Journal of Information Technology and Computer Science (INTECOMS)
Volume 7 Nomor 5, Tahun 2024
e-ISSN : 2614-1574
p-ISSN : 2621-3249

KLASIFIKASI SPAM EMAIL BERBASIS SEMANTIK MENGGUNAKAN METODE BERT

SEMANTIC-BASED EMAIL SPAM CLASSIFICATION USING BERT METHOD

Yunita Renta Hutagaol[1], Yulyani Arifin[2]
Del Institute of Technology[1], Universitas Bina Nusantara[2]
yunita.hutagaol@binus.ac.id

ABSTRACT

The development of technology encourages many people around the world, including in Indonesia, to be able to take advantage of the sophistication of this technology. One of these technologies is the internet and gadgets. The rapid development of smartphones has not changed the function of one of its service providers, namely a text messaging service called email. Email is currently still used to send messages to users who already know each other or to people who do not know each other, with various purposes including offering products or services. This is a problem for classifying incoming emails as spam or non-spam (ham). Email classification in this study uses the BERT and Long Short-Term Memory (LSTM) algorithms. The purpose of this study is to evaluate and determine the most effective algorithm for categorizing spam emails and also to find out whether emails received are spam or not spam. The results of the study show that the XL Net algorithm has a higher accuracy compared to the Bert Algorithm, Roberta Algorithm, and LSTM algorithm, with a value of 1.00. The precision, recall, f1-score, and accuracy values of the Bert algorithm also have the best performance compared to the LSTM algorithm.

Keywords: *Email Spam, Bert Algorithm, Roberta Algorithm, XL Net, LSTM*

ABSTRAK

Perkembangan teknologi mendorong banyak orang di seluruh dunia, termasuk di Indonesia, untuk dapat memanfaatkan kecanggihan teknologi tersebut. Salah satu teknologi tersebut adalah internet dan gadget. Perkembangan smartphone yang begitu pesat ternyata tidak mengubah fungsi dari salah satu penyedia layanannya, yaitu layanan pesan teks yang disebut email. Email saat ini masih digunakan untuk mengirimkan pesan kepada pengguna yang sudah saling mengenal maupun kepada orang yang belum saling mengenal, dengan berbagai tujuan termasuk untuk menawarkan produk atau jasa. Hal ini menjadi masalah untuk mengklasifikasikan Email yang masuk sebagai Email spam atau bukan spam (ham). Klasifikasi email pada penelitian ini menggunakan algoritma BERT dan Long Short-Term Memory (LSTM). Tujuan penelitian ini adalah untuk mengevaluasi dan menentukan algoritma yang paling efektif untuk mengkategorikan Email spam dan juga untuk mengetahui Email yang diterima sebagai spam atau bukan spam. Hasil penelitian menunjukkan bahwa algoritma XL Net memiliki akurasi yang lebih tinggi dibandingkan dengan Algoritma Bert, Algoritma Roberta, dan algoritma LSTM, dengan nilai 1.00. Nilai precision, recall, f1-score, dan akurasi dari algoritma Bert juga memiliki performa yang paling baik dibandingkan dengan algoritma LSTM.

Kata Kunci*: Spam Email, Bert Algorithm, Roberta Algorithm, XL Net, LSTM*

PENDAHULUAN

Email adalah metode komunikasi populer yang digunakan oleh dua pihak. Email merupakan sarana komunikasi yang penting bagi orang-orang untuk berbagi data dan informasi. Hasilnya, komunikasi dalam bisnis menjadi lebih produktif dan efektif, memberikan kemudahan dan menyederhanakan akses dan replikasi untuk pemeriksaan. Komunikasi email telah menjadi semakin populer melalui penggunaan komputer, laptop, dan ponsel dalam beberapa tahun terakhir. Email adalah bentuk komunikasi yang menggabungkan kemampuan beradaptasi dan transmisi informasi yang cepat (Srinivasan dkk., 2021). V. A. Shiva Ayyadurai membuat perangkat lunak email pada tahun 1978 saat berusia 14 tahun, dengan menggabungkan elemen-

elemen dari semua aplikasi perangkat lunak email yang ada sekarang. Kotak Masuk, Memo (Kepada, Dari, Tanggal, Perihal, Cc, Bcc), Kotak Keluar, Buku Alamat, Sampah, Folder, Lampiran, dan lainnya. Dengan tujuan untuk menggantikan sistem pos pneumatika di fasilitas medis kecil, yang digunakan untuk mengangkut surat di antara karyawan kantor, dengan email. Dalam perspektif yang baru, Shiva Ayyadurai menjelaskan bagaimana para ahli salah memprediksi lenyapnya email sejak awal dengan mencampuradukkannya dengan bentuk media lain seperti chatting, papan buletin online, SMS, pesan instan, dan blog. Namun, jika kita merefleksikan asal-usul email, sistem surat internal yang mendorong komunikasi bisnis, jelas bahwa email akan tetap menjadi alat penting untuk berbagai ukuran bisnis di masa mendatang. Singkatan "Email" berasal dari Electronic Mail. Awalnya diciptakan pada tahun 1960-an dan 1970-an untuk memungkinkan para peneliti berkomunikasi jarak jauh dan dengan cepat mendapatkan popularitas pada tahun 1990-an ketika bisnis dan individu dengan mulai mengenal Internet (Altulaihan dkk., 2023).

Email adalah sarana komunikasi yang penting di era digital. Email memiliki banyak keuntungan, seperti komunikasi yang efisien, kenyamanan, dan aksesibilitas. Meskipun email nyaman, efektif, dan berdampak, email tidak dapat diandalkan sebagai pengganti komunikasi tatap muka ketika berhadapan dengan situasi seperti menyampaikan berita negatif atau membicarakan topik yang sensitif. Selain itu, etiket email yang tepat dan profesionalisme untuk interaksi harus dijaga. Namun demikian, para pelaku scam sering menggunakan email untuk melakukan aktivitas penipuan atau bekerja sama dengan rekan-rekannya dalam kejahatan. Contoh kejahatan dunia maya yang menggunakan email dapat berupa spoofing, phishing, dan penawaran palsu. Keamanan ada jika tidak ada ancaman, bahaya, ketakutan, atau kecemasan. Keamanan email melibatkan pengamanan akun email dan pesan dari akses yang tidak sah, kehilangan, atau kompromi (Baafi, 2022).

Bisnis dapat meningkatkan keamanan email dengan menerapkan kebijakan dan memanfaatkan perangkat untuk melindungi dari ancaman berbahaya seperti malware, spam, dan upaya phishing. Metode ini biasanya digunakan untuk meretas jaringan perusahaan dan mencuri informasi rahasia. Oleh karena itu, sangat penting untuk mengamankan akun email, data, dan komunikasi dari akses yang melanggar hukum, kehilangan data, dan potensi ancaman lainnya (Karim dkk., 2020).

Pertukaran email berisi berbagai bukti seperti informasi pengirim, asal, lampiran data, ID pesan, stempel waktu, jenis pesan, dan pesan yang dibagikan. Oleh karena itu, para ahli forensik komputer menggunakan alat dan metode yang sesuai untuk memeriksa dan mengambil bukti dari email. Pemeriksa atau penyelidik penipuan harus mengumpulkan semua bukti yang relevan, yang mungkin termasuk data elektronik atau digital. Selain bukti fisik, penyelidik harus memiliki kemampuan untuk mengakses dan menganalisis data digital sebagai bukti. Satu email memiliki potensi untuk menyertakan banyak email dan kontak. Oleh karena itu, sangat penting untuk memanfaatkan alat forensik digital untuk memeriksa konten email (Atlam & Oluwatimilehin, 2023).

Kecerdasan Buatan (AI), dikombinasikan dengan Internet of Things (IoT) yang disebut sebagai AIoT, dapat membuat keputusan cerdas dengan analisis mandiri. Karena penggunaannya yang luas dalam berbagai situasi, perangkat IoT menghasilkan volume data yang besar yang dimanipulasi oleh peretas untuk mengganggu operasi dan layanan reguler. Oleh karena itu, klasifikasi data secara proaktif diperlukan untuk menghentikan kejahatan siber. Sangat penting untuk menganalisis header dan isi email sebagai

maksud dari komunikasi yang dapat membantu menentukan sumber bukti potensial. Karena meningkatnya volume data yang dibagikan melalui email, para penyelidik kini berhadapan dengan tugas yang sulit untuk mengekstrak informasi penting dari pertukaran email yang ekstensif, yang menyebabkan penundaan dalam prosedur penyelidikan. Hal ini memberikan keuntungan bagi para penjahat untuk menghilangkan bukti perbuatan jahat mereka.

Penelitian terbaru telah menemukan berbagai karakteristik dalam header email yang dapat digunakan untuk mengategorikan spam. Penelitian ini menunjukkan berbagai karakteristik yang dapat digunakan dalam sistem deteksi untuk mendapatkan hasil yang efisien dalam mencegah email spam. Webmail populer yang memanfaatkan fitur yang tersedia di header masing-masing adalah Yahoo mail, Gmail dan Hotmail. Setiap perubahan yang terjadi pada fitur-fitur ini dapat diasumsikan sebagai perilaku spam. Ilmu forensik digital mengandalkan teknik ilmiah yang telah terbukti untuk mengumpulkan dan menganalisis data atau jejak digital (Hina dkk., 2021).

Memanfaatkan metode forensik digital untuk mengumpulkan dan memeriksa data email menghadirkan aspek baru dalam memerangi spam. Menambahkan konsep kesiapan forensik digital pada email akan membuat pengumpulan jejak digital menjadi lebih mudah. Teknik forensik digital dapat memverifikasi data dalam header jejak email. Menambahkan kesiapan forensik digital dilakukan pada bagian “Receive:” yang terdapat pada jejak header SMTP. Melalui integrasi forensik digital, data jejak header dapat melayani fungsi lain seperti mengembangkan alat anti-spam atau melacak sumber spam. Kesiapan forensik digital diintegrasikan ke dalam sampul email, memastikan bahwa konten email tetap tidak terpengaruh. Sehingga, kontennya tidak berubah. Dengan begitu, pada akhirnya dapat disimpulkan bahwa meningkatkan kesiapan forensik digital dapat meningkatkan integritas jejak header SMTP, yang pada akhirnya dapat meningkatkan kepercayaan pengguna (Hina dkk., 2021).

Beberapa penelitian terkait klasifikasi spam email menjadi perhatian besar bagi para peneliti, hal ini mengingat besarnya kerugian dalam penipuan spam email ini. Selain itu, dalam perkembangan penelitian spam email, deep learning dianggap sebagai pendekatan yang cenderung sangat baik dalam mengklasifikasikan spam email. Penelitian sebelumnya telah mengungkapkan bahwa deep learning mengungguli pendekatan machine learning klasik, dan bahkan para pelaku scam telah mengembangkan metodenya untuk menembus sistem yang dibangun dengan menggunakan machine learning klasik. Hal ini menunjukkan bahwa ada kekurangan dalam arsitektur pembelajaran mesin ini. Oleh karena itu, pada penelitian ini kami akan merancang dan meneliti spam email dengan menggunakan pendekatan deep learning (Hina dkk., 2021; Pan dkk., 2022).

Penelitian sebelumnya telah mengembangkan deep learning untuk masalah spam email dan banyak metode/arsitektur yang telah diusulkan untuk pengembangan penelitian lebih lanjut. Email Spam memiliki banyak bentuk pesan, selain itu juga memiliki berbagai macam domain topik untuk email spam itu sendiri. Venkatraman. dkk. Mengungkapkan bahwa spammer modern dapat melewati sistem yang dirancang dengan menggunakan metode machine learning yang didasarkan pada pengetahuan konseptual. Oleh karena itu, perlu dilakukan pendekatan berbasis semantik untuk menyediakan model yang dapat menganalisis isi email. Penelitian berbasis semantik telah banyak dilakukan. Pan dkk. telah melakukan penelitian tentang email spam berbasis semantik dengan menggunakan pendekatan deep learning, penelitian ini membawa masalah klasifikasi ini ke dalam sebuah grafik

klasifikasi, artinya penelitian ini mengembangkan penelitian deep learning berbasis grafik. Grafik yang dirancang cenderung sangat baik dalam memeriksa konteks kalimat. Metode yang diusulkan tidak membutuhkan proses penanaman kata seperti yang biasanya dijelaskan, tetapi grafik itu sendiri yang akan membangun hubungan antar kata (Pan dkk., 2022). Hina, Maryam. dkk. Telah melakukan penelitian berbasis semantik dalam analisis forensik untuk masalah klasifikasi email. Penelitian ini menggabungkan LSTM dan GRU untuk menginterpretasikan email. Penelitian ini menggabungkan subjek email beserta isi email untuk kebutuhan fitur. Penelitian ini juga mengembangkan sistem pendeteksi spam yang dapat diadaptasi untuk email teks pendek. Dimana, model gabungan ini dapat mengidentifikasi email dengan panjang 1000+ karakter.

Saidani dkk., (2020) telah berhasil melakukan penelitian untuk klasifikasi spam email dengan berfokus pada ekstraksi fitur yang kuat. Penelitian ini menyatakan bahwa memisahkan domain email memungkinkan model untuk menganalisis semantik email dengan lebih baik. Oleh karena itu, penelitian ini memisahkan domain topik dari email. Penelitian ini merancang sebuah sistem dengan menerapkan 3 metode ekstraksi fitur untuk memberikan hasil fitur yang kuat ketika dilatih oleh model. Penelitian sebelumnya menggunakan arsitektur Word2Vec + eTVSM + CN2-SD. Srinivasan, Sriram. dkk. Telah melakukan penelitian tentang klasifikasi spam email berbasis semantik. Penelitian ini menyatakan bahwa representasi kata/ representasi vektor/ penyematan kata pada domain natural language processing (NLP) sangat sulit untuk dipilih, karena sifatnya yang sangat bergantung pada lingkungan dan ketersediaan data. Oleh karena itu, penelitian ini berfokus pada analisis mendalam tentang rekayasa fitur. Penelitian ini juga mengungkapkan bahwa spam email saat ini sangat mirip dengan email asli. Oleh karena itu, analisis rekayasa fitur perlu dilakukan. Penelitian ini mempertimbangkan metode representasi kata seperti Bag-of-Words (BOW), Vector Space Model (VSM), Term Document Matrix (TDM), Term Frequency-Inverse Document Frequency (TF-IDF), Latent Semantic Analysis (LSA), Word2Vec, FastText, Keras Word Embedding, dan Neural Bag-of-Words (NBOW). Untuk model klasifikasi, penelitian ini mempertimbangkan metode klasik dari machine learning dan Deep Learning. Penelitian ini mengungkapkan bahwa, Word2vec + CNN-LSTM mengungguli semua arsitektur lainnya. Hal ini semakin membuktikan bahwa pendekatan deep learning lebih baik sebagai solusi untuk klasifikasi spam email.

Spammer juga memiliki cara untuk menghindari sistem deteksi spam email. Venkatraman. dkk. Mengungkapkan bahwa kata-kata polisemi dan ambigu dapat mengelabui sistem pendeteksi spam email yang dibangun dengan menggunakan machine learning biasa. Oleh karena itu, penelitian ini merancang sebuah pendekatan machine learning yang menggunakan Conceptual Semantic + Semantic Similarity + Combination Conceptual Semantic sebagai solusi dari permasalahan spam email. Arsitektur yang diusulkan dapat menangani masalah polisemi kata dan ambiguitas kata. Hal ini merupakan pergerakan yang cukup besar dalam rekayasa fitur. Penelitian ini juga secara tidak langsung menyatakan bahwa ekstraksi fitur untuk email dan spam email benar-benar sangat penting dalam mendeteksi spam email (AbdulNabi & Yaseen, 2021).

AbdulNabi dkk. telah membuka jalan untuk mempertimbangkan pendekatan Transformer yang sudah terlatih untuk solusi masalah spam email ini. Penelitian ini berfokus pada penyematan kata yang efektif untuk masalah ini. Penelitian ini mengusulkan 2 model arsitektur, yaitu BERTBase dan Keras Embedding + Bi-LSTM. Penelitian ini berhasil membangun

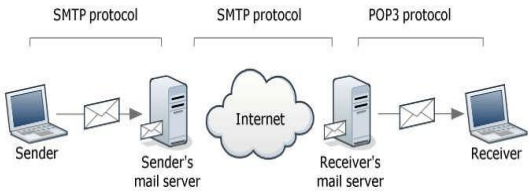
kedua arsitektur tersebut. Pendekatan BERTBase mengungguli arsitektur model lainnya. Namun, penelitian ini hanya terbatas pada 1 bentuk BERT itu sendiri atau 1 bentuk yang sudah terlatih. Padahal, saat ini ada banyak model lain yang sudah terlatih.

Pada penelitian sebelumnya, kami melihat bahwa masalah spam email ini perlu difokuskan pada basis semantik. Karena, pada penelitian sebelumnya, kami melihat bahwa ada banyak bentuk pesan email spam, sehingga dalam beberapa kasus, spammer dapat menghindari sistem pendeteksian spam. Oleh karena itu, penelitian ini akan berfokus pada basis semantik untuk masalah klasifikasi email. Pada penelitian sebelumnya, telah banyak penelitian spam email berbasis semantik, dan beberapa penelitian berfokus pada rekayasa fitur. Abdul Nabi. dkk. Mengungkapkan bahwa penggunaan word embedding berbasis Transformer efektif dalam memberikan representasi semantik yang dapat dipelajari oleh model. Namun, pada penelitian sebelumnya, mereka tidak melakukan analisis mendalam terhadap jenis transformer terlatih lainnya, seperti BERTLarge, AlBERT, RoBERTa, XLNet, dan lainnya, yang mungkin memiliki performa yang lebih baik dibandingkan metode lainnya (Ganesan dkk., 2021). Oleh karena itu, dalam penelitian ini, kami akan mengatasi kekurangan tersebut dengan mempertimbangkan berbagai model transformator yang telah dilatih sebelumnya. Model tersebut tidak hanya bertindak sebagai pengklasifikasi. Selain itu, analisis mendalam terhadap model transformer terlatih akan dilakukan sebagai metode untuk ekstraksi fitur dengan menambahkan algoritma deep learning lainnya. Kami juga akan mempertimbangkan untuk menggunakan embedding kata yang tidak berasal dari transformer sebagai pembanding.

TINJAUAN PUSTAKA

Arsitektur Email

Sistem email pada umumnya terdiri dari 2 (dua) komponen inti dan protokol komunikasi yang memungkinkan terjadinya pertukaran pesan elektronik di antara keduanya. Elemen awal adalah email client, berfungsi sebagai antarmuka untuk menerima, membaca, menulis, dan mengirim email. Elemen berikutnya adalah server email atau SMTP server atau Message Transfer Agent (MTA), yang berfungsi sebagai pengirim pesan dari sumber ke tujuan. Protokol sistem email meliputi SMTP untuk mengirim pesan dari server email sumber ke server email tujuan, dan POP3 atau IMAP untuk mengambil pesan dari server email tujuan ke klien email (Altulaihan et al., 2023).
Interaksi antar komponen dan protokol dalam sistem email dapat dilihat seperti pada Gambar 1.



Gambar 1. Protokol dan Interaksi Komponen dalam Sistem E-mail.

Deep Learning

Model DL menjadi populer di kalangan spesialis NLP karena kemampuannya dalam mengatasi masalah yang menantang. DL berakar pada pelatihan jaringan saraf dalam yang dipengaruhi oleh cara kerja otak dan bergantung pada data dalam jumlah besar. Mereka memiliki kemampuan untuk mengatasi masalah skalabilitas dan secara otomatis mengambil karakteristik data. Para peneliti NLP paling menyukai model-model seperti CNN dan jaringan LSTM dalam penelitian DL mereka. CNN adalah salah satu yang paling signifikan dan merupakan algoritma DL yang paling banyak digunakan sebagai solusi untuk masalah NLP. CNN telah digunakan secara efektif dalam analisis sentimen, gambar, klasifikasi teks, pengenalan pola,

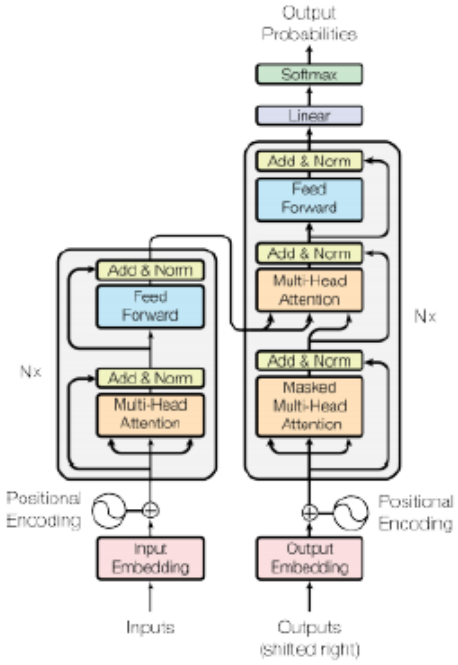
dan berbagai aplikasi lainnya (Iqbal dkk., 2023).

Pada tahun 2023, Elijah John-Africa dan Victor T. Emmah menggunakan unit berulang dan LSTM untuk mengambil data berurutan dari teks, dan menyimpulkan bahwa LSTM dan unit berulang adalah yang paling efektif untuk tugas-tugas tekstual. Selain data teks, RNN/LSTM juga unggul dalam menganalisis data deret waktu berurutan. Jaringan semacam ini banyak digunakan dalam berbagai aplikasi NLP (Sharmeen dkk., 2020).

Souad Larabi-Marie-Sainte dkk. meneliti 34.519 catatan dalam korpus Enron dengan jaringan LSTM dan model GRU untuk menunjukkan spam. Pada penelitian sebelumnya yang terinspirasi oleh LSTM dan dari GRU, forensik email yang detail mencapai akurasi 96%. Jika mereka bisa mendapatkan skor F1 sekitar 92,8, mereka berpotensi melampaui metode kategorisasi tradisional. Penelitian tambahan telah menggunakan algoritma DL untuk menganalisis teks, memungkinkan ekstraksi detail kontekstual untuk mendeteksi spam.

Representasi Encoder Dua Arah dari Transformer (BERT)

BERT merupakan model representasi bahasa yang sudah terlatih yang pertama kali dipopulerkan oleh Google pada tanggal 11 Oktober 2018. Tidak seperti model bahasa lainnya, BERT diciptakan sebagai pre-trained model atau model yang telah dilatih dua arah dari data teks yang tidak berlabel dengan menggabungkan konteks dari sisi kiri dan kanan layer. Dengan cara ini, model BERT dapat disempurnakan dengan penambahan satu lapisan saja (Devlin dkk., 2019)



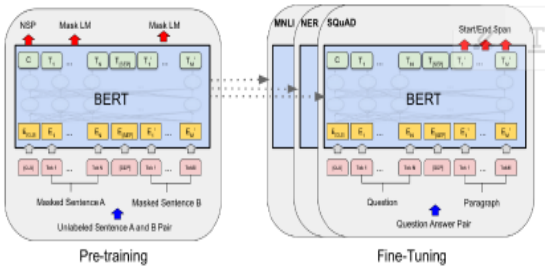
Gambar 2. Arsitektur Transformers
Sumber: Vaswani et al., (2017)

Desain BERT mencakup enkoder Transformer dua arah dengan beberapa lapisan. Transformer adalah mekanisme untuk memperhatikan yang mempelajari bagaimana kata atau sub-kata dalam teks terkait dalam konteks. (Vaswani dkk., 2017) Transformer menggunakan mekanisme self-attention untuk memahami representasi input dan output. Transformator memiliki 2 mekanisme yang terpisah, yaitu encoder untuk membaca teks input dan decoder yang menentukan prediksi. Setiap lapisan encoder berisi dua komponen, mekanisme multi head self-attention dan jaringan feed forward yang terhubung penuh. Setiap input akan melalui lapisan self-attention di dalam encoder. Tingkat ini memungkinkan encoder untuk mengamati kata-kata tambahan di dalam kalimat. Output dari lapisan attention kemudian digunakan sebagai input dari jaringan syaraf feed forward. BERT menggunakan penyematan WordPiece dengan 30.000 token dalam kosakata. Tujuan awal dari penyematan WordPiece adalah untuk mengatasi masalah segmentasi dalam bahasa Jepang

dan Korea dalam sistem pengenalan suara Google. (Gupta, 2024) Segmentasi akan menyebabkan banyak kata yang tidak termasuk dalam kosakata. WordPiece dirancang untuk secara otomatis memahami unit kata dari kumpulan data yang luas tanpa menyebabkan masalah Out of Vocabulary (OOV). Kata-kata yang sering digunakan akan tetap disimpan sebagai satu kata yang utuh, sedangkan kata-kata yang jarang digunakan akan dipisahkan menjadi sub-kata dan bahkan karakter.

Gambar 3. Representasi Masukan BERT
Sumber: Devlin et al., (2019)

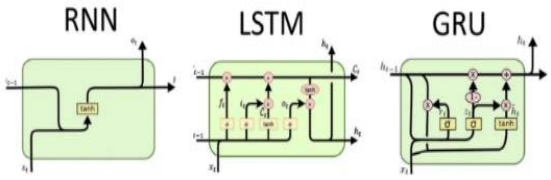
Token pertama dari setiap urutan adalah token klasifikasi kelas khusus ([CLS]). Untuk membedakan setiap kalimat, sebuah token terpisah ([SEP]) ditambahkan di bagian akhir. Kemudian penyisipan segmen disertakan untuk setiap token untuk membedakan antara kata-kata dari kalimat A dan kata-kata dari kalimat B. Selanjutnya, penyisipan posisi ditambahkan untuk menandai token dalam kalimat. Input untuk encoder BERT adalah jumlah dari Token Embedding, Segment Embedding, dan Positional Embedding. Representasi input BERT ditunjukkan pada Gambar 3. Dua tahap dalam kerangka kerja BERT adalah pretraining dan fine-tuning. Tahap pre-training adalah tahap ketika model mempelajari bahasa dan konteks dengan melakukan Mask Language Modeling (MLM) dan Next Sentence Prediction (NSP) secara bersamaan. MLM memungkinkan representasi untuk menggabungkan konteks kiri dan konteks kanan, yang memungkinkan kita untuk melatih transformator dua arah yang dalam dan NSP yang menggabungkan representasi pasangan teks yang telah dilatih sebelumnya.



Gambar 4. Ilustrasi Tahapan Pra-pelatihan dan Penyempurnaan
Sumber: Fajri et al. (2022).

LSTM, dan GRU

Gated recurrent units (GRU) adalah salah satu bentuk recurrent neural network (RNN), sebuah jaringan saraf tiruan yang memungkinkan dinamika temporal dalam sekuens dengan membentuk koneksi antar node.



Gambar 5. RNN, LSTM dan GRU Sequence Model.

Dalam hal melupakan gerbang untuk memori jangka pendek, GRU mirip dengan LSTM, tetapi lebih ringkas dalam parameter karena tidak memiliki gerbang output. Ditemukan bahwa GRU memiliki kinerja yang sebanding dengan LSTM dalam tugas-tugas seperti pemodelan musik polifonik, pemodelan sinyal ucapan, dan pemrosesan bahasa alami. GRU telah menunjukkan kinerja yang lebih baik pada dataset yang lebih kecil dan jarang.

Model yang akan kami gunakan meliputi Embedding Layer, Dropout layer untuk mengatasi overfitting, GRU layer, dan output layer, yang direpresentasikan dalam diagram berikut.

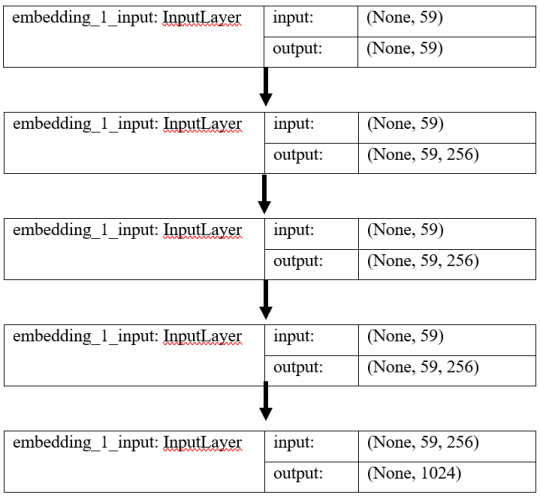


Figure 6. Embedding Layer in GRU.

Confusion Matrix

Menilai kualitas klasifikasi dapat dilakukan dengan menggunakan confusion matrix. Confusion Matrix mengevaluasi seberapa baik hasil klasifikasi sistem sesuai dengan hasil klasifikasi data yang sebenarnya dengan mengukur tingkat akurasi, presisi, dan recall. Akurasi mengukur proporsi prediksi yang akurat dibandingkan dengan keseluruhan dataset yang dinilai. Presisi mengukur seberapa akurat jawaban atau hasil sistem dibandingkan dengan data yang diminta. Sementara itu, recall mengukur proporsi data yang diklasifikasikan dengan benar dalam kelas tertentu terhadap jumlah total data yang seharusnya termasuk dalam kelas tersebut. Ilustrasi berikut ini menunjukkan confusion matrix untuk klasifikasi multi-kelas (Riehl dkk., 2023).

Tabel 1. Contoh Confusion Matrix untuk multi-klasifikasi

Kelas		**Kelas prediksi**		
		A	B	C
Klasifikasi actual	A	TN	FP	TN
	B	FN	TP	FN
	C	TN	FP	TN

Sebagai contoh, prediksi yang benar untuk kelas B adalah kondisi True Positive (TP), sedangkan prediksi yang salah untuk kelas B adalah kondisi False Negative (FN). Kondisi False Positive (FP) terjadi ketika kelas lain diprediksi sebagai kelas B, dan True Negative (TN) jika kelas lain tidak diprediksi sebagai kelas B [5]. Rumus untuk menghitung akurasi, presisi, dan recall untuk kelas *k* disajikan di bawah ini:

Akurasi: Persentase dari semua prediksi yang benar.

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Presisi: Persentase prediksi observasi positif dari semua observasi yang diprediksi positif.

$$Presisi = \frac{TP}{TP + FP} \quad (2)$$

Recall (Sensitivitas): Persentase pengamatan positif yang diprediksi dengan benar dari semua pengamatan positif yang sebenarnya.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Selain akurasi, presisi, dan recall, pada tahap evaluasi, F1-Score juga dihitung untuk menentukan kinerja model dengan menggabungkan nilai presisi dan recall. Nilai 1 menandakan kinerja prediksi yang efektif dari model, sedangkan nilai 0 menunjukkan kinerja yang tidak memadai. Dalam klasifikasi biner, F-1 Score dirumuskan sebagai berikut.

$$F1 - Score = 2 * \frac{Presisi * Recall}{Presisi + Recall} \quad (4)$$

Perhitungan F1-Score harus mencakup seluruh kelas dengan menghitung Makro F1-Score dan Mikro F1-Score (Fadlila Nurwanda dkk., 2023).

METODE

Dataset

Pada tahap Pembahasan, dilakukan pemahaman terhadap subjek penelitian. Pemahaman terhadap objek penelitian

dilakukan dengan menggali informasi melalui beberapa akun email dan melihat folder spam dan inbox. Motivasi pada tahap ini adalah subjek email berbahasa Inggris pada folder spam dan inbox. Pada tahap ini, pemahaman diperlukan untuk mengidentifikasi metode klasifikasi yang optimal, yang akan membantu dalam pengolahan data dengan membandingkan hasil algoritma dan meningkatkan efisiensi metode klasifikasi. Pada tahap Dataset Retrieval, dilakukan proses pengambilan data mentah (subjek email) sesuai dengan atribut yang dibutuhkan. Data diperoleh dari beberapa akun email. Data primer yang didapatkan adalah setelah dilakukan proses cleaning, dataset berjumlah 5157 data, dimana 87% merupakan spam, dan 13% merupakan spam. Semua data tersebut disimpan menjadi satu, baik itu spam maupun non spam dan disimpan dalam bentuk dokumen teks berekstensi (.csv) (John-Africa & Emmah, 2022).

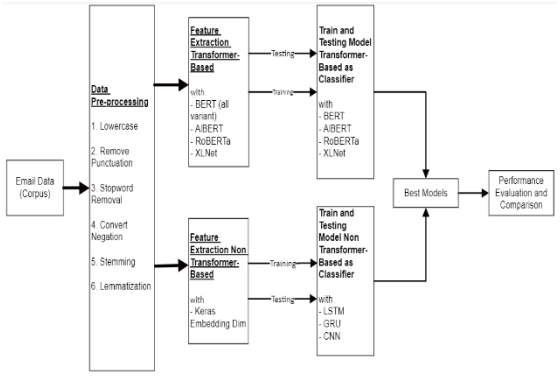
Tabel 2. Contoh Dataset spam and ham emails.

Kategori	Pesan
ham	"Go until jurong point, crazy... Available only in bugis n great world la e buffet... Cine there got a more wat..."
ham	Ok lar... Joking wif u oni...
spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question (std txt rate) T&C's apply 08452810075over18's
ham	U dun say so early hor... U c already then say...
ham	"Nah I don't think he goes to usf, he lives around here though"
spam	"FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to send, Â£1.50 to rcv"
ham	Even my brother is not like to speak with me. They treat me like aids patent.
ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
Spam	WINNER!! As a valued network customer you have been selected to receivea Â£900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
Spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030

Pengembangan Model Klasifikasi Email

Email yang digunakan dalam penelitian dibagi menjadi Ham dan Spam, dimana Ham untuk non-spam dan spam sendiri adalah email yang tidak relevan dan cenderung merusak yang mirip dengan email normal. Arsitektur alur kerja penelitian untuk mengklasifikasikan email ke dalam beberapa kelas ditunjukkan pada Gambar 7 menyiapkan data dan teks, mengekstraksi fitur, melakukan tuning parameter, dan melakukan klasifikasi dengan model deep learning yang diusulkan. Bagian email terdiri dari subjek-email dan isi-email, dalam penelitian ini kami menggunakan keduanya sebagai fitur. Kami memasukkan bagian dari set validasi ke dalam proses pelatihan model. Pada akhirnya, sebuah set pengujian digunakan untuk memvalidasi model kinerja. Bahasa Python digunakan untuk implementasi dan eksperimen di lingkungan Google Colab.

Identifikasi dan kategorisasi data dalam email. Algoritma menguraikan proses metode yang disarankan untuk menyortir email. Kami akan melakukan analisis mendalam terhadap model-model pretrained transformer (Seperti. BERTLarge, Bert-Base-Uncased, AlBERT, RoBERTa, dan XLNet), sebagai kata sisipan + pengklasifikasi dan juga kami akan mempertimbangkan pretrained transformer ini hanya sebagai metode representasi teks, sehingga akan terlihat bagaimana model-model ini dapat memberikan semantik pada model-model deep learning lainnya, di sini kami memilih LSTM, CNN, RNN, GRU, dan ANN. Selain itu, kami juga mempertimbangkan analisis mendalam tentang metode representasi teks transformator yang tidak terlatih seperti Word2Vec, GloVe, dan FastText.



Gambar 7. Model yang diusulkan untuk mendeteksi dan mengklasifikasikan email.

Persiapan Data

Tahap awal persiapan data melibatkan penggunaan langkah-langkah berbasis bahasa alami yang menstandarkan teks dan mempersiapkannya untuk analisis. Tahap ini terdiri dari beberapa tahap, seperti yang diuraikan di bawah ini.

1) Tokenisasi. Memecah dokumen menjadi kata-kata mengikuti aturan yang telah ditetapkan. Proses tokenisasi dijalankan dalam bahasa pemrograman Python dengan menggunakan pustaka SpaCy.
2) Penyaringan. Kata-kata yang umum digunakan seperti “a” dan “the” tidak menambah nilai pada E-mail dan menciptakan gangguan yang tidak perlu pada data teks. Kata-kata ini, yang dikenal sebagai stop words, dapat dikecualikan dari teks selama pemrosesan. Kami menggunakan fitur "NLTK" dari Python Library untuk menghilangkan stop words dalam teks.
3) Penghapusan Tanda Baca. Tanda baca disertakan (misalnya, titik (.), koma (,), tanda kurung) untuk memisahkan kalimat dan memperjelas makna. Untuk penghapusan tanda baca, kami menggunakan pustaka ``NLTK".

Ekstraksi Fitur

Setelah melakukan preprocessing pada teks, proses ekstraksi fitur dilakukan. Dimana proses ini akan merepresentasikan teks ke dalam bentuk numerik, sehingga dapat dilakukan pelatihan untuk setiap algoritma yang diusulkan. (Sharmeen dkk., 2020).

a. Ekstraksi Fitur Berbasis Pre-Trained Transformer

Pada penelitian ini, kami akan melakukan analisis mendalam mengenai model pre-trained transformer sebagai metode ekstraksi ciri. Di sini kita akan melihat bagaimana model ini akan memberikan analisis berbasis semantik untuk pelatihan model. Model yang diusulkan adalah BERT, AlBERT, RoBERTa, dan XLNet (AbdulNabi & Yaseen, 2021).

b. Ekstraksi Fitur Berbasis Transformator yang Tidak Terlatih

Pada penelitian ini kami juga mempertimbangkan metode ekstraksi fitur lain sebagai pembanding. Hal ini dimaksudkan agar dapat memberikan analisis yang mendalam tentang rekayasa fitur. Di sini kami mengusulkan Keras, yaitu Embedding Dim.

Pelatihan Model Klasifikasi Email Spam

Pada tahap ini kami mengusulkan model klasifikasi, untuk melakukan analisis mendalam tentang hal ini.

a. Pelatihan Model Klasifikasi Berbasis Transformator Deep Learning.
   Setiap model berbasis transformator yang telah dilatih sebelumnya memiliki lapisan klasifikasi sendiri. Oleh karena itu, model ini dapat bertindak sebagai ekstraksi fitur dan pengklasifikasi (Brindha dkk., 2023)
b. Pelatihan Model Klasifikasi Berbasis Non-Trafo Deep Learning.
   Setelah ekstraksi fitur dilakukan dengan menggunakan model berbasis transformer dan non transformer. Analisis ekstraksi fitur perlu dilakukan untuk metode deep learning lainnya. Hal ini dikarenakan, kita perlu memberikan analisis mendalam terhadap model berbasis transformator yang telah dilatih sebelumnya sebagai ekstraksi fitur dan melakukan pelatihan menggunakan fitur yang diekstraksi

oleh model berbasis transformator. Pada penelitian ini pengklasifikasi yang digunakan adalah LSTM, GRU, CNN, RNN, dan ANN (Bagui dkk., 2021)

Metode dan Hasil Evaluasi

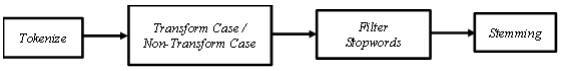
Pada tahap Iterasi dan Evaluasi, peneliti melakukan iterasi berulang melalui percobaan hyperparameter. Untuk setiap kombinasi hyperparameter, melatih model pada data pelatihan, mengevaluasi pada data validasi, dan mencatat metrik performa (akurasi, presisi, recall, dan F1-score). Para peneliti menggunakan alat visualisasi untuk melihat bagaimana kinerja berubah seiring dengan perubahan hyperparameter. Berbagai metrik digunakan untuk mengevaluasi seberapa baik kinerja pengklasifikasi, seperti akurasi, presisi, recall, dan F1-score. Pengukuran evaluasi ini dihitung dengan menggunakan metode confusion matrix

	Category	Message	Label
5382	ham	make squeezed bucks dad	0
955	spam	filthy stories girls waiting	1
5310	ham	yeah thinking	0
4311	spam	Someone know asked dating service contact cant…	1
5281	ham	Princess xcx	0
4360	ham	send contents page	0
5307	ham	leave	0
3098	ham	hungry like mofo	0
2967	ham	good baby	0
4782	ham	yup hey one day fri ask miwa jiayin take leave.	0

HASIL DAN PEMBAHASAN

Tahap Persiapan Data

Tahap persiapan data pada proses persiapan data bertujuan untuk mendapatkan data yang sudah dibersihkan dan siap untuk diteliti. Tahap preprocessing merupakan langkah awal dalam teks mining yang akan dilakukan. Berikut ini adalah tahap perancangan model preprocessing yang digunakan oleh peneliti dalam tahap awal pengolahan data:



Gambar 8. Desain Model Preprocessing

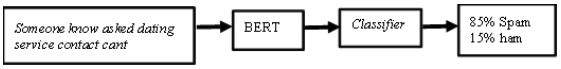
Gambar di atas menjelaskan alur proses preprocessing. Tahapan yang digunakan dalam proses preprocessing adalah, Tokenization: token linguistik, Transform Cases, Filtering stopwords (bahasa Inggris), Stemming bahasa Inggris (snowball) (Saidani dkk., 2020).

Pemodelan

Tahap ini terdiri dari pemanfaatan teknik data mining, khususnya pemilihan teknik dan pemilihan algoritma yang akan diterapkan. Sebelum menentukan algoritma yang akan digunakan, perlu dilakukan pelabelan data.

Tabel 3. Contoh data set yang mengandung email spam dan ham setelah dilakukan pelabelan.

Setelah melakukan pelabelan data, selanjutnya dilakukan proses pemodelan data dengan menentukan tool yang digunakan. Pada tahap pemodelan Bert penelitian ini menggunakan BertTokenizer, TFBertModel, dan BertConfig. Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan spam dan ham dengan menggunakan algoritma Bert-Base-Uncased.



Gambar 9. Langkah-langkah klasifikasi teks dengan model Bert

Dataset dibagi menjadi dua bagian, yaitu dataset yang digunakan untuk pelatihan dan dataset untuk validasi. Perbandingan antara dataset pelatihan dan dataset validasi menggunakan rasio 80:20.

Tabel 4. Perhitungan Kinerja Model

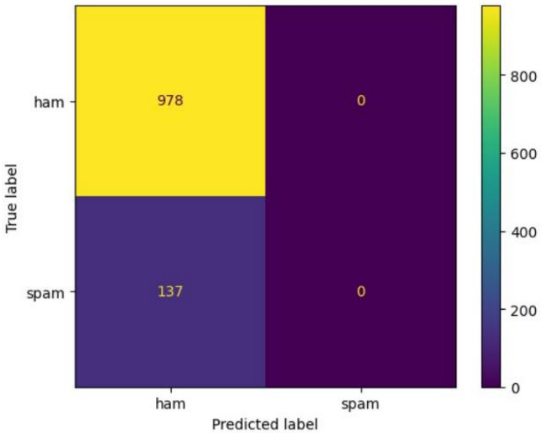
Model	Layer (Type)	Output Shape	Param #
	bert	multiple	109482240

Bert	(TFBertMainLayer)		
	dropout_37 (Dropout)	multiple	0 (unused)
	classifier (Dense)	multiple	1538
Roberta	roberta (TFRobertaMainLayer)	Multiple	124055040
	classifier (TFRobertaClassificationHead)	Multiple	592130
XL Net	transformer (TFXLNetMainLayer)	Multiple	116718336
	sequence_summary (TFSequenceSummary)	Multiple	590592
	logits_proj (Dense)	Multiple	1538
Albert	albert (TFAlbertMainLayer)	multiple	11683584
	dropout_4 (Dropout)	multiple	0 (unused)
	classifier (Dense)	multiple	1538
LSTM	embedding (Embedding)	(None, 50, 16)	8000
	spatial_dropout1d (Spatial Dropout1D)	(None, 50, 16)	0
	lstm (LSTM)	(None, 128)	74240
	dropout (Dropout)	(None, 128)	0
	dense (Dense)	(None, 1)	129

Evaluasi Model

Tabel 5. Evaluasi Model

Model	Class	precision	recall	f1-score	support
Bert	ham	0.98	1.00	0.99	968
	spam	0.98	0.89	0.94	147
	accuracy			0.98	1115
Roberta	ham	0.98	1.00	0.99	968
	spam	0.98	0.89	0.94	147
	accuracy			0.98	1115
Albert	ham	0.99	0.95	0.97	962
	spam	0.75	0.92	0.82	153
	accuracy			0.95	1115
XL Net	ham	1.00	1.00	1.00	968
	spam	0.98	0.99	0.99	147
	accuracy			1.00	1115
LSTM	ham	0.98	0.99	0.99	953
	spam	0.97	0.86	0.91	162
	accuracy			0.98	1115



Gambar 10. Hasil Matriks Confusion

KESIMPULAN

Penelitian ini telah berhasil menyelesaikan sebuah tantangan baru untuk mengatasi masalah yang disebabkan oleh email spam dan penipuan phishing, yang terus mempengaruhi pengguna email di seluruh dunia. Kami telah mempresentasikan sebuah solusi untuk memfilter pembuatan spam dengan memanfaatkan teknologi canggih seperti BERT, Albert, RoBERTa, dan XL Net, dan menyesuaikannya dengan tujuan pembelajaran tertentu seperti klasifikasi spam.

Penelitian ini mencakup tinjauan menyeluruh terhadap literatur, eksperimen, evaluasi yang dilengkapi dengan analisis model komparatif dan grafik garis dari kumpulan data, kami dapat menentukan keefektifan XL Net, yang dibandingkan secara langsung dengan metode konvensional seperti LSTM. Para pelaku spam selalu menyempurnakan strategi mereka; hal ini juga berlaku untuk sistem pendeteksi spam. Penelitian di masa lalu dapat meningkatkan dan memperkuat ketahanan model terhadap masalah.

Penelitian di masa depan dapat memperluas model kategorisasi untuk menangani email dalam berbagai bahasa, penelitian kami berkonsentrasi pada email berbahasa Inggris. Hal ini akan memerlukan penggunaan model seperti mBERT (multilingual BERT) atau melatih model multilingual BERT untuk

mengklasifikasikan email spam dalam berbagai bahasa dan konteks budaya.

REFERENCE

AbdulNabi, I., & Yaseen, Q. (2021). Spam email detection using deep learning techniques. *Procedia Computer Science*, *184*(October), 853–858. https://doi.org/10.1016/j.procs.2021.03.107

Altulaihan, E., Alismail, A., Hafizur Rahman, M. M., & Ibrahim, A. A. (2023). Email Security Issues, Tools, and Techniques Used in Investigation. *Sustainability (Switzerland)*, *15*(13). https://doi.org/10.3390/su151310612

Atlam, H. F., & Oluwatimilehin, O. (2023). Business Email Compromise Phishing Detection Based on Machine Learning: A Systematic Literature Review. *Electronics (Switzerland)*, *12*(1), 1–28. https://doi.org/10.3390/electronics12010042

Baafi, P. O. (2022). Tools For Cyber Forensics. *Advances in Multidisciplinary and Scientific Research Journal Publication*, *1*(1), 285–290. https://doi.org/10.22624/aims/crp-bk3-p46

Bagui, S., Nandi, D., Bagui, S., & White, R. J. (2021). Machine Learning and Deep Learning for Phishing Email Classification using One-Hot Encoding. *Journal of Computer Science*, *17*(7), 610–623. https://doi.org/10.3844/jcssp.2021.610.623

Brindha, R., Nandagopal, S., Azath, H., Sathana, V., Joshi, G. P., & Kim, S. W. (2023). Intelligent Deep Learning Based Cybersecurity Phishing Email Detection and Classification. *Computers, Materials and Continua*, *74*(3), 5901–5914. https://doi.org/10.32604/cmc.2023.030784

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*(Mlm), 4171–4186.

Fadlila Nurwanda, Winita Sulandari, Yuliana Susanti, & Zakya Reyhana. (2023). Comparative Analysis Of Performance Levels Of Svm And Naïve Bayes Algorithm For Lifestyle Classification On Twitter Social Media. *International Conference On Digital Advanced Tourism Management And Technology*, *1*(1 SE-Articles), 215–230. https://doi.org/10.56910/ictmt.v1i1.65

Fajri, F., Tutuko, B., & Sukemi, S. (2022). Membandingkan Nilai Akurasi BERT dan DistilBERT pada Dataset Twitter. *JUSIFO (Jurnal Sistem Informasi)*, *8*(2), 71–80. https://doi.org/10.19109/jusifo.v8i2.13885

Ganesan, A. V., Matero, M., Ravula, A. R., Vu, H., & Schwartz, H. A. (2021). Empirical Evaluation of Pre-trained Transformers for Human-Level NLP: The Role of Sample Size and Dimensionality. *NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, 4515–4532. https://doi.org/10.18653/v1/2021.naacl-main.357

Gupta, R. (2024). Bidirectional encoders to state-of-the-art: a review of BERT and its transformative impact on natural language processing. *Информатика. Экономика. Управление - Informatics. Economics. Management*, *3*(1), 0311–0320. https://doi.org/10.47813/2782-

5280-2024-3-1-0311-0320

Hina, M., Ali, M., Javed, A. R., Srivastava, G., Gadekallu, T. R., & Jalil, Z. (2021). Email Classification and Forensics Analysis using Machine Learning. *Proceedings - 2021 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of People, and Smart City Innovations, SmartWorld/ScalCom/UIC/ATC/IoP/ SCI 2021*, *July 2022*, 630–635. https://doi.org/10.1109/SWC50871.2021.00093

Iqbal, F., Javed, A. R., Jhaveri, R. H., Almadhor, A., & Farooq, U. (2023). Transfer Learning-based Forensic Analysis and Classification of E-Mail Content. *ACM Transactions on Asian and Low-Resource Language Information Processing*. https://doi.org/10.1145/3604592

John-Africa, E., & Emmah, V. T. (2022). Performance Evaluation of LSTM and RNN Models in the Detection of Email Spam Messages. *European Journal of Information Technologies and Computer Science*, *2*(6), 24–30. https://doi.org/10.24018/compute.2022.2.6.80

Karim, A., Azam, S., Shanmugam, B., & Kannoorpatti, K. (2020). Efficient Clustering of Emails into Spam and Ham: The Foundational Study of a Comprehensive Unsupervised Framework. *IEEE Access*, *8*, 154759–154788. https://doi.org/10.1109/ACCESS.2020.3017082

Pan, W., Li, J., Gao, L., Yue, L., Yang, Y., Deng, L., & Deng, C. (2022). Semantic Graph Neural Network: A Conversion from Spam Email Classification to Graph Classification. *Scientific Programming*, *2022*(ii). https://doi.org/10.1155/2022/6737080

Riehl, K., Neunteufel, M., & Hemberg, M. (2023). Hierarchical confusion matrix for classification performance evaluation. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *72*(5), 1394–1412. https://doi.org/10.1093/jrsssc/qlad057

Saidani, N., Adi, K., & Allili, M. S. (2020). A semantic-based classification approach for an enhanced spam detection. *Computers and Security*, *94*, 101716. https://doi.org/10.1016/j.cose.2020.101716

Sharmeen, S., Ahmed, Y. A., Huda, S., Kocer, B. S., & Hassan, M. M. (2020). Avoiding future digital extortion through robust protection against ransomware. *IEEE Access*, *8*, 24522–24534.

Srinivasan, S., Ravi, V., Alazab, M., Ketha, S., Al-Zoubi, A. M., & Kotti Padannayil, S. (2021). Spam Emails Detection Based on Distributed Word Embedding with Deep Learning. *Studies in Computational Intelligence*, *919*(December), 161–189. https://doi.org/10.1007/978-3-030-57024-8_7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 5999–6009.