

**IMPLEMENTASI K-MEANS CLUSTERING PADA TERJEMAHAN
AL-QUR'AN BERDASARKAN KETERKAITAN TOPIK**

**Diajukan sebagai salah satu syarat memperoleh gelar strata satu Program
Studi Teknik Informatika**



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

Disusun Oleh :
Ahmad Salam Wahid Faizin
NIM:12650026

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA
2018**



Nomor : B-166/UIN.02/D.ST/PP.01.1/05/2018

Dr. Murtono, M.Si
NIP. 19691212 200003 1 001



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal :

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Ahmad Salam Wahid Faizin

NIM : 12650026

Judul Skripsi : Implementasi K-Means Clustering Pada Terjemahan

Al-Qur'an Berdasarkan Keterkaitan Topik

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Teknik Informatika.

Dengan ini kami mengharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 08 Mei 2018

Pembimbing

Muhammad Didik Rohmad Wahyudi, S.T., MT.

NIP. 197608122009011015

PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini:

Nama : Ahmad Salam Wahid Faizin

NIM : 12650026

Program Studi : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi dengan judul **"Implementasi K-Means Clustering Pada Terjemahan Al-Qur'an Berdasarkan Keterkaitan Topik"** tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 08 Mei 2018



Menyatakan

Ahmad Salam Wahid Faizin

NIM. 12650026

KATA PENGANTAR

Alhamdulillahirabilalamin, puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah serta inayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi *K-Means Clustering* Pada Terjemahan *Al-Qur'an* Berdasarkan Keterkaitan Topik” ini dengan baik sesuai dengan kewajiban dalam memenuhi gelar Strata 1 Komputer (S.Kom) di Jurusan Teknik Informatika Fakultas Sains dan teknologi Universitas Islam Negeri Sunan Kalijaga Yogyakarta. Tidak lupa shalawat serta salam tetap tercurah kepada junjungan Nabi Muhammad SAW dan semoga kelak kita mendapat syafaat darinya.

Penulis menyadari bahwa dalam penyusunan skripsi ini tidak lepas dari dukungan baik semangat maupun materiil yang diberikan sehingga skripsi dapat terselesaikan. Oleh karena itu, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Prof. Drs. K.H. Yudian Wahyudi, M.A., Ph.D. selaku Rektor Universitas Islam Negeri Sunan Kalijaga Yogyakarta.
2. Bapak Dr. Murtono, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga.
3. Bapak M. Didik Rohmad Wahyudi, ST., MT sebagai Dosen Pembimbing yang telah dengan sabar membimbing penulis dalam penyusunan skripsi, dan terima kasih pula karena telah memberikan arahan, saran, waktu serta masukan kepada penulis dalam menyusun skripsi.
4. Bapak Dr. Bambang Sugiantoro, M.T., selaku Ketua Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
5. Bapak Aulia Faqih Rifa'i, M.Kom., selaku dosen pembimbing akademik kelas reguler Teknik Informatika 2012.
6. Dosen Program Studi Informatika UIN Sunan Kalijaga yang telah memberikan banyak bekal ilmu kepada penulis.

7. Teman-teman Teknik Informatika yang tidak dapat disebutkan satu persatu yang telah sedikit banyak memberikan bantuan, dukungan, serta motivasi kepada penulis.
8. Semua pihak yang telah memberikan bantuan dan dukungan selama menempuh strata satu teknik informatika khususnya dalam penyusunan skripsi ini yang tidak dapat disebut satu persatu. Terima kasih.

Akhirnya penyusun hanya dapat bersyukur kepada Allah semoga semua yang telah dilakukan selama ini menjadi amal dan bekal di akhirat nanti. Penyusun menyadari sepenuhnya masih banyak kesalahan dan kekurangan dalam skripsi ini, maka dari itu berbagai saran dan kritik sangat diharapkan demi perbaikan. Semoga skripsi ini dapat bermanfaat bagi penyusun sendiri pada khususnya dan bagi para pembaca pada umumnya. Terima kasih.

Yogyakarta, 08 Mei 2018

Penulis

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

Laporan skripsi ini saya persembahkan kepada :

1. Prodi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.
2. Orang tua tercinta, Bapak Amat Khasani dan Ibu Siti Zulaiha, terimakasih atas doa, dukungan serta kasih sayang yang tiada terkira.
3. Adik Muhammad Malik Mubarak, yang memberikan dukungan dalam penyusunan skripsi.
4. Terima kasih kepada teman-teman yang memberikan dukungan penuh dalam pembuatan skripsi Aziz, Alfian, Mustafid, Puguh, Fauzi, Pamuji, Anwar, Wahib, Yaumi, Mia, Sasti.
5. Teman-teman Teknik Informatika 2012 yang tidak bisa disebut satu persatu.
6. Teman-teman kos Bapak Sudarto.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN MOTTO

من جدّ وجد
من صبر ظفر
من سار على الدرب وصل

Barang Siapa Bersungguh-sungguh, maka dia akan mendapat (kesuksesan).

Barang Siapa Bersabar, maka dia akan beruntung.

Barang Siapa Berjalan pada jalannya, maka dia akan sampai (pada tujuannya).

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

PENGESAHAN SKRIPSI/TUGAS AKHIR	ii
SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR	iii
PERNYATAAN KEASLIAN SKRIPSI.....	iv
KATA PENGANTAR	v
HALAMAN PERSEMBAHAN	vii
HALAMAN MOTTO	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR PERSAMAAN	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan Masalah.....	2
1.4. Tujuan Penelitian.....	2
1.5. Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA DAN LANDASAN TEORI	4
2.1. Tinjauan Pustaka	4
2.2. Landasan Teori	5
2.2.1. Data Mining	5
2.2.2. Clustering.....	6
2.2.3. Algoritme K-Means	7
2.2.3.1. Penentuan Jumlah Cluster	16
2.2.3.2. Seed	17
2.2.4. Text mining.....	18
2.2.5. Preprocessing	20
2.2.5.1. Pemecahan Kalimat.....	20
2.2.5.2. Case Folding	21

2.2.5.3.	Tokenizing.....	21
2.2.5.4.	Stoplist / Stop Word Removal	22
2.2.5.5.	Stemming	22
2.2.5.6.	Feature Selection.....	25
2.2.6.	Classifier	26
2.2.6.1.	Term Frequency-Inverse Document Frequency (TF-IDF)	26
2.2.6.2.	Similarity.....	28
2.2.6.2.1.	Cosine Similarity	28
2.2.6.2.2.	Jacard Simmilarity.....	28
2.2.7.	Waikato Environment For Knowledge Analysis (WEKA).....	29
BAB III METODE PENELITIAN.....		31
3.1.	Penyiapan Data.....	31
3.2.	Uji Coba dan Analisis Hasil	32
3.2.1.	Text Preprocessing.....	32
3.2.2.	Pembobotan	33
3.2.3.	K-Means Clustering.....	33
3.3.	Penyusunan Kesimpulan dan Saran Pengembangan Penelitian	33
BAB IV HASIL DAN PEMBAHASAN		34
4.1.	Text Preprocessing	34
4.1.1.	Data Terjemahan Ayat Al-Qur'an	34
4.1.2.	Case Folding	35
4.1.3.	Tokenisasi	37
4.1.4.	Stopword Removal	39
4.1.5.	Stemming.....	40
4.2.	Term Frequency-Inverse Document Frequency (TF-IDF).....	41
4.3.	Clustering	46
BAB V PENUTUP.....		55
5.1.	Kesimpulan.....	55
5.2.	Saran	55
DAFTAR PUSTAKA		57
BIODATA.....		61

DAFTAR TABEL

Tabel 2.1 Daftar objek yang akan diolah dalam clustering	13
Tabel 2.2 Hasil clustering	16
Tabel 2.3 Contoh Pemecah Kalimat	20
Tabel 2.4 Contoh Case Folding.....	21
Tabel 2.5 Contoh Tokenizing.....	21
Tabel 2.6 Contoh Stopword	22
Tabel 2.7 Aturan pemenggalan Awalan Stemmer Nazief dan Adriani.....	24
Tabel 4.1 Data Terjemahan Ayat	34
Tabel 4.2 Hasil Case Folding	35
Tabel 4.3 Hasil Tokenisasi	37
Tabel 4.4 Hasil Stopword Removal	39
Tabel 4.5 Hasil Stemming.....	40
Tabel 4.6 Hasil Perhitungan Bobot TF-IDF (tanpa stemming) setelah dinormalisasi	42
Tabel 4.7 Hasil Perhitungan Bobot TF-IDF (stemming) setelah dinormalisasi....	44
Tabel 4.8 Hasil Clustering	47
Tabel 4.9 Daftar Topik Berdasarkan Hasil Cluster	52
Tabel 4.10 Daftar Topik Berdasarkan Hasil Cluster menggunakan WEKA	52
Tabel 4.11 Daftar Kemunculan Topik Dalam QS. Al-Baqarah	53
Tabel 4.12 Daftar idTopik.....	54

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

Gambar 2.1 Diagram alir algoritme K-Means.....	12
Gambar 2.2 Ilustrasi algoritme K-Means	12
Gambar 3.1 Metode Penelitian.....	32



DAFTAR PERSAMAAN

Persamaan 2.1	10
Persamaan 2.2	10
Persamaan 2.3	26
Persamaan 2.4	27
Persamaan 2.5	27
Persamaan 2.6	27
Persamaan 2.7	28



IMPLEMENTASI K-MEANS CLUSTERING PADA TERJEMAHAN AL-QUR'AN BERDASARKAN KETERKAITAN TOPIK

Ahmad Salam Wahid Faizin

12650026

INTISARI

Proses *clustering* dapat melakukan pengelompokan data sehingga data yang memiliki kesamaan tinggi akan dikelompokkan ke dalam *cluster* yang sama. Salah satu algoritme *clustering* yang sering digunakan adalah K-Means. Pengelompokan ayat yang berhubungan akan memungkinkan pengguna untuk menemukan tema dalam *Al-Qur'an*. Penelitian ini bertujuan untuk melihat akurasi dari algoritme K-Means untuk melakukan *clustering* pada ayat-ayat *Al-Qur'an*.

Penelitian ini dilakukan dengan tahapan pra pemrosesan untuk ayat *Al-Qur'an* yang berupa teks bahasa indonesia, pembobotan dengan TF-IDF, normalisasi menggunakan *cosine normalization* kemudian dilakukan pengelompokan data dengan K-Means. Berdasarkan hasil pengujian menggunakan terjemahan Qs. *Al-Baqarah* yang dilakukan *Clustering* menggunakan algoritme *K-Means* didapatkan *cluster* dengan akurasi sebesar 43%. Untuk meningkatkan nilai pengujian diperlukan algoritme pemilihan centroid untuk nilai awal, pengurangan dimensi data, dan algoritme untuk pengukuran jarak dan *similarity*.

Kata kunci : *Clustering, K-Means Clustering, Al-Qur'an, Text Mining.*

**IMPLEMENTASI K-MEANS CLUSTERING PADA TERJEMAHAN
AL-QUR'AN BERDASARKAN KETERKAITAN TOPIK**

Ahmad Salam Wahid Faizin

12650026

ABSTRACT

The clustering process can perform grouping of data, so data which have high similarity will be grouped into the same cluster. One of the most commonly used clustering algorithms is K-Means. Grouping related paragraph will allow the user to find a theme in the Qur'an. This study aims to see the accuracy of the K-Means algorithm for clustering the verses of the Qur'an

This research was conducted with pre-processing text of Qur'an verse, term-weighting with TF-IDF, normalize using cosine normalization, and then clustering using K-Means algorithm. Based on the test result using K-Means algorithm successfully perform clustering on Al-Baqarah verses with accuracy of 43%. To increase the value of testing required centroid selection algorithms for initial values, reduction of data dimensions, and algorithms for distance measurement and similarity.

Keywords : Clustering, K-Means Clustering, Al-Qur'an, Text Mining.

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

BAB I

PENDAHULUAN

1.1. Latar Belakang

Ayat dalam kitab suci *Al-Qur'an* merupakan objek menarik bagi ilmuwan komputer untuk menunjukkan pengetahuan, kearifan dan hukum dari ayat *Al-Qur'an* di dalam sistem komputer. Dengan membangun sistem cerdas yang dapat menjawab berbagai macam pertanyaan berdasarkan pengetahuan dari ayat dan dapat membantu masyarakat, baik muslim maupun non muslim untuk memahami dan mengerti ayat dalam *Al-Qur'an*.

Memahami maksud ayat dengan membaca tafsir (penjelasan detil dari maksud ayat) akan sangat membantu tetapi belum cukup memberikan gambaran utuh dari pesan yang kitab ini coba untuk sampaikan kepada pembaca. Hal ini dikarenakan *Al-Qur'an* mencakup satu tema di banyak surat yang berbeda dan untuk mendapat gambaran utuh, pembaca harus merujuk semua bagian yang saling berhubungan.

Proses clustering dapat melakukan pengelompokan data sehingga data yang memiliki kesamaan tinggi akan dikelompokkan ke dalam cluster yang sama. Salah satu algoritme clustering yang sering digunakan adalah K-Means. Pengelompokan ayat yang berhubungan akan memungkinkan pengguna untuk menemukan tema dalam *Al-Qur'an*.

1.2. Rumusan Masalah

Berdasar latar belakang tersebut dirumuskan masalah yaitu penggunaan metode *Clustering* dengan Algoritme *K-Means* untuk pengelompokan ayat-ayat *Al-Qur'an* pada terjemahan Bahasa Indonesia. Dokumen akan dilakukan tahapan pra-pemrosesan dan ditentukan bobot berdasarkan frekuensi muncul pada proses *clustering*. Sehingga diperoleh *cluster* yang berisi ayat-ayat yang memiliki kemiripan dengan tema tertentu. Judul yang diambil adalah Implementasi *K-Means Clustering* pada Terjemahan *Al-Qur'an* Berdasarkan Keterkaitan Topik.

1.3. Batasan Masalah

Adapun batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini tidak dilakukan menyeluruh pada 30 Juz *Al-Qur'an*, tetapi hanya pada QS. *Al-Baqarah* Ayat 1-286.
2. Dalam penelitian ini hanya menggunakan terjemahan *Al-Qur'an* dalam bentuk latin, tidak dalam bentuk arabnya.

1.4. Tujuan Penelitian

Berdasar latar belakang serta rumusan masalah di atas, maka tujuan yang hendak dicapai dalam penelitian ini adalah untuk mengetahui akurasi algoritme *K-Means* untuk melakukan pengelompokan terjemahan ayat *Al-Qur'an*.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah untuk melihat kecocokan algoritme *K-Means* untuk mengelompokkan terjemahan ayat *Al-Qur'an*.



BAB V

PENUTUP

5.1. Kesimpulan

Proses *clustering* terjemahan ayat-ayat *Al-Qur'an* menggunakan *K-Means clustering* dilakukan menggunakan hasil dari TF-IDF yang kemudian dilakukan normalisasi menggunakan *cosine normalization*. Dan hasil yang didapatkan sudah menggambarkan keterkaitan antar ayat. Hasil *clustering* menunjukkan adanya ayat-ayat sejenis, yang mempresentasikan kemiripan antar dokumen.

Berdasarkan penelitian yang telah dilakukan, maka diperoleh kesimpulan bahwa penelitian menggunakan algoritme *K-Means* untuk melakukan *clustering* terjemahan ayat-ayat *Al-Qur'an* mempunyai akurasi sebesar 43%. Hasil tersebut dikarenakan pada beberapa ayat *Al-Qur'an* terdapat lebih dari satu topik, dan algoritme *K-Means* hanya melakukan *clustering* berdasarkan satu topik saja.

5.2. Saran

Peneliti menyarankan adanya pengembangan penelitian di masa yang akan datang :

1. Penelitian menambahkan algoritme lain untuk melakukan proses *preprocessing* dan saat melakukan perhitungan bobot antara lain dengan melakukan normalisasi data, atau dengan algoritme *Jacard Similarity*.
2. Penggunaan algoritme lain untuk melakukan proses *clustering*, antara lain *K-Medoids*, *K-Modes*, *Fuzzy K-Means* dll.

3. Penelitian dikembangkan dengan menggunakan seluruh terjemahan ayat *Al-Qur'an* bahasa indonesia.
4. Melakukan validasi hasil penelitian dengan sumber lain.



DAFTAR PUSTAKA

- _____, <http://www.qurandownload.com/> diakses pada tanggal 10 April 2017.
- _____, <http://www.qurandatabase.org/> diakses pada tanggal 10 April 2017.
- _____, www.pengobatan.com/ajaran_islam/IndeksQuranTercanggih.xls diakses pada tanggal 10 April 2017
- Agusta, Y. (2007). K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, 3(1), 47-60.
- Agustian, S., & Wulandari, I. S. SISTEM QUR'AN RETRIEVAL TERJEMAHAN BAHASA INDONESIA BERBASIS WEB DENGAN REORGANISASI KORPUS.
- Alfina, T., Santosa, B., & Barakbah, A. R. (2012). Analisa Perbandingan Metode Hierarchical Clustering, K-Means dan Gabungan Keduanya dalam Cluster Data (Studi Kasus: Problem Kerja Praktek Teknik Industri ITS). *Jurnal Teknik ITS*, 1(1), A521-A525.
- Arai, K., & Barakbah, A. R. (2007). Hierarchical K-means: an algorithm for centroids initialization for K-means. *Reports of the Faculty of Science and Engineering*, 36(1), 25-31.
- Berry, M. W., & Kogan, J. (Eds.). (2010). *Text mining: applications and theory*. John Wiley & Sons.
- Christopher, D. M., Prabhakar, R., & Hinrich, S. (2008). Introduction to information retrieval. *An Introduction To Information Retrieval*, 151(177), 5.
- Cios, K. J., Pedrycz, W., Swiniarski, R. W., & Kurgan, L. A. (2007). *Data mining: a knowledge discovery approach*. Springer Science & Business Media.
- Defiyanti, S., & Jajuli, M. (2015). Integrasi Metode Klasifikasi Dan Clustering dalam Data Mining. *Konferensi Nasional Informatika (KNIF)*, 39-44.
- Even-Zohar, Y. (2002). Introduction to text mining. *Automated Learning Group National Center for Supercomputing Applications University of Illinois*.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
- Ghanem, O., & Alhanjouri, M. (2014). Evaluating the effect of preprocessing in arabic documents clustering. *Islamic University*.
- Godara, S., & Verma, A. (2013). Analysis of various clustering algorithms. *International Journal of Innovative Technology and Exploring Engineering*, 3(1), 186-9.

- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Han, P., Wang, D. B., & Zhao, Q. G. (2011, July). The research on Chinese document clustering based on WEKA. In *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on* (Vol. 4, pp. 1953-1957). IEEE.
- Handoyo, R., Mangkudjaja, R., & Nasution, S. M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means Pada Pengelompokan Dokumen. *JSM (Jurnal SIFO Mikroskil)*, 15(2), 73-82.
- Harianja, H. (2008). Visualisasi K-Means clustering pada data potensi pertanian desa di Bogor menggunakan MapServer.
- Hartigan, J. A. (1975). Clustering algorithms.
- Hastuti, N. F. (2013). Pemanfaatan Metode K-Means Clustering Dalam Penentuan Penerima Beasiswa.
- Herdianto, A., Sa'adah, U., & Muhtada'i, N. R. (2010). PENCARIAN AYAT-AYAT AL-QUR'AN BERDASARKAN KONTEN MENGGUNAKAN TEXT MINING BERBASIS APLIKASI DESKTOP. *EEPIS Final Project*.
- Indranandita, A., Susanto, B., & Rahmat, A. (2011). Sistem Klasifikasi dan Pencarian Jurnal dengan Menggunakan Metode Naive Bayes dan Vector Space Model. *Jurnal Informatika*, 4(2).
- Jenitha, G., & Vennila, V. (2014, July). Comparing the partitional and density based clustering algorithms by using WEKA tool. In *Current Trends in Engineering and Technology (ICCTET), 2014 2nd International Conference on* (pp. 328-331). IEEE.
- Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons.
- Kao, A., & Poteet, S. R. (Eds.). (2007). *Natural language processing and text mining*. Springer Science & Business Media.
- Kendall, K. E., & Kendall, J. E. (2010). *Systems analysis and design*. Prentice Hall Press.
- Khan, I., Ghauri, T. A., & Majeed, S. (2012). Impact of brand related attributes on purchase intention of customers: a study about the customers of Punjab, Pakistan. *Interdisciplinary journal of contemporary research in business*, 4(3), 194-200.
- MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).

- Madhulatha, T. S. (2012). An overview on clustering methods. *arXiv preprint arXiv:1205.1117*.
- Muflikhah, Lailil. (2013). Penggalan Data Dalam Penentuan Keterkaitan Topik Pada Terjemahan Ayat-Ayat Al-Qur'an.
- Mustaqhfiri, M., Abidin, Z., & Kusumawati, R. (2012). Peringkasan Teks Otomatis Berita Berbahasa Indonesia Menggunakan Metode Maximum Marginal Relevance. *MATICS*.
- Nazief, B., & Adriani, M. (1996). Confix Stripping: Approach to Stemming Algorithm for Bahasa Indonesia. *Internal publication, Faculty of Computer Science, University of Indonesia, Depok, Jakarta*.
- Noor, M. H., & Hariadi, M. (2015, June). Image Cluster Berdasarkan Warna untuk Identifikasi Kematangan Buah Tomat dengan Metode Valley Tracing. In *Seminar Nasional Informatika (SEMNASIF)* (Vol. 1, No. 1).
- Nuningsih, S. (2010). K-Means Clustering: Studi Kasus pada Data Pengujian Kualitas Susu di Koperasi Peternakan Bandung Selatan. *Skripsi. FPMIPA UPI. Bandung*.
- Nurwidodo, W. R. S., & Hariadi, M. (2013). Sistem Pendukung Keputusan Multidimensi Menggunakan K-means Clustering Berbasis Mahalobis Distance. In *Prosiding Seminar Nasional Manajemen Teknologi* (Vol. 17, pp. 20-4).
- Nurwidodo, Wahid RS & Hariadi, M. (2012). Sistem Pendukung Keputusan Multidimensi Berbasis Mean Shift Clustering Untuk Penetapan Calon Peserta Diklat.
- Pena, J. M., Lozano, J. A., & Larranaga, P. (1999). An empirical comparison of four initialization methods for the k-means algorithm. *Pattern recognition letters*, 20(10), 1027-1040.
- Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, 21(6), 759-772.
- Rahmawati, L., Sihwi, S. W., & Suryani, E. (2016). ANALISA CLUSTERING MENGGUNAKAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING (STUDI KASUS: DOKUMEN SKRIPSI JURUSAN KIMIA, FMIPA, UNIVERSITAS SEBELAS MARET). *ITSMART: Jurnal Teknologi dan Informasi*, 3(2), 66-73.
- Rencher, A. C. (2003). *Methods of multivariate analysis* (Vol. 492). John Wiley & Sons.
- Repi, A. L., & Baramuli, D. N. (2011). PENGARUH KEPERCAYAAN DAN KEPUASAN TERHADAP LOYALITAS NASABAH Pada PT Bank Mandiri Tbk Manado. *Jurnal Inovasi Manajemen*, 8(2), 64-74.

- Rohmawati, N., Defiyanti, S., & Jajuli, M. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jurnal Ilmiah Teknologi Informasi Terapan*, 1(2).
- Salton, G. (1989). Automatic text processing: The transformation, analysis, and retrieval of. Reading: Addison-Wesley.
- Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. *Yogyakarta: Graha Ilmu*, 978(979), 756.
- Sarwono, Y. T. (2010). Aplikasi Model Jaringan Syaraf Tiruan Dengan Radial Basis Function Untuk Mendeteksi Kelainan Otak (Stroke Infark). *Sekolah Tinggi Manajemen Informatika & Teknik Komputer Surabaya*.
- Sharmila & Misra. (2013). K-Means Clustering, Expectation Maximization (EM) Algoritma Dan Hierarchical Clustering.
- Sri, A. (2007). Pembentukan Cluster dalam Knowledge Discovery in Database dengan Algoritma K-means. *SEMNAS Matematika dan Pendidikan Matematika 2007 dengan tema "Trend Penelitian Matematika dan Pendidikan Matematika di Era Global*.
- Sugiyanto, S., Surarso, B., & Sugiharto, A. (2014). Analisa Performa Metode Cosine dan Jacard pada Pengujian Kesamaan Dokumen. *Jurnal Masyarakat Informatika*, 5(10), 1-8.
- Tekmono, K., 2007, K-Means Clustering Tutorial. <http://people.revolude.com/Kardi/Tutorial/Kmean.html> diakses pada tanggal 21 Desember 2017.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- Wu, X., Wu, B., Sun, J., Qiu, S., & Li, X. (2015). A hybrid fuzzy K-harmonic means clustering algorithm. *Applied Mathematical Modelling*, 39(12), 3398-3409.
- Yao, Y., Liu, Y., Yu, Y., Xu, H., Lv, W., Li, Z., & Chen, X. (2013). K-SVM: An Effective SVM Algorithm Based on K-means Clustering. *JCP*, 8(10), 2632-2639.
- Zaman, B., & Winarko, E. (2011). Analisis Fitur Kalimat untuk Peringkat Teks Otomatis pada Bahasa Indonesia. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 5(2).

BIODATA

Nama : Ahmad Salam Wahid Faizin
Tempat Lahir : Purworejo
Tanggal Lahir : 08 Agustus 1993
Golongan Darah : B
Agama : Islam
Kewarganegaraan : Indonesia
Alamat Asal : Krajan Lor, RT.02/02, Rimun, Loano, Purworejo
Email : faizin2q@hotmail.com
No. HP : 085729892644



Riwayat Pendidikan :

1999-2005	SD N Rimun, Loano, Purworejo
2005-2008	SMP N 2 Purworejo
2008-2011	SMK TKM Teknik Tamansiswa Purworejo
2012-2018	S1 Teknik Informatika. Universitas Islam Negeri Sunan Kalijaga Yogyakarta