**Sistem Deteksi Hoax Email Menggunakan**

**Metode Jarak Levenshtein**

Yoke Yie Chen , Suet-Peng Yong and Adzlan Ishak

Computer & Information Sciences Department

Universiti Teknologi PETRONAS

Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia

chenyokeyie@petronas.com.my, yongsuetpeng@petronas.com.my, [adzlan89@gmail.com](mailto:adzlan89@gmail.com)

**Abstrak**

Hoax adalah virus yang tidak berbahaya. Mereka hidup dengan menipu persepsi manusia dengan menyampaikan klaim palsu sebagai kebenaran. Sepanjang sejarah, hoax sebenarnya telah mampu mempengaruhi banyak orang hingga mencoreng citra dan kredibilitas korbannya. Apalagi informasi yang salah dan menyesatkan selalu menjadi distorsi bagi pertumbuhan manusia. Beberapa hoax dibuat sedemikian rupa sehingga mereka bahkan dapat memperoleh data pribadi dengan meyakinkan korban bahwa data tersebut diperlukan untuk tujuan resmi. Hoax berbeda dari spam dengan cara mereka menyamar melalui alamat orang-orang yang terkait baik secara langsung maupun tidak langsung dengan kami. Sebagian besar waktu, mereka muncul sebagai pesan yang diteruskan dan terkadang dari perusahaan yang sah. Makalah ini membahas masalah ini dengan mengembangkan sistem deteksi hoax dengan menggabungkan metode pencocokan teks menggunakan ukuran Levenshtein Distance. Model yang diusulkan digunakan untuk mengidentifikasi email hoax berbasis teks. Sensitivitas dan spesifisitas digunakan untuk mengevaluasi keakuratan sistem dalam mengidentifikasi email hoax.

1. **Pengantar**

Hoax email diketahui tidak berbahaya terhadap sistem yang dibuat. Hoax dianggap sebagai sampah yang dapat memberikan informasi menyesatkan kepada pengguna atau pembaca email. Menurut Blanzieri dan Bryl [1], spam adalah surat sampah yang sebagian besar tidak diinginkan atau tidak diminta yang dikirim baik secara langsung maupun tidak langsung oleh personel yang tidak memiliki hubungan saat ini dengan pengguna email. Hoax adalah email yang tidak diminta dan tidak diinginkan yang secara langsung atau tidak langsung dikirim oleh personel yang memiliki hubungan saat ini dengan pengguna email [2]. Singkatnya, hoax adalah spam versi 'pintar' yang menyamar dengan baik melalui personel yang hadir sebagai kontak seseorang. Biasanya hoax datang dalam bentuk penerusan pesan dari berbagai sumber. Informasi menyesatkan dari hoax dapat menyebabkan kerugian finansial dan menjengkelkan pengguna individu. Lebih buruk lagi, tipuan memiliki kemampuan untuk mengumpulkan informasi dan mungkin dapat meyakinkan penerima email untuk bertindak atas peristiwa yang tidak ada.

Beberapa email hoax telah menimbulkan banyak false alarm yang berdampak negatif bagi pihak-pihak tertentu. Misalnya, PayPal harus merusak dan mengambil tanggung jawab setelah penipuan Nigeria yang dianggap sebagai jenis tipuan diteruskan ke sebagian besar penerimanya oleh sumber yang tidak dikenal yang menyamar dengan menggunakan alamat email yang mirip dengan PayPal pada tahun 2003. Kepalsuan lainnya Kewaspadaan akan adanya hujan asam akibat ledakan reaktor nuklir di Jepang juga sempat menimbulkan kekacauan di Asia Tenggara dimana hoax ini diteruskan ke banyak orang setelah seminggu dari kejadian gempa di Jepang. Meskipun hoax ini belum terbukti sebagai penipuan, isinya menyesatkan [2-4].

Meskipun beberapa e-mail hoaks tidak berbahaya, namun seringkali meresahkan masyarakat. Lebih jauh, berdasarkan teori Meme [5], ketika pikiran seseorang terinfeksi oleh informasi yang salah, akan sulit bagi orang yang terinfeksi untuk diyakinkan sebaliknya tentang kebenaran yang sebenarnya. Masalahnya hanya dapat diselesaikan setelah pengguna terpapar pada fakta sebenarnya dari tipuan ini. Meskipun demikian, tidak akan mudah untuk meyakinkan seseorang tentang kebenaran yang hakiki kecuali jika sumber kebenaran yang dikemukakan itu sahih menurut pandangan orang tersebut [5]. Oleh karena itu, hoaks sangat penting untuk diidentifikasi terutama untuk menghindari kesalahpahaman lebih lanjut yang dapat membawa kerugian yang signifikan serta dampak negatif bagi pertumbuhan suatu partai. Berkaitan dengan hal tersebut, pengembangan sistem pendeteksi hoax secara otomatis bertujuan untuk melindungi penerima email dari penyesatan informasi palsu. Dalam literatur, belum banyak sistem pendeteksi hoax yang dibuat selama ini. Pekerjaan yang terkait dengan sistem deteksi email hoax dinyatakan dalam [2-4]. Sebagian besar pola deteksinya sama tetapi berbeda menurut definisi [1].

Karya-karya sebelumnya di bidang ini menderita kemampuan untuk memperbarui database hoax mereka untuk secara otomatis mengidentifikasi keberadaan hoax terbaru. Dalam makalah ini, kami mempresentasikan sistem deteksi hoax yang kami usulkan dengan menggunakan pendekatan pencocokan teks untuk menentukan tingkat kecurigaan hoax dari email yang masuk. Tujuan utama dari proyek ini adalah untuk mengembangkan Sistem Deteksi Hoax yang mampu mendeteksi email palsu baru dan lama yang disajikan di dunia saat ini. Perhatian utamanya adalah untuk meningkatkan kesadaran akan adanya email hoax dan memberikan pencerahan kepada pengguna tentang validitas dari setiap konten email yang telah diteruskan.

Pada sistem yang diusulkan, e-mail baru yang masuk akan melalui algoritma penyaringan berbasis statistik. Berdasarkan hasil statistik, email yang diidentifikasi sebagai hoax yang tidak ditemukan di database hoax akan disimpan sebagai hoax baru di database. Dengan kata lain, sistem yang kami usulkan mampu melakukan pembaruan otomatis saat mendeteksi hoax.

1. **Karya Terkait**
2. Sistem Deteksi Hoax vs. Filter Spam

Sistem pendeteksi hoax tidak dikomersialkan sebaik pendeteksi spam [1, 2]. Detektor spam dan filter spam mirip satu sama lain [6]. Namun, konsep spam cukup sederhana dan mudah dimengerti. Jika spam tidak dapat dilacak, maka spam tersebut merupakan spam yang efektif sedangkan jika spam tersebut dapat dikenali, maka filter spam sudah pasti merupakan filter spam yang efektif [6]. Filter spam memiliki banyak kemiripan dengan sistem pendeteksi hoax. Misalnya, konten email hoax dan spam berisi petunjuk seperti kata-kata huruf kapital dan tanda seru. Mungkin, satu-satunya perbedaan dalam kedua subjek ini adalah tanda tangan dari subjek filter yang ditargetkan, definisi mereka sendiri dan juga metode ekstraksi. Oleh karena itu, hal ini menunjukkan bahwa pendekatan yang digunakan dalam sistem penyaringan spam juga dapat diterapkan pada sistem pendeteksi hoax. Proses persetujuan apakah sebuah email adalah spam atau bukan adalah sebagai berikut [1, 2]:

* Bandingkan email dengan database spam
* Jika basis data spam berisi spam yang terdeteksi dalam email, tandai sebagai spam dan hapus ke folder sampah secara otomatis (digunakan oleh semua penyedia layanan email saat ini)
* Spam yang tidak ditandai dapat ditandai sebagai netral (aman) atau mencurigakan tergantung pada probabilitas spam dan disajikan kepada pengguna untuk memutuskan email mana yang harus mereka hapus sebagai spam secara manual nanti.

Pendekatan paling umum dalam sistem penyaringan spam adalah algoritma Naive Bayes, pendekatan berbasis cluster, metode crossregulation, dan teknik text mining. Naif Pengklasifikasi Bayes adalah metode yang paling umum digunakan dalam sistem penyaringan spam hari ini. Menurut Paliouras et Al. [7], pemfilteran spam cocok dengan pembelajaran mesin klasifikasi seperti Naive Bayes (NB) seperti yang telah terbukti dapat mengatasi tugas ini dengan baik meskipun sederhana. Di sisi lain, klasifikasi berbasis cluster pendekatan adalah pendekatan lain yang menarik untuk spam penyaringan. Seringkali spam didistribusikan lebih dari surat sah biasa yang menangani miring konsep distribusi kelas sedangkan konsep drift didasarkan pada preferensi pengguna yang dapat berubah seiring waktu atau topik spam dapat bervariasi sesuai dengan tren mode [8]. Pendekatan klasifikasi berbasis cluster mampu menangani konsep drift dan distribusi kelas miring dengan cukup baik. Metode cross-regulation merupakan metode yang baru ditemukan untuk penyaringan spam [9]. Itu diperkenalkan dengan tujuan untuk lebih mendeteksi spam berdasarkan konsep sistem kekebalan. Sejauh ini, penelitian tersebut telah membuktikan hasil positif tetapi memiliki sedikit kesamaan dengan sistem pendeteksi hoax berdasarkan arsitekturnya. Teknik text mining memiliki kemampuan untuk menangani frase kunci sebagai bagian dari teks yang ditambang [8]. Penelitian tentang text mining [10-12] menunjukkan upaya yang menjanjikan dalam membandingkan antar teks. Konsep penambangan teks adalah alat penambangan data yang mampu mengekstrak konten berbasis teks berdasarkan hasil yang diinginkan. Namun, ada beberapa keterbatasan pada konsep ini dan metode pemrosesannya mungkin sedikit lebih lambat.

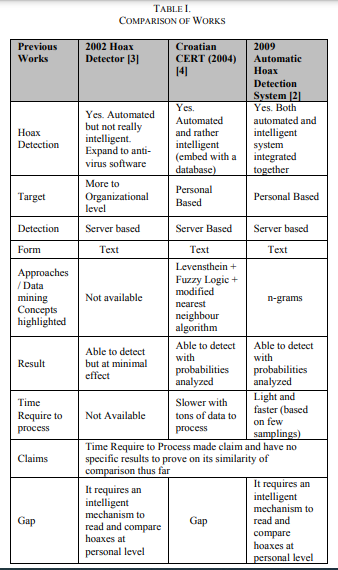
1. Perbandingan Pekerjaan Sebelumnya pada Sistem Deteksi Hoax

Penelitian telah menunjukkan bahwa adalah mungkin untuk mengotomatisasi sistem deteksi hoax. Hoax memiliki ciri-ciri tertentu di mana sebagian besar isinya dapat ditemukan tebal, bergaris bawah, banyak tanda seru dan sering menggunakan kata-kata yang membutuhkan keputusan instan untuk dibuat. Oleh karena itu, tidak menutup kemungkinan hoax dapat terindentifikasi secara otomatis [13].

Penelitian pendeteksian hoax mulai aktif pada tahun 2002. Model ini memulai penelitiannya setelah melihat bahwa hoax memang menjadi hal yang meresahkan di masyarakat. Menurut Hernandez dkk. [3], pengetahuan sangat penting untuk pertumbuhan manusia dan informasi yang menyesatkan bisa mahal. Karena itu, mereka mulai dengan menggunakan metode berbasis aturan sebagai awal untuk penelitian khusus ini. Pada tahun 2002, metode berbasis aturan adalah salah satu metode ekstraksi teks terkemuka yang digunakan untuk anti-spam. Metode berbasis aturan adalah metode di mana ekstraksi dilakukan berdasarkan aturan dan kondisi tertentu. Metode filter tradisional ini menggunakan metode sederhana dan mudah yang mengklasifikasikan spam dengan mencocokkan kolom email dengan kata kunci. Kegagalan metode ini disebabkan oleh kenyataan bahwa kompleksitas spam saat ini telah menggantikan manfaat yang belum ditawarkan metode ini.

Pada tahun 2004, CERT Kroasia mengambil langkah lebih jauh dengan mengembangkan versi lain dari sistem pendeteksi hoax [4]. Sistem ini telah terbukti berhasil karena lalu lintas Internet yang rendah di negara ini. Namun, sistem ini telah menunjukkan beberapa kelemahan di area tertentu terutama pada ide penggunaan logika fuzzy dan metode Levenshtein. Kedua metode ini telah menunjukkan hasil yang menjanjikan dalam mendeteksi hoaks tetapi kecepatan pemrosesan hoaks dianggap lambat ditambah dengan lalu lintas Internet yang padat dan volume data yang besar. Apalagi database hoax tidak bisa update sendiri.

Pada tahun 2009, studi untuk sistem deteksi hoax otomatis menjadi jauh lebih layak dari sebelumnya [2]. Alih-alih menggunakan logika berbasis aturan dan logika fuzzy, penelitian ini menggunakan n-gram sebagai metode ekstraksi teks. N-gram adalah salah satu alat penambangan data yang memprediksi kata-kata dalam email. N-gram memungkinkan sistem untuk memprediksi kata-kata berikutnya dalam email yang dianalisis di mana ia dapat memproses kata-kata lebih cepat (diasumsikan dan dianggap oleh penemu) Namun, mengingat studi untuk n-gram dispam [14, 15], kompleksitas probabilitas dalam mengeksekusi konsep ini cukup rumit untuk sistem pendeteksi hoax. Menurut Vukovic dkk. [2], pendekatan ngrams mampu mempercepat proses pendeteksian hoax. Namun kelemahan dari pendekatan n-grams adalah tidak dapat melatih pengidentifikasi hoax pada hoax terbaru yang ditemukan berdasarkan ciri-ciri hoax. Tabel I menunjukkan persamaan dan perbedaan dari karya-karya yang ada yang kami susun dari literatur.



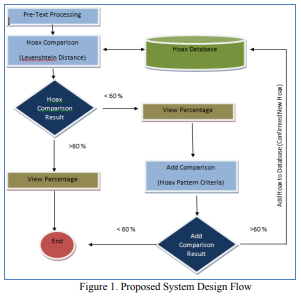
Berdasarkan analisis seperti yang ditunjukkan pada Tabel I, masih ada ruang untuk perbaikan pada pendekatan saat ini dalam mengembangkan sistem deteksi hoax. Jenis subjek terkait lainnya sebagai bagian dari investigasi terkait deteksi hoax adalah spam dan alat penambangan teks. Arsitektur filter spam mirip dengan sistem pendeteksi hoax yang dibuat selama bertahun-tahun dengan definisi yang berbeda. Spam dapat dilacak berdasarkan alamat pengirim dari email tertentu [1, 6] sedangkan hoax perlu diatasi dengan isinya [2-4]. Metode yang digunakan dalam deteksi hoax [2-4] juga telah digunakan oleh filter spam sebelumnya dalam literatur [7, 8, 14, 16].

Sepengetahuan kami, tidak ada sistem pendeteksi hoax yang mampu memperbarui databasenya sendiri tentang keberadaan hoax baru. Artinya, sistem tidak dapat mengenali hoax yang baru dibuat. Inilah tantangan terbesar yang dihadapi dalam penelitian ini. Oleh karena itu, sistem yang kami usulkan adalah untuk menutup celah dalam pendeteksian hoaks dengan mengidentifikasi karakteristik hoaks berdasarkan data hoaks online.

Tujuan akhir dari proyek ini adalah untuk meningkatkan kesadaran akan adanya hoaks dan untuk menghindari informasi menyesatkan yang tidak perlu yang dapat menyebabkan kesalahpahaman di kalangan masyarakat. Oleh karena itu, untuk menutup kesenjangan dengan penelitian sebelumnya, sistem yang kami usulkan mencakup fitur untuk mengidentifikasi ciri-ciri hoaks dalam membantu mengklasifikasikan e-mail yang masuk ke dalam tingkat kecurigaan hoaks yang berbeda.

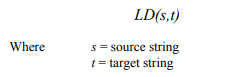
1. Metodologi
2. Sistem Arsitektur

Sistem deteksi hoax yang diusulkan terdiri dari tiga komponen utama: pra-pemrosesan teks, deteksi hoax dan deteksi hoax baru. Pra-pemrosesan teks mengumpulkan email untuk diuji validitasnya menjadi email asli atau tipuan. Header dan tanda tangan email akan dihapus dari email asli. Sistem menggunakan Levenshtein Distance untuk mengidentifikasi potensi hoax dari konten email dengan membandingkannya dengan database hoax. Basis data hoax terdiri dari email hoax yang dikumpulkan dari Hoax-Slayers.com, yang diakui oleh Perpustakaan Nasional Australia (PANDORA) sebagai sumber hoax yang baik. Alur perancangan sistem yang diusulkan ditunjukkan pada Gambar 1.



1. Jarak Levenshtein

Levenshtein Distance (LD) adalah metrik untuk mengukur jumlah perbedaan antara dua string. Berikut ini adalah notasi untuk Jarak Levenshtein:



Jarak antar string diukur dengan jumlah penyisipan, penghapusan atau penggantian yang diperlukan untuk mengubah string s menjadi string t. Sebagai contoh:

* Jika string sumber (s) "tebal" dan string target (t) juga "tebal", maka LD(s,t) = 0, karena tidak diperlukan transformasi. String identik satu sama lain.
* Jika string sumber asli (s) "tebal" dan string target (t) adalah "bonk", maka LD(s,t) = 2, karena dua substitusi (perubahan "l" menjadi "n" dan "d menjadi “k”) diperlukan untuk mengubah string asli menjadi string yang dibandingkan.

Oleh karena itu, semakin tinggi jarak Levenshtein, semakin berbeda stringnya dan semakin kecil kemungkinan email itu hoax. Isi email yang diinput oleh pengguna akan diperlakukan sebagai string sumber dan string target adalah email hoax yang disimpan dalam database hoax. Hitungan jarak akhir akan diubah menjadi persentase menggunakan rumus berikut:

**Jarak = (100 – jumlah jarak total) \* 100**

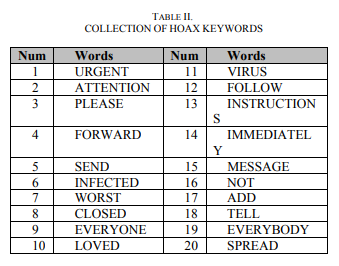
Seperti yang telah disebutkan di bagian sebelumnya, sebagian besar sistem pendeteksi hoax yang disebutkan dalam literatur tidak mampu mendeteksi hoax baru. Untuk mengatasi masalah ini, sistem yang diusulkan menambahkan komponen lain untuk mengidentifikasi potensi hoax baru. Oleh karena itu, aturan berikut diterapkan dalam sistem.

1. Jarak lebih besar dari 60%, email adalah kebohongan
2. Jarak kurang dari 60%, emailnya adalah diduga hoax. Isi email akan dikirim untuk penyaringan akhir.
3. Penyaringan Akhir

Berdasarkan pengamatan kami, hoax memiliki pola serupa yang mudah dideteksi seperti:

* Sering menggunakan huruf besar
* Judul terdiri dari kata-kata seperti URGENT! Mohon FORWARD pesan ini dan lain-lain
* Termasuk kata-kata ekstrim seperti TERBURUK, SANGAT SERIUS dan lain-lain

Kami menyimpan sekantong kata-kata yang terdiri dari kata-kata umum yang sering ditemukan di email hoax. Tabel II adalah kumpulan sampel kantong kata-kata yang ditemukan dari hoax.



Oleh karena itu, kami menghitung persentase kata yang ditulis dalam huruf besar atas jumlah total karakter yang ditemukan dalam email.

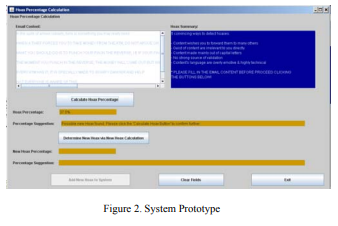
*Persentase Huruf Besar = (Jumlah Total Huruf Besar Karakter / Jumlah Karakter dalam email) \* 100*

*Kata Kunci Persentase = (Jumlah Kata Kunci ditemukan / Jumlah kata dalam email) \* 100*

Kami menerapkan aturan 60% yang disebutkan di atas pada rumus untuk menentukan apakah email tersebut baru atau tidak tipuan email. Hoax email baru yang teridentifikasi akan ditambahkan ke dalam database hoax.

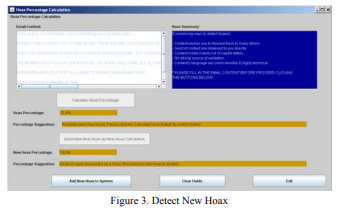
1. Hasil dan diskusi
2. Prototipe Sistem

Seperti ditunjukkan pada Gambar. 2, pengguna yang ingin menguji apakah konten email adalah tipuan perlu menyalin email konten ke sistem. Setelah pengguna mengklik 'Hitung' Tombol Persentase Hoax, level mencurigakan tipuan akan dihitung. Jika level mencurigakan kurang dari 60%, sistem akan menyarankan pengguna untuk menentukan apakah ada tipuan baru terdeteksi.



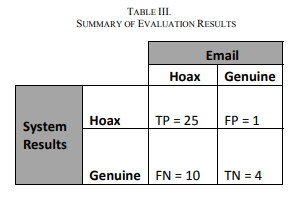
Dalam kasus ketika sistem menyarankan untuk menentukan apakah tipuan baru terdeteksi, pengguna diminta untuk mengklik Dalam kasus ketika sistem menyarankan untuk menentukan apakah tipuan baru terdeteksi, pengguna diminta untuk mengklik tentang ‘Menentukan Hoax Baru melalui Perhitungan Hoax Baru’ tombol. Setelah diklik, level mencurigakan tipuan baru akan dihitung. Seperti yang ditunjukkan pada Gambar. 3, tingkat mencurigakan tipuan adalah 74%. Oleh karena itu, tipuan baru telah terdeteksi dan

konten email akan ditambahkan ke database hoax.



1. Sistem Evaluasi

Keakuratan sistem dalam mengidentifikasi email hoax perlu dinilai untuk memastikan sistem yang diusulkan efektif dalam mengidentifikasi email hoax. Ukuran kinerja yang digunakan untuk mengevaluasi keakuratan sistem adalah nilai prediksi positif, nilai prediksi negatif, sensitivitas dan spesifisitas. Sebanyak 40 sampel email diuji menggunakan sistem. Tabel III merangkum hasil evaluasi.



true positive (TP) = email hoax yang benar diidentifikasi oleh system

false positive (FP) = email asli diidentifikasi sebagai email hoax

true negative (TN) = email asli teridentifikasi

false negative (FN) = email hoax diidentifikasi sebagai email asli

Positive Predictive Value (PPV) = TP / (TP + FP)

= 0.96

Negative Predictive Value(NPV) = TN / (FN + TN)

= 0.29

Sensitivity = TP / (TP + FN)

= 0.71

Specificity = TN / (FP + TN)

= 0.80

Nilai prediksi positif yang tinggi (PPV = 0,96) menunjukkan bahwa banyak email hoax yang benar telah diidentifikasi oleh sistem. Nilai prediksi negatif yang rendah (NPV = 0,29) menunjukkan bahwa sistem lemah dalam mengidentifikasi email asli. Oleh karena itu, kemungkinan besar sistem akan memperlakukan email hoax sebagai email asli. Hal ini sesuai dengan nilai sensitivitas yang tinggi (0,71) yang menunjukkan bahwa kemungkinan sistem mengidentifikasi email hoax sebagai email asli tinggi. Sistem memiliki nilai spesifisitas tinggi (0,80); ini menunjukkan bahwa kemungkinan sistem mengidentifikasi email asli sebagai hoax adalah tinggi. Meskipun sistem mampu memberikan hasil yang baik dalam mengidentifikasi email hoax, sistem ini kurang mampu mengidentifikasi email hoax yang ditulis dengan format email asli. Selama pengujian sistem, kami dapat mengamati bahwa sebagian besar email hoax yang telah ditulis secara profesional belum teridentifikasi sebagai hoax oleh sistem. Hal ini karena email yang ditulis secara profesional membantu meningkatkan kepercayaan diri pembaca dan membuat mereka percaya bahwa cerita yang mereka sampaikan adalah benar. Oleh karena itu, dengan menggunakan format email (huruf besar) sebagai salah satu faktor untuk mengidentifikasi email hoax mungkin bukan metode yang mutlak efektif.

1. Kesimpulan Dan Rekomendasi

Kesadaran akan adanya hoaks menciptakan kesadaran yang lebih besar untuk menemukan pengetahuan yang tepat yang dapat digunakan untuk lebih mengembangkan persepsi manusia saat mereka tumbuh. Informasi yang menyesatkan hanya dapat merugikan kelompok tertentu termasuk mencoreng citra mereka di mata publik. Dalam makalah ini telah dikembangkan sistem pendeteksi hoax dengan menggunakan Levenshtein Distance. E-mail yang diekstraksi dibandingkan dengan database hoax untuk mendukung keputusan identifikasi hoax. Di sisi lain, fitur nilai tambah disarankan dalam sistem yang diusulkan dengan menghadirkan kemungkinan hoaks kepada penerima. Ini membuat perbedaan dari sistem sebelumnya. Sistem ini mampu memberikan nilai prediksi positif tinggi 0,96 tetapi tidak memiliki kemampuan untuk mengidentifikasi email asli. Email hoax tidak hanya dikirim dalam format teks tetapi juga dengan gambar. Koleksi kata kunci hoax terbatas. Oleh karena itu, ini adalah keterbatasan sistem dan masih ada ruang untuk perbaikan.

Untuk pekerjaan di masa depan, sistem dapat ditingkatkan dengan mengintegrasikan teknik kecerdasan buatan seperti logika fuzzy. Dengan teknik ini, sistem seharusnya dapat lebih mengklasifikasikan email hoax dan email asli. Karya-karya sebelumnya dalam pendeteksian hoax tidak mengadopsi teori pemrosesan teks kecuali n-gram. Namun, diyakini bahwa alat penambangan teks lain yang menggunakan algoritma pembelajaran mesin [10] dapat digunakan sebagai pra-pemrosesan teks untuk mengekstrak teks dari email.