P-ISSN: | E-ISSN:
Vol. 2, No. 1 Februari 2026, p.5-10

Journal of SONIC

Research Article [OPEN ACCESS]

Kriptanalisis: Analisis Frekuensi Pada Dokumen Berbahasa Indonesia

Astri Hijratul Rakhmah[1]
Teknologi Rekayasa Perangkat Lunak, Politeknik Bisnis Digital Indonesia, Indonesia
Email: astrihijratul@polbis.ac.id

ABSTRAK

Dalam kriptografi klasik, beberapa metode kriptanalisis digunakan untuk memecahkan pesan tersandi (*chipertext*). Salah satunya adalah dengan menggunakan metode Analisis Frekuensi yang digunakan untuk menghitung banyaknya kemunculan suatu huruf atau beberapa huruf (n-gram) dalam pesan tersandi. Misalnya dalam dokumen berbahasa Inggris, hasil implementasi Analisis Frekuensi menemukan bahwa trigram (jumlah kemunculan 3 huruf) yang paling banyak muncul Adalah THE. Pada penelitian ini penulis menggunakan metode Analisis Frekuensi untuk mengetahui trigram yang paling banyak muncul dalam dokumen berbahasa Indonesia, sehingga dapat digunakan untuk menjadi dasar dalam melakukan kriptanalisis pada pesan tersandi dalam Bahasa Indonesia. Analisis Frekuensi diterapkan pada sejumlah dokumen dengan total 39.422 kata (212.883 huruf) yang bersumber dari novel berbahasa Indonesia. Selain itu peneliti juga menggunakan *tool Frequency Analysis* yang tersedia pada cryptool.org. Hasil penelitian menunjukkan bahwa, trigram terbanyak muncul dalam dokumen berbahasa Indonesia Adalah ANG, yaitu sebanyak 2.050 kemunculan. Temuan dari hasil penelitian ini merupakan Langkah awal yang dapat dijadikan sebagai landasan dalam melakukan kriptanalisis pada pesan tersandi dalam Bahasa Indonesia.

Kata Kunci: Kriptanalisis, Analisis, Frekuensi, Bahasa, Indonesia

1. PENDAHULUAN

Kriptanalisis Adalah sebuah seni dalam memecahkan pesan tersandi (*chipertext*). Dalam literatur yang ada saat ini sudah banyak metode-metode yang digunakan dalam memecahkan *chipertext* terutama *chipertext* yang dihasilkan dari proses enkripsi menggunakan kriptografi klasik. Dalam kriptografi klasik, umumnya metode yang digunakan adalah dengan cara merubah satu huruf *plaintext* menjadi satu huruf *ciphertext* (cipher substitusi). Atau dengan cara merubah urutan huruf *plaintext* tanpa menggantinya (cipher transposisi). Akan tetapi cipher-cipher yang digunakan pada kriptografi klasik, baik dengan cara subtitusi maupun transposisi memiliki kelemahan, yaitu dapat dilihatnya hubungan statistik yang dimiliki *plaintext* dan *ciphertextnya*, dalam artian, huruf yang paling banyak muncul dalam *plaintext* memiliki kemungkinan besar akan banyak muncul juga dalam *ciphertext*nya.

Salah satu metode yang digunakan untuk mendeteksi banyaknya kemunculan huruf dalam sebuah dokumen, atau dalam hal ini adalah *ciphertext* yaitu dengan menggunakan metode Analisis Frekuensi. Dalam Analisis Frekuensi banyaknya kemunculan huruf dapat dicari dengan menggunakan Teknik n-gram yang memecah teks menjadi urutan item yang berdekatan dan kemudian menghitung seberapa sering setiap urutan tersebut muncul. Item-item ini dapat berupa karakter, suku kata, atau kata. Dalam dokumen atau pesan tersandi berbahasa Inggris, Teknik n-gram menemukan bahwa huruf yang paling banyak muncul Adalah E, dan bigram (2 huruf) yang paling banyak muncul adalah TH, kemudian untuk trigram (3 huruf) yang paling banyak muncul Adalah THE. Hal ini kemudian dijadikan landasan oleh kriptanalis dalam memecahkan chipertext berbahasa Inggris.

Penelitian terkait penggunaan metode Frekuensi Analisis dalam Kriptografi dilakukan oleh Faizah (2022), dalam penelitiannya ia memanfaatkan metode tersebut untuk menguji ketahanan *ciphertext* hasil enkripsi pesan berbahasa Inggris yang dilakukan dengan Kriptografi DNA. Hasil penelitiannya menemukan bahwa hasil *plaintext* berbahasa Inggris terenkripsi secara acak dan susah ditebak menggunakan metode

Analisis Frekuensi. Pramudya (2025) dalam penelitiannya memanfaatkan metode Analisis Frekuensi untuk membandingkan hasil sebaran dan jumlah kemunculan huruf-huruf *plaintext* dan *ciphertext* berbahasa Inggris yang disandikan dengan algoritma Caesar Cipher dan DES. Hasil dari penelitiannya meyimpulkan bahwa algoritma Caesar cipher sudah tidak lagi layak untuk digunakan dalam konteks keamanan modern dan hanya untuk tujuan edukasi, sedangkan DES, meskipun lebih kompleks dan memiliki keamanan lebih baik dibanding Caesar cipher tetapi tetap tergolong lemah karena penggunaan Panjang kunci 56-bit yang sudah tidak memadai dalam menghadapi kemampuan komputasi saat ini. Sejauh ini, penelitian-penelitian terkait kriptografi yang memanfaatkan metode Analisis Frekuensi dilakukan pada konteks pesan berbahasa Inggris, sedangkan penelitian terkait pada konteks pesan berbahasa Indonesia sangat terbatas. Disinilah letak Scientific gap dari penelitian ini.

Penelitian ini bertujuan untuk mengetahui Tingkat kemunculan huruf dan pasangan huruf (bigram atau trigram) dalam Bahasa Indonensia dengan memanfaatkan metode Analisis Frekuensi. Sehingga dapat menambah kekayaan keilmuan baik di bidang Kriptografi khususnya Kriptanalisis maupun bidang lainnya (*text processing*) yang juga memanfaatkan metode Analisis Frekuensi dalam prosesnya.

2. METODE PENELITIAN

Penelitian ini menggunakan metode studi literatur dan eksperimen, yaitu dengan mempelajari buku-buku, karya ilmiah dan studi kasus terkait kelemahan kriptografi klasik dan Teknik kriptanalisis untuk memecahkan *ciphertext* yang disandikan menggunakan kriptografi klasik, serta melakukan eksperimen terhadap dokumen berbahasa Indonesia yang terdiri dari 39.422 kata (212.883) huruf yang bersumber dari Novel.

2.1 Tinjauan Pustaka

Dooley (2022) dalam bukunya menyatakan bahwa “Frekuensi kemunculan setiap huruf merupakan karakteristik sebuah bahasa, tidak mungkin menyembunyikan frekuensi kemunculan ini jika seseorang mengganti satu huruf dengan huruf lain dalam sebuah pesan tersandi. Khususnya, frekuensi huruf akan terlihat melalui substitusi seperti suar yang mengarahkan kriptanalis kepada huruf-huruf tersembunyi dalam teks biasa”. Dalam Bahasa Inggris disebutkan bahwa huruf-huruf yang paling sering muncul biasanya diberikan dalam urutan ETAOINSHRDLWU. Hal ini berdasarkan pada perhitungan terhadap sekitar 95.512 kata atau sekitar 450.583 huruf.

Munir (2019) dalam bukunya mengulas teknik kriptanalisis pada *ciphertext* yang dihasilkan melalui proses enkripsi menggunakan kriptografi klasik. Kriptografi klasik sudah tidak relevan lagi untuk digunakan dalam kriptografi modern karena kelemahannya yang mudah ditebak. Dalam bukunya disebutkan, salah satu kelemahan kriptografi klasik adalah terlihatnya hubungan statistik antara *plaintext* dan *ciphertext*nya. Dalam artian, huruf yang paling banyak muncul dalam *plaintext* memiliki kemungkinan besar akan muncul juga dalam *ciphertext*nya. Hal ini tentunya menjadi celah yang dapat dimanfaatkan oleh kriptanalis untuk memecahkan isi pesan dalam *ciphertext*. Salah satu metode yang sering digunakan untuk memecahkan *ciphertext* yang dihasilkan dari enkripsi kriptografi klasik Adalah metode Analisis Frekuensi. Metode Analisis Frekuensi dapat diterapkan pada *ciphertext* yang dienkripsi menggunakan diantaranya Caesar Cipher, dan Vigenere cipher.

Faizah, Widya N., dkk (2022) dalam penelitiannya, menerapkan metode analisis frekuensi terhadap hasil enkripsi pesan (*plaintext*) dengan Algoritma Kriptografi DNA dan Transformasi Digraf. Proses enkripsi dilakukan dengan cara mengubah pesan *plaintext* menggunakan Transformasi Digraf, cara kerjanya Adalah dengan mengubah pasangan huruf (digraph) *plaintext* (misalnya XY) menjadi pasangan huruf *ciphertext*. Selanjutnya, hasil dari transformasi Digraf tersebut dienkripsi menggunakan algoritma Kriptografi DNA. Algoritma Kriptografi DNA Adalah metode enkripsi data yang menggunakan urutan DNA (A, T, C, G) untuk menyandikan informasi atau pesan, cara kerjanya Adalah dengan mengubah data teks pesan menjadi urutan DNA untuk menghasilkan ciphertext yang acak dan tidak mudah dibaca. Untuk menguji ketahanan hasil enkripsi terhadap serangan kriptanalis, maka digunakan metode Analisis Frekuensi untuk melihat

hubungan statistic yang mungkin muncul antara *plaintext* dan *ciphertext*nya. Dari hasil eksperimen yang dilakukan, diperoleh hasil bahwa *plaintext* masih sulit ditebak.

Pramudya, Farhan A., dkk (2025) melakukan komparasi terhadap keamanan 2 algoritma kriptografi, yaitu Caesar cipher dan DES dalam konteks kebutuhan keamanan modern. Teknik komparatif yang digunakan meliputi implementasi kedua algoritma dengan parameter uji meliputi waktu eksekusi, *avalanche effect* (prinsip penting dalam kriptografi di mana perubahan kecil pada data input (seperti *plaintext* atau kunci) akan menghasilkan perubahan yang sangat besar dan tidak terduga pada data output (teks tersandi)), ruang kunci, dan ketahanan terhadap kriptanalisis. Dalam hal ini, pengujian terhadap ketahanan terhadap kriptanalisis dilakukan dengan menerapkan serangan brute force dan Analisis Frekuensi. Kesimpulan dari hasil eksperimen dan komparasi menyatakan bahwa Caesar Cipher tidak layak digunakan dalam konteks keamanan modern karena ruang kunci yang sangat kecil, efek avalanche rendah, serta kerentanannya terhadap brute force dan analisis frekuensi. Sementara itu, DES meskipun lebih kompleks dan memiliki keamanan lebih baik dibanding Caesar Cipher, tetap tergolong lemah (*deprecated*) akibat panjang kunci 56-bit yang sudah tidak memadai menghadapi kemampuan komputasi saat ini.

Herawati, Riska., dkk. (2019) dalam penelitiannya mencoba untuk mendeskripsikan bentuk kata berimbuhan (Afiksasi) dalam kata-kata Mutiara berbahasa Indonesia pada caption di media sosial Instagram. Penelitian dilakukan dengan menggunakan pendekatan kualitatif deskriptif, dan Teknik untuk menganalisis datanya Adalah dengan menggunakan metode agih dengan Teknik dasar yaitu Teknik bagi unsur langsung. Pengumpulan data dilakukan dengan cara simak, dokumentasi dan catat. Dan keabsahan data menggunakan triangulasi penyidik. Hasil dari penelitiannya menemukan bahwa dari 87 kemunculan Afiks, frekuensi kemunculan prefix adalah yang tertinggi, yaitu sebanyak 31 kemunculan, disusul 18 buah kemunculan sufiks, 18 buah kemunculan klofiks, 16 buah kemunculan konfiks dan 4 buah kemunculan infiks.

3. HASIL DAN PEMBAHASAN

Penerapan metode Analisis Frekuensi untuk mengetahui seberapa tinggi tingkat kemunculan suatu huruf, pasangan huruf dan *triple* huruf pada teks berbahasa Indonesia merupakan focus utama dari penelitian ini. Sehingga mempermudah kriptanalis untuk menerka bagian *chipertext* berbahasa Indonesia. Ada beberapa tahapan yang dilakukan, yaitu pemilihan data, *preprocessing*, dan analisis frekuensi.

3.1. Pemilihan dokumen

Pemilihan dokumen merupakan tahap awal yang dilakukan dalam penelitian ini. Dokumen berbahasa Indonesia sendiri sangatlah beragam, seperti puisi, kata Mutiara, naskah berita, laporan atau penulisan ilmiah, novel dan lain-lain. Dalam penelitian ini penulis memilih novel sebagai sumber teks yang akan diuji. Hal ini didasari oleh penggunaan Bahasa dalam novel yang kaya dengan deskripsi, kiasan dan naratif serta kecenderungannya untuk semirip mungkin dengan penggunaan Bahasa sehari-hari. Berbeda dengan sumber teks lainnya yang cenderung terkesan kaku dan formal.

3.2. *Preprocessing*

Preprocessing Adalah tahap esensial dalam beberapa bidang seperti *text processing, natural Language Processing, Data Mining* dan lain-lain. dalam *text processing* sendiri, *preprocessing* Adalah tahap membersihkan dan menata (mempersiapkan) data teks mentah agar siap diolah lebih lanjut oleh algoritma dengan tujuan meningkatkan akurasi dan efektivitas analisis. Seperti halnya dalam Analisis Frekuensi, *preprocessing* perlu dilakukan untuk mempersiapkan data teks agar siap untuk dianalisis Tingkat kemunculan huruf yang dicari. Beberapa tahapan yang dilakukan dalam *preprocessing* ini Adalah:

a. Remove Punctuation

Remove punctuation Adalah proses penghapusan tanda baca atau karakter-karakter non-alfanumerik dari sebuah teks. *Punctuation* meliputi beberapa karakter seperti titik (.), koma (,), tanta tanya (?), tanda seru (!) dan karakter-karakter lainnya yang tidak termasuk dalam kategori huruf atau angka. Penghapusan tanda baca dapat membantu meningkatkan konsistensi dan keakuratan hasil pengolahan teks, terutama dalam kasus-kasus di mana tanda baca tidak relevan dengan analisis atau pemrosesan yang dilakukan (Hendrawan & Ema Utami, 2023).

b. Numeric Removal

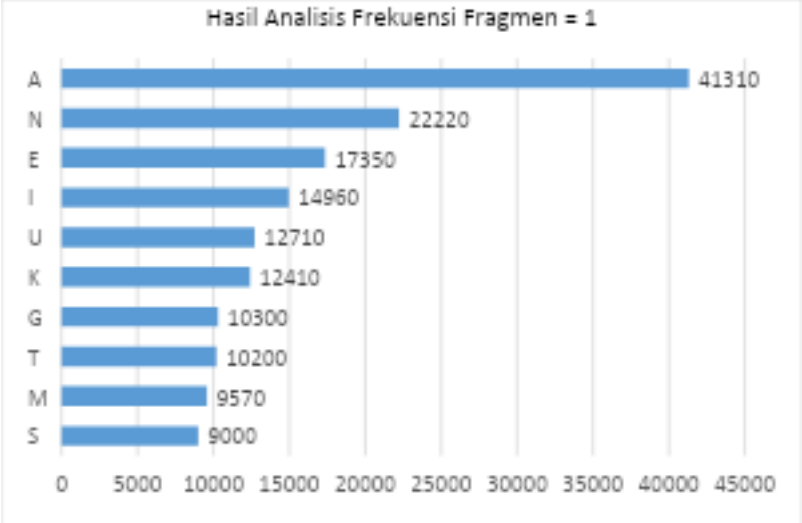
Angka atau karakter numerik yang muncul dalam teks dalam penelitian ini tidak relevan. Sehingga menghapus angka dapat membantu menyederhanakan teks dan memfokuskan pada informasi yang lebih penting.

c. Case Folding

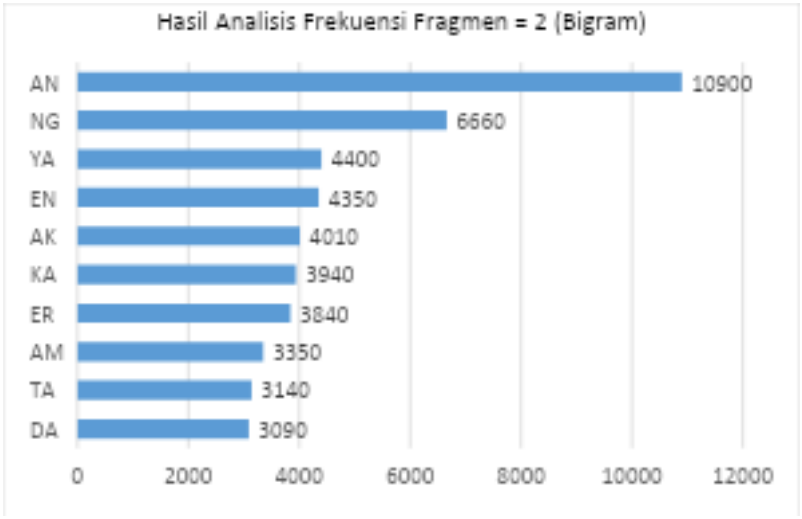
Case folding merupakan proses dalam *text preprocessing* yang dilakukan untuk menyeragamkan karakter pada data. Dalam hal ini *case folding* dilakukan dengan menyeragamkan huruf menjadi *lower case* (huruf kecil) (Masruroh & Etna Syirfa Qorina, 2022).

3.3. Analisis Frekuensi

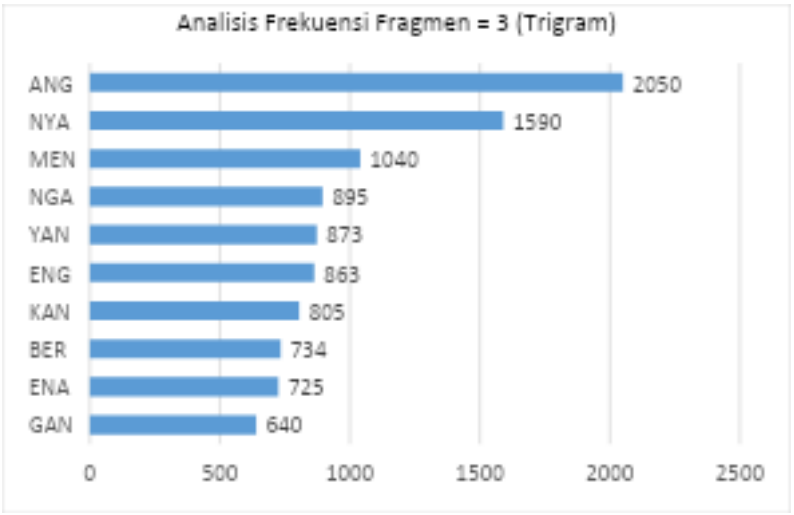
Setelah tahap *preprocessing* pada teks yang dipilih selesai dilakukan, maka Langkah selanjutnya Adalah menerapkan analisis frekuensi. Analisis frekuensi dilakukan dengan menggunakan *tools Frequency Analysis* pada cryptool.org. Fragmen yang diatur menjadi 3 kriteria, yaitu 1 huruf, 2 huruf dan 3 huruf. Dari Analisis frekuensi pada teks dengan fragmen 1 hingga 3 diperoleh hasil masing-masing sebagai berikut:



Gambar 1. Hasil Analisis Frekuensi Fragmen = 1



Gambar 2. Hasil Analisis Frekuensi Fragmen = 2 (Bigram)



Gambar 3. Hasil Analisis Frekuensi Fragmen = 3 (Trigram)

Dari grafik yang disajikan dapat dilihat bahwa frekuensi kemunculan huruf terbanyak dalam teks novel berbahasa Indonesia Adalah huruf A, N, E, I, U, K, G, T, M,S, dan kemunculan pasangan huruf (Bigram) terbanyak Adalah AN, NG, YA, EN, AK, KA, ER, AM, TA, DA serta kemunculan 3 huruf (Trigram) terbanyak Adalah ANG, NYA, MEN, NGA, YAN, ENG, KAN, BER, ENA, dan GAN.

4. KESIMPULAN

Penelitian ini bertujuan untuk mengetahui Tingkat kemunculan huruf dan pasangan huruf (bigram dan trigram) dalam Bahasa Indonensia dengan memanfaatkan metode Analisis Frekuensi. Sumber teks yang digunakan dalam Analisis Frekuensi bersumber dari Novel dengan total 39.422 kata (212.883 huruf). Dalam prosesnya, analisis frekuensi dilakukan dalam 3 fragmen, yaitu fragmen 1, 2 dan 3 untuk mengetahui Tingkat kemunculan tertinggi 1 huruf, 2 huruf dan 3 huruf. Hasil yang diperoleh dari Fragmen 1 huruf Adalah huruf A merupakan huruf yang paling tinggi kemunculannya dalam data teks, yaitu sebanyak 41.310 kemunculan, disusul N, E, I, U, K, G, T, M, S. Kemudian pada fragmen = 2 ditemukan bahwa Tingkat kemunculan pasangan huruf (Bigram) yang tertinggi Adalah pada pasangan huruf AN yaitu sebanyak 10.900 kemunculan, disusul NG, YA, EN, AK, KA, ER, AM, TA, dan DA. Sedangkan pada Fragmen = 3 (Trigram)

Tingkat kemunculan tertinggi Adalah pada kata ANG yaitu sebanyak 2.050 kemunculan, disusul NYA, MEN, NGA, YAN, ENG, KAN, BER, ENA dan GAN.

Dengan diketahuinya Tingkat kemunculan huruf-huruf ini maka dapat menjadi peluang bagi kriptanalis untuk menebak dan mencoba memecahkan pesan tersandi dalam Bahasa Indonesia yang disandikan dengan menggunakan Kriptografi Klasik. Sehingga diperlukan algoritma yang lebih kuat untuk dapat menghindari metode kriptanalisis dengan cara Analisis Frekuensi.

5. REFERENSI

Dooley, John F. (2022). *History of Cryptography and Cryptanalysis: Codes, ciphers, and Their Algorithms (2nd Edition)*. Springer. 10.1007/978-3-3 19-90443-6

Faizah, Widya N., dkk. (2022). Analisis Frekuensi Ciphertext dengan Algoritma Kriptografi DNA dan Transformasi Digraf. *JRMM (Jurnal Riset Mahasiswa Matematika)*, *1*(6), 283-287. http://dx.doi.org/10.18860/jrmm.v1i6.14591

Hendrawan, Ivan Rifky & Ema Utami (2023). *Natural Language Processing*. Penerbit Andi

Herawati, Riska., dkk. (2019). Analisis Afiksasi Dalam Kata-kata Mutiara Pada Cation Di Media Sosial Instagram Dan Implikasinya Terhadap Pembelajaran Bahasa Indonesia Di SMP. *Jurnal Membaca: Bahasa & Sastra Indonesia, 4*(1), 45-50

Masruroh, Siti Ummi., & Etna Syirfa Qorina. (2022). *Fast Text dan Word2Vec Pada Query: Kesamaan Semantik Sistem Temu Kembali Informasi.* Deepublish Publisher

Munir, R (2019). *Kriptografi: Edisi kedua*. Penerbit Informatika

Pramudya, Farhan A. (2025). Analisis Keamanan Komparatif Caesar Cipher DES dalam Konteks Teknik Kriptografi Modern. *Cosmic Jurnal Teknik*, *2*(3), 96-105. 10.55537/cosmic