

**KLASTERISASI DOKUMEN PENELITIAN PERGURUAN TINGGI
MENGUNAKAN *K-MEANS CLUSTERING*, SEBAGAI ANALISA
PENERAPAN SISTEM TEMU KEMBALI**



LAPORAN PENELITIAN

Sebagai Salah Satu Bentuk Pengamalan Tri Dharma Perguruan Tinggi

Oleh:

Nama	NIPY
1. Very Kurnia Bakti, M. Kom	09.008.044
2. Arif Rahman, S.E, S.Pd, M.Kom	05.016.291
3. Muhammad Rizal Arfani	18040004

**PROGRAM STUDI DIPLOMA III TEKNIK KOMPUTER
POLITEKNIK HARAPAN BERSAMA
TEGAL
AGUSTUS 2021**

**SK Direktur Nomor : 098.05/PHB/V/2021 Tanggal 31 Mei 2021
Nomor Kontrak : 032.16/P3M.PHB/V/2021 Tanggal 6 Mei 2021**

**HALAMAN PERSETUJUAN
LAPORAN PENELITIAN**

**KLASTERISASI DOKUMEN PENELITIAN PERGURUAN TINGGI
MENGUNAKAN *K-MEANS CLUSTERING*, SEBAGAI ANALISA
PENERAPAN SISTEM TEMU KEMBALI**

Sebagai Salah Satu Bentuk Pengamalan Tri Dharma Perguruan Tinggi

Oleh :

Nama	NIPY
1. Very Kurnia Bakti, M. Kom	09.008.044
2. Arif Rahman, S.E, S.Pd, M.Kom	05.016.291
3. Muhammad Rizal Arfani	NIM 18040004

Tegal, Agustus 2021

Mengetahui,
Ketua Program Studi DIII Teknik Komputer
POLITEKNIK HARAPAN BERSAMA


Rais, S. Pd., M. Kom
NIPY 07.011.083

Menyetujui,
Ketua Pusat Penelitian dan
Pengabdian Masyarakat
POLITEKNIK HARAPAN BERSAMA


Kusnadi, M. Pd
NIPY 014.015.217

**HALAMAN PENGESAHAN
LAPORAN PENELITIAN**

1. Judul

: KLASTERISASI DOKUMEN PENELITIAN PERGURUAN TINGGI
MENGUNAKAN K-MEANS CLUSTERING, SEBAGAI ANALISA
PENERAPAN SISTEM TEMU KEMBALI

2. Ketua Peneliti

a. Nama Lengkap : Very Kurnia Bakti, M.Kom
b. NIDN : 0625118301
c. NIPY : 09.008.044
d. Jabatan Fungsional : Lektor
e. Program Studi : DIII Teknik Komputer
f. Alamat e-mail :

3. Jumlah Anggota

3

Nama Anggota 1 : Arif Rakhman, SE, S.Pd, M. Kom
Nama Mahasiswa 1 : Mohammad Rijal Arfani

Biaya Penelitian

: Rp.3,314,000

Tegal, 31 Agustus 2021

Reviewer 1


Arif Rakhman, SE, S.Pd, M. Kom

NIPY. 05.016.291

Menyetujui,


Ketua Prodi DIII Teknik Komputer
Politeknik Harapan Bersama


Rais, S.Pd, M.Kom

NIPY. 07.011.083


Mengetahui,

Wakil Direktur 1
Politeknik Harapan Bersama


api, Heru Nurcahyo, S.Farm., M.Sc

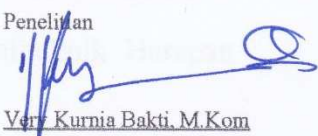
NIPY. 10.007.038

Reviewer 2


M. FIKRI HIDAYATTULLAH,
M. Kom

NIPY. 09.016.307

Ketua Tim Pelaksana
Penelitian


Very Kurnia Bakti, M. Kom

NIPY. 09.008.044

Mengesahkan,

Ketua P3M
Politeknik Harapan Bersama


Kusnadi, M.Pd

NIPY. 04.015.217

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadirat Allah SWT, Tuhan Yang Maha Pengasih dan Maha Penyayang yang telah melimpahkan segala rahmat, hidayah dan inayah-Nya hingga terselesaikannya penyusunan laporan penelitian yang dengan judul Klasterisasi Dokumen Penelitian Perguruan Tinggi Menggunakan *K-Means Clustering*, Sebagai Analisa Penerapan Sistem Temu Kembali “.

Laporan Penelitian merupakan dibuat untuk sebagai salah satu kewajiban yang harus dilaksanakan oleh setiap dosen untuk memenuhi salah satu dharma pada Tri Dharma Perguruan Tinggi, yakni pada Politeknik Harapan Bersama. Selama penyusunan proposal penelitian dan kemudian tersusun banyak pihak yang telah memberikan bantuan dan dukungan.

Pada kesempatan ini, peneliti mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Nizar Suhendra, S E, MPP selaku Direktur Politeknik Harapan Bersama.
2. Bapak Kusnadi, M.Pd Selaku Kepala P3M Politeknik Harapan Bersama Tegal.
3. Bapak Rais S, Pd, M.Kom selaku Ketua Program Studi DIII Teknik Komputer yang telah memberikan dukungan sarana maupun prasarana untuk penyusunan proposal penelitian.

Semoga proposal ini dapat diterima dan memberikan sumbangan untuk pengembangan ilmu pengetahuan dan teknologi.

Tegal, Agustus 2021

Peneliti

PERNYATAAN

Dengan ini kami menyatakan bahwa :

1. Laporan Penelitian ini tidak pernah dibuat oleh peneliti lain dengan tema, judul, isi, metode, objek penelitian yang sama.
2. Laporan Penelitian ini bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik di suatu perguruan tinggi.
3. Dalam laporan penelitian ini juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Tegal , Agustus 2021

Ketua Tim Peneliti



Very Kurnia Bakti, M.Kom
NIPY. 09.008.044

Anggota Peneliti



Arif Rahman, S.E, S.Pd, M.Kom
NIPY 05.016.291

Anggota Peneliti



Muhammad Rizal Arfani
18040004

DAFTAR ISI

COVER	i
HALAMAN PERSETUJUAN LAPORAN PENELITIAN.....	ii
HALAMAN PENGESAHAN LAPORAN PENELITIAN	iii
KATA PENGANTAR	iv
PERNYATAAN.....	v
DAFTAR ISI.....	vi
DAFTAR TABEL	ix
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan.....	4
1.5. Manfaat.....	4
BAB II TINJAUAN PUSTAKA.....	6
BAB III METODE PENELITIAN.....	7
3.1. Bahan Penelititan.....	7
3.2. Alat Penelititan	7
3.3. Prosedur Penelitian	7
BAB IV HASIL DAN PEMBAHASAN	10
4.1 Proses Pengambilan Data	10
4.2 Ekstraksi Dokumen	10
4.3 Proses Klasterisasi dengan K-Means Clustering	14
4.4 Hasil Pengukuran Davies Bouldin Index (DBI)	20
4.5 Luaran.....	20
BAB V KESIMPULAN	21
5.1 Kesimpulan.....	21
5.2 Saran	21
DAFTAR PUSTAKA	22
ORGANISASI PENGUSUL.....	23

REALISASI ANGGARAN	24
LAMPIRAN.....	25

DAFTAR GAMBAR

Gambar 1. Prosedur Penelitian.....	9
------------------------------------	---

DAFTAR TABEL

Tabel 4. 1 Hasil kata yang telah melalui case folding.....	11
Tabel 4. 2 Hasil dari proses ekstraksi dokumen dan pembobotan TF-IDF.....	13
Tabel 4. 3 Jumlah anggota dari masing-masing kluster	15
Tabel 4. 4 Jumlah data judul penelitian dari tiap-tiap program studi.....	16
Tabel 4. 5 Index dari masing-masing kluster hasil pengolahan menggunakan metode k-means clustering.....	16
Tabel 4. 6 Kesimpulan data penelitian pada tiap cluster.....	20

BAB I

PENDAHULUAN

1.1. Latar Belakang

Bertambahnya jumlah data dokumen Penelitian Perguruan Tinggi dari berbagai program studi dapat memberi kontribusi besar dalam sulitnya proses pencarian suatu dokumen. Pencarian dokumen yang ada saat ini yaitu menampilkan hasil pencarian berurut berdasarkan peringkat kecocokan (*document ranking*). Hal tersebut menyebabkan penemuan data dokumen tidak secara akurat terkelompok pada masing – masing tema. Dengan adanya pengelompokan dokumen, maka tidak harus membuka halaman terlalu banyak, karena dokumen hasil pencarian telah dikelompokkan berdasarkan kategori yang dapat menggambarkan isi dari suatu dokumen, hal tersebut tentunya dapat mempermudah dalam menemukan beberapa dokumen yang diinginkan, oleh karenanya sebelum proses tersebut dilakukan maka, proses tersebut perlu dianalisis sebelum benar-benar diaplikasikan ke dalam aplikasi yang sifatnya *executable*. Oleh karena itu diperlukan metode pengelompokan / *clustering* yang nantinya dapat dipastikan keberhasilan pengelompokan suatu dokumen Penelitian Perguruan Tinggi dengan baik.

Clustering dapat digunakan dalam pengkategorian atau pengelompokan dokumen. Caranya adalah dengan mengelompokkan dokumen-dokumen ke dalam *clusters* berdasarkan kedekatan atau kemiripan antar dokumen (*similarity*) [1, 5], sehingga dokumen yang berhubungan dengan suatu tema tertentu secara otomatis ditempatkan pada cluster yang sama. Saat ini ada beberapa algoritma *clustering* diantaranya *partitional* (K Means) dan hierarchical. [1] *Clustering* dengan metode *K means* adalah algoritma sederhana yang dikembangkan Mac Queen pada tahun 1967. Algoritma tersebut terkenal dengan kemampuannya untuk

mengklaster data yang besar dan dapat menangani data *outlier*. *K-means* merupakan metode pengklasteran yang memisahkan data kedalam k kelompok yang berbeda artinya sebelum dilakukan klasterisasi maka perlu menentukan jumlah k yang diinginkan. Selain itu *k means* merupakan *center based clustering* yang menentukan setiap klaster dari titik pusat klasternya [1, 3].

K-means clustering salah satu metode data *clustering* non-hirarki dimana metode ini memiliki kemampuan mengelompokkan data kedalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik sama akan dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil [8]. Metode *K-means clustering* ini memungkinkan diterapkan sebagai metode yang bisa digunakan dalam *text mining* khususnya dalam pengklasteran judul Penelitian Perguruan Tinggi.

Sebelum melakukan pengkategorian dokumen perlu proses *preprocessing*, yaitu pembuangan dengan melalui *stopwords*, dan *stemming*. Hal ini diperlukan untuk mengurangi jumlah kata yang diproses. Selain itu diperlukan juga suatu pembobotan dari setiap kata yang ada dalam suatu dokumen sebagai representasi dari dokumen. Pada penelitian ini, pembobotan pada sebuah *term* yang digunakan adalah TF-IDF yaitu kombinasi antara TF (*Term Frequency*) dengan IDF (*Invers Document Frequency*).

Pada saat pembobotanpun dibutuhkan normalisasi disebabkan karena setiap dokumen mempunyai panjang kata yang tidak sama. Salah satu normalisasi yang dapat digunakan adalah normalisasi dengan *Maximum Frequency*, nilai frekuensi paling tinggi dari seluruh atribut pada sebuah dokumen. Sedangkan untuk menentukan nama kategori dari

suatu kluster adalah dengan mencari *term* yang memiliki frekuensi kemunculan terbesar ($\text{Max } \{f(t,d)\}$) dari tiap kluster.

1.2. Rumusan Masalah

1. Setiap dokumen mempunyai panjang jumlah kata yang berbeda, maka diperlukan normalisasi pembobotan pada setiap term. Maka diperlukan normalisasi dengan *Maximum Frequency*. Untuk mengelompokkan dokumen, kemudian setiap kata yang telah memiliki bobot direpresentasikan kedalam *vector space*.
2. Tiap kata yang didapatkan dari abstrak Penelitian Perguruan Tinggi memiliki dimensi kata yang tinggi dan tidak semua kata merupakan kata pembeda (dan, di, ke, me, an, yang, dan sebagainya), sehingga perlu dilakukan *pre processing* untuk mengatasi jumlah dimensi kata yang tinggi dan mengurangi waktu dalam pemrosesan setiap katanya atau biasa disebut dengan *stopwords*.
3. Penamaan kategori pada masing-masing kluster akan memudahkan proses pada saat mencari dokumen yang dibutuhkan. dikarenakan dengan melihat nama kategori, dapat mendeskripsikan isi dokumen yang ada dalam tiap kluster.
4. Mengklasterkan dokumen abstrak Penelitian Perguruan Tinggi berbahasa Indonesia dengan menerapkan *Algoritma K Means*. Menganalisis nilai *convergence* dan tingkat akurasi tiap kluster yang dihasilkan dan relevansi nama kategori dengan tiap dokumen yang ada dalam tiap kluster.

1.3. Batasan Masalah

Batasan-batasan dalam penelitian ini adalah sebagai berikut

1. Dokumen yang digunakan adalah dokumen abstrak Penelitian Perguruan Tinggi berbahasa Indonesia pada Politeknik Harapan Bersama.

2. Pembahasan hanya pada analisa *text mining* dan salah satu metode data mining yaitu *clustering*.
3. Proses stemming mengacu pada Ejaan Yang Disempurnakan (EYD).
4. Penulisan kata yang ada pada dokumen yang tidak sesuai dengan EYD. Tidak melalui tahapan *pre processing*.
5. Algoritma yang digunakan adalah K Means *clustering*.
6. Kata yang merupakan istilah bahasa asing tidak termasuk dalam proses *preprocessing*
7. Studi kasus yang digunakan adalah dokumen Abstrak Penelitian Perguruan Tinggi Dosen Politeknik Harapan Bersama Program Studi D3. Akutansi, D3 Teknik Komputer, D3. Farmasi, D3. Kebidanan, D3 Teknik Mesin, D3 Teknik Elektronika, D4 Teknik Informatika

1.4. Tujuan

Dari rumusan masalah diatas maka tujuan dalam penelitian ini adalah:

1. Menganalisa Algoritma K Means *Clustering* sebagai suatu metode dalam mengelompokan dokumen Penelitian Perguruan Tinggi.
2. Menganalisis kesesuaian nama kategori dengan mengamati tingkat relevansi nama kategori dengan dokumen yang ada tiap kluster.
3. Menganalisis hasil klasterisasi yang didapat dengan melihat nilai *convergence* tiap kluster dan tingkat relevansinya dengan cara membandingkan hasil *clustering* dengan hasil pengelompokan secara manual. Maka dari hasil analisa tersebut dapat dijadikan pertimbangan dalam membangun sebuah Sistem temu kembali Penelitian Perguruan Tinggi.

1.5. Manfaat

1. Dapat menjadi bahan pertimbangan untuk perlu atau tidaknya dibuat sebuah aplikasi yang dapat membantu dalam proses pencarian dokumen Penelitian Perguruan Tinggi dengan metode *k-means clustering*.

2. Menjadi bahan kajian dan referensi dalam penelitian secara berkelanjutan dalam bidang *data mining* atau *text mining*

BAB II TINJAUAN PUSTAKA

Penelitian yang dilakukan Vidya Ayuningtias¹, M. Arif Bijaksana², Rimba Widhiana Ciptasari³ mengungkapkan Algoritma K Means pada *clustering* dapat diterapkan pada pengelompokan dokumen. Hal tersebut dibuktikan dari pengujian, dimana kluster tersebut menghasilkan maksimum convergence 0,0026 pada inputan jumlah kluster dua, pembobotan TF IDF, dan jarak antar objek dengan kesamaan cosine. Kualitas hasil kluster pengelompokan dokumen tidak dipengaruhi oleh banyaknya jumlah dokumen.

Haryo Guritno, dalam penelitiannya dengan melakukan klasterisasi menggunakan metode k-means *clustering* untuk menentukan jenis / genre cerpen. Didapat hasil pengklasteran yang memiliki nilai maximum convergence 0,001 dengan capaian *stemming* dan *stopwords* mencapai 8000 kata.

Eric Budiman Gosno menyimpulkan dalam penelitiannya dari segi waktu, algoritma *KD-TreeK-Means Clustering* memiliki waktu *training* yang cukup lama dalam klasterisasi dokumen. Namun menghasilkan nilai distorsi 3×10^5 lebih rendah dibandingkan dengan nilai rerata distorsi dari *K-Means Clustering*.

BAB III

METODE PENELITIAN

3.1. Bahan Penelititan

Bahan Penelitian yang digunakan dalam penelitian berupa, data Penelitian Perguruan Tinggi yang diambil pada bagian abstraknya saja. Dari masing – masing 7 program studi di Politeknik Harapan Bersama.

3.2. Alat Penelititan

Alat yang digunakan pada saat penelitian adalah komputer/ laptop dengan sistem *operasi windows 10* dan aplikasi *rapid miner* sebagai perangkat lunak tambahan untuk melakukan analisis.

3.3. Prosedur Penelitian

1. Pengumpulan data sampel

Dokumen yang digunakan dalam penelitian ini diambil dari sampel abstrak Penelitian Perguruan Tinggi dari P3M Politeknik Harapan Bersama dari 7 program studi.

2. Ekstraksi Dokumen / *Pre Processing*

Proses ekstraksi ini yaitu proses *stemming* dan *stopwords* bertujuan untuk menghasilkan term-term yang akan digunakan sebagai *prototype* bagi setiap dokumen. Tiap term tersebut dicari bentuk kata dasar-nya berdasarkan kamus kata dasar Bahasa Indonesia. Hal ini untuk menghindari tersimpannya kata-kata yang memiliki kata dasar yang sama namun berimbuhan berbeda. Disamping itu dilakukan penyaringan (*filtering*) terhadap kata-kata yang tidak layak untuk dijadikan sebagai pembeda. Kelompok kata ini biasanya disebut sebagai stoplist [6] atau ada juga yang menyebutnya dengan *stopwords*.

3. Penghitungan Pembobotan

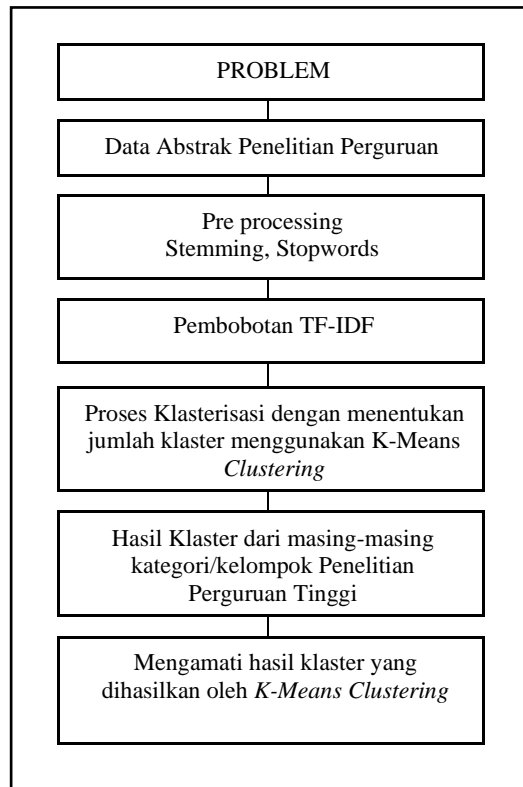
Metode TF-IDF merupakan integrasi antar *term frequency* (tf), dan *inverse document frequency* (idf) [1][9]. Dengan rumus :

$$w(t,d) = tf(t,d) * \log_2(N/nt)$$

Simbol $w(t,d)$ merupakan bobot dari term t dalam sebuah dokumen Penelitian Perguruan Tinggi sedangkan $tf(t,d)$ adalah frekuensi term dalam dokumen (tf) dimana N merupakan ukuran data training yang digunakan untuk penghitungan IDF. Adapun nt adalah jumlah dari dokumen yang ditraining yang mengandung nilai t . Fungsi metode ini adalah untuk mencari representasi nilai dari tiap-tiap dokumen dari suatu kumpulan data *training* (*training set*). Dari sini akan dibentuk suatu vektor antara dokumen dengan kata (documents with terms) yang kemudian untuk kesamaan antar dokumen dengan *cluster* akan ditentukan oleh sebuah *prototype* vektor yang disebut juga dengan cluster centroid [6, 7].

4. Algoritma K-Means Clustering

Algoritma tersebut mampu untuk mengklaster data yang besar seperti data Penelitian Perguruan Tinggi yang sangat banyak, dan dapat menangani data *outlier*. *K-means* merupakan metode pengklasteran yang memisahkan data kedalam k kelompok yang berbeda artinya sebelum dilakukan klasterisasi maka perlu menentukan jumlah k . jumlah k disini dapat dijadikan penentu berapa program studi yang akan menjadi kategori tema Penelitian Perguruan Tinggi yang diinginkan. Selain itu k *means* merupakan *center based clustering* yang menentukan setiap klaster dari titik pusat klasternya, data Penelitian Perguruan Tinggi yang memiliki nilai saling mendekati akan menyatu dan membentuk klaster. Hasil klaster tersebut dapat diamati apakah klaster yang terbentuk sesuai yang diinginkan.



Gambar 1. Prosedur Penelitian

BAB IV

HASIL DAN PEMBAHASAN

4.1 Proses Pengambilan Data

Pengambilan data diperoleh dengan mengumpulkan data-data penelitian yang diperoleh dari Bagian Penelitian dan Pengabdian Masyarakat (P3M) Politeknik Harapan Bersama dari enam program studi diantaranya program studi D3-Kebidanan, D3-Teknik Komputer, D3-Akutansi, D3-Farmasi, D4 Teknik Informatika, D3 Teknik Mesin. Data yang diperoleh didapat dari kurun waktu tahun 2017 – 2020 secara acak dan masing-masing program studi diambil 30 judul penelitian. Sehingga diperoleh total data 180 judul penelitian.

Dari hasil data yang diperoleh dilakukan pensortiran dengan cara hanya mengambil abstrak dari masing-masing judul dan dirubah kedalam format *.txt. Dari data abstrak tersebut tersebut dianggap isi teks mewakili isi keseluruhan penelitian dan nantinya tiap kata dari data tersebut sebagai bahan untuk diolah menggunakan komputer agar diperoleh informasinya dalam menentukan klaster.

4.2 Ekstraksi Dokumen

Dalam tahap ini data abstrak dari masing-masing penelitian yang berbentuk dokumen berformat *.txt diperlukan proses *pre-processing*, yaitu *case folding*, *Tokenizing*, *stopwords* dan pembobotan.

4.2.1 Case Folding

Dari dokumen penelitian yang ada selanjutnya masuk kedalam proses *case folding* dimana proses ini melakukan penyamaan antar kata dengan cara merubah huruf besar menjadi seluruhnya huruf kecil. Sehingga seluruh kata yang diproses semuanya menjadi huruf kecil. Dari hasil pemrosesan tersebut dapat diambil contoh hasil pemrosesan *case folding* seperti pada tabel 4.1.

Tabel 4. 1 Hasil kata yang telah melalui case folding

Alat	Bahas	Citrus
Alba	Bahasa	citrushystrix
Albicans	bahasan	City
Album	bahaya	Ckb
Alergi	bahwa	ckhl
alfiani	bahwa	class
Alir	baic	client
Aliran	baitul	closely
Alkali	bakar	club
alkalinya	baketri	cluster
alkaloid	bakteri	Cmc
alkohol	baku	cocok
Aloe	banding	cocor
alokasi	bandingkan	cocos
alpinia	bangsa	coefficient
Alri	bangun	coklat
alternatif	bangunan	Coli
alternative	bani	coliform
Altilis	bank	collected
Alun	banteng	collection
Alur	bantu	coloni
amalia	bantuan	coming
amaliyah	banyak	commerce
aman	banyak	communis
amandel	bapak	community
amaryllifolius	basa	compare
ambil	basah	compensastion
ambon	based	compensation
amdk	basic	completeness

4.2.2 Tokenizing

Proses selanjutnya setelah melalui tahap *case folding* adalah proses Pemisahan rangkaian term (*tokenization*). Term dapat berupa kata atau frasa di dalam dokumen. Namun, kata-kata yang tidak memberikan perbedaan seperti ini, itu, saya, kamu, serta tanda-tanda baca dihilangkan atau dianggap bukan *term*. Hal ini bertujuan untuk mendapatkan hanya kata – kata tertentu saja yang nantinya didapat dan berkontribusi sebagai

ciri – ciri dari masing-masing jenis judul penelitian. Selain itu masih dalam proses *tokenizing* dilakukan pula *filter token* yaitu dengan memberikan batasan jumlah karakter dari masing-masing kata minimal 3 karakter dan maksimal 25 karakter dari setiap kata. Hal ini dilakukan untuk mensortir atau menghilangkan kata salah pengetikan karena terlalu pendek atau terlalu panjang dalam tiap kata.

4.2.3 Stopwords

Stopwords dilakukan dalam penelitian ini untuk mendapatkan kata dasar, dikarenakan di tiap dokumen penelitian banyak terdapat kata yang memiliki banyak imbuhan. Hal tersebut akan mempengaruhi hasil kluster nantinya. Dalam proses *stopwords* inilah nantinya tiap kata-yang memiliki kata dasar yang sama akan dihilangkan imbuhanannya sehingga didapat hanya kata dasar saja. Proses *stopwords* ini dilakukan dengan cara manual yaitu dengan membuat kamus *stopwords* sendiri sebanyak 8093 kata yang tentunya masih sedikit jika dibandingkan dengan jumlah kata pada bahasa Indonesia.

Dalam proses pencarian kata dasar ini, mengingat kata yang terdapat dalam dokumen penelitian tidak hanya kata berbahasa Indonesia saja, melainkan juga terdapat kata dengan bahasa Inggris, maka dalam penelitian ini dilakukan pemrosesan sebanyak dua kali *stopwords* yaitu dengan *stopwords* bahasa Indonesia dan *stopwords* bahasa Inggris. Dengan harapan hasil kata dasar yang dihasilkan memiliki kontribusi dalam menentukan hasil kluster nantinya.

4.2.4 Pembobotan TF-IDF

Tahap pembobotan ini menggunakan Metode TF-IDF dimana terdapat integrasi antar term frequency (tf), dan inverse document frequency (idf) Dengan rumus :

$$w(t,d) = tf(t,d) * \log_2(N/nt)$$

Simbol $w(t,d)$ merupakan bobot dari term t dalam sebuah dokumen penelitian d sedangkan $tf(t,d)$ adalah frekuensi term dalam dokumen penelitian (tf) dan N merupakan ukuran data training yang digunakan untuk penghitungan IDF. Adapun nt adalah jumlah dari dokumen yang ditraining yang mengandung nilai t .

Tabel 4. 2 Hasil dari proses ekstraksi dokumen dan pembobotan TF-IDF

Word	in documents	total	in class (komputer)	in class (akutansi)	in class (kebidanan)	in class (farmasi)
Aaffecting	1,0	1,0	,0	1,0	,0	,0
Abad	1,0	2,0	2,0	,0	,0	,0
Abdurrahman	1,0	1,0	,0	1,0	,0	,0
Abortus	1,0	1,0	,0	,0	1,0	,0
abrasivernya	1,0	1,0	,0	,0	,0	1,0
Abstract	8,0	8,0	1,0	5,0	2,0	,0
Abstrak	138,0	138,0	49,0	42,0	47,0	,0
Acak	2,0	2,0	,0	,0	1,0	1,0
Acc	2,0	2,0	,0	2,0	,0	,0
Accepted	3,0	8,0	,0	8,0	,0	,0
Access	3,0	3,0	3,0	,0	,0	,0
accessories	1,0	1,0	1,0	,0	,0	,0
accidental	1,0	1,0	,0	,0	1,0	,0
accommodated	1,0	1,0	,0	1,0	,0	,0
accompanied	1,0	1,0	,0	,0	1,0	,0
acconting	1,0	1,0	,0	1,0	,0	,0
accordance	2,0	2,0	,0	2,0	,0	,0
accountants	1,0	5,0	,0	5,0	,0	,0
accounting	4,0	4,0	,0	4,0	,0	,0
accuracy	1,0	1,0	,0	1,0	,0	,0
achievement	1,0	1,0	,0	1,0	,0	,0
Action	2,0	2,0	2,0	,0	,0	,0
actionsript	2,0	3,0	3,0	,0	,0	,0
Active	1,0	1,0	,0	,0	1,0	,0
activities	1,0	3,0	,0	3,0	,0	,0
activity	2,0	2,0	2,0	,0	,0	,0
Acuan	1,0	1,0	,0	,0	,0	1,0
adakah	1,0	1,0	,0	,0	,0	1,0
adapun	6,0	10,0	4,0	2,0	1,0	3,0
address	1,0	2,0	2,0	,0	,0	,0
Ade	2,0	2,0	,0	2,0	,0	,0
Adeps	1,0	5,0	,0	,0	,0	5,0

Adi	1,0	1,0	,0	,0	,0	1,0
Adie	1,0	1,0	,0	1,0	,0	,0
adiknya	1,0	1,0	1,0	,0	,0	,0
Aditya	1,0	1,0	,0	1,0	,0	,0
adiwarna	4,0	14,0	1,0	,0	13,0	,0
adjusted	1,0	1,0	,0	1,0	,0	,0
Admin	1,0	1,0	1,0	,0	,0	,0
adminiatriasi	1,0	1,0	1,0	,0	,0	,0
administrasi	4,0	10,0	2,0	8,0	,0	,0
Adobe	1,0	1,0	1,0	,0	,0	,0
advantage	2,0	6,0	,0	6,0	,0	,0
Advice	1,0	1,0	,0	,0	1,0	,0
Aena	1,0	1,0	,0	1,0	,0	,0
Affect	1,0	1,0	,0	1,0	,0	,0
Afip	1,0	1,0	1,0	,0	,0	,0
Agak	1,0	1,0	,0	,0	,0	1,0
agama	2,0	2,0	1,0	1,0	,0	,0
Agen	1,0	1,0	1,0	,0	,0	,0
agency	1,0	1,0	,0	1,0	,0	,0
Agent	3,0	11,0	,0	,0	,0	11,0
Agung	1,0	2,0	,0	2,0	,0	,0
agustina	1,0	1,0	,0	1,0	,0	,0

4.3 Proses Klasterisasi dengan K-Means Clustering

Dari tahapan-tahapan yang telah dilalui dengan proses ekstraksi dokumen, maka langkah selanjutnya adalah proses mengklaster dokumen. Dalam penelitian ini proses klaster dokumen dilakukan dengan menggunakan algoritma *k-means clustering* dengan pertimbangan beberapa referensi dari penelitian sebelumnya dengan tema yang sama yaitu, *information retrieval* dengan *text mining*. Pada pengklasteran menggunakan *k-means clustering* dengan dasar algoritmanya adalah sebagai berikut:

1. Langkah pertama adalah dengan mengelompokan atau Inisialisasi klaster dalam penelitian ini dibuat empat klaster disesuaikan dengan jumlah dokumen empat program studi, D 3- Teknik Komputer, D 3- Kebidanan, D3 -Akutansi, D3-Farmasi, D4 Teknik Informatika, D3 Teknik Mesin Masing-masing program studi ada 30 judul penelitian.

2. Memasukan semua dokumen ke klaster yang paling cocok dengan berdasarkan pusat klaster. Dengan persamaan sebagai berikut:

$$d(P, Q) = \sqrt{\sum_j^p (x_j(P) - x_j(Q))^2}$$

Dimana :

$D(ij)$ = Jarak data ke i ke pusat cluster j

$X(ij)$ = data ke i pada atribut data ke k

$X(ij)$ = Titik pusat ke j pada atribut ke k

3. Tahap untuk mendapatkan klaster yang terdekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap cluster (*centroid*). Maka data yang terdekat dengan *centroid* akan dijadikan satu klaster.
4. Setelah semua data telah terkumpul dipusat klaster yang paling dekat, kemudian dilakukan perhitungan kembali ke *centroid* yang baru dengan rata-rata anggota yang ada pada cluster tersebut.
5. Setelah diperoleh *centroid* yang baru dari tiap cluster, kemudian dilakukan kembali dari tahap ketiga sampai *centroid* dari setiap klaster tidak berubah dan tidak ditemukan lagi data yang berpindah-pindah dari satu klaster ke klaster yang lain.

4.3 Hasil Klaster

Dari penerapan algoritma K-means Clustering didapat hasil klaster dengan pengelompokan seperti pada tabel 4.3.

Tabel 4. 3 Jumlah anggota dari masing-masing klaster

Jumlah Klaster	Anggota Klaster
Cluster 0:	24 items
Cluster 1:	13 items
Cluster 2:	23 items

Cluster 3:	76 items
Cluster 4:	11 items
Cluster 5:	33 items

Dari masing-masing klaster yang terbentuk, terdapat anggota – anggota klaster yang seharusnya tidak termasuk dalam klaster tersebut mengingat pengelompokan data yang sebenarnya adalah seperti pada tabel 4.4

Tabel 4. 4 Jumlah data judul penelitian dari tiap-tiap program studi

Nama Program Studi	Jumlah Judul Penelitian
D3 Teknik Komputer	30
D3 Kebidanan	30
D3 Akutansi	30
D3 Farmasi	30
D4 Teknik Informatika	30
D4 Teknik Mesin	30

Namun hasil pengelompokan yang diolah dengan algoritma k-means, memperlihatkan jumlah klaster yang bisa dikatakan cukup merata dimana hampir tiap-tiap klaster yang terbentuk hampir keseluruhannya mendekati angka 30.

Masing masing klaster dapat beranggotakan tidak hanya dari satu program studi saja, melainkan bisa lebih dari satu judul program studi yang masuk didalamnya. Untuk lebih jelasnya anggota dari tiap-tiap klaster dapat diperlihatkan dari tabel 4.6

Tabel 4. 5 Index dari masing-masing klaster hasil pengolahan menggunakan metode k-means clustering

Cluster	Jumlah Anggota	Anggota		
Cluster 0:	24 items	Akuntansi 17 2019.txt	Farmasi 9 2020.txt	Teknik Mesin 18 2020.txt
		Akuntansi 14 2019.txt	Kebidanan 22.2018.txt	Teknik Mesin 19 2020.txt

		Akuntansi 19 2019.txt	Kebidanan 23.2018.txt	Teknik Mesin 25 2020.txt
		Akuntansi 20 2019.txt	Kebidanan 27.2018.txt	Teknik Mesin 28 2020.txt
		Akuntansi 21 2019.txt	Kebidanan 8.2020.txt	
		Akuntansi 24 2019.txt	komputer 10.2018.txt	
		Akutansi 4 2020.txt	komputer 22. 2018.txt	
		Farmasi 1 2020.txt	komputer 27.2017.txt	
		Farmasi 6 2020.txt	TI 7.2020.txt	
		Farmasi 7 2020.txt	Teknik Mesin 13 2020.txt	
Cluster 1:	13 items	Farmasi 10 2020.txt	Teknik Mesin 16 2020.txt	Teknik Mesin 8 2020.txt
		Farmasi 11 2020.txt	Teknik Mesin 17 2020.txt	Teknik Mesin 9 2020.txt
		Teknik Mesin 10 2020.txt	Teknik Mesin 20 2020.txt	
		Teknik Mesin 14 2020.txt	Teknik Mesin 24 2020.txt	
		Teknik Mesin 15 2020.txt	Teknik Mesin 29 2020.txt	
		Teknik Mesin 16 2020.txt	Teknik Mesin 3 2020.txt	
Cluster 2:	23 items	Akuntansi 11 2020.txt	Kebidanan 2.2020.txt	TI 27.2017.txt
		Akuntansi 29 2018.txt	Kebidanan 21.2019.txt	TI 28.2017.txt
		Akutansi 5 2020.txt	Kebidanan 24.2018.txt	TI 9.2020.txt
		Farmasi 21 2018.txt	Kebidanan 26.2018.txt	
		Kebidanan 1.2020.txt	Kebidanan 28.2018.txt	
		Kebidanan 10.2020.txt	Kebidanan 29.2018.txt	
		Kebidanan 13.2020.txt	Kebidanan 3.2020.txt	
		Kebidanan 16.2019.txt	Kebidanan 5.2020.txt	
		Kebidanan 17.2019.txt	komputer 14.2018.txt	

		Kebidanan 19.2019.txt	TI 16.2018.txt	
Cluster 3:	76 items	Akuntansi 16 2019.txt	Kebidanan 18.2019.txt	TI 1.2020.txt
		Akuntansi 12 2019.txt	Kebidanan 20.2019.txt	TI 10.2019.txt
		Akuntansi 13 2019.txt	Kebidanan 25.2018.txt	TI 11.2019.txt
		Akuntansi 22 2019.txt	Kebidanan 30.2018.txt	TI 12.2019.txt
		Akuntansi 23 2019.txt	Kebidanan 4.2020.txt	TI 14.2018.txt
		Akuntansi 25 2019.txt	Kebidanan 9.2020.txt	TI 17.2018.txt
		Akuntansi 26 2019.txt	komputer 1 2020.txt	TI 18.2018.txt
		Akuntansi 27 2018.txt	komputer 13.2018.txt	TI 19.2017.txt
		Akuntansi 3 2020.txt	komputer 16. 2018.txt	TI 2.2020.txt
		Akuntansi 30 2018.txt	komputer 17.2018.txt	TI 20.2017.txt
		Farmasi 12 2019.txt	komputer 19.2018.txt	TI 21.2017.txt
		Farmasi 13 2019.txt	komputer 20.2018.txt	TI 22.2017.txt
		Farmasi 17 2018.txt	komputer 23.2018.txt	TI 23.2017.txt
		Farmasi 19 2018.txt	komputer 24.2018.txt	TI 24.2017.txt
		Farmasi 2 2020.txt	komputer 25.2017.txt	TI 25.2017.txt
		Farmasi 22 2018.txt	komputer 29. 2017.txt	TI 26.2017.txt
		Farmasi 24 2018.txt	komputer 3. 2020.txt	TI 29.2017.txt
		Farmasi 27 2018.txt	komputer 30.2017.txt	TI 3.2020.txt
		Farmasi 28 2018.txt	komputer 5.2020.txt	TI 30.2017.txt
		Farmasi 29 2018.txt	komputer 6.2019.txt	TI 4.2020.txt
		Farmasi 3 2020.txt	komputer 8.2019.txt	TI 5.2020.txt
		Farmasi 30 2018.txt	komputer 9.2020.txt	TI 6. 2020.txt

		Farmasi 5 2019.txt		Teknik Mesin 1 2019.txt
		Farmasi 8 2020.txt		Teknik Mesin 21 2020.txt
		Kebidanan 12.2020.txt		Teknik Mesin 22 2020.txt
		Kebidanan 14.2019.txt		Teknik Mesin 6 2018.txt
		Kebidanan 15.2019.txt		Teknik Mesin 7 2017.txt
Cluster 4:	11 items	Akuntansi 1 2020.txt	komputer 2.2020.txt	Teknik Mesin 2 2020.txt
		Akuntansi 8 2020.txt	komputer 21.2018.txt	Teknik Mesin 23 2020.txt
		komputer 15.2018.txt	Teknik Mesin 11 2020.txt	Teknik Mesin 30 2017.txt
		komputer 18.2018.txt	Teknik Mesin 12 2020.txt	
Cluster 5:	33 items	Akuntansi 10 2020.txt	Farmasi 20 2018.txt	komputer 12. 2018.txt
		Akuntansi 15 2019.txt	Farmasi 23 2018.txt	komputer 26.2017.txt
		Akuntansi 18 2019.txt	Farmasi 25 2018.txt	komputer 28.2017.txt
		Akuntansi 2 2020.txt	Farmasi 26 2018.txt	komputer 4. 2020.txt
		Akuntansi 28 2018.txt	Farmasi 4 2020.txt	komputer 7.2019.txt
		Akuntansi 6 2020.txt	Kebidanan 11.2020.txt	TI 13.2019.txt
		Akuntansi 7 2020.txt	Kebidanan 6.2020.txt	TI 15.2018.txt
		Akuntansi 9 2020.txt	Kebidanan 7.2020.txt	TI 8.2020.txt
		Farmasi 14 2019.txt	komputer 11.2018.txt	Teknik Mesin 26 2020.txt
		Farmasi 15 2019.txt		Teknik Mesin 27 2020.txt
		Farmasi 16 2019.txt		Teknik Mesin 4 2018.txt
		Farmasi 18 2018.txt		Teknik Mesin 5 2018.txt

Dari hasil tabel diatas dapat disimpulkan bahwa data penelitian pada tiap tiap cluster dihuni oleh beberapa anggota dengan data seperti pada tabel 4.6

Tabel 4. 6 Kesimpulan data penelitian pada tiap cluster

Cluster	Akuntansi	Farmasi	Teknik Mesin	Teknik Komputer	Teknik Informatika	Kebidanan	Jumlah
cluster 0	7	5	4	3	1	4	24
cluster 1	0	2	11	0	0	0	13
cluster 2	3	1	0	1	4	14	23
cluster 3	10	14	5	16	22	9	76
cluster 4	2	0	5	4	0	0	11
cluster 5	8	9	4	6	3	3	33

4.4 Hasil Pengukuran Davies Bouldin Index (DBI)

Davies-Bouldin Index merupakan metode validasi cluster dari hasil clustering. Pendekatan pengukuran DBI yaitu memaksimalkan jarak inter cluster serta meminimalkan jarak intra cluster. Semakin kecil nilai DB Index menunjukkan skema cluster yang paling optimal. Semakin besar nilai purity (mendekati 1) semakin baik kualitas cluster. Dalam penelitian ini hasil klaster dari k-means clustering mendapatkan nilai DBI sebesar -6.186

4.5 Luaran

Luaran yang akan dicapai dalam Penelitian ini antara lain:

- Pemodelan Metode K-Means
- Jurnal Internasional scopus Informatika pada Univeersitas Ahmad Dahlan (draf)

BAB V

KESIMPULAN

5.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan pengklasteran dokumen abstrak penelitian berbahasa Indonesia dengan menerapkan Algoritma K Means, klaster yang dihasilkan cukup baik, sehingga dapat dijadikan rekomendasi bahwa metode K-Means klastering baik digunakan dalam penerapan sistem temu kembali.

Pre processing pada tiap data abstrak harus tetap dilakukan karena proses, *case folding*, *stemming*, *stopwords*, dan term mempengaruhi hasil klaster yang dibentuk.

5.2 Saran

Isi dari kamus stopwords berbahasa Indonesia sebaiknya diperbanyak, sebagai kontribusi pengaruh terhadap proses pre-processing.

Hasil klaster Tugas Akhir masih dapat ditingkatkan dengan menerapkan algoritma klaster yang lain, dengan tetap menggunakan proses pre-processing.

Hasil penelitian ini memungkinkan menjadi pertimbangan dibuatnya sebuah sistem temu kembali maupun sebagai sistem deteksi plagiat.

DAFTAR PUSTAKA

- [1] Akaresti, Amalia Kusuma, Mochammad Ali Fauzi, and Fitra Abdurrachman Bachtiar. "Peringkasan Multi-Dokumen Berbasis Clustering pada Sistem Temu Kembali Berita Online Menggunakan Metode K-Means." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN 2548 (2018): 964X.
- [2] Janani, R., and S. Vijayarani. "Text document clustering using spectral clustering algorithm with particle swarm optimization." *Expert Systems with Applications* 134 (2019): 192-200..
- [3] Abualigah, Laith Mohammad, Ahamad Tajudin Khader, and Essam Said Hanandeh. "A new feature selection method to improve the document clustering using particle swarm optimization algorithm." *Journal of Computational Science* 25 (2018): 456-466..
- [4] Chakraborty, Saptarshi, and Swagatam Das. "k– Means clustering with a new divergence-based distance metric: Convergence and performance analysis." *Pattern Recognition Letters* 100 (2017): 67-73.
- [5] Hajeer, Safaa I., et al. "A new stemming algorithm for efficient information retrieval systems and web search engines." *Multimedia Forensics and Security*. Springer, Cham, 2017. 117-135.
- [6] Arifin, Agus Zainal, and Ari Novan Setiono. "Klasifikasi Dokumen Berita Kejadian Berbahasa Indonesia dengan Algoritma Single Pass *Clustering*." *Prosiding Seminar on Intelligent Technology and its Applications (SITIA), Teknik Elektro, Institut Teknologi Sepuluh Nopember Surabaya.*[This page intentionally left blank]. 2002.
- [7] Wang, Zhongqing, and Yue Zhang. "A Neural Model for Joint Event Detection and Summarization." *IJCAI*. 2017.
- [8] Agusta, Y. 2007. K-means - Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika* Vol. 3 (Februari 2007): 47-60.
- [9] Gosno, Eric Budiman, Isye Arieshanti, and Rully Soelaiman. "Implementasi KD-Tree K-Means Clustering untuk Klasterisasi Dokumen." *Jurnal Teknik ITS* 2.2 (2013): A432-A437.
- [10] Haryo Guritno. "Klasterisasi Dokumen Cerpen Dengan Metode K-Means Clustering" thesis Udinus (2015).
- [11] Duan, Yanling, Qun Liu, and Shuyin Xia. "An improved initialization center k-means clustering algorithm based on distance and density." *AIP Conference Proceedings*. Vol. 1955. No. 1. AIP Publishing LLC, 2018.

ORGANISASI PENGUSUL

Ketua

Nama : Very Kurnia Bakti, M.Kom
NIPY : 09008004
NIDN : 0625118301
Pangkat / Golongan : IIIC
Jabatan Fungsional : Lektor
Jabatan Struktural : Kepala Bagian TIK
Bidang Ilmu : Teknik Informatika
Pengalaman Penelitian :

- a. Segmentasi Dan Perbaikan Citra Untuk Proses Pengukuran Dimensi Beras
- b. Upaya Peningkatan Pendapatan Industri Kecil Menengah Di Kota Tegal Dengan Pembuatan Toko Daring Berbasis Framework Codeigniter
- c. Upaya Peningkatan Responden Tracer Study Dengan Metode Hibrid Gis, Lbs Dan Media Sosial Pada Politeknik Harapan Bersama
- d. Sistem Informasi Rekam Medik Pasien Sebagai Implementasi Big Data Dengan NIK Di Pelayanan Kesehatan Kota Tegal

Anggota

Nama : Arif Rakhman, SE, S.Pd, M.Kom
NIPY : 05.016.291
NIDN : 0623118301
Pangkat / Golongan : IIIC
Jabatan Fungsional : Lektor
Jabatan Struktural : Dosen
Bidang Ilmu : Teknik Informatika
Pengalaman Penelitian :

Pengalaman Penelitian :

- a. Sistem Informasi Stok Kebutuhan Darah Pada Palang Merah Indonesia Dengan Metode Weighted Moving Average
- b. Analisis Sentimen Review Media Massa Menggunakan Metode C 4.5 berbasis forward selection
- c. Upaya Peningkatan Pendapatan Industri Kecil Menengah Di Kota Tegal Dengan Pembuatan Toko Daring Berbasis Framework Codeigniter

REALISASI ANGGARAN

LAMPIRAN



SURAT KEPUTUSAN
DIREKTUR POLITEKNIK HARAPAN BERSAMA
NOMOR: 098 .05/PHB/V/2021
TENTANG
PENERIMA PENDANAAN HIBAH KOMPETITIF PENELITIAN DAN
PENGABDIAN MASYARAKAT OLEH INSTITUSI
BAGI DOSEN POLITEKNIK HARAPAN BERSAMA
TAHUN ANGGARAN 2020/2021 SEMESTER GENAP

- Menimbang : a. bahwa untuk meningkatkan kualitas dan kuantitas pelaksanaan penelitian dan pengabdian kepada masyarakat bagi Dosen di Politeknik Harapan Bersama, maka perlu menetapkan kebijakan dalam bidang pendanaan penelitian dan pengabdian kepada masyarakat;
- b. bahwa untuk tertib administrasi keuangan dalam pendanaan penelitian dan pengabdian kepada masyarakat, maka perlu ditetapkan tahapan penyerahan pendanaan oleh institusi untuk hibah kompetitif penelitian dan pengabdian masyarakat kepada Dosen Politeknik harapan Bersama;
- c. bahwa nama-nama yang tercantum dalam lampiran telah lolos kualifikasi untuk menerima pendanaan hibah kompetitif dari Institusi;
- d. berdasarkan pertimbangan sebagaimana dimaksud pada huruf a dan b, dipandang perlu menetapkan Surat Keputusan Direktur Politeknik Harapan Bersama;
- Mengingat : 1. Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional (Lembaran Negara Republik Indonesia Tahun 2003 Nomor 78, Tambahan Lembaran Negara Republik Indonesia Tahun 2003 Nomor 4301);
2. Undang-Undang Republik Indonesia Nomor 28 Tahun 2004 tentang Perubahan Undang-Undang Nomor 16 Tahun 2001 tentang Yayasan (Lembaran Negara Republik Indonesia Tahun 2004 Nomor 115, Tambahan Lembaran Negara Republik Indonesia Tahun 2004 Nomor 4430);
3. Undang-Undang Republik Indonesia Nomor 14 Tahun 2005 tentang Guru dan Dosen (Lembaran Negara Republik Indonesia Tahun 2005 Nomor 157, Tambahan Lembaran Negara Republik Indonesia Tahun 2005 Nomor 4586);
4. Undang-Undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi (Lembaran Negara Republik Indonesia Tahun 2012 Nomor 158, Tambahan Lembaran Negara Republik Indoneisa Tahun 2012 Nomor 5336);

5. Peraturan Pemerintah..

5. Peraturan Pemerintah Republik Indonesia Nomor 4 Tahun 2014 tentang Penyelenggaraan Pendidikan Tinggi dan Pengelolaan Perguruan Tinggi (Lembaran Negara Republik Indonesia Tahun 2014 Nomor 16, Tambahan Lembaran Negara Republik Indonesia Tahun 2014 Nomor 5500);
 6. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 3 Tahun 2020 tentang Standar Nasional Pendidikan Tinggi (Berita Negara Republik Indonesia Tahun 2020 Nomor 47);
 7. Keputusan Menteri Pendidikan Nasional Republik Indonesia Nomor: 128/D/0/2002 tentang Pemberian Ijin Penyelenggaraan Program-Program Studi dan Pendirian Politeknik Harapan Bersama di Tegal yang Diselenggarakan oleh Yayasan Pendidikan Harapan Bersama di Tegal;
 8. Keputusan Menteri Hukum dan Hak Asasi Manusia Republik Indonesia Nomor: AHU-2674.AH.01.04 Tahun 2012 tentang pengesahan Yayasan Pendidikan Harapan Bersama (Tambahan Berita Negara Republik Indonesia Tanggal 20/6-2014 No. 49);
 9. Keputusan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor: 231/KPT/1/2018 tentang Yayasan Pendidikan Harapan Bersama sebagai Badan Penyelenggara Politeknik Harapan Bersama;
 10. Surat Keputusan Yayasan Pendidikan Harapan Bersama Nomor 114.05/YPHB/XII/2020 tentang Statuta Politeknik Harapan Bersama;
- Memperhatikan Surat Pemberitahuan Pusat Penelitian dan Pengabdian Masyarakat (P3M) Nomor: 064.03/P3M.PHB/111/2021 tentang pengajuan dan penerimaan proposal Penelitian dan Pengabdian Kepada Masyarakat Politeknik Harapan Bersama Semester Genap Tahun Akademik 2020/2021.

MEMUTUSKAN:

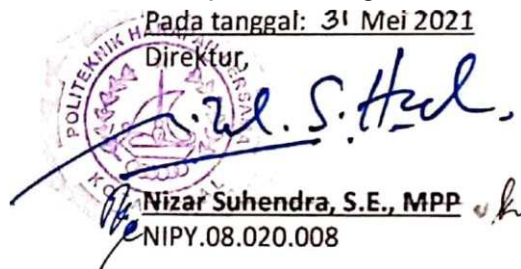
- Menetapkan Surat Keputusan Direktur Politeknik Harapan Bersama tentang Penerima Pendanaan Oleh Institusi Untuk Hibah Kompetitif Penelitian dan Pengabdian Masyarakat Bagi Dosen Politeknik Harapan Bersama Tahun Anggaran 2020/2021.
- Pertama Menetapkan nama yang tercantum dalam lampiran Keputusan ini sebagai Penerima Pendanaan Oleh Institusi Untuk Hibah Kompetitif Penelitian dan Pengabdian Masyarakat Bagi Dosen Politeknik Harapan Bersama Tahun Anggaran 2020/2021.
- Kedua
1. Pemberian bantuan dana penelitian minimal Rp. 2.000.000,- (Dua juta rupiah) per judul;
 2. Pemberian bantuan dana pengabdian kepada masyarakat minimal Rp. 2.000.000,- (Dua juta rupiah) per judul;
 3. Pembayaran dilakukan dengan 2 (dua) tahap, yaitu:
 - a. Pembayaran tahap I sebesar 60% dari total dana yang didapatkan setelah menyerahkan proposal dan perjanjian yang telah ditandatangani oleh Direktur Politeknik Harapan Bersama;
 - b. Pembayaran Tahap II sebesar 30% dari total dana yang didapatkan setelah menyerahkan laporan hasil; dan
 - c. 10% dari total dana yang didapatkan diserahkan kepada P3M.

- Ketiga Dosen yang melaksanakan Penelitian dan/atau Pengabdian Kepada Masyarakat wajib menyerahkan laporan hasil kepada Direktur dan Wakil Direktur I melalui Pusat Penelitian dan Pengabdian Masyarakat (P3M), meliputi:
- a. Laporan penelitian sebanyak 2 (dua) eksemplar;
 - b. Softcopy Jurnal;
 - c. Softcopy.
- Keempat Semua produk hasil penelitian dan pengabdian masyarakat termasuk Paten menjadi hak milik Politeknik Harapan Bersama.
- Kelima Surat Keputusan ini berlaku sejak tanggal ditetapkan dan apabila di kemudian hari terdapat kekeliruan akan diadakan perbaikan sebagaimana mestinya.

Ditetapkan di: Tegal

Pada tanggal: 31 Mei 2021

Direktur,



Nizar Suhendra, S.E., MPP

NIPY.08.020.008

Lampiran: Surat Keputusan Direktur Politeknik
Harapan Bersama

Tentang : Penerima Pendanaan Oleh Institusi
Untuk Hibah Kompetitif Penelitian dan
Pengabdian Masyarakat Bagi Dosen
Politeknik Harapan Bersama Tahun
Anggaran 2020/2021 Semester Genap

Nomor : og6 .05/PHB/V/2021

Tanggal : 21 Mei 2021

26	Nllatul liah, S.ST, M.Web. Umriaty, S.ST., M.ites.	Pengaruh Stunting Terhadap Perkembangan Anak Usia 2—6 Tahun Di Wilayah Kota Tegal	DIII Kebidanan	Penelitian	Rp. 3,100,000
27	Ulfatul Latifah, S.KM, M. Kes. Riska Arsita Harnawati, S. ST., M.M.	Perbedaan Perilaku Penerapan Protokol Kesehatan Pada Ibu Menyusui Yang Sudah Divaksinasi Dan Belum Divaksinasi Covid-19 Di Wilayah Kecamatan Margadana Tegal	DIII Kebidanan	Penelitian	Rp. 3,028,500
28	Iroma Maulida, SxM, M.Epid. Reny Eka Saputri, S.Tr.Keb., M.Tr.Keb.	Studi Kasus Penanganan Keluarga Terhadap Penderita Covid-19 Dengan Penyakit Penyerta	DIII Kebidanan	Penelitian	Rp. 2,957,000
29	UIII Albab, S. ST., M.T. Qirom, S.Pd, MT. Muhammad Irfan Fauzi.	Analisa Efektivitas Alat Pengusir Nyamuk Komersil Dengan Alat Pengusir Nyamuk Berdasarkan Gelombang Ultrasonik Dan UV Light Trap	DIII Teknik Elektronika	Penelitian	Rp. 2,900,000
30	Bahrn Niam, M.T. Rony Darpono, M.T. Rofi Irfanto.	Pengembangan Deteksi Suhu Dan Kelembaban Laboratorium Elektronika Dengan Menggunakan Metode <i>Fuzzy logic</i>	DIII Teknik Elektronika	Penelitian	Rp. 2,571,000
31	Rony Darpono, M.T. Ratri Wikaningtya S, M.Pd.	Robot Sterilisasi Ruang Covid- 19 Menggunakan Sinar UV Type-C Berbasis Arduino Dengan Kendali Android	DIII Teknik Elektronika	Penelitian	Rp. 2,828,500
32	Very Kurnia Bakti, M.Kom. Arif Rakhman, S.E., S.Pd., M. Kom. Mohammad Rij al Arfani	Klasterisasi Dokumen Penelitian Perguruan Tinggi Menggunakan K-Means <i>Clustering</i> , Sebagai Analisa Penerapan Sistem Temu Kembali	DIII Teknik Komputer	Penelitian	Rp. 3,314,000
33	Dani Fitria Brilianti, M.Pd. Ariel Zul Fauzi, M.Pd. Dinda Ayu Ningrum.	Penerapan Metode Voice Over (VO) Pada Pengajaran <i>Listening</i> Untuk Meningkatkan Kemandirian Belajar Mahasiswa Saat Pembelajaran Daring	DIII Teknik Komputer	Penelitian	Rp. 2,728,500

Lampiran: Surat Keputusan Direktur Politeknik

Harapan Bersama

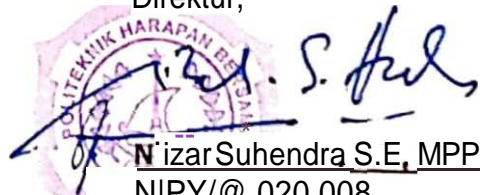
Tentang : Penerima Penrlanaan Oleh Institansi
Untuk Hibah Kompetitif Penelitian dari
Pengabdian Masyarakat Bagi Dosen
Politeknik Harapan Bersama Tahun
Anggaran 2020/2021 Semester Genap

Nomor 05/PHB/V/2021

Tanggal : 08 Mei 2021

13	Slamet Wlyono, S. Pd., M. Eng Dega Surono Wibowo, S.T., M. Kom, Riszki Wijayatun Pratiwi, S.Kom., M.Cs. Naimatul Maulidiyah Getar Dewantara Agung Iswanto	Pemanfaatan Teknik <i>Scraping</i> ! Data Untuk Perencanaan Usaha Jualan Online Menggunakan <i>Marketplace</i>	Sarjana Terapan Teknik Informatika	PKM	Rp. 2,900,000
14	Muhammad Fikri Hidayattullah, S.T., M.Kom. Dega Surono Wibowo, S.T., M. Kom. Ardi Susanto, S.Kom., M.Cs. Alfin Auzikri Wildan Sania Alfiansyah	Pengenalan <i>Sofware</i> Al- Mausu'ah Al-Hadits Bagi Santr\ Madrasah Fiqih Sumber Ilmu Dalam Melakukan Studi Takhrij Hadits	Sarjana Terapan Teknik Informatika	PKM	Rp. 2,787,500

Direktur,



Nizar Suhendra S.E., MPP \

N|PY/@.020.008

KLASTERISASI DOKUMEN PENELITIAN PERGURUAN TINGGI MENGUNAKAN *K-MEANS CLUSTERING*, SEBAGAI ANALISA PENERAPAN SISTEM TEMU KEMBALI

Very Kurnia Bakti¹, Arif Rakhman²

verykurniabakti@gmail.com

^{1,2} DIII Teknik Komputer Politeknik Harapan Bersama
Jalan Mataram No 9 Tegal

Abstrak

Pencarian dokumen yang ada saat ini yaitu menampilkan hasil pencarian berurut berdasarkan peringkat kecocokan (document ranking). Hal tersebut menyebabkan penemuan data dokumen tidak secara akurat terkelompok pada masing – masing tema. *Clustering* dapat digunakan dalam pengkategorian atau pengelompokan dokumen. *Clustering* dengan metode K means adalah algoritma sederhana yang dikembangkan Mac Queen pada tahun 1967. Dari hasil penelitian yang telah dilakukan pengklasteran dokumen abstrak penelitian dosen berbahasa Indonesia dengan menerapkan Algoritma K Means, klaster yang dihasilkan cukup baik, sehingga dapat dijadikan rekomendasi bahwa metode K-Means klastering baik digunakan dalam penerapan sistem temu kembali. Dengan nilai Davies Bouldin Index sebesar -6.186.

Kata Kunci: *dokumen, k-means, clustering*

1. PENDAHULUAN

Banyaknya jumlah data dokumen penelitiandari berbagai program studi dapat memberi kontribusi besar dalam sulitnya proses pencarian suatu dokumen. Pencarian dokumen yang ada saat ini hanya menampilkan hasil pencarian berurut berdasarkan peringkat kecocokan (document ranking). Hal tersebut menyebabkan penemuan data dokumen tidak secara akurat terkelompok pada masing – masing tema.

Dengan adanya pengelompokan dokumen, maka tidak harus membuka halaman terlalu banyak, karena dokumen hasil pencarian telah dikelompokkan berdasarkan kategori yang dapat menggambarkan isi dari suatu dokumen, hal tersebut tentunya dapat mempermudah dalam menemukan beberapa dokumen yang diinginkan, oleh karenanya sebelum proses tersebut dilakukan maka, proses tersebut perlu dianalisis sebelum benar-benar diaplikasikan ke dalam aplikasi yang sifatnya *executable*. Oleh karena itu diperlukan metode pengelompokan/ *clustering* yang nantinya dapat dipastikan keberhasilan pengelompokan suatu dokumen penelitiandengan baik.

Clustering dapat digunakan dalam pengkategorian atau pengelompokan dokumen. Caranya adalah dengan mengelompokkan dokumen-dokumen ke dalam *clusters* berdasarkan kedekatan atau kemiripan antar dokumen (similarity) ^[1,5], sehingga dokumen yang berhubungan dengan suatu tema tertentu

secara otomatis ditempatkan pada cluster yang sama. Saat ini ada beberapa algoritma *clustering* diantaranya partitional (K Means) dan hierarchical. ^[1] *Clustering* dengan metode K means adalah algoritma sederhana yang dikembangkan Mac Queen pada tahun 1967. Algoritma tersebut terkenal dengan kemampuannya untuk mengklaster data yang besar dan dapat menangani data *outlier*. *K-means* merupakan metode pengklasteran yang memisahkan data kedalam *k* kelompok yang berbeda artinya sebelum dilakukan klasterisasi maka perlu menentukan jumlah *k* yang diinginkan. Selain itu *k means* merupakan *center based clustering* yang menentukan setiap klaster dari titik pusat klasternya ^[1, 3].

2. METODE PENELITIAN

A. Pengumpulan Data

Bahan Penelitian yang digunakan dalam penelitian berupa, data Penelitian Perguruan Tinggi yang diambil pada bagian abstraknya saja. Dari masing – masing 7 program studi di Politeknik Harapan Bersama.

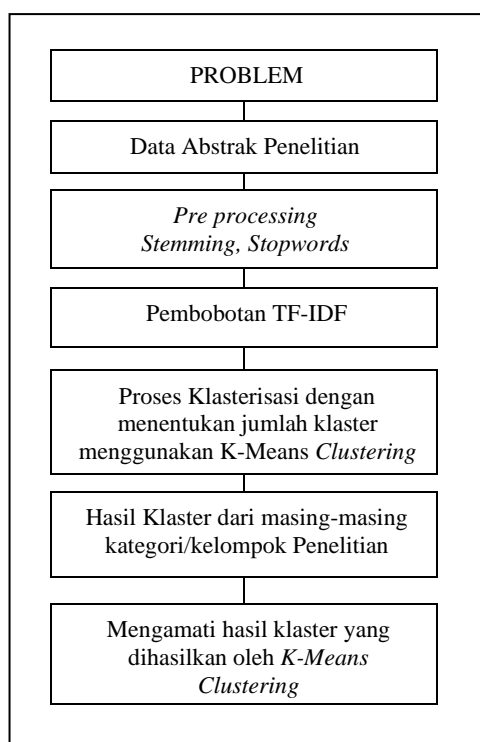
B. Alat Penelitian

Alat yang digunakan pada saat penelitian adalah perangkat keras dan perangkat lunak komputer. Perangkat keras yang dibutuhkan berupa komputer/laptop Sedangkan perangkat lunak yang dibutuhkan adalah:

- 1) Sistem Operasi Windows yang digunakan sebagai sistem untuk menjalankan aplikasi pemrograman
- 2) Aplikasi Rapidminer 5.3 untuk menganalisa

C. Prosedur Penelitian

Dokumen yang digunakan dalam penelitian ini diambil dari sampel abstrak Penelitian Perguruan Tinggi dari P3M Politeknik Harapan Bersama dari 7 program studi.



Gambar 1. Prosedur Penelitian

3. HASIL DAN PEMBAHASAN

A. Ekstraksi Dokumen

Dalam tahap ini data abstrak dari masing-masing penelitian yang berbentuk dokumen berformat *.txt diperlukan proses *pre-processing*, yaitu *case folding*, *Tokenizing*, *stopwords* dan pembobotan.

B. Case Folding

Dari dokumen penelitian yang ada selanjutnya masuk kedalam proses *case folding* dimana proses ini melakukan penyamaan antar kata dengan cara merubah huruf besar menjadi seluruhnya huruf kecil. Sehingga seluruh kata yang diproses semuanya menjadi huruf kecil. Dari hasil pemrosesan tersebut dapat diambil contoh

hasil pemrosesan *case folding* seperti pada tabel 1.

Tabel 1 Hasil kata proses *case folding*

alpinia	Bangsa	Coefficient
alri	Bangun	Coklat
alternatif	Bangunan	Coli
alternative	Bani	Coliform
altilis	Bank	Collected
alun	Banteng	Collection
alur	Bantu	Coloni
amalia	Bantuan	Coming
amaliyah	Banyak	Commerce

C. Tokenizing

Proses selanjutnya setelah melalui tahap *case folding* adalah proses Pemisahan rangkaian term (*tokenization*). Term dapat berupa kata atau frasa di dalam dokumen. Namun, kata-kata yang tidak memberikan perbedaan seperti ini, itu, saya, kamu, serta tanda-tanda baca dihilangkan atau dianggap bukan *term*. Hal ini bertujuan untuk mendapatkan hanya kata – kata tertentu saja yang nantinya didapat dan berkontribusi sebagai ciri – ciri dari masing-masing jenis judul tugas akhir. Selain itu masih dalam proses *tokenizing* dilakukan pula *filter token* yaitu dengan memberikan batasan jumlah karakter dari masing-masing kata minimal 3 karakter dan maksimal 25 karakter dari setiap kata. Hal ini dilakukan untuk mensortir atau menghilangkan kata salah pengetikan karena terlalu pendek atau terlalu panjang dalam tiap kata.

D. Stopwords

Stopwords dilakukan dalam penelitian ini untuk mendapatkan kata dasar, dikarenakan di tiap dokumen penelitian banyak terdapat kata yang memiliki banyak imbuhan. Hal tersebut akan mempengaruhi hasil kluster nantinya. Dalam proses *stopwords* inilah nantinya tiap kata yang memiliki kata dasar yang sama akan dihilangkan imbuhanannya sehingga didapat hanya kata dasar saja. Proses *stopwords* ini dilakukan dengan cara manual yaitu dengan membuat kamus *stopwords* sendiri sebanyak 8093 kata yang tentunya masih sedikit jika dibandingkan dengan jumlah kata pada bahasa Indonesia.

Dalam proses pencarian kata dasar ini, mengingat kata yang terdapat dalam dokumen penelitian tidak hanya kata berbahasa Indonesia saja, melainkan juga terdapat kata

dengan bahasa Inggris, maka dalam penelitian ini dilakukan pemrosesan sebanyak dua kali *stopwords* yaitu dengan *stopwords* bahasa Indonesia dan *stopwords* bahasa Inggris. Dengan harapan hasil kata dasar yang dihasilkan memiliki kontribusi dalam menentukan hasil kluster nantinya.

E. Pembobotan TF-IDF

Tahap pembobotan ini menggunakan Metode TF-IDF dimana terdapat integrasi antar term frequency (tf), dan inverse document frequency (idf) Dengan rumus :

$$w(t,d) = tf(t,d) * \log_2(N/nt) \quad (1)$$

Simbol $w(t,d)$ merupakan bobot dari term t dalam sebuah dokumen penelitian d sedangkan $tf(t,d)$ adalah frekuensi term dalam dokumen penelitian (tf) dan N merupakan ukuran data training yang digunakan untuk penghitungan IDF. Adapun nt adalah jumlah dari dokumen yang ditraining yang mengandung nilai t .

F. Proses Klasterisasi dengan K-Means Clustering

Dari tahapan-tahapan yang telah dilalui dengan proses ekstraksi dokumen, maka langkah selanjutnya adalah proses mengkluster dokumen. Dalam penelitian ini proses kluster dokumen dilakukan dengan menggunakan algoritma *k-means clustering* dengan pertimbangan beberapa referensi dari penelitian sebelumnya dengan tema yang sama yaitu, *information retrieval* dengan *text mining*. Pada pengklasteran menggunakan *k-means clustering* dengan dasar algoritmanya adalah sebagai berikut:

- 1) Langkah pertama adalah dengan mengelompokan atau Inisialisasi kluster dalam penelitian ini dibuat empat kluster disesuaikan dengan jumlah dokumen empat program studi, D 3- Teknik Komputer, D 3-Kebidanan, D3-Akutansi, D3-Farmasi. Masing-masing program studi ada 50 judul tugas akhir.
- 2) Memasukan semua dokumen ke kluster yang paling cocok dengan berdasarkan pusat kluster. Dengan persamaan sebagai berikut:

Tabel 2. Dokumen yang paling cocok berdasarkan pusat kluster

word	in doc ument s	to tal	in class (ko mpu ter)	in class (aku tans i)	in class (keb idan an)	in class (far masi)
affecting	1,0	1,0	,0	1,0	,0	,0
abad	1,0	2,0	2,0	,0	,0	,0
abortus	1,0	1,0	,0	,0	1,0	,0
abrasiver	1,0	1,0	,0	,0	,0	1,0

4. PENUTUP

Dari hasil penelitian yang telah dilakukan pengklasteran dokumen abstrak penelitian berbahasa Indonesia dengan menerapkan Algoritma K Means, kluster yang dihasilkan cukup baik, sehingga dapat dijadikan rekomendasi bahwa metode K-Means klustering baik digunakan dalam penerapan sistem temu kembali. *Pre processing* pada tiap data penelitian harus tetap dilakukan karena proses, *case folding*, *stemming*, *stopwords*, dan term mempengaruhi hasil kluster yang dibentuk. Dalam penelitian ini hasil kluster dari k-means clustering mendapatkan nilai DBI sebesar - 6.186.

5. DAFTAR PUSTAKA

- [1] Agusta, Y. 2007. K-means - Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika Vol. 3 (Februari 2007): 47-60.
- [2] Arifin, Agus Zainal, and Ari Novan Setiono. "Klasifikasi Dokumen Berita Kejadian Berbahasa Indonesia dengan Algoritma Single Pass Clustering." *Prosiding Seminar on Intelligent Technology and its Applications (SITIA), Teknik Elektro, Institut Teknologi Sepuluh Nopember Surabaya.* [This page intentionally left blank]. 2002.
- [3] Cui, Xiaohui, Thomas E. Potok, and Paul Palathingal. "Document clustering using particle swarm optimization." *Swarm Intelligence Symposium, 2005. SIS 2005. Proceedings 2005 IEEE.* IEEE, 2005.
- [4] Gosno, Eric Budiman, Isye Ariesanti, and Rully Soelaiman. "Implementasi KD-Tree K-Means Clustering untuk Klasterisasi Dokumen." *Jurnal Teknik ITS* 2.2 (2013): A432-A437.

- [5] Haryo Guritno. "Klasterisasi Dokumen Cerpen Dengan Metode K-Means Clustering" thesis Udinus (2015).
- [6] Huang, Anna. "Similarity measures for text document clustering." *Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand*. 2008.
- [7] Selim, Shokri Z., and Mohamed A. Ismail. "K-means-type algorithms: a generalized convergence theorem and characterization of local optimality." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 1 (1984): 81-87.
- [8] Tala, Fadillah Z. "A study of stemming effects on information retrieval in Bahasa Indonesia." *Institute for Logic, Language and Computation Universeit Van Amsterdam* (2003).
- [9] Vidya Ayuningtias, M. Arif Bijaksana, Rimba Widhiana Ciptasari "Pengkategorian hasil Pencarian Dokumen dengan klastering" *tugas akhir, Universitas telkom university*. 2008
- [10] Yang, Yiming, et al. "Learning approaches for detecting and tracking news events." *IEEE Intelligent Systems* 4 (1999): 32-43.
- [11] Yi, B., Qiao, H., Yang, F., & Xu, C. (2010). An Improved Initialization Center Algorithm for K-Means Clustering. 2010 *International Conference on Computational Intelligence and Software Engineering, IEEE* (1), 1-4.



Yayasan Pendidikan Harapan Bersama
PoliTekniK Harapan Bersama

SmartComp (Jurnalnya Orang Pintar Komputer)

Campus I: Jl. Mataram No.9 Tegal 52142 Telp. 0283-352000 Fax. 0283-353353

Website: www.yeyekids.id

Email: snia@yeyekids.id

Tegal, 30 Agustus 2021

Nomor : 06/04/Smartcoop.PHBfVIII/2021
Lampiran -
Hal : **Surat Keterangan Submit Naskah Jurnal**

Kepada :

Politeknik Harapan Bersama

Terimakasih telah mengirimkan artikel ilmiah untuk diterbitkan pada Jurnal SmartComp: Jurnalnya Orang pintar komputer Volume 10 Nomor 3 Oktober 2021 (P—ISSN 2089—676X, E—ISSN 2549—0796) dengan Judul:

**“ KLASTERISASI INFORMASI PENELITIAN PERGURUAN TINGGI
SISTEM TEMU KEMBALI”**

Demikian informasi ini disampaikan, agar dapat digunakan sebagaimana mestinya. Atas perhatiannya, kami ucapkan terimakasih.

Hormat kami,
Managing Editor



Arif Rahman, SE, & Pd, M.Kom