proses dan direpresentasikan dengan FastText, proses klasifikasi dilakukan menggunakan algoritma Bernoulli Naïve Bayes, dengan tahapan sebagai berikut::

- Binarisasi, vektor FastText dikonversi ke dalam bentuk biner (0/1).
- Pelatihan, model dilatih pada data latih.
- Klasifikasi, model memprediksi label pada data uji.

IV. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil eksperimen dari klasifikasi email spam yang diusulkan menggunakan representasi kata FastText dan pengklasifikasi Bernouli Naïve Bayes. Evaluasi dilakukan dalam tiga skenario untuk menilai pegaruh rasio pembagian data latih-uji, evaluasi inerja model, serta analisis perbandingan antara Bernoulli Naïve Bayes dengan dan tanpa representasi FastText.

A. Skenario I – Pengaruh Rasio Data Latih-Uji

Skenario pertamo mengevaluasi pengaruh rasio pembagia data latih-uji yang berbeda (70:30, 80:20, dan 90:10) terhadap kinerja klasifikasi. Tabel 1 menunjukkan akurasi, presisi, recall dan f1-score yang diperoleh untuk setiap rasio.

Tabel 1. Perbandingan Kinerja dengan Rasio Data Latih-Uji

Rasio	Akurasi	Presisi	Recall	F1-Score
70:30	0.93	0.90	0.96	0.92
80:20	0.92	0.90	0.96	0.92
90:10	0.93	0.92	0.96	0.93

Berdasarkan Tabel 1, rasio 70:30 membrikan kinerja paling seimbang dan konsisten, sehingga dipilih sebagai konfigurasi utama untuk eksperimen selanjutnya..

B. Skenario II – Evaluasi FastText + Bernoulli Naïve Bayes

Pada skenario kedua, model dilatih dengan rasio 70:30, kemudian dievaluasi menggunakan metric akurasi, presisi, recall dan f1-score untuk kelas spam dan non-spam.

Tabel 2. Klasifikasi FastText + Bernoulli Naïve Bayes

Kelas	Presisi	Recall	F1-Score	Akurasi
Spam	0.93	0.98	0.96	0.95
Non-Spam	0.98	0.92	0.95	0.95

Table 3. Confusion Matrix FastText + Bernoulli Naïve Bayes

Actual	Prediction	
	Spam	Non-Spam
Spam	401	8
Non-Spam	28	345

Confusion matrix pada Tabel 3 menunjukkan jumlah *true positive* (TP) dan *true negative* (TN) yang tinggi, yang berarti

model mampu mengklasifikasikan email spam dan non-spam dengan baik.

Selain itu, kurva ROC (fig 2) menunjukkan nilai AUC mendekati 1, yang menandakan kemamuan diskriminasi yang sangat baik. Kurva Precision-Recall (fig 3) juga mengonfirmasi kinerja yang stabil meskipun terdapat ketidakseimbangan kelas ringan.

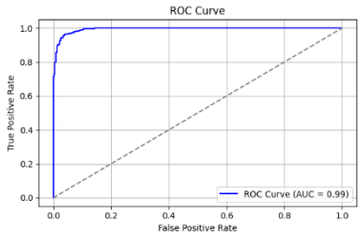

Fig. 2. Kurva ROC FastText + Bernoulli Naïve Bayes

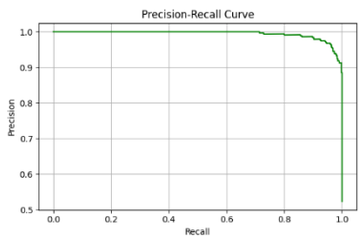

Fig. 3. Kurva Precision-Recall FastText + Bernoulli Naïve Bayes

C. Skenario III – Analisis Perbandingan dengan atau tanpa FastText

Skenario ketiga membandingkan Bernoulli Naïve Bayes dengan dan tanpa representasi kata FastText. Hasil pada Tabel 4 menunjukkan bahwa meskipun Bernoulli Naïve Bayes tanpa FastText mencapai recall sempurna untuk spam (1.00), model tersebut juga menghasilkan lebih banyak *false positive* dengan mengklasifikasikan email non-spam sebagai spam. Sebaliknya, kombinasi Bernoulli Naïve Bayes dengan FastText menunjukkan keseimbangan yang lebih baik antara presisi dan recall pada kedua kelas, sehingga lebih andal untuk penyaringan spam secara umum.

Tabel 4. Perbandingan Kinerja Model

Model	Kelas	Presisi	Recall	F1-Score	Akurasi
Dengan FastText	Spam	0.93	0.98	0.96	0.95
	Non-Spam	0.98	0.92	0.95	0.95
Tanpa FastText	Spam	0.92	1.00	0.96	0.95
	Non-Spam	1.00	0.91	0.95	0.95

D. Pembahasan

Hasil eksperimen menunjukkan bahwa integrasi embedding kata FastText dengan Bernoulli Naïve Bayes memberikan keseimbangan antara presisi dan recall pada kelas spam maupun non-spam. Sementara model tanpa FastText mencapai recall sempurna untuk spam, model tersebut juga meningkatkan tingkat *false positive*, sehingga mengurangi keandalannya dalam penerapan nyata. Model dengan FastText memiliki keuntungan dalam menangani variasi morfologi dan bentuk kata informal dalam bahasa Indonesia, sehingga lebih sesuai untuk tugas deteksi spam pada email berbahasa Indonesia.

V. KESIMPULAN

Penelitian ini menguji kinerja klasifikasi deteksi spam email berbahasa Indonesia menggunakan model embedding FastText yang dikombinasikan dengan algoritma Bernoulli Naïve Bayes (BNB). Eksperimen dilakukan pada dataset berjumlah 2.604 email, terdiri dari 1.362 spam dan 1.242 non-spam.

Tiga konfigurasi percobaan dilakukan. Pada konfigurasi pertama, variasi rasio latih–uji (70:30, 80:20, 90:10) menunjukkan bahwa rasio 70:30 memberikan kinerja paling seimbang dan konsisten, dengan akurasi 93% serta nilai presisi dan recall yang seimbang. Konfigurasi kedua, yang berfokus pada evaluasi detail dengan rasio 70:30, mencapai akurasi keseluruhan 95% untuk FastText + BNB, dengan kurva ROC menghasilkan AUC mendekati 1 dan kurva Precision–Recall menunjukkan kinerja stabil pada semua kelas. Konfigurasi ketiga membandingkan BNB dengan dan tanpa FastText. Hasil menunjukkan bahwa BNB tanpa FastText sedikit lebih unggul dalam recall spam (1.00) namun menghasilkan lebih banyak *false positive*.

Temuan ini menegaskan bahwa FastText efektif dalam menangani variasi morfologi dan istilah informal yang umum pada bahasa Indonesia, sehingga memberikan keseimbangan yang baik antara presisi dan recall. Namun, pada skenario dengan dominasi spam yang tinggi, BNB tanpa FastText dapat mencapai recall lebih tinggi, meskipun dengan konsekuensi meningkatnya jumlah *false positive*.

A. Keterbatasan

Penelitian ini dilakukan menggunakan dataset berjumlah 2.604 sampel, sehingga generalisasi hasil mungkin masih terbatas. Selain itu, prapemrosesan dan pelatihan model juga dibatasi oleh persyaratan kompatibilitas (misalnya, versi *numpy* untuk FastText < 2.0).

B. Penelitian Selanjutnya

Penelitian di masa depan dapat mengeksplorasi penggunaan dataset yang lebih besar dan beragam, teknik representasi kata tambahan (misalnya *contextual embeddings* seperti BERT), serta pendekatan klasifikasi hibrida untuk lebih meningkatkan kinerja dan kemampuan adaptasi terhadap pola spam yang terus berkembang.

REFERENSI

[1]] AV-Test, “Spam statistics and trens,” AV-Test. [Online]. Available: https://www.avtest.org/en/statistics/spam/.

[2] Y. Vernanda, S. Hansun, and M. B. Kristanda, “Indonesian language email spam detection using n-gram and naïve Bayes algorithm,” Bulletin of Electrical Engineering and Informatics, vol. 9, no. 5, 2020. [Online]. Available: https://doi.org/10.11591/eei.v9i5.2444.

[3] G. Singh, B. Kumar, L. Gaur, and A. Tyagi, “Comparison between Multinomial and Bernoulli Naïve Bayes for text classification,” in 2019 International Conference on Automation, Computational and Technology Management (ICACTM), 2019. [Online]. Available: https://doi.org/10.1109/ICACTM.2019.8776800.

[4] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in Proc. 1st Int. Conf. on Learning Representations (ICLR) – Workshop Track, 2013.

[5] T. Yao, Z. Zhai, and B. Gao, “Text classification model based on fastText,” in Proc. 2020 IEEE International Conference on Artificial

Intelligence and Information Systems (ICAIIS), 2020. [Online]. Available: https://doi.org/10.1109/ICAIIS49377.2020.9194939.

[6] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017. [Online]. Available: https://doi.org/10.1162/tacl_a_00051.

[7] A. Nurdin, B. Anggo Seno Aji, A. Bustamin, and Z. Abidin, "Perbandingan kinerja word embedding Word2Vec, GloVe, dan fastText pada klasifikasi teks," Jurnal Tekno Kompak, vol. 14, no. 2, 2020. [Online]. Available: https://doi.org/10.33365/jtk.v14i2.732.

[8] E. N. Putra, "Pengiriman e-mail spam sebagai kejahatan cyber di Indonesia," Jurnal Cakrawala Hukum, vol. 7, no. 2, 2016. [Online]. Available: https://doi.org/10.26905/idjch.v7i2.1906.

[9] G. V. Cormack, "Email spam filtering: A systematic review," Foundations and Trends in Information Retrieval, vol. 1, no. 4, pp. 335–455, 2006. [Online]. Available: https://doi.org/10.1561/1500000006.

[10] J. Ha, M. Kambe, and J. Pe, Data Mining: Concepts and Techniques. 2011. [Online]. Available: https://doi.org/10.1016/C2009-0-61819-5.

[11] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in Proc. 15th Conf. of the European Chapter of the Association for Computational Linguistics (EACL), vol. 2, 2017, pp. 427–431. [Online]. Available: https://doi.org/10.18653/v1/e17-2068.

[12] S. Xu, "Bayesian Naïve Bayes classifiers to text classification," Journal of Information Science, vol. 44, no. 1, pp. 48–59, 2018. [Online]. Available: https://doi.org/10.1177/0165551516677946