PEMILIHAN KOMBINASI FITUR UNTUK *AUTHOR VERIFICATION* DENGAN METODE *CLUSTERING* K-MEANS PADA TEKS BERBAHASA INDONESIA

Skripsi

Diajukan oleh:

THOMAS WIDIARYA BUDIMAN

71160018

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS KRISTEN DUTA WACANA

YOGYAKARTA

2020

PEMILIHAN KOMBINASI FITUR UNTUK *AUTHOR VERIFICATION* DENGAN METODE *CLUSTERING* K-MEANS PADA TEKS BERBAHASA INDONESIA

Skripsi

Diajukan kepada Fakultas Teknologi Informasi Program Studi Informatika
Universitas Kristen Duta Wacana
Sebagai salah satu syarat dalam memperoleh gelar Sarjana Komputer

Diajukan oleh:
THOMAS WIDIARYA BUDIMAN
71160018

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS KRISTEN DUTA WACANA
YOGYAKARTA
2020

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI/TESIS/DISERTASI UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Kristen Duta Wacana, saya yang bertanda tangan di bawah ini:

Nama	: Thomas Widiarya Budiman
NIM	: 71160018
Program studi	: Informatika
Fakultas	: Teknologi Informasi
Jenis Karya	: Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Kristen Duta Wacana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

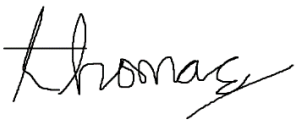
"PEMILIHAN KOMBINASI FITUR UNTUK AUTHOR VERIFICATION DENGAN METODE CLUSTERING K-MEANS PADA TEKS BERBAHASA INDONESIA"

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Kristen Duta Wacana berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama kami sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Yogyakarta
Pada Tanggal : 12 April 2020

Yang menyatakan



Thomas Widiarya Budiman
NIM.71160018

PERNYATAAN KEASLIAN SKRIPSI

Saya menyatakan dengan sesungguhnya bahwa skripsi dengan judul:

PEMILIHAN KOMBINASI FITUR UNTUK AUTHOR VERIFICATION DENGAN METODE CLUSTERING K-MEANS PADA TEKS BERBAHASA INDONESIA

yang saya kerjakan untuk melengkapi sebagian persyaratan menjadi Sarjana Komputer pada pendidikan Sarjana Program Studi Informatika Fakultas Teknologi Informasi Universitas Kristen Duta Wacana, bukan merupakan tiruan atau duplikasi dari skripsi kesarjanaan di lingkungan Universitas Kristen Duta Wacana maupun di Perguruan Tinggi atau instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.

Jika dikemudian hari didapati bahwa hasil skripsi ini adalah hasil plagiasi atau tiruan dari skripsi lain, saya bersedia dikenai sanksi yakni pencabutan gelar kesarjanaan saya.

Yogyakarta, 10 Juli 2020

THOMAS WIDIARYA BUDIMAN
71160018

iii

HALAMAN PERSETUJUAN

Judul	:	Pemilihan Kombinasi Fitur untuk *Author Verification* dengan Metode *Clustering K-Means* pada Teks Berbahasa Indonesia
Nama	:	Thomas Widiarya Budiman
NIM	:	71160018
Mata Kuliah	:	Skripsi
Kode	:	TI0366
Semester	:	Genap
Tahun akademik	:	2019/2020

Telah diperiksa dan disetujui
Di Yogyakarta,
Pada Tanggal 7 Juli 2020

Dosen Pembimbing I	Dosen Pembimbing II
Dr. Phil. Lucia Dwi K., SS., M.A.	Laurentius Kuncoro P. S., S.T., M.Eng

HALAMAN PENGESAHAN

PEMILIHAN KOMBINASI FITUR UNTUK AUTHOR VERIFICATION DENGAN METODE CLUSTERING K-MEANS PADA TEKS BERBAHASA INDONESIA

Oleh: THOMAS WIDIARYA BUDIMAN / 71160018

Dipertahankan di depan Dewan Penguji Skripsi
Program Studi Informatika Fakultas Teknologi Informasi
Universitas Kristen Duta Wacana - Yogyakarta
Dan dinyatakan diterima untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer
pada tanggal 23 Juli 2020

Yogyakarta, 28 Juli 2020
Mengesahkan,

Dewan Penguji:

1. Lucia Dwi Krisnawati, Dr. Phil.
2. Laurentius Kuncoro Probo Saputra, S.T., M.Eng.
3. Willy Sudiarto Raharjo, S.Kom.,M.Cs.
4. Aditya Wikan Mahastama, S.Kom., M.Cs.

Dekan

(Restyandito,S.Kom.,MSIS.,Ph.D.)

Ketua Program Studi

(Gloria Virginia,Ph.D.)

UCAPAN TERIMA KASIH

Puji syukur senantiasa peneliti panjatkan kepada Tuhan Yang Maha Esa atas berkat, penyertaan, dan rahmat-Nya, sehingga peneliti bisa menyelesaikan skripsi. Peneliti juga berterimakasih atas dukungan, saran dan bimbingan selama proses pengerjaan skripsi ini kepada:

1. Bapak Restyandito, S.Kom., MSIS, Ph.D. selaku Fakultas Teknologi Informasi UKDW.
2. Ibu Gloria Virginia, S.Kom., MAI, Ph.D. selaku Ketua Program Studi Informatika Fakultas Teknologi Informasi UKDW.
3. Ibu Dr. Phil. Lucia Dwi K.,SS., M.A. selaku dosen pembimbing I yang bersedia membantu dan memberikan arahan serta dukungan dari awal perencanaan penelitian hingga skripsi ini selesai. Saya sangat berterimakasih karena sudah menyempatkan waktu untuk konsultasi secara luring walaupun sedang terjadi pandemic Covid-19.
4. Bapak Laurentius Kuncoro Probo Saputro, S.T., M.Eng. selaku dosen pembimbing II yang bersedia membantu dan memberikan arahan serta dukungan dari awal perencanaan penelitian hingga skripsi ini selesai.
5. Seluruh dosen di Fakultas Teknologi Informasi, khususnya dosen Program Studi Informatika yang telah memberikan bekal pengetahuan selama penulis menempuh pendidikan di Universitas Kristen Duta Wacana.

Peneliti berharap semoga pengorbanan dan segala sesuatunya yang dengan tulus dan ikhlas telah diberikan akan selalu mendapat limpahan rahmat-Nya.

INTISARI

Pemilihan Kombinasi Fitur untuk *Author Verification*
dengan Metode *Clustering K-Means* pada Teks Berbahasa Indonesia

Penipuan dengan menggunakan identitas orang lain umumnya dilakukan melalui media tertulis, karena pelaku tidak perlu memperlihatkan fisik maupun suaranya. Maka dari itu diperlukan sistem yang bisa membedakan identitas tulisan seseorang. Identitas tulisan seseorang bisa dianalisa menggunakan fitur *stylometry*, fitur ini merupakan salah satu faktor sistem verifikasi penulis bisa berjalan dengan baik. Penelitan ini mengasumsikan jika kombinasi fitur *stylometry* berhasil merepresentasikan gaya penulisan seseorang maka proses klasterisasi atau klasifikasi dokumen berdasarkan gaya penulisan seseorang akan menghasilkan nilai evaluasi klasterisasi (purity) dan klasifikasi (akurasi, presisi, sensitivitas, *FScore*) yang memuasakan. Hasil penelitian dengan dokumen berbahasa Indonesia pada *klasifier* MKNN, KNN, dan SVM menunujukan bahwa nilai purity bisa menggambarkan nilai evaluasi klasifikasi walaupun tidak terlalu tepat, kombinasi fitur mendapatkan nilai purity kurang dari 0,5 memiliki kemungkinan besar untuk mendapatkan nilai evaluasi klasifikasi kurang dari 0,5 juga, hal ini bisa menghemat proses jika kombinasi fitur dengan nilai purity kurang dari 0,5 tidak digunakan untuk proses klasifikasi. Kombinasi fitur yang mengandung fitur frekuensi relatif tanda baca (Fitur Sintaksis), frekuensi relatif stopword (Fitur Sintaksis) cenderung memiliki nilai evaluasi klasifikasi maupun klasterisasi yang lebih baik dari pada kombinasi fitur yang tidak mengandung kedua fitur tersebut dan ketika ditambah dengan fitur rata-rata panjang paragraf (Fitur Struktural) kombinasi fitur ini menjadi kombinasi fitur terbaik pada semua *klasifier*.

Kata kunci: identitas tulisan seseorang, verifikasi penulis, fitur *stylometry*

ABSTRACT

*Selection of Feature Combinations for Author Verification with K-Means Clustering Method in Indonesian Language Tex*t

Fraud by using other people's identities is generally done through written media, because the perpetrators do not need to show physical or voice. Therefore a system that can distinguish someone's writing identity is needed. The identity of a person's writing can be analyzed using the stylometry feature, this feature is one of the factors the author's verification system can work well. This research assumes that if a combination of stylometry features successfully represents a person's writing style, the clustering or classification of documents based on one's writing style will result in a satisfactory classification metric (purity) and classification metric (accuracy, precision, sensitivity, FScore). The results of research with Indonesian language documents in the MKNN, KNN, and SVM classifiers show that the value of purity can describe the value of classification evaluation even though it is not very precise, a combination of features getting a purity value of less than 0.5 has a high likelihood to get a classification evaluation value of less than 0.5 also, this can save the process if a combination of features with a purity value of less than 0.5 is not used for the classification process. Combinations of features that contain punctuation relative frequency features (Syntactic Features), stopword relative frequencies (Syntax Syntactic) tend to have better classification and classification evaluation values than feature combinations that do not contain both features and when added to the average length feature paragraph (Structural Features) this combination of features becomes the best combination of features in all classifiers.

keywrods: *author writing identity*, *author verification*, *stylometry feature*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1 Latar Belakang

Penggunaan identitas palsu atau identitas orang lain merupakan sebuah tindakan yang sudah tidak asing lagi, hal ini dikarenakan kemudahan seseorang untuk mendapatkan identitas orang lain melalui dunia maya. Karena kemudahan ini banyak tindakan kriminal yang bisa saja terjadi di dunia maya dengan menggunakan identitas orang lain. Contoh kasus tindakan kriminal yang dapat dilakukan dengan menggunakan identitas orang lain adalah kasus penipuan melalui aplikasi sosial media dengan menggunakan identitas palsu (Amiruddin, 2019)

Penipuan dengan menggunakan identitas orang lain umumnya dilakukan melalui media tertulis. Media tertulis dianggap aman karena pelaku tidak perlu memperlihatkan fisik maupun suaranya. Namun faktanya gaya penulisan juga dapat dijadikan sebagai penanda identitas tulisan seseorang. J. K. Rowling yang menggunakan nama samaran Robert Galbraith ketika menulis novel The Cuckoo's Calling terungkap identitas tulisanya ketika gaya penulisannya dianalisis dengan menggunakan sistem verifikasi penulis (*Author Verification*) oleh Juloa (2013). Sistem verifikasi penulis merupakan sistem yang memiliki tugas untuk menjawab pertanyaan ‘apakah penulis *x* yang menulis dokumen ini?’ (Luyckx & Daelemans, 2008). Salah satu faktor yang membuat sistem verifikasi penulis ini bekerja dengan baik adalah fitur gaya penulisan *stylometry* yang digunakan, fitur *stylometry* yang bisa merepresentasikan gaya penulisan seseorang akan membuat sistem verifikasi penulis bekerja lebih baik.

Berangkat dari kasus J. K. Rowling, penelitian ini berfokus membangun sistem verifikasi penulis pada teks berbahasa Indonesia dengan melakukan uji coba berbagai kombinasi fitur gaya penulisan (*stylometry*) dan ingin melihat hubungan antara nilai purity (hasil klasterisasi dengan KMeans) dengan nilai akurasi, presisi, sensitivitas dan *Fscore* (hasil klasifikasi dengan MKNN) pada masing-masing kombinasi fitur yang di teliti. Penelitan ini mengasumsikan jika kombinasi fitur *stylometry* yang dipilih merupakan kombinasi fitur yang merepresentasikan gaya

penulisan seseorang maka proses klasterisasi maupun klasifikasi dokumen berdasarkan penulis akan menghasilkan nilai evaluasi yang memuasakan.

1.2 Rumusan Masalah

Melihat latar belakang yang ada, berikut adalah rumusan masalah dalam penelitian ini:

1.2.1 Kombinasi fitur gaya penulisan (*stylometry*) apa saja yang dapat digunakan untuk memverifikasi penulis dalam dokumen berbahasa Indonesia, sehingga ketika dilakukan klasterisasi dengan K-means, dokumen berbahasa Indonesia ini akan terklaster dengan tepat berdasarkan gaya penulisan seseorang ?

1.2.2 Apakah kombinasi fitur gaya penulisan (*stylometry*) memiliki nilai purity yang berbanding lurus dengan nilai metode evaluasi klasifikasi pada metode MKNN ?

1.3 Batasan Masalah

Dalam penelitian ini terdapat batasan-batasan masalah seperti berikut:

a Sumber dokumen berbentuk opini dan diambil dari website Seword.com yang di publikasikan pada tahun 2019 sampai 2020.

b Dokumen yang digunakan merupakan dokumen berbahasa Indonesia.

c Setiap dokumen hanya ditulis oleh 1 penulis.

d Jumlah penulis 11 orang dengan 25 dokumen masing-masing penulis.

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menemukan kombinasi fitur yang dapat merepresentasikan gaya penulisan seseorang pada dokumen berbahasa Indonesia dan untuk melihat hubungan nilai purity dengan nilai f1score yang dimiliki oleh masing-masing kombinasi fitur gaya penulisan.

1.5 Manfaat Penelitian

Manfaat yang ingin didapat dari penelitian ini adalah sebagai berikut:

1.5.1 Penelitian ini diharapkan mampu mendorong terbentuknya sistem verifikasi penulis yang dapat membantu pemerintah Indonesia untuk mengatasi kejahatan di dunia maya.

1.5.2 Diharapkan sistem yang dibangun ini dapat membantu penelitian di bidang teks forensik selanjutnya untuk mendapatkan kombinasi fitur yang dapat merepresentasikan gaya penulisan seseorang sehingga proses pengerjaan penelitian selanjutnya dapat dipercepat dengan adanya penelitian ini.

1.6 Metodologi Penelitian

Berikut merupakan metode-metode yang peneliti lakukan untuk memecahkan permasalah yang peneliti tulis di latar belakang.

1.6.1 Studi Pustaka

Tahap ini adalah merupakan tahap awal dari penelitian yang dilakukan oleh peneliti, dimana peneliti membaca jurnal dan situs-situs yang mengangkat tentang verifikasi penulis.

1.6.2 Pengumpulan Data

Tahap selanjutnya adalah tahap pengumpulan data, pengumpulan data dilakukan secara semi otomatis, yaitu dengan bantuan Bahasa pemrograman python. Mula-mula peneliti mengambil kunci API yang dimiliki oleh website seword.com, selanjutnya peneliti mengambil beberapa *username* dari beberapa penulis seword.com, lalu peneliti membuat script untuk menembak API yang disediakan oleh seword.com untuk mengambil data tulisan dari masing-masing penulis berdasarkan *username* yang dipilih oleh peneliti. Data yang diambil merupakan data yang berbentuk opini, berbahasa Indonesia, bisa saja memiliki topik yang berbeda dan ditulis oleh penulis tunggal. Data yang diambil adalah data memiliki genre yang sama yaitu opini karena akan lebih susah untuk melakukan verifikasi penulis pada banyak genre (Kestemont, Luyckx, Daelemans, & Crombez, 2012).

1.6.3 Pengembangan Sistem

Setelah data terkumpul tahap selanjutnya adalah membangun sistem, tahap ini meliputi pembagian data, normalisasi data, prapremrosesan data, ektraksi fitur, kombinasi fitur, klasterisasi dan klasifikasi. Secara garis besar tahap ini merupakan tahap pembangunan sistem verifikasi.

1.6.4 Evaluasi dan Validasi Sistem

Setelah sistem verifikasi terbentuk tahap selanjutnya adalah melakukan verifikasi dan validasi terhadap sistem tersebut. Peneliti melakukan beberapa skenario percobaan untuk menilai kinerja dari sistem yang sudah dibangun ini. Tahap validasi pada penelitian ini menggunakan KFold yang ter-*stratified*. Sedangkan tahap evaluasi klasterisasi menggunakan purity dan klasifikasi menggunakan akurasi, presisi, sensitivitas, dan *FScore*.

1.6.5 Analisa Hasil Evaluasi

Tahap ini merupakan tahap terakhir dalam penelitian ini, tahap ini merupakan tahap dimana peneliti melakukan analisa berdasarkan hasil evaluasi yang didapatkan pada pengujian yang dilakukan pada penelitian ini.

1.7 Sistematika Penulisan

BAB I PENDAHULUAN membahas tentang gambaran umum yang akan dilakukan pada penelitian ini. Gambaran umum pada bab ini meliputi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian dan sistematika penulisan. Pada sub bab 1.1 menjelaskan tentang kasus nyata mengenai verifikasi penulis dan rencana penelitian secara umum. Sub bab 1.2 menjelaskan permasalah apa yang menjadi fokus dalam penelitian ini, pada sub bab 1.3 berisi tentang ruang lingkup dari penelitian ini. Tujuan penelitian tertulis pada sub bab 1.4 sedangkan manfaat dari penelitian ini ditulis pada sub bab 1.5. Pada sub bab 1.6 dijelaskan metode apa saja yang digunakan dalam penelitian ini, sedangkan pada sub bab 1.7 berfokus pada menjelaskan seluruh isi bab yang ada dalam penelitian ini secara singkat.

BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI mencakup 2 bagian utama yaitu tinjauan pustaka dan landasan teori. Sub bab 2.1 berisi tentang tinjauan dari berbagai penelitian sebagai pendukung penelitian yang akan dilakukan dan perbedaan penelitian ini dengan penelitian yang sudah dilakukan sebelumnya. Sub bab 2.2 berisi tentang landasan dasar teori yang menjadi prinsip utama ataupun konsep dilakukannya penelitian ini.

BAB III PERANCANGAN SISTEM membahas tentang perancangan sistem verifikasi penulis yang dibangun selama proses penelitian. pada sub bab 3.1 akan menjelaskan tentang kebutuhan perangkat keras maupun perangkat lunak dalam membangun sistem. Sub bab selanjutnya yaitu sub bab 3.2 menjelaskan tentang perancangan sistem verifikasi penulis. Pada sub bab 3.3 akan dijelaskan mengenai alur sistem beserta desain antarmukanya. Pada sub bab 3.4 akan dijelaskan rancangan pengujian sistem.

BAB IV IMPLEMENTASI DAN ANALISIS DATA mencakup tentang detail implementasi dari perancangan sistem penelitian. Sub bab 4.1 akan memaparkan hasil implementasi sistem dari penelitian ini. Sub bab 4.2 memaparkan hasil implementasi desain antarmuka yang dibangun. Pada sub bab berikutnya yaitu sub bab 4.3 akan dijelaskan hasil dan analisis pengujian dari penelitian.

BAB V KESIMPULAN DAN SARAN membahas tentang penarikan kesimpulan dari hasil penelitian ini yang akan ditulis pada sub bab 5.1 dan sub bab 5.2 berisi saran-saran yang dapat digunakan pada penelitian yang berkaitan di masa yang akan datang.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil pengujian dalam penelitian ini beserta analisa yang peneliti buat, peneliti mendapatkan beberapa kesimpulan sebagai berikut

a. Kombinasi fitur *stylometry* dari fitur frekuensi relatif tanda baca (Fitur Sintaksis), frekuensi relatif *stopword* (Fitur Sintaksis), rata-rata panjang paragraf (Fitur Struktural) merupakan kombinasi fitur *stylometry* terbaik yang ditemukan pada penelitian ini, pernyataan ini dibuktikan dengan kombinasi fitur *stylometry* ini mendapatkan nilai voting terbanyak pada percobaan pengujian pertama, ketiga dan keempat. Selain itu kombinasi fitur ini mendapatkan nilai *FScore* terbaik pada pengujian pertama, keempat dengan *klasifier* KNN, dan keempat dengan *klasifier* SVM dengan nilai 0,86, 0,91, dan 0,88.

b. Kombinasi fitur *stylometry* yang mengandung fitur frekuensi relatif tanda baca (Fitur Sintaksis), frekuensi relatif *stopword* (Fitur Sintaksis) cenderung memiliki nilai evaluasi klasifikasi maupun klasterisasi yang lebih baik dari pada kombinasi fitur *stylometry* yang tidak mengandung kedua fitur tersebut. Hal ini ditunjukan pada kombinasi fitur *stylometry* dengan kode adalah ‘KF1’, ‘KF2’, ‘KF3’, dan ‘KF4’ hampir selalu menempati peringkat 1 sampai 4 pada proses voting di percobaan pertama, percobaan ketiga, dan percobaan keempat. Hal ini disebabkan karena keempat kombinasi fitur ini berhasil menghasilkan nilai *FScore* dan nilai *purity* yang tinggi.

c. Nilai purity memberikan gambaran mengenai hasil evaluasi klasifikasi walaupun tidak terlalu tepat hal ini dapat dilihat ketika nilai purity pada suatu kombinasi fitur *stylometry* lebih besar dari kombinasi fitur *stylometry* lain, tidak menjamin kombinasi fitur *stylometry* tersebut memiliki nilai evaluasi klasifikasi yang lebih besar juga.

d. Kombinasi fitur *stylometry* yang memiliki nilai purity kurang dari 0,5 memiliki kemungkinan yang besar akan mengasilkan nilai evaluasi

klasifikasi kurang dari 0,5 juga. Tetapi Kombinasi fitur *stylometry* yang memiliki nilai purity lebih dari sama dengan 0,5 memiliki kemungkinan yang besar akan mengasilkan nilai evaluasi klasifikasi lebih dari sama dengan 0,5 juga.

e. Kombinasi fitur yang memiliki nilai purity tertinggi hampir selalu mendapatkan voting terbanyak dan memiliki nilai evaluasi klasifikasi yang tertinggi juga.

f. Kombinasi fitur *stylometry* yang dijelaskan pada penelitian Halvani (2015) dan Luyckx dan Daelemans (2008) bisa digunakan untuk teks bahasa Indonesia.

g. Fitur sintaksis juga bisa bekerja lebih baik daripada fitur pada level token lainya pada teks bahasa Indonesia.

5.2 Saran

Saran dari peneliti yang peneliti dapatkan setelah melakukan penelitian ini adalah sebagai berikut.

a. Mencoba fitur *Part-of-speech* dan frekuensi relatif *Part-of-speech.*

b. Mencoba fitur pola kemunculan *stopword*.

c. Mencoba kombinasi fitur *stylometry* yang isinya keseluruhan merupakan fitur dengan jenis sintaksis.

d. Mencoba model klasifikasi lain apakah hasil dari klasifikasi juga berbanding lurus dengan hasil dari klasterisasi.

e. Mencoba verifikasi penulis dengan paradigma berbasis profil.

f. Mengatur nilai H pada MKNN untuk meningkatkan hasil klasifikasi.

Daftar Pustaka

Amiruddin, H. (2019, Februari 21). *Polisi Tangkap 3 Penipu Online Jaringan Internasional*. Diambil kembali dari www.news.okezone.com: https://news.okezone.com/read/2019/02/21/609/2021204/polisi-tangkap-3-penipu-online-jaringan-internasional

Atmoko, L. (2018). *Sistem Analisis Gaya Penulisan Untuk Pengelompokan Dokumen Dengan Metode K-means.* Diambil kembali dari http://sinta.ukdw.ac.id.

Diez, P. (2018). *Smart Wheelchairs and Brain-Computer Interfaces.* Argentina: Elsevier.

Gonçalves, D. N., Gonçalves, C. d., Assis, T. F., & Silva, M. A. (2014). Analysis of the difference between the Euclidean distance and the actual road distance in Brazil. *Transportation Research Procedia 3*, 876 – 885.

Halvani, O. (2015). Register & Genre Seminar: Towards Intrinsic Plagiarism Detection.

Indraloka, D. S., & Santosa, B. (2017). Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia. *JURNAL SAINS DAN SENI ITS*, 2337-3520.

Juola, P. (2013, Agustus 20). *How a Computer Program Helped Show J.K. Rowling write A Cuckoo's Calling*. Diambil kembali dari www.scientificamerican.com: https://www.scientificamerican.com/article/how-a-computer-program-helped-show-jk-rowling-write-a-cuckoos-calling/

Kestemont, M., Luyckx, K., Daelemans, W., & Crombez, T. (2012). Cross-Genre Authorship Verification Using Unmasking. *English Studies*, 340–356.

Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia 25*, 1104 – 1109.

Luyckx, K., & Daelemans, W. (2008). Authorship Attribution and Verification with Many Authors and Limited. *Proceedings of the 22nd International Conference on Computational Linguistics*, 513-520.

Parvin, H., Alizadeh, H., & Minaei-Bidgoli, B. (2008). MKNN: Modified K-Nearest Neighbor. *Proceedings of the World Congress on Engineering and Computer Science*, 22 - 24.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825-2830.

Potha, N., & Stamatatos, E. (2014). A Profile-Based Method for Authorship Verification. *Springer International Publishing Switzerland*, 313–326.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management 45*, 427–437.

Sripada, S. C., & Rao, D. M. (2011). COMPARISON OF PURITY AND ENTROPY OF K-MEANS CLUSTERING AND FUZZY C MEANS CLUSTERING. *Indian Journal of Computer Science and Engineering (IJCSE)*, 343-346.

Stamatatos, E., Tschuggnall, M., Verhoeven, B., Daelemans, W., Specht, G., Stein, B., & Potthast, M. (2016). Clustering by Authorship Within and Across Documents. *CLEF*, 691-715.

Vartapetiance, A., & Gillam, L. (2016). A Big Increase in Known Unknowns: from Author Verification to Author Clustering. *CLEF*, 1008-1013.

Wafiyah, F., Hidayat, N., & Perdana, R. S. (2017). Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 1210-1219.

Zainuddin, N., & Selamat, A. (2014). Sentiment Analysis Using Support Vector Machine. *International Conference of Computer, Communications, and Control Technology* (hal. 333-337). Langkawi: IEEE.