

# Implementasi Algoritma Text Mining TF-IDF Untuk Fitur Autoresponder

Pratama Djodi

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Param, Medan, Indonesia  
Email: pratamajodi460@gmail.com  
Email Penulis Korespondensi: pratamajodi460@gmail.com

**Abstrak**— Era industri 4.0 teknologi semakin menjadi kebutuhan manusia, baik dalam kegiatan sehari-hari maupun dan kegiatan bisnis. Kegiatan mengirim surat dulu dianggap sangat efektif menggunakan layanan POS namun saat ini kegiatan tersebut telah diambil alih oleh elektronik atau sering disebut dengan e-mail atau elektornik mail. Email adalah sarana kirim mengirim surat melalui jalur jaringan komputer seperti internet, kelebihan yang ditawarkan oleh e-mail yaitu proses mengirim dan menerima surat yang lebih cepat, serta tersedia fasilitas mengirim gambar, video dan audio, tentunya jika dibandingkan dengan layanan POS maka e-mail akan menjadi sarana yang sangat tepat untuk bisnis dalam hal kegiatan surat menyurat. Seiring penggunaan email yang kini semakin banyak digunakan oleh perusahaan besar maupun perusahaan kecil dalam melayani pelanggannya untuk memudahkan dalam kegiatan mengirim dan menerima surat, maka munculah masalah baru yaitu keterlambatan dalam membalas e-mail masuk yang dapat disebabkan oleh banyaknya email masuk atau butuhnya koordinasi atas surat masuk sehingga butuh waktu yang lama dalam menanggapi setiap e-mail yang masuk, dengan demikian akan muncul rasa kekecewaan pelanggan karena dianggap tidak dilayani atau bahkan merasa diabaikan.

**Kata Kunci:** Implementasi; Algoritma Text Mining; TF-IDF; Autoresponder; Email;

**Abstract**—The industrial era 4.0 technology is increasingly becoming a human need, both in daily activities and in business activities. The activity of sending letters used to be considered very effective using POS services, but now this activity has been taken over by electronics or often referred to as e-mail or electronic mail. Email is a means of sending letters through computer networks such as the internet, the advantages offered by e-mail are the process of sending and receiving letters that are faster, and facilities for sending images, video and audio are available, of course when compared to POS services, e-mail will be a very appropriate tool for businesses in terms of correspondence activities. As the use of email is now increasingly being used by large companies and small companies in serving their customers to make it easier to send and receive letters, a new problem arises, namely delays in replying to incoming e-mails which can be caused by the large number of incoming emails or the need for coordination of letters. logged in so that it takes a long time to respond to every incoming e-mail, thus there will be a sense of customer disappointment because it is considered not served or even feels neglected.

**Keywords:** Implementation; Text Mining Algorithm; TF-IDF; Autoresponder; Email;

## 1. PENDAHULUAN

Era industri 4.0 teknologi semakin menjadi kebutuhan manusia, baik dalam kegiatan sehari-hari maupun dan kegiatan bisnis. Kegiatan mengirim surat dulu dianggap sangat efektif menggunakan layanan POS namun saat ini kegiatan tersebut telah diambil alih oleh elektronik atau sering disebut dengan e-mail atau elektornik mail. Email adalah sarana kirim mengirim surat melalui jalur jaringan komputer seperti internet, kelebihan yang ditawarkan oleh e-mail yaitu proses mengirim dan menerima surat yang lebih cepat, serta tersedia fasilitas mengirim gambar, video dan audio, tentunya jika dibandingkan dengan layanan POS maka e-mail akan menjadi sarana yang sangat tepat untuk bisnis dalam hal kegiatan surat menyurat.

Seiring penggunaan email yang kini semakin banyak digunakan oleh perusahaan besar maupun perusahaan kecil dalam melayani pelanggannya untuk memudahkan dalam kegiatan mengirim dan menerima surat, maka munculah masalah baru yaitu keterlambatan dalam membalas e-mail masuk yang dapat disebabkan oleh banyaknya email masuk atau butuhnya koordinasi atas surat masuk sehingga butuh waktu yang lama dalam menanggapi setiap e-mail yang masuk, dengan demikian akan muncul rasa kekecewaan pelanggan karena dianggap tidak dilayani atau bahkan merasa diabaikan.

Text Mining adalah prose ekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks. Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan. Penambangan teks dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan penambangan data. Proses yang umum dilakukan oleh penambangan teks di antaranya adalah perangkuman otomatis, kategorisasi dokumen, penggugusan teks, deteksi plagiarisme, [1]. *Term Frequency — Inverse Document Frequency* atau TF — IDF adalah suatu metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen [2]

Penulis memilih menggunakan algoritma text mining dan TF-IDF berdasarkan peneliti sebelumnya yang dilakukan oleh Abdul Azis Ma'arif dengan judul PENERAPAN ALGORITMA TF-IDF UNTUK PENCARIAN KARYA

ILMIAH menyimpulkan bahwa Penerapan algoritma TF-IDF untuk pencarian karya ilmiah menghasilkan karya ilmiah yang dapat dilihat atau diunduh dalam bentuk format pdf. Dan penelitian yang dilakukan oleh Moh. Mahdi Sya'bani dan Reni Umilasari dengan judul "Penerapan Metode *Cosine Similarity* dan Pembobotan TF/IDF pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember" menyimpulkan bahwa pengujian aplikasi terdapat 120 dokumen sinopsis dengan 10 kategori dan menghasilkan nilai *precision* sebesar 90,91% pada *threshold* 0,1 dan nilai *recall* sebesar 100% pada *threshold* 0,1 dan 0,2. Ketepatan akurasi pada sistem aplikasi yang diuji adalah 80,83% [3].

Berdasarkan penjelasan diatas maka penulis ingin menyelesaikan masalah membalas surat secara otomatis dengan menerapkan algoritma text mining dan algoritma *Term Frequency – Inverse Document Frequency* (TF-IDF), dengan demikian maka surat akan dapat dijawab secara otomatis sesuai dengan hasil dari kedua algoritma tersebut..

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Dalam melakukan penelitian agar mendapat hasil seperti yang diharapkan, maka diperlukan tahapan penelitian. Dimana tahapan penelitian yang akan dilakukan sebagai berikut:

a. Pengumpulan data (*data collection*)

Pengumpulan data yang dilakukan adalah studi *literatur*, yaitu dengan mempelajari sumber-sumber pustaka yang dapat dijadikan referensi. Dan melakukan pengumpulan data-data dengan memanfaatkan buku-buku, dan bahan bacaan lainnya yang berhubungan dengan Algoritma *Grille Chiper* dan *Cardan Grille*.

b. Studi Pustaka (Library Reseach)

Melakukan studi kepustakaan, mengumpulkan dan mempelajari berbagai referensi yang berkaitan dengan penelitian.

c. Analisa dan Perancangan Sistem (Analitsis and System Development)

Tahap ini merupakan kegiatan untuk menganalisa kebutuhan sistem yang akan dibuat dan menjadi dasar untuk perancangan sistem, seperti analisa algoritma, data yang dijadikan input user interface design.

d. Implementasi (Implementation)

Tahap ini merupakan kegiatan untuk membuat perangkat lunak yang sesuai dengan analisa dan perancangan sistem yang telah ditentukan.

e. Pengujian (Testing)

Tahap ini merupakan kegiatan untuk melakukan pengujian sistem apakah hasil penelitian yang telah dilakukan sesuai dengan tujuan penelitian dan menjawab rumusan masalah yang telah ditentukan.

f. Implementasi sistem (*Implementation system*)

Implementasi ini akan dilakukan dengan algoritma *Grille Chiper* dan *Cardan Grille*.

### 2.2 Autoresponder

Autoresponder merupakan suatu program teknologi yang dirancang untuk merespon email, komentar atau pesan secara otomatis. Autoresponder berfungsi untuk menangkap data dari pengunjung yang biasanya berupa nama dan alamat emali, kemudian melakukan follow up dengan mengirimkan email secara berkala sesuai dengan tujuan dan strategi pengirim email. Biasanya autoresponder ini digunakan untuk menjalin hubungan dengan konsumen atau orang lain, memberikan informasi terbaru, dan masih banyak lagi [4].

### 2.3 Text Mining

Text mining dapat didefinisikan secara luas sebagai proses pengetahuan intensif di mana pengguna berinteraksi dengan koleksi dokumen dari waktu ke waktu dengan menggunakan seperangkat alat analisis. Text mining berusaha untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi pola yang menarik. Text mining banyak mengarah pada bidang penelitian data mining. Oleh karena itu, tidak mengherankan bahwa text mining dan data mining akan berada pada tingkat arsitektur yang sama [5].

### 2.4 Term Frequency- Inverce Document Frequency(TF-IDF)

Dalam perhitungan bobot menggunakan TF-IDF, dihitung terlebih dahulu nilai TF perkata dengan bobot masing-masing kata adalah 1. Sedangkan nilai IDF diformulasikan pada persamaan (1).

$$IDF(word) = \log \frac{td}{df} \quad (1)$$

IDF (word) adalah nilai IDF dari setiap kata yang akan di cari, td adalah jumlah keseluruhan dokumen yang ada, df jumlah kemunculan kata pada semua dokumen [6].

Cara kerja dalam mencari nilai *term-frequency* melalui beberapa persamaan 1 yaitu:

$$T_{ft,d} = 1 + {}^{10}\text{Log } tf \quad (2)$$

dimana setiap variabel dijelaskan sebagai berikut :

tf : *term* frekuensi atau banyaknya kata pada dokumen  
 Tft,d : *term* frekuensi atau banyaknya kata t pada dokumen d atau pembobotan local  
 Mencari nilai *inverse document-frequency* melalui persamaan 2 :

$$Idf_t = 10 \log^n / dft \tag{3}$$

Idft : *inverse document-frequency* atau pembobotan global  
 n : banyaknya dokumen  
 dft : banyaknya dokumen yang memiliki kata t.

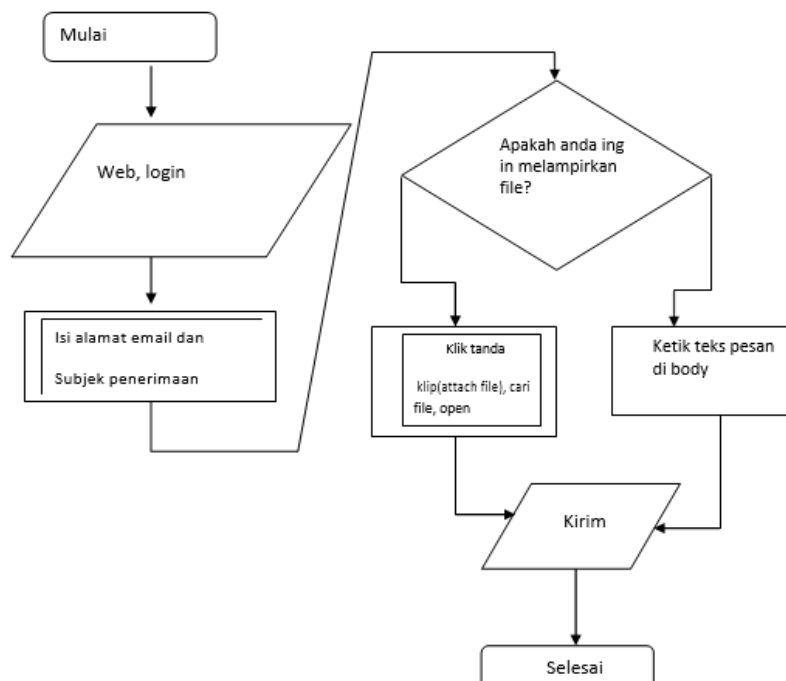
### 3. HASIL DAN PEMBAHASAN

#### 3.1 Analisa Masalah

Era industri 4.0 teknologi semakin menjadi kebutuhan manusia, baik dalam kegiatan sehari-hari maupun dan kegiatan bisnis. Kegiatan mengirim surat dulu dianggap sangat efektif menggunakan layanan POS namun saat ini kegiatan tersebut telah diambil alih oleh elektronik kata usering disebut dengan e-mail atau elektornik mail. Email adalah sarana kirim mengirim surat melalui jalur jaringan komputer seperti internet, kelebihan yang ditawarkan oleh e-mail yaitu proses mengirim dan menerima surat yang lebih cepat, serta tersedia fasilitas mengirim gambar, video dan audio, tentunya jika dibandingkan dengan layanan POS maka e-mail akan menjadi sarana yang sangat tepat untuk bisnis dalam hal kegiatan surat menyurat.

Seiring penggunaan email yang kini semakin banyak digunakan oleh perusahaan besar maupun perusahaan kecil dalam melayani pelanggannya untuk memudahkan dalam kegiatan mengirim dan menerima surat, maka munculah masalah baru yaitu keterlambatan dalam membalas e-mail masuk yang dapat disebabkan oleh banyaknya email masuk atau butuhnya koordinasi atas surat masuk sehingga butuh waktu yang lama dalam menanggapi setiap e-mail yang masuk, dengan demikian akan muncul rasa kekecewaan pelanggan karena dianggap tidak dilayani atau bahkan merasa diabaikan.

Text Mining adalah prose sekstraksi pola berupa informasi dan pengetahuan yang berguna dari sejumlah besar sumber data teks, seperti dokumen Word, PDF, kutipan teks, dll. Jenis masukan untuk penambangan teks ini disebut data tak terstruktur dan merupakan pembeda utama dengan penambangan data yang menggunakan data terstruktur atau basis data sebagai masukan. Penambangan teks dapat dianggap sebagai proses dua tahap yang diawali dengan penerapan struktur terhadap sumber data teks dan dilanjutkan dengan ekstraksi informasi dan pengetahuan yang relevan dari data teks terstruktur ini dengan menggunakan teknik dan alat yang sama dengan penambangan data. Proses yang umum dilakukan oleh penambangan teks di antaranya adalah perangkuman otomatis, kategorisasi dokumen, penggugusan teks, deteksi *plagiarisme* [1]. *Term Frequency — Inverse Document Frequency* atau TF — IDF adalah suatu metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan. Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen[2]



**Gambar 1.** Skema Penerapan Algoritma *Text Mining* dan TF-IDF

### 3.2 Penerapan Algoritma Text Mining

Algoritma *text mining* adalah salah satu cara untuk melakukan pengklasifikasi dengan cara menemukan pola yang berbeda di dalam data tekstual yang berjumlah besar. Berikut ini adalah contoh kasus untuk email Autoresponder.

$$\text{Similarity} (S_i, S_j) = \frac{|WK_{S_i} \cap WK_{S_j}|}{\log(|S_i|) + \log(|S_j|)}$$

Keterangan :

WK = Jumlah kata(term) yang sama antara kalimat  $S_i$  dan  $S_j$

**Tabel 1.** Data Komentar E-mail Responding

| Komentar E-mail Responding  |
|---|
| barangnya ready gak kak?  |
| ukuran 42 harganya 27.900 itu beneran kak?                              |
| bener buat orang dewasa kak?  |
| tolong kok ini barang saya di prosesnya lama banget udah hamper 3 hari. |
| ini ada talinya apa beli sendiri yah kak?                               |
| bagus barangnya makasih ya lazada.                                      |
| Kak itu sepatunya beneran harga aslinya tidak kak?                      |
| Pengiriman dari mana kak?   |
| Barang ready apa tidak min?   |
| Warna sepatunya semua ready gak kak?                                    |

Terdapat beberapa langkah dalam penerapan *text mining* pada email. Berikut langkah-langkah dari *text mining*:

a. *Tokenizing*

*Tokenizing* merupakan langkah awal dalam penguraian deskripsi yang berupa suatu kalimat menjadi potongan kata. Berikut proses *tokenizing* pada kalimat-kalimat yang terdapat didalam abstrak.

**Tabel 3.** Hasil *Tokenizing*

| No | Kalimat   | Hasil Tokenezing   |
|----|---|--|
| 1  | barangnya ready gak kak sudah di buka                                       | Barangnya<br>ready<br>gak<br>kak<br>sudah<br>di<br>buka                                    |
| 2  | Ukuran 42 harganya 27.900 itu beneran kak?                                  | Ukuran<br>42<br>Harganya<br>27.900<br>Itu<br>Beneran<br>Kak                                |
| 3  | Itu beneran buat orang dewasa kak?  | Itu<br>beneran<br>buat<br>orang<br>dewasa<br>kak   |
| 4  | Kak tolong kok ini barang saya di prosesnya lama banget udah hamper 3 hari. | Kak<br>Tolong<br>Kok<br>Ini<br>barang<br>saya<br>di<br>prosesnya<br>lama<br>banget<br>udah |

|    |  |   |   |
|----|--|---|---|
|    |  |   | hamper<br>3<br>Hari<br>Kak<br>ini<br>ada<br>talinya<br>apa<br>beli<br>sendiri<br>yah<br>kak |
| 5  |  | Kak ini ada talinya apa beli sendiri yah kak? |   |
| 10 |  | Warna sepatunya semua ready gak kak?          | Warna<br>Sepatunya<br>Semua<br>Ready<br>Gak<br>Kak  |

b. *Filtering (wordlist/stoplist)*

Tahap *filtering* merupakan tahap menghilangkan kata yang tidak penting (*stoplist*) atau menyimpan kata yang penting (*wordlist*). *Stoplist/stopword* adalah kata-kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *stopwords* adalah “yang”, “dan”, “di”, “dari” dan seterusnya. Data *stopword* dapat diambil dari suatukamus kata-kata yang menyimpan kata-kata yang bisa dihilangkan atau dengan kata lain kata-kata yang tidak penting disebut dengan kamus tala.

**Tabel 4.** Hasil *Filtering*

| No | Hasil Tokenizing  | Hasil Filtering   |
|----|---|---|
| 1  | barangnya<br>ready<br>gak<br>kak<br>sudah<br>di<br>buka   | barangnya<br>ready<br>kak<br>buka                           |
| 2  | Ukuran<br>42<br>Harganya<br>27.900<br>Itu<br>Beneran<br>kak   | ukuran<br>42<br>Harganya<br>27900<br>beneran<br>kak         |
| 3  | Itu<br>beneran<br>buat<br>orang<br>dewasa<br>kak  | beneran<br>orang<br>dewasa<br>kak                           |
| 4  | Kak<br>Tolong<br>Kok<br>Ini<br>Barang<br>saya<br>di<br>prosesnya<br>lama<br>banget<br>udah<br>hampir<br>3 | kak<br>tolong<br>barang<br>prosesnya<br>banget<br>udah<br>3 |

|       |  |  |
|-------|--|--|
| 5     | hari<br>Kak<br>ini<br>ada<br>talinya                     | Kak<br>Talinya                                     |
| ----- | -----  | -----  |
| 10    | Warna<br>sepatunya<br>semua r<br>eady<br>gak<br>kak<br>? | Warna<br>Sepatunya<br>Semua<br>Ready<br>Gak<br>Kak |

c. *Stemming*

Mencari bentuk awal/*root* dari tiap kata lampau atau kata hasil *stemming* Tahap *stemming* merupakan tahap mengubah kata imbuhan menjadi kata dasar atau dengan kata lain adalah mengembalikan kata ke dalam bentuk aslinya.

**Tabel 5.** Hasil *Stemming*

| Kalimat | Hasil Filtering   | Hasil Stemming  |
|---------|---|---|
| 1       | Barangnya<br>ready<br>kak<br>buka   | Barang<br>Ready<br>Kak<br>Buka  |
| 2       | Ukuran<br>42<br>harganya<br>27.900<br>Itu<br>beneran<br>kak   | Ukur<br>42<br>Harga<br>27900<br>Itu<br>Bener<br>Kak   |
| 3       | Itu<br>beneran<br>buat<br>orang<br>dewasa<br>kak  | Kak<br>Bener<br>Buat<br>Orang<br>Dewasa<br>Kak  |
| 4       | Kak<br>Tolong<br>Kok<br>Ini<br>Barang<br>Saya<br>Di<br>Prosesnya<br>lama<br>banget<br>udah<br>hampir<br>3 | Kak<br>Tolong<br>Kok<br>Ini<br>Barang<br>Saya<br>Di<br>Proses<br>Lama<br>Banget<br>Uda<br>Hampir<br>3 |
| 5       | Hari<br>Kak<br>ini<br>ada<br>talinya  | Hari<br>Kak<br>Ini<br>Ada<br><br>Tali   |
| -----   | -----   | -----   |
| 10      | Warna<br>Sepatunya<br>Semua<br>Ready<br>Gak   | Warna<br>Sepatu<br>Semua<br>Ready<br>Gak  |

Kak

Kak

### 3.4 Penerapan Algoritma TF-IDF

Pembobotan TF/IDF adalah suatu hubungan kata (*term*) yang berada pada dokumen yang akan diberikan suatu nilai bobot. Metode ini menggabungkan dua konsep untuk perhitungan bobot yaitu, frekuensi kemunculan sebuah kata di dalam sebuah dokumen hasil dari pemrosesan teks abstrak yang dibandingkan. Frekuensi kemunculan sebuah kata dalam sebuah dokumen disebut dengan *Term Frequency* (TF) dan *inverse frekuensi* dokumen yang mengandung kata yang disebut *Inverse Document Frequency* (IDF). Tujuan dari algoritma ini adalah untuk melihat seberapa penting kata (*term*) tersebut di dalam tabel gejala printer.

Setelah dilakukannya proses *text mining* dan menghasilkan kalimat yang memiliki peran penting di dalam perhitungan, selanjutnya yaitu perhitungan TF-IDF. Proses menghitung nilai bobot, terlebih dahulu hitung jumlah kemunculan *term* pada setiap dokumen, kemudian menjumlahkan setiap kemunculan kata yang sama pada semua dokumen letakkan pada kolom DF. Selanjutnya, mencari IDF yaitu hitung jumlah dokumen dibagi DF kemudian hasil pembagian di-log-kan. Langkah selanjutnya, mencari WDT yaitu nilai TF dikalikan nilai IDF sehingga memperoleh nilai bobot *term*.

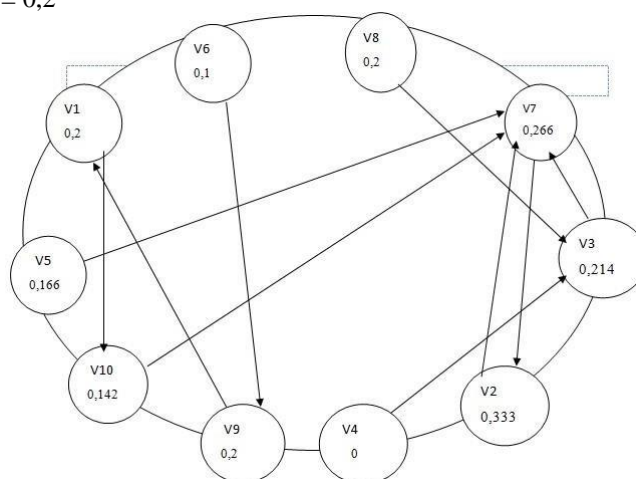
Berdasarkan contoh diatas juga dilakukan untuk mencari *term* yang lainnya, berikut hasil keseluruhan proses dari perhitungan TF-IDF yang disajikan pada tabel 7

**Tabel 7. Vertex 1TF-IDF**

| Vertex 1                      | Perbandingan  |  |
|-------------------------------|---|--|
| 1. <b>BarangReadyKak</b> Buka | <ul style="list-style-type: none"> <li>• Ukur 42 Harga 27900 Itu Bener <b>Kak</b></li> <li>• <b>Kak</b> Bener Buat Orang Dewasa <b>Kak</b></li> <li>• Di Proses Lama Banget Uda Hampir 3 Hari</li> <li>• <b>Kak</b> Ini Ada Tali</li> </ul> | Vertex2<br>Vertex3<br>Vertex4<br>Vertex5 |
| -----                         |   |  |
|                               | <ul style="list-style-type: none"> <li>• Warna Sepatu Semua <b>Ready</b> Gak <b>Kak</b></li> </ul>  | Vertex10                                 |

$$\text{Similarity} (S_i, S_j) = \frac{|W_k \cap W_l \cap S_i \cap S_j|}{\log(|S_i|) + \log(|S_j|)}$$

$$\text{Similarity} (S_1, S_{10}) = \frac{4+6}{2 \dots \dots [2.2]} = \frac{2}{10} = 0,2$$



**Gambar 2. Lingkaran vertex**

Berdasarkan 2 metode yang digunakan yaitu metode Text Mining dan TF-IDF, maka diperoleh hasil sebagai berikut :

**Tabel 8. Hasil pembobotan kata**

| Vertex  | Nilai |
|---------|-------|
| Vertex2 | 0,333 |
| Vertex7 | 0,266 |
| Vertex3 | 0,214 |
| Vertex1 | 0,2   |
| Vertex8 | 0,2   |

|          |       |
|----------|-------|
| Vertex9  | 0,2   |
| Vertex5  | 0,166 |
| Vertex10 | 0,142 |
| Vertex6  | 0,1   |
| Vertex4  | 0     |

Dari permasalahan yang dialami dalam kerusakan printer dapat kita temukan hasil tahapan yang sudah kita kerjakan, diperoleh 10 vertex yang sangat berpengaruh terhadap ganggian kerusakan pada printer. Dimana vertex 2 adalah nilai yang tertinggi dalam gejala kerusakan printer dengan nilai 0,333.

| W=TF*IDF |    |         |         |         |         |             |         |     |
|----------|----|---------|---------|---------|---------|-------------|---------|-----|
| D5       | D6 | D7      | D8      | D9      | D10     |             |         |     |
| 0        | 0  | 0       | 0       | 0       | 0       |             |         |     |
| 0        | 0  | 0       | 0       | 0       | 0       |             |         |     |
| 0        | 0  | 0       | 0       | 0       | 0       |             |         |     |
| 0        | 0  | 0       | 0       | 0,69897 | 0,69897 |             |         |     |
| 0        | 0  | 0       | 0       | 0       | 1       |             |         |     |
| 0        | 0  | 0       | 0       | 0,69897 | 0,69897 |             |         |     |
| 0,39794  | 0  | 0,39794 | 0,39794 | 0       | 0       |             |         |     |
| 0        | 0  | 0       | 0       | 0       | 0       |             |         |     |
| 0        | 1  | 0       | 0       | 0       | 0       |             |         |     |
| 0        | 0  | 1       | 0       | 0       | 0       |             |         |     |
| 0        | 0  | 0       | 1       | 0       | 0       |             |         |     |
| 0        | 0  | 0       | 0       | 2,39794 | 1,39794 | PERSENTASE  | TOTAL   |     |
| 0,39794  | 1  | 1,39794 | 1,39794 | 0       | 0       | 7,193820026 | 56,2651 | 56% |
| 0,39794  | 1  | 1,39794 | 1,39794 | 0       | 0       | 5,591760035 | 43,7349 | 44% |
|          |    |         |         |         |         |             | 100%    |     |

**Gambar 3.** Hasil tabel TF-IDF

Pada metode ini menjelaskan tentang berapa persentase yang dihasilkan dari gejala-gejala yang terjadi pada kerusakan printer. Maka berdasarkan perhitungannya didapatkan total persentase sifat positif bernilai 56.2651 atau 56%, sedangkan persentase sifat negatif bernilai 43.7349 atau 44%.

## 4. KESIMPULAN

Adapun kesimpulan yang diperoleh dari penelitian ini dimana proses menerapkan algoritma teks mining dan TF-IDF dapat membalas pesan secara langsung dilakukan dengan mengolah kata yang terdapat di dalam email. Menerapkan algoritma teks mining dan TF-IDF pada email autoresponder yaitu dengan cara menghitung berapa jumlah kata yang sama dalam sebuah pesan email. Kata yang tidak penting akan dibuang dan hal ini akan mempermudah dalam membalas pesan secara langsung. Merancang aplikasi email dengan menggunakan bahasa pemrograman Visual Basic 2008 dan dapat digunakan dengan baik..

## REFERENCES

- [1] K. Br.Rajaguguk and 15110347, "Penerapan Algoritma Text Mining Pada Aplikasi Ujian Berbasis LAN Pada Pada Sekolah Dasar," 2019.
- [2] M. Mahdi, "Penerapan Metode Cosine Similarity dan Pembobotan TF / IDF pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember," pp. 31– 42.
- [3] C. S. Sitti Munifah, Abdul Syukur, "PENGELOMPOKAN ARSIP UNIVERSITAS MENGGUNAKAN ALGORITMA K-MEANS DENGAN FEATURE SELECTION CHI SQUARE," J. Teknol. Inf., vol. 11, pp. 160–171, 2015.
- [4] N. B. Batubara and 5110695, "IMPLEMENTASI METODE EVEN-RODEH CODE UNTUK KOMPRESI KITAB UNDANG-UNDANG HUKUM PIDANA (KUHP) BERBASIS ANDROID," 2019. [1] W.Purbo, Onno, Text Mining, Yogyakarta : Andi. 2019.
- [5] Manalu, Boy Utomo, Analisis Sentimen Pada Twitter Menggunakan Text Mining : Ilmu Komputer. 2014.
- [6] Buntoro, Ghulam Asrofi, Penggunaan Text Mining pada Analisis Sentimen Calon Gubernur Jawa Timur 2018 : Journal Of Informatics Pelita Nusantara. 2019
- [7] Moch.Ali Ramdhani, Rahim Oki Nandoko, Analisis Sentimen Untuk Mengukur Popularitas Tokoh Publik Berdasarkan Data Pada Media Sosial Twitter Menggunakan Algoritma Data Mining dengan teknik klasifikasi : Jurnal Informasi. 2014.
- [8] Rahim, Oki Nandoko, Moch.Ali Ramdhani, Analisis Sentimen Untuk Mengukur Popularitas Tokoh Publik Berdasarkan Data Pada Media Sosial Twitter Menggunakan Algoritma Data Mining dengan teknik klasifikasi : Jurnal Informasi. 2014.