

BUKU MANUAL APLIKASI



EVALUASI *KMEANS CLUSTERING* PADA *PREPROCESSING* SISTEM TEMU KEMBALI INFORMASI

Oleh:
Yudha Pradana Putra
NIM. 1641720044

**JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
SEPTEMBER 2020**

KATA PENGANTAR

Dengan mengucapkan puji syukur kehadirat Allah SWT, karena atas limpahan rahmat & hidayah-Nya, penulis dapat menyelesaikan buku manual untuk dengan judul:

“Evaluasi *Kmeans Clustering* Pada *Preprocessing* Sistem Temu Kembali Informasi”

Dalam buku manual ini berisi bagaimana tahapan yang dilakukan dalam menerapkan metode *Kmeans Clustering* pada sistem temu kembali informasi. Tahapan tersebut mulai dari spesifikasi perangkat komputer yang digunakan sampai implementasi dan pengujian efektifitas dan efisiensi penelitian. Mudah-mudahan dengan mempelajari buku manual ini diharapkan dapat memberikan tambahan ilmu pengetahuan bagi pembaca dan penelitian dalam bidang sistem cerdas.

Pada kesempatan ini penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Yth.

1. Bapak Drs. Awan Setiawan, MMT., selaku Direktur di Politeknik Negeri Malang.
2. Bapak Erfan Rohadi, ST., M.Eng. Ph.D, selaku Kepala P2M Politeknik Negeri Malang.
3. Bapak Rudy Ariyanto, ST., M.Cs., selaku Ketua Jurusan Teknologi Informasi Politeknik Negeri Malang.
4. Bapak Imam Fahrur Rozi, ST., MT selaku ketua Program Studi Teknik Informatika.
5. Bapak Dr. Eng. Faisal Rahutomo, ST., M.Kom. dan Bapak Yoppy Yunhasnawa, S.ST., M.Sc. selaku dosen pembimbing skripsi, yang telah memberikan waktu, kesempatan, petunjuk dan bimbingan.
6. Dosen-dosen pengajar Program Studi Teknik Informatika, Jurusan Teknologi Informasi, Politeknik Negeri Malang yang telah memberikan bimbingan dan ilmunya.
7. Teman-teman yang selalu meluangkan waktunya untuk membantu dan mendukung penulis sehingga penyusunan skripsi ini dapat diselesaikan dengan baik,
8. Semua pihak yang tidak dapat disebutkan satu persatu, yang telah memberikan bantuan hingga penyusunan skripsi ini selesai

Penulis menyadari bahwa buku manual ini masih jauh dari kesempurnaan. Akhir kata, semoga dengan terselesaikannya buku manual ini akan memberikan perubahan dan

perbaikan ke arah yang lebih baik bagi dunia pendidikan, khususnya di Politeknik Negeri Malang.

Malang, 29 September 2020

Penulis

DAFTAR ISI

KATA PENGANTAR (CONTOH).....	ii
DAFTAR ISI.....	iv
BAB I PENDAHULUAN.....	2
BAB II PERSIAPAN HARDWARE DAN SOFTWARE.....	3
BAB III PETUNJUK INSTALASI APLIKASI.....	4
BAB IV PETUNJUK FITUR-FITUR APLIKASI.....	5
BAB V PETUNJUK PENGELOLAAN APLIKASI.....	6
BAB VI PETUNJUK PENGGUNAAN APLIKASI.....	7

BAB I PENDAHULUAN

Sistem Temu Kembali Informasi digunakan untuk menemukan kembali informasi-informasi yang relevan terhadap kebutuhan pengguna dari suatu kumpulan informasi secara otomatis. Jadi secara konsep ada beberapa dokumen yang berisi informasi yang diorganisasikan ke dalam sebuah media penyimpanan untuk mempermudah ditemukannya kembali dokumen tersebut berdasarkan *query* yang dimasukkan oleh pengguna. Teknik pencarian pada sistem temu kembali informasi ini berbeda dengan sistem manajemen basis data, dimana pada sistem temu kembali informasi ini memiliki dua bagian yaitu bagian pengindexan dan pencarian.

Namun terkadang muncul permasalahan pada efisiensi sewaktu sistem memproses data yang sangat besar. Kurang efisiensi tersebut dikarenakan waktu tunggu sistem yang menjadi lebih lama karena diperlukan waktu untuk menghitung tingkat kemiripan *query* dengan dokumen. Sehingga pada penelitian ini diterapkan metode *Kmeans Clustering* untuk mengurangi jumlah proses penghitungan kemiripan *query* dengan dokumen. Dengan diterapkan metode *kmeans* pada sistem temu kembali informasi, diharapkan mampu mengatasi permasalahan tersebut.

BAB II PERSIAPAN HARDWARE DAN SOFTWARE

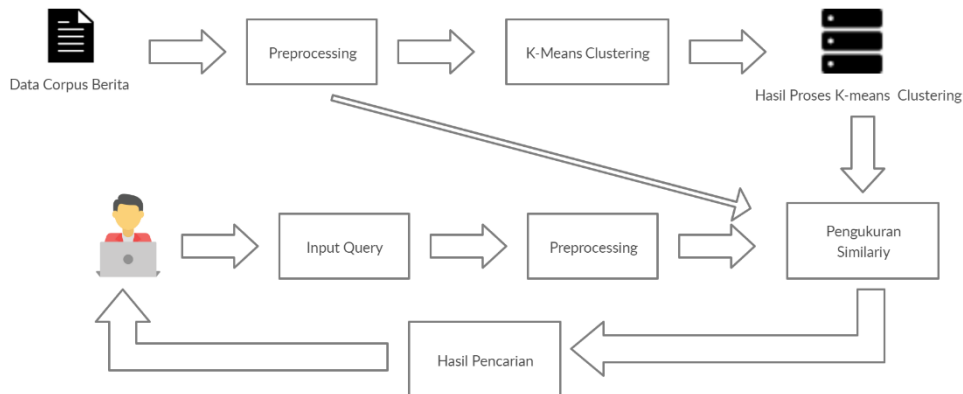
Spesifikasi perangkat keras minimal yang dibutuhkan untuk merancang, mengimplementasikan dan mengoperasikan sistem temu kembali informasi yang menggunakan *kmeans clustering* pada *preprocessing* ini adalah laptop atau komputer dengan spesifikasi sebagai berikut:

- a. *Processor* Intel(R) Core(TM) i5-6200U CPU 2.30 GHz
- b. *Memory* RAM 8 GB
- c. *Harddisk drive* dengan minimal *space* 500 GB
- d. Intel(R) HD *Graphics* 4000

Untuk dapat merancang, menerapkan dan menjalankan sistem temu kembali informasi yang menggunakan *kmeans clustering* pada *preprocessing* ini secara tepat dan lancar, perlu disiapkan komponen perangkat lunak berikut:

- a. Sistem Operasi Windows 10 64-bit
- b. *Pycharm*
- c. Web Browser Google Chrome
- d. *Python 3.7.7*
- e. *Package python* yang dibutuhkan: framework flask, math, datetime, os, csv, json, minidom, sastrawi, numpy, matplotlib, sklearn, genism
- f. MySQL, Web Server Apache, XAMPP

BAB III PETUNJUK INSTALASI APLIKASI



Pada tahap perencanaan ini, peneliti menentukan arsitektur sistem dan perancangan antarmuka sistem yang akan dibuat. Pada penelitian ini diterapkan tahap *preprocessing* pada corpus yang berisi berita untuk diklasterkan terlebih dahulu menggunakan metode *kmeans clustering*. Sebelum dikelompokkan, berita diproses terlebih dahulu pada *text processing* (*case folding, filtering, stemming, tokenizing*) dengan pembobotan kata menggunakan *TF – IDF* dan *Doc2vec*. Hasil dari proses klaster tersebut digunakan untuk proses sistem temu kembali informasi berdasarkan query yang diinputkan dan telah dilakukan tahap *preprocessing* sebelumnya. Terakhir dilakukan proses pengukuran similarity menggunakan metode cosine similarity antara *vector query* dan *vector centroid cluster*.

BAB IV PETUNJUK FITUR-FITUR APLIKASI

No	Menu	Penjelasan
1	Dashboard	Halaman yang menampilkan data hasil pengujian <i>recall</i> , <i>precision</i> dan <i>fscore</i>
2	News	Halaman yang menampilkan data berita yang tersimpan di database. Di halaman ini dapat menambah, mengubah dan menghapus berita yang akan diolah di sistem
3	Text Preprocessing	Halaman yang memproses dan menampilkan hasil text preprocessing setiap berita
4	Word Weighting	Halaman yang memproses dan menampilkan proses pembobotan kata menggunakan metode doc2vec dan tfidf setiap berita
5	Kmeans	Halaman yang menampilkan hasil proses pembentukan cluster. Pada halaman ini dapat membuat process kmeans baru dengan memasukkan jumlah cluster yang ingin dibentuk
6	Information Retrieval	Halaman yang digunakan untuk melakukan pengujian sistem temu kembali informasi dengan form pencarian berita berdasarkan kata kunci
7	Testing	Halaman yang menampilkan hasil detail pengujian sistem temu kembali informasi

BAB V PETUNJUK PENGELOLAAN APLIKASI

Aplikasi ini menggunakan *database* MySQL. Untuk konfigurasi *python* dengan MySQL, diharuskan menginstall *package* Flask-MySQL dengan cara “*pip install Flask-MySQL*”.

Setelah itu menambahkan *code* pada file *route* sebagai berikut :

```
from flaskext.mysql import MySQL

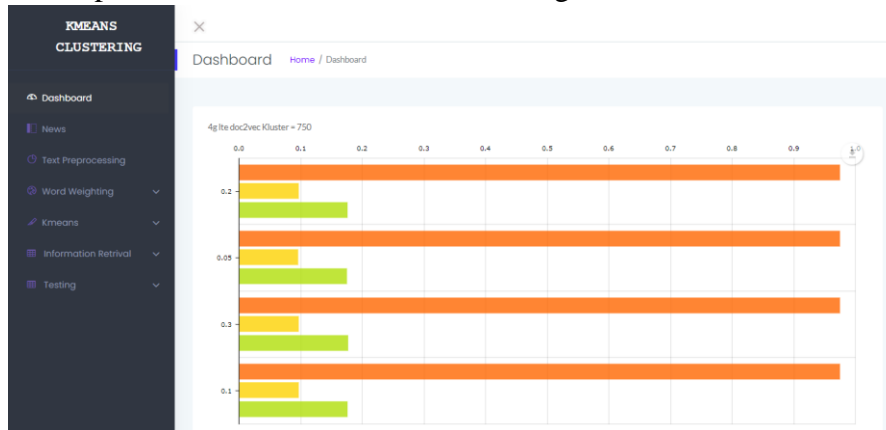
mysql = MySQL()
app.config['MYSQL_DATABASE_USER'] = 'root'
app.config['MYSQL_DATABASE_PASSWORD'] = ''
app.config['MYSQL_DATABASE_DB'] = 'skripsi'
app.config['MYSQL_DATABASE_HOST'] = 'localhost'
mysql.init_app(app)
```

BAB VI PETUNJUK PENGGUNAAN APLIKASI

Sebelum mengakses aplikasi, jalankan sistem terlebih dahulu melalui terminal dengan perintah “set FLASK_APP=skripsi.py”, “python -m flask run”.

1. Halaman Home

- Akses halaman web melalui <http://127.0.0.1:5000/>.
- Akan tampil halaman home dimana halaman ini menampilkan hasil pengujian recall, precision dan fscore dalam bentuk grafik



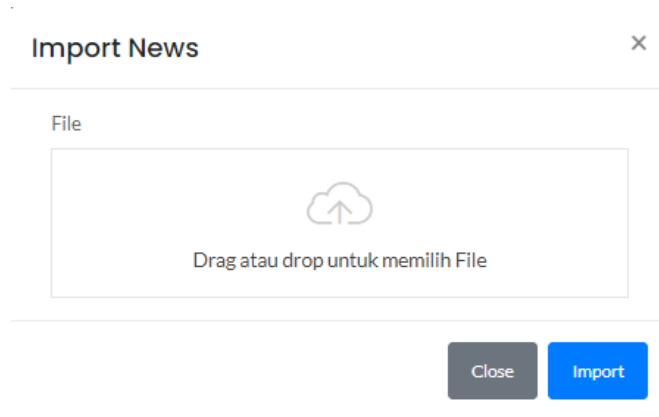
2. Halaman News

Pilih menu News pada menu bar. Dan akan muncul tampilan data berita yang akan diolah.

No	Title	Source	Action
0	Ponsel Huawei Honor 4C Dibanderol Rp 2,2 Juta	kompas.com	[Add] [Edit] [Delete]
1	Asosiasi RPP E-commerce Tidak Sesuai Hasil Diskusi	kompas.com	[Add] [Edit] [Delete]
2	Pemesan iPhone Jadi Sabun? Karyawan Pesaing Lazada?	kompas.com	[Add] [Edit] [Delete]
3	"Autofeather Failure", Momok bagi Pesawat Baling-baling	kompas.com	[Add] [Edit] [Delete]
4	Laptop Bezel Tertipis di Dunia Masuk Indonesia	kompas.com	[Add] [Edit] [Delete]
5	"Ok Google" Bakal Bisa Diakses Tanpa Internet	kompas.com	[Add] [Edit] [Delete]
6	Di Indonesia, Pasar Kamera Mirrorless Masih Kecil	kompas.com	[Add] [Edit] [Delete]
7	Menkominfo: E-commerce Tak Perlu Izin, Cukup Akreditasi	kompas.com	[Add] [Edit] [Delete]

a) Tambah Berita

- Untuk menambahkan berita dalam jumlah besar secara bersamaan, dapat melalui fitur import pada menu News. Klik pada ikon import. Jika sudah pilih file (hanya mendukung ekstensi .xml, .csv dan xls) yang akan di import dan kemudian klik tombol import.



- Untuk menambahkan berita secara satu persatu, klik tombol Add News. Kemudian isi form data berita mulai dari judul, deskripsi dan sumber berita. Jika sudah benar, klik tombol save.

b) Ubah Berita

- Untuk mengubah berita, klik tombol ikon edit berwarna kuning, kemudian akan muncul form untuk mengubah data berita. Ubah data berita mulai dari judul, deskripsi dan sumber berita. Jika sudah sesuai, klik tombol save changes.

c) Hapus Berita

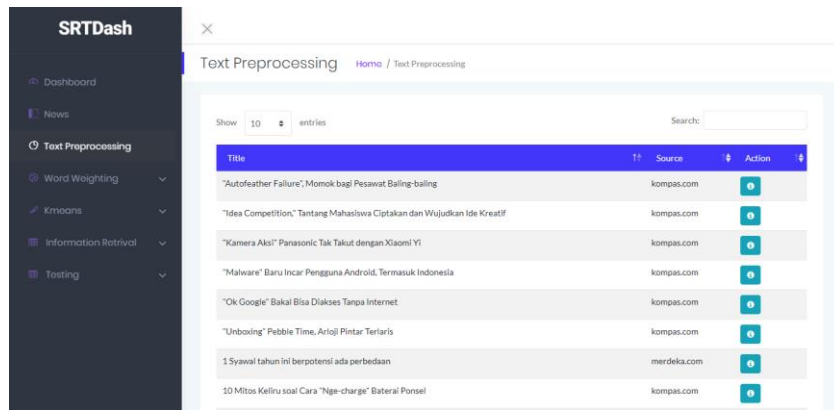
- Untuk menghapus berita, klik tombol ikon hapus berwarna merah pada berita yang ingin dihapus.

d) Detail Berita

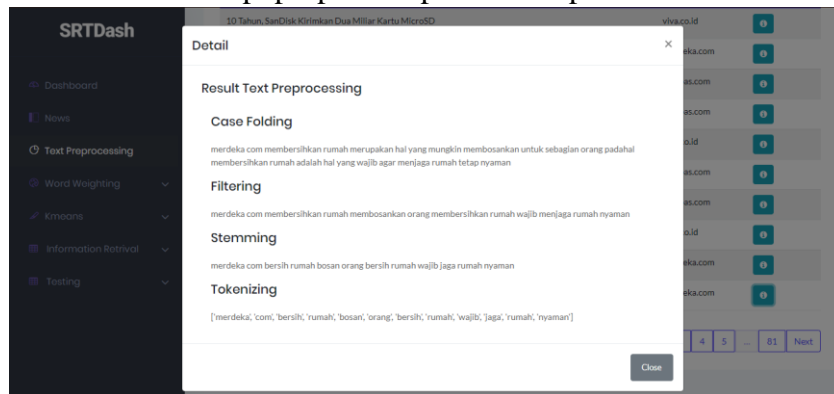
- Untuk melihat detail berita, klik tombol ikon info berwarna biru pada berita yang ingin dilihat.

3. Halaman Text Preprocessing

- Pilih menu Text Preprocessing pada menu bar. Halaman ini akan menampilkan data berita yang telah dipreprocessing.

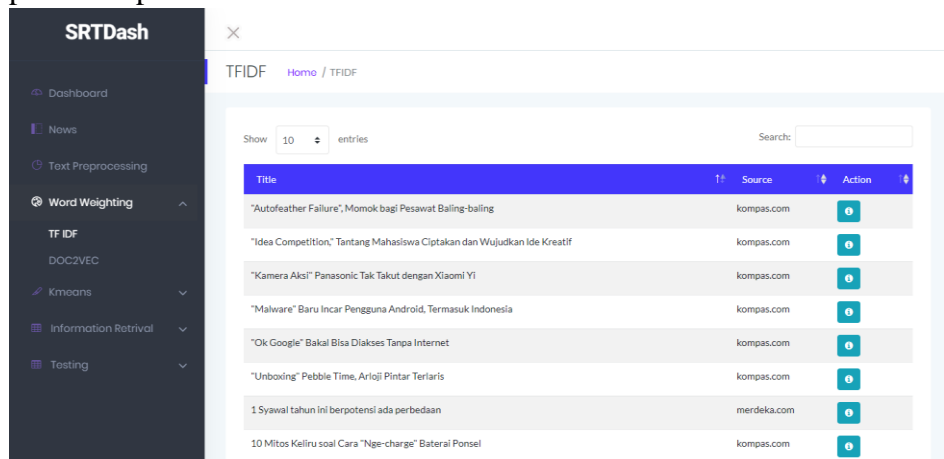


- Untuk melihat detail proses, klik tombol ikon info pada setiap berita. Dan akan muncul pop-up detail proses setiap berita.

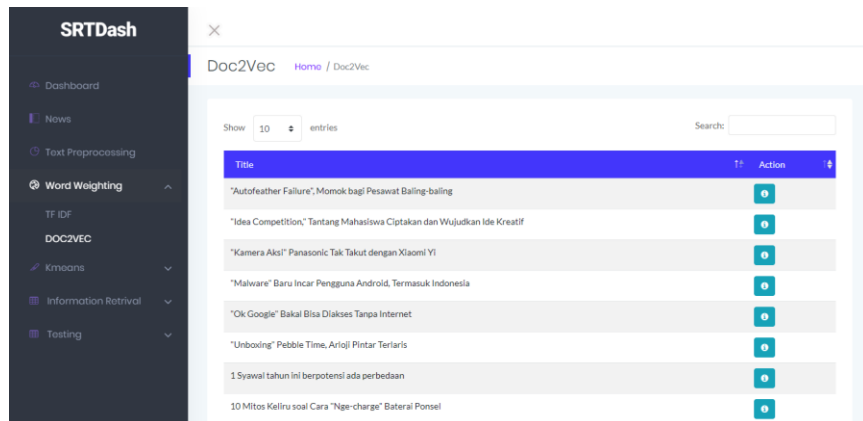


4. Halaman Word Weighting

- Pilih menu Word Weighting → TFIDF pada menu bar. Halaman ini akan menampilkan pembobotan kata hasil dari proses text preprocessing pada setiap berita.



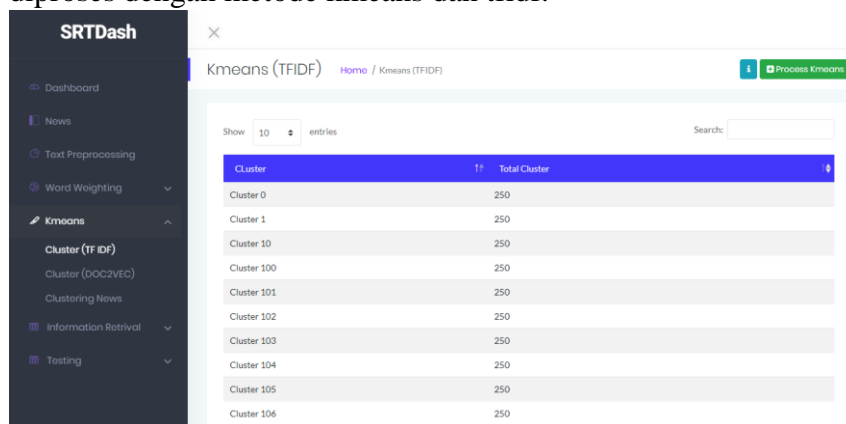
- Pilih menu Word Weighting → Doc2Vec pada menu bar. Halaman ini akan menampilkan pembobotan kata hasil dari proses text preprocessing pada setiap berita.



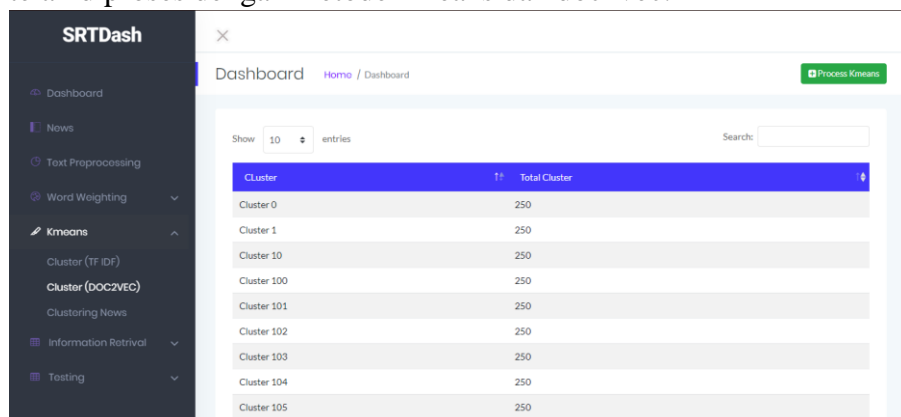
5. Halaman Kmeans

Pada halaman kmeans ini, terdapat tiga submenu yaitu Cluster TFIDF, Cluster Doc2Vec dan Clustering News :

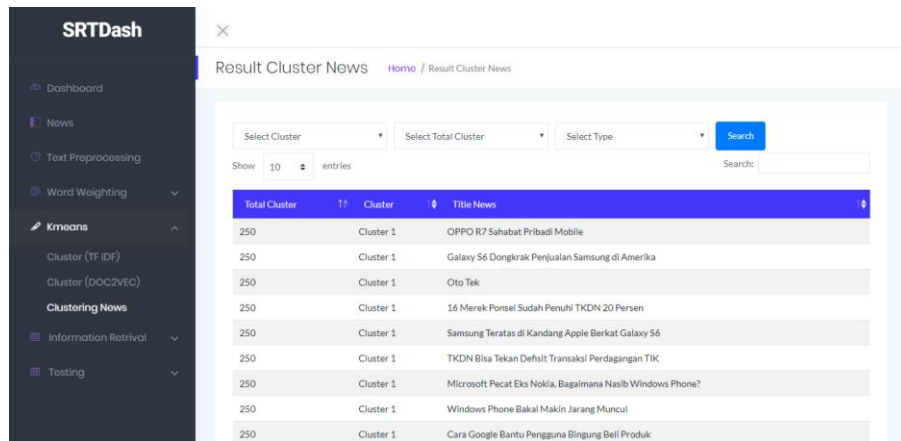
- Pada Cluster TFIDF, halaman ini menampilkan hasil cluster yang telah diproses dengan metode kmeans dan tfidf.



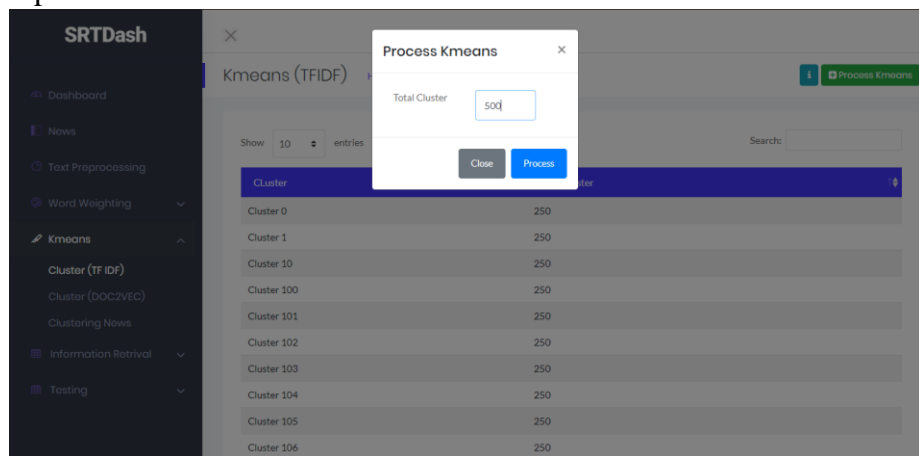
- Pada Cluster Doc2Vec, halaman ini menampilkan hasil cluster yang telah diproses dengan metode kmeans dan doc2vec.



- Pada Clustering News, halaman ini menampilkan berita sesuai dengan form pencarian *cluster* berdasarkan *cluster*, jumlah *cluster* dan tipe *cluster* dan pembobotan kata (kmeans-tfidf, kmeans-doc2vec).



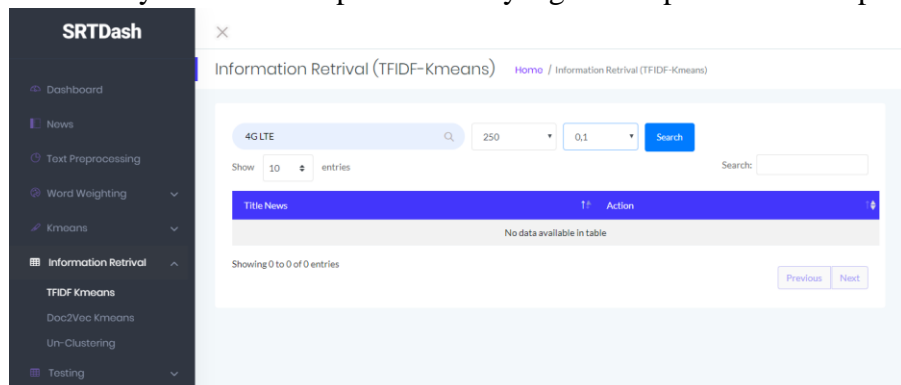
- Untuk menambahkan proses cluster baru, klik tombol process kmeans berwarna hijau. Kemudian masukkan jumlah cluster yang ingin diproses. Jika sudah klik tombol Process.



6. Halaman Information Retrieval

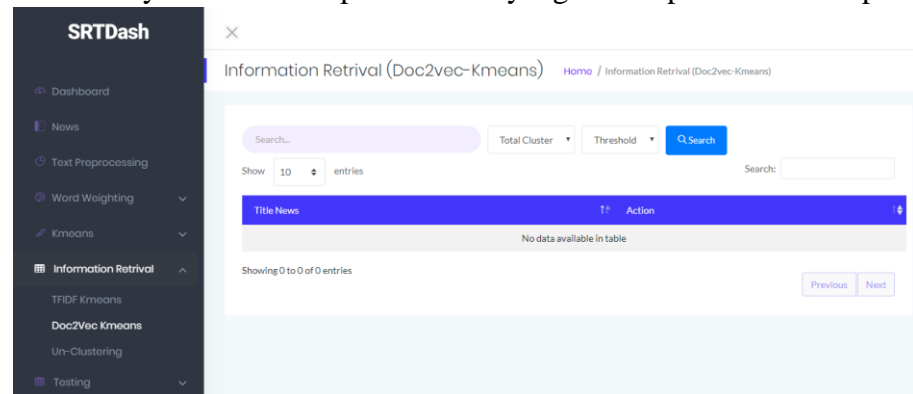
Pada halaman ini ada tiga submenu yaitu TFIDF Kmeans, Doc2Vec Kmeans dan Un-Clustering.

- Pada TFIDF Kmeans, menunjukkan tampilan halaman sistem temu kembali informasi yang menggunakan metode *kmeans clustering* dan pembobotan kata *tfidf*. Pada halaman ini *query* dimasukkan untuk dilakukan proses pencarian dengan *threshold* dan jumlah *cluster* terpilih dan hasilnya akan menampilkan berita yang berada pada *cluster* terpilih.

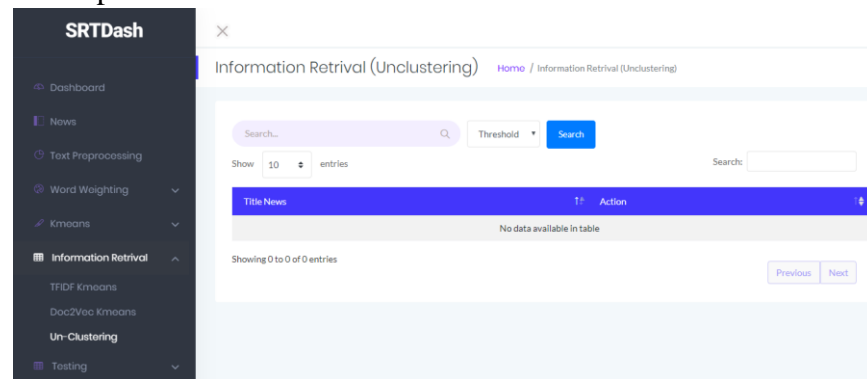


- Pada Doc2vec Kmeans, menunjukkan tampilan halaman sistem temu kembali informasi yang menggunakan metode *kmeans clustering* dan pembobotan kata *doc2vec*. Pada halaman ini *query* dimasukkan untuk

dilakukan proses pencarian dengan *threshold* dan jumlah *cluster* terpilih dan hasilnya akan menampilkan berita yang berada pada *cluster* terpilih.



- Pada Unclustering, menunjukkan tampilan halaman sistem temu kembali informasi yang tidak menggunakan metode *clustering*. Pada halaman ini *query* dimasukkan untuk dilakukan proses pencarian dengan *threshold* dan hasilnya akan menampilkan berita yang mempunyai nilai kemiripan.



7. Halaman Testing

Pada halaman ini ada tiga submenu yaitu TFIDF Kmeans, Doc2Vec Kmeans dan Un-Clustering.

- Pada TFIDF Kmeans, menampilkan data pengujian *recall*, *precision* dan *fscore* untuk sistem temu kembali informasi yang menggunakan metode *kmeans clustering* dan pembobotan kata tfidf.

Query	K	Threshold	Recall	Precision	Fscore	Time
4g lte	250	0.1	0.113924	0.155172	0.131387	8.0
android	250	0.1	0.168067	0.571429	0.25974	4.0
android	250	0.3	0.0420168	1.0	0.0806452	5.0
aplikasi whatsapp	250	0.1	1.0	0.0263158	0.0512821	5.0
game	250	0.1	0.162162	1.0	0.27907	4.0
iphone	250	0.1	0.686567	0.119481	0.20354	5.0

- Pada Doc2Vec Kmeans, menampilkan data pengujian *recall*, *precision* dan *fscore* untuk sistem temu kembali informasi yang menggunakan metode *kmeans clustering* dan pembobotan kata doc2vec.

SRTDash

Testing (Doc2vec-Kmeans) Home / Testing (Doc2vec-Kmeans)

Show 10 entries Search:

Query	K	Threshold	Recall	Precision	Fscore	Time
4g lte	250	0.1	1.0	0.0978934	0.17833	2.0
android	250	0.1	0.983193	0.14625	0.254625	1.0
android	250	0.3	0.991597	0.147685	0.257081	1.0
game	250	0.1	1.0	0.04625	0.088411	2.0
iphone	250	0.1	1.0	0.08375	0.154556	1.0

Showing 1 to 5 of 5 entries

Previous 1 Next

- Pada Unclustering, menampilkan data pengujian *recall*, *precision* dan *fscore* untuk sistem temu kembali informasi yang tidak menggunakan metode *clustering*.

SRTDash

Testing Unclustering Home / Testing Unclustering

Show 10 entries Search:

Query	K	Threshold	Recall	Precision	Fscore	Time
4g lte	0	0.1	0.810127	0.831169	0.820513	21.0
android	0	0.1	0.294118	1.0	0.454545	16.0
iphone	0	0.1	0.552239	1.0	0.711538	29.0

Showing 1 to 3 of 3 entries

Previous 1 Next