# DETEKSI KESALAHAN EJA KATA LULUH PADA BERITA DENGAN ALGORITMA JACCARD SIMILARITY (STUDI KASUS : TRIBUNNEWS)


**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

**Nicholas Evan**
**00000027900**

**PROGRAM STUDI INFORMATIKA**
**FAKULTAS TEKNIK DAN INFORMATIKA**
**UNIVERSITAS MULTIMEDIA NUSANTARA**
**TANGERANG**
**2023**

# DETEKSI KESALAHAN EJA KATA LULUH PADA BERITA DENGAN ALGORITMA JACCARD SIMILARITY (STUDI KASUS : TRIBUNNEWS)


**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

**Nicholas Evan**
**00000027900**



**PROGRAM STUDI INFORMATIKA**
**FAKULTAS TEKNIK DAN INFORMATIKA**
**UNIVERSITAS MULTIMEDIA NUSANTARA**
**TANGERANG**
**2023**

## HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

| | | |
|---|---|---|
| Nama | : | Nicholas Evan |
| Nomor Induk Mahasiswa | : | 00000027900 |
| Program Studi | : | Informatika |

Skripsi dengan judul:

**Deteksi Kesalahan Eja Kata Luluh pada Berita dengan Algoritma Jaccard Similarity (Studi Kasus : Tribunnews)**

merupakan hasil karya saya sendiri bukan plagiat dari karya ilmiah yang ditulis oleh orang lain, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/ penyimpangan, baik dalam pelaksanaan Skripsi maupun dalam penulisan laporan Skripsi, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk Tugas akhir yang telah saya tempuh.

Tangerang, 3 Januari 2023

(Nicholas Evan)

## HALAMAN PENGESAHAN

Skripsi dengan judul

**DETEKSI KESALAHAN EJA KATA LULUH PADA BERITA DENGAN ALGORITMA JACCARD SIMILARITY (STUDI KASUS : TRIBUNNEWS)**

oleh

| | | |
|---|---|---|
| Nama | : | Nicholas Evan |
| NIM | : | 00000027900 |
| Program Studi | : | Informatika |
| Fakultas | : | Fakultas Teknik dan Informatika |

Telah diujikan pada hari Senin, 9 Januari 2023
Pukul 10.00 s/s 11.30dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut

Ketua Sidang

(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)
NIDN: 0311106903

Penguji

(Angga Aditya Permana, S.Kom., M.Kom.)
NIDN: 0407128901

Pembimbing

Digitally signed by Marlinda Vasty Overbeek
Date: 2023.01.18 12:50:45 +07'00'

(Marlinda Vasty Overbeek, S.Kom, M.Kom)
NIDN: 0818038501

Ketua Program Studi Informatika,

Digitally signed by Marlinda Vasty Overbeek
Date: 2023.01.18 12:50:59 +07'00'

(Marlinda Vasty Overbeek, S.Kom., M.Kom.)
NIDN: 0818038501

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Multimedia Nusantara, saya yang bertanda tangan di bawah ini:

| | | |
|---|---|---|
| Nama | : | Nicholas Evan |
| NIM | : | 00000027900 |
| Program Studi | : | Informatika |
| Fakultas | : | Teknik dan Informatika |
| Jenis Karya | : | Skripsi |

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada **Universitas Multimedia Nusantara** hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

**DETEKSI KESALAHAN EJA KATA LULUH PADA BERITA DENGAN ALGORITMA JACCARD SIMILARITY (STUDI KASUS : TRIBUNNEWS)**

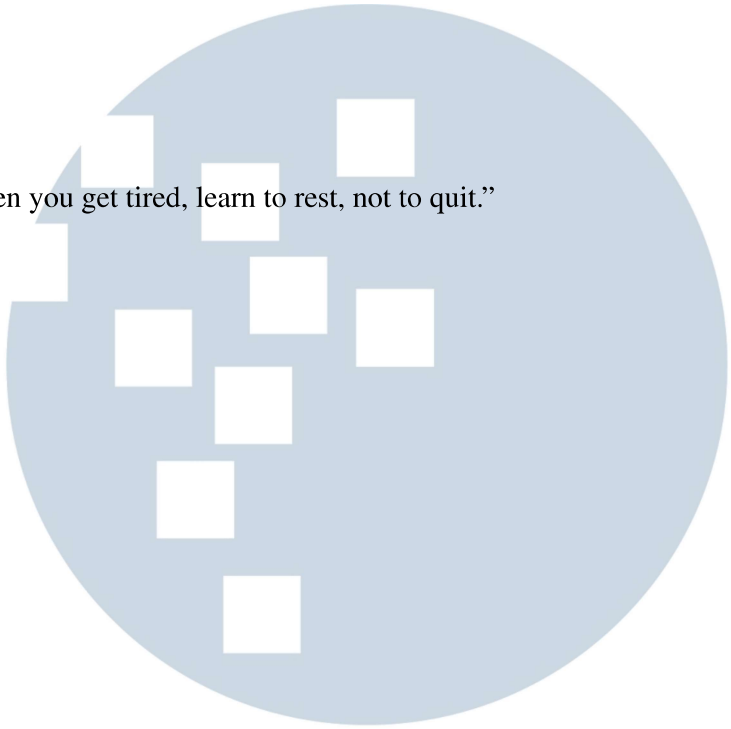
Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Multimedia Nusantara berhak menyimpan, mengalih media / format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 3 Januari 2023

Yang menyatakan

Nicholas Evan

**Halaman Persembahan / Motto**



”When you get tired, learn to rest, not to quit.”

Banksy

# KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Deteksi Kesalahan Eja Kata Luluh pada Berita dengan Algoritma Jaccard Similarity (Studi Kasus : Tribunnews) dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Marlinda Vasty Overbeek, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara dan Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
4. Keluarga dan teman-teman yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.
5. Jessica Augustine S. yang telah memberikan dukungan dan bantuan moral, selama pengerjaan tugas akhir ini.
6. Jerico Olwen, selaku teman seperjuangan selama pengerjaan tugas akhir.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 3 Januari 2023

Nicholas Evan

**DETEKSI KESALAHAN EJA KATA LULUH PADA BERITA DENGAN ALGORITMA JACCARD SIMILARITY (STUDI KASUS : TRIBUNNEWS)**

Nicholas Evan

## ABSTRAK

Bahasa Indonesia merupakan bahasa nasional yang digunakan dalam kehidupan sehari-hari, namun kesalahan berbahasa kerap terjadi dalam di sekitar kita, salah satunya pada portal berita *online*. Kesalahan berbahasa merupakan penyimpangan bahasa dari kaidah tata bahasa dan salah satunya adalah peluluhan fonem. Hal ini terjadi akibat penulisan dilakukan secara *manual* sehingga memungkinkan untuk terjadinya kesalahan pengetikan. Dengan terjadinya kesalahan peluluhan ini, dilakukanlah penelitian yaitu pembuatan sistem dengan menggunakan algoritma *Jaccard Similarity* untuk mendeteksi kesalahan eja pada kata terluluh. *Jaccard Similarity* merupakan algoritma yang digunakan untuk membandingkan dokumen untuk menghitung kesamaan nilai dari dua dokumen. Evaluasi dilakukan dengan menggunakan *confusion matrix* yang kemudian diambil F-1 *score*-nya selain itu efisiensi sistem juga diperhitungkan. Hasil deteksinya memiliki F-1 *score* sebesar 66.6% dan efisiensi sistem dipengaruhi oleh jumlah kalimat dan jumlah kata yang terluluh. Sistem yang dibangun dapat mendeteksi kesalahan eja pada kata terluluh saat dihadapkan dengan berita dari portal berita Tribun.

**Kata kunci**: berita, *Jaccard Similarity*, kesalahan eja, peluluhan fonem, sistem deteksi

**Detection of Spelling Errors in Indonesian Language News with Jaccard Similarity Algorithm (Case : Tribun News )**

Nicholas Evan

## ABSTRACT

Language is an organized communication tool in the form of units such as words, groups of words, clauses and sentences. Indonesian is the national language which should be used in accordance with Enhanced Indonesian Spelling, but language errors often occur accidentally on online news portals. Language errors are language deviations from grammatical rules and one of them is phoneme decay. This may occur due to writing done manually so that it allows typing errors. With the occurrence of this error, research was carried out, namely making a system using the Jaccard Similarity algorithm to detect misspellings in melted words. Jaccard Similarity is an algorithm used to compare documents to calculate the similarity of values from two documents. Evaluation is done by using the confusion matrix which then takes the F-1 score besides that system efficiency is also taken into account. The detection results have an F-1 score of 66.6% and system efficiency is influenced by the number of sentences and the number of words decayed. The system built can detect misspelled words when dealing with news from the Tribun news portal.

**Keywords**: *detection system, Jaccard Similarity, misspelling, news, phoneme decay*

# DAFTAR ISI

# DAFTAR GAMBAR

# DAFTAR TABEL

# DAFTAR KODE

## DAFTAR LAMPIRAN