

**PROGRAM STUDI INFORMATIKA**

**PEMANFAATAN TEXT MINING PADA ARTIKEL BERITA  
UNTUK MENETUKAN TIPE BERITA MENGGUNAKAN  
METODE FUZZY C-MEANS**

**ANUGRAH OKTO DIO  
1414.200.40**

**Skripsi Ini Telah Diterima Sebagai Syarat Memperoleh Gelar Sarjana  
Komputer Di Universitas Bina Darma**



**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BINADARMA  
2020**



**PEMANFAATAN TEXT MINING PADA ARTIKEL BERITA  
UNTUK MENENTUKAN TIPE BERITA MENGGUNAKAN  
METODE FUZZY C-MEANS**

**ANUGRAH OKTO DIO  
1414.200.40**

**Skripsi Ini Telah Diterima Sebagai Syarat Memperoleh Gelar Sarjana  
Komputer Di Universitas Bina Darma**

**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BINADARMA  
2020**

**HALAMAN PENGESAHAN**

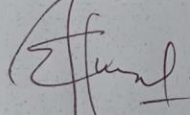
**PEMANFAATAN TEXT MINING PADA ARTIKEL BERITA  
UNTUK MENENTUKAN TIPE BERITA MENGGUNAKAN  
METODE FUZZY C-MEANS**

**ANUGRAH OKTO DIO**

**1414.200.40**

**Telah Diterima Sebagai Syarat Memperoleh Gelar Sarjana Komputer Pada  
Program Studi Informatika**

**Pembimbing I**



**Dr. H. Jemakmun, M.Si**

**Pembimbing II**



**Ade Putra M. Kom.**

**Palembang, Maret 2020  
Fakultas Ilmu Komputer  
Universitas bina darma**



**Dedy Syamsuar, S.kom., M.I.T., Ph.D**

## HALAMAN PERSETUJUAN

Skripsi Berjudul "PEMANFAATAN TEXT MINING PADA ARTIKEL BERITA UNTUK MENENTUKAN TIPE BERITA MENGGUNAKAN METODE FUZZY C-MEANS" Oleh "ANUGRAH OKTO DIO" telah dipertahankan didepan komisi penguji pada hari jum'at tanggal 21 february 2020

### Komisi Penguji

- |                       |                            |         |
|-----------------------|----------------------------|---------|
| 1. Ketua Penguji      | Dr. H Jemakmun, M.si       | (.....) |
| 2. Sekretaris penguji | Ade Putra M.Kom            | (.....) |
| 3. Anggota Penguji    | Diana, S,Si., M.Kom        | (.....) |
| 4. Anggota Penguji    | Febriyanti Panjaitan M.Kom | (.....) |

Mengetahui  
Program Studi Teknik Informatika  
Fakultas Ilmu Komputer  
Universitas Bina Darma  
Ketua Program Studi

Universitas Bina Darma  
Fakultas Ilmu Komputer  
Dr. Widya Cholil, S.Kom., M.I.T

## SURAT PERNYATAAN

Saya yang bertanda tangan dibawah ini :

Nama : Anugrah Okto Dio

Nim : 141420040

Dengan ini menyatakan bahwa :

1. Skripsi ini asli dan belum pernah di ajukan untuk mendapatkan gelar sarjana di universitas bina darma atau perguruan tinggi lain;
2. Skripsi ini murni gagasan, rumusan dan penelitian saya sendiri dengan arahan tim pembimbing
3. Di dalam skripsi ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas di kutip dengan mencantumkan nama pengarang dan memasukan ke dalam daftar rujukan;
4. Saya bersedia skripsi yang dihasilkan di cek keasliannya menggunakan plagiarism checker serta diunggah ke internet, sehingga dapat di akses public secara daring;
5. Surat pernyataan ini saya buat sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidak benaran dalam pernyataan ini, saya bersedia menerima sanksi sesuai dengan peraturan dan perundang-undangan yang berlaku.

Demikianlah surat pernyataan ini saya buat agar dapat di pergunakan sebagaimana semestinya.

Palembang , 09 Maret 2020

Yang membuat pernyataan,

  
Anugrah Okto Dio  
NIM : 141420040

## MOTTO DAN PERSEMBAHAN

*Motto :*

- *“Allah SWT tidak akan memberikan cobaan kepada umat-nya melebihi batas kemampuan manusia itu sendiri”(QS.Al-Baqarah: 286)*
- *“hidup tidak menghadikan barang sesuatupun kepada manusia tanpa bekerja keras”*
- *“Jangan pernah putus asa, berdoa dan minta tolong lah kepada Allah SWT dan tetap yakin”*

*Persembahan :*

- *Allah SWT yang telah memberi jalan dan kelancaran serta sehat jasmani dan rohani dalam penulisan skripsi ini.*
- *Ayah dan Ibu yang telah mendukung baik secara moral dan material selama penulisan skripsi ini.*
- *Saudara dan teman-teman saya*
- *Bapak Dr.H.Jemakmun, M.Si selaku pembimbing I dan bapak Ade Putra M.Kom selaku pembimbing II yang telah memberikan bimbingan selama proses penulisan skripsi*
- *Ibu Dr.Widya Cholil, S.Kom, M.I.T, Ketua Program Studi Informatika*

## ABSTRAK

Internet telah mengubah bagaimana suatu informasi ataupun berita dapat disebar dengan mudah. Berita tiap hari volumenya semakin hari bertambah besar sehingga perlu ditangani agar mudah diakses di kemudian hari. *Clustering* teks dokumen adalah salah satu operasi pada *text mining* untuk mengelompokkan dokumen yang memiliki kemiripan. *Clustering* dapat digunakan untuk membantu menganalisis berita dengan mengelompokkan secara otomatis berita yang memiliki kemiripan (*similarity*). Berbagai macam metode *clustering* salah satunya adalah Fuzzy C-Means. Metode Fuzzy C-Means bekerja dengan baik dan mampu mengelompokkan atribut-atribut kata atau token dengan masing-masing nilainya pada masing-masing cluster yang dominan baik dominan ekonomi, kesehatan, kriminal, olahraga dan politik.

**Kata kunci :** *Clustering, Text Mining, similarity, Fuzzy C-Means*

## **ABSTRACT**

The internet has changed how information or news can be spread easily. News every day the volume gets bigger so it needs to be handled so that it is easily accessible later on. Document text clustering is one of the operations in text mining to group documents that have similarities. Clustering can be used to help analyze news by automatically grouping news that has similarity (similarity). Various kinds of clustering methods, one of which is Fuzzy C-Means. Fuzzy C-Means method works well and is able to classify the attributes of words or tokens with each value in each dominant cluster of the dominant economy, health, crime, sports and politics.

Keywords :. Clustering, Text Mining, similarity, Fuzzy C-Means



## KATA PENGATAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Puji dan syukur kehadiran Allah SWT atas berkat, rahmat dan Karunia-Nya lah, Skripsi ini dapat diselesaikan guna untuk memenuhi syarat dalam menyelesaikan tugas akhir dan bisa lulus dari perkuliaan agar bisa memperoleh gelar sarjana. Disadari penulisan skripsi ini masih jauh untuk bisa dikatakan sempurna, telah dilakukan revisi dan perbaikan lebih dari satu kali agar akhirnya bisa diterima benar, hal tersebut dikarenakan penulis memiliki kemampuan yang terbatas, oleh karena itulah dibutuhkan bimbingan agar bisa menyelesaikan penulisan skripsi ini.

Pada kesempatan yang baik ini, penulis ingin menyampaikan rasa syukur dan ucapan terima kasih kepada pihak – pihak yang telah memberikan dukungan, bimbingan dan nasihat – nasihat selama penulisan skripsi berlangsung. Penulis mengucapkan terima kasih kepada:

1. Dr. Sunda Ariana, M.Pd, M.M. Selaku Rektor Universtas Bina Darma Palembang.
2. Dedy Syamsuar, S.Kom, MIT. Selaku Dekan Fakultas Ilmu Komputer.
3. Dr. Widya Cholil, S.Kom, MIT. Selaku Ketua Program Studi Informatika.
4. Dr. H. Jemakmun, M.si. Selaku Pembimbing I dan bapak Ade Putra M.Kom yang telah memberikan arahan dan dukungan dalam penulisan skripsi.
5. Bapak dan Ibu saya yang telah memberikan dukungan moril dan materil kepada saya selama proses penulisan skripsi. Dan juga teman – teman dan saudara saya.

Palembang Maret 2020

Penulis



Anugrah Okto Dio



## DAFTAR ISI

|   | <b>Halaman</b> |
|---|----------------|
| Cover Dalam .....                       | i              |
| Halaman Pengesahan .....                | ii             |
| Halaman Persetujuan.....                | iii            |
| Pernyataan .....                        | iv             |
| Motto Dan Persembahan .....             | v              |
| Abstrak .....                           | vi             |
| Kata Pengantar .....                    | viii           |
| Daftar Isi.....                         | ix             |
| Daftar Gambar.....                      | xi             |
| Daftar Table .....                      | xiii           |
| <b>BAB I PENDAHULUAN.....</b>           | <b>1</b>       |
| 1.1 Latar Belakang .....                | 1              |
| 1.2 Rumusan Masalah .....               | 3              |
| 1.3 Batasan Masalah.....                | 3              |
| 1.4 Tujuan dan Manfaat Penelitian ..... | 3              |
| 1.4.1 Tujuan Penelitian .....           | 3              |
| 1.4.2 Manfaat penelitian.....           | 4              |
| 1.5 Metodologi Penelitian .....         | 4              |
| 1.5.1 Waktu penelitian .....            | 4              |
| 1.5.2 Alat Dan Bahan Penelitian .....   | 4              |
| 1.5.3 Data Penelitian .....             | 5              |
| 1.5.4 Metode pengumpulan Data .....     | 5              |
| 1.6 Sistematika Penulisan .....         | 6              |
| <b>BAB II LANDASAN TEORI .....</b>      | <b>8</b>       |
| 2.1 Tinjauan Umum .....                 | 8              |
| 2.1.1 Definisi Berita .....             | 8              |

|  |           |
|--|-----------|
| 2.2 Landasan Teori.....  | 8         |
| 2.2.1 <i>Text Mining</i> .....                                     | 8         |
| 2.2.2 <i>Text Processing</i> .....                                 | 9         |
| 2.2.3 <i>Term Frequency Dan Document Frequency (TF-IDF)</i> .....  | 9         |
| 2.2.4 <i>Data Mining</i> .....                                     | 12        |
| 2.2.5 Pengelompokan <i>Data Mining</i> .....                       | 15        |
| 2.2.6 Tahapan-Tahapan <i>Data Mining</i> .....                     | 16        |
| 2.2.7 Teknik Pengelompokan.....                                    | 17        |
| 2.2.8 Algoritma Fuzzy C-Means .....                                | 18        |
| 2.2.9 Rapid Miner .....  | 20        |
| 2.3 Penelitian Sebelumnya .....                                    | 21        |
| <b>BAB III ANALISA DATA MINING.....</b>                            | <b>23</b> |
| 3.1 <i>Data Selection</i> .....                                    | 23        |
| 3.2 <i>Pre Processing</i> .....                                    | 27        |
| 3.3 <i>Transformation</i> .....                                    | 31        |
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b>                           | <b>32</b> |
| 4.1 <i>Text Mining Data Mining</i> .....                           | 32        |
| 4.1.1 Memasang Fitur <i>Text Mining</i> Dan Upgrade Tool .....     | 33        |
| 4.1.2 <i>Import</i> Data Ke Rapid Miner .....                      | 33        |
| 4.1.3 <i>Pre Processing</i> (Pra Proses .....                      | 37        |
| 4.1.4 Penerapan Metode Fuzzy C-Means Menggunakan Rapid Miner ..... | 44        |
| 4.1.5 Hasil Evaluasi ( <i>Evaluation Result</i> .....              | 47        |
| 4.1.6 Hasil Analisis .....   | 52        |
| <b>BAB V PENUTUP.....</b>  | <b>54</b> |
| 5.1 Kesimpulan .....   | 54        |
| 5.2 Saran.....   | 55        |
| <b>DAFTAR PUSTAKA .....</b>  | <b>56</b> |
| <b>LAMPIRAN.....</b>   | <b>57</b> |

## DAFTAR GAMBAR

|  | <b>Halaman</b> |
|--|----------------|
| <b>Gambar 2.1</b> Classification – <i>Decison Tree</i> .....                                       | 12             |
| <b>Gambar 2.2</b> <i>Clustering</i> .....  | 13             |
| <b>Gambar 2.3</b> Product Association .....  | 13             |
| <b>Gambar 2.4</b> Time Series .....  | 14             |
| <b>Gambar 2.5</b> Web Navigation Sequence.....   | 14             |
| <b>Gambar 2.6</b> Proses KDD .....   | 16             |
| <b>Gambar 4.1</b> <i>Marketplace</i> Untuk Menambahkan <i>extensi</i> .....                        | 33             |
| <b>Gambar 4.2</b> Folder Yang Mengelompokan judul berita yang tersimpan difile ...                 | 34             |
| <b>Gambar 4.3</b> File-File Dalam Masing-Masing <i>Folder</i> Kelas/Kategori Berita.....           | 35             |
| <b>Gambar 4.4</b> Contoh Isi File Berita .....   | 35             |
| <b>Gambar 4.5</b> Bagan Proses Data <i>Mining</i> .....  | 36             |
| <b>Gambar 4.6</b> <i>Entry</i> lokasi <i>Folder</i> Data .....                                     | 37             |
| <b>Gambar 4.7</b> Bagan Operator Untuk Proses Text Mining .....                                    | 38             |
| <b>Gambar 4.8</b> Bagan Proses <i>Input Output</i> pembacaan dokumen .....                         | 39             |
| <b>Gambar 4.9</b> Hasil Proses Tokenisasi Dan Perhitungan <i>Term Frequency</i> ( <i>TF</i> )..... | 40             |
| <b>Gambar 4.10</b> Hasil <i>IDF</i> .....  | 41             |
| <b>Gambar 4.11</b> Terdapat ontoh atribut tak Penting(Belum Dilakukan <i>Cleaning</i> .....        | 41             |
| <b>Gambar 4.12</b> Isi File <i>Stopword</i> Sebelum Ditambahkan Lagi.....                          | 42             |
| <b>Gambar 4.13</b> Isi File <i>Stopword</i> Setelah Ditambahkan .....                              | 43             |
| <b>Gambar 4.14</b> Penambahan Bagan Operator ( <i>Filter Token By Length</i> ).....                | 44             |
| <b>Gambar 4.15</b> Token Terbaru Setelah Perbaharuan <i>Stopword</i> .....                         | 44             |
| <b>Gambar 4.16</b> Data Atribut Yang Sudah Melalui Proses Dibersihkan( <i>Cleaning</i> ).45        | 45             |
| <b>Gambar 4.17</b> Bagan Operator Proses Fuzzy C-Means Di Rapidmier .....                          | 45             |
| <b>Gambar 4.18</b> Parameter <i>Setting</i> Fuzzy C-Means.....                                     | 46             |

|  |    |
|--|----|
| <b>Gambar 4.19</b> Hasil <i>Cluster Model</i> .....                      | 47 |
| <b>Gambar 4.20</b> Data Yang Tercluster.....                             | 48 |
| <b>Gambar 4.21</b> Detail <i>Cluster 0</i> .....                         | 48 |
| <b>Gambar 4.22</b> Detail <i>Cluster 1</i> .....                         | 49 |
| <b>Gambar 4.23</b> Detail <i>Cluster 2</i> .....                         | 49 |
| <b>Gambar 4.24</b> Detail <i>Cluster 3</i> .....                         | 50 |
| <b>Gambar 4.25</b> Detail <i>Cluster 4</i> .....                         | 50 |
| <b>Gambar 4.26</b> Jarak <i>Matrix</i> Partisi Yang Relatif Pendek ..... | 52 |

## DAFTAR TABLE

|   | <b>Halaman</b> |
|---|----------------|
| <b>Tabel 2.1</b> <i>Term Frequency</i> (TF) .....                                   | 10             |
| <b>Table 2.2</b> <i>Document Frequency</i> (DF).....                                | 11             |
| <b>Tabel 2.3</b> <i>Inverse Document Frequency</i> (IDF).....                       | 11             |
| <b>Tabel 3.1</b> Data Judul Berita Ekonomi .....                                    | 24             |
| <b>Tabel 3.2</b> Data Judul Berita Kesehatan .....                                  | 25             |
| <b>Tabel 3.3</b> Data Judul Berita Kriminal.....                                    | 25             |
| <b>Tabel 3.4</b> Data Judul Berita Olahraga .....                                   | 26             |
| <b>Tabel 3.5</b> Data Judul Berita Politik .....                                    | 26             |
| <b>Tabel 3.1</b> kejadian kemunculan term dan nilai TF( <i>Term Frequency</i> ..... | 29             |
| <b>Tabel 3.6</b> Nilai <i>Inversed Document Frequency</i> (IDF).....                | 30             |