

ANALISIS DATA STATUS DI FACEBOOK UNTUK MEREKOMENDASIKAN PRODUK MAKANAN RINGAN MENGGUNAKAN TF-IDF DAN DOCUMENT WEIGHTING

Rohmad Abidin

Magister Sistem Informasi
Universitas Kristen Satya Wacana Salatiga
Email: rohmad@stekom.ac.id

Eko Sedyono

Magister Sistem Informasi
Universitas Kristen Satya Wacana Salatiga
Email: eko@staff.uksw.edu

Sri Yulianto Joko Prasetyo

Magister Sistem Informasi
Universitas Kristen Satya Wacana Salatiga
Email: sri.yulianto@staff.uksw.edu

ABSTRAK

Tujuan dari penelitian ini adalah mengembangkan sistem analisis status pada akun media sosial untuk mendukung kegiatan promosi Usaha Mikro Kecil khususnya produk makanan ringan. Penelitian menggunakan teknik yang terdapat pada text mining, analisis Term Frequency- Invers Document Frequency dan Document Weighting untuk merekomendasikan produk. Objek penelitian akun media sosial facebook dengan sample random yang tergabung dalam grup Warga Demak serta produk hasil UMK yang terdaftar pada Dinas Perdagangan dan Koperasi Kabupaten Demak. Hasil akhir penelitian ini adalah prototype sistem analisis status media sosial untuk mengklasifikasikan pelanggan produk UMK yang sesuai dengan hasil analisa status facebook. Dari hasil analisa didapatkan rata-rata tingkat akurasi prediksi sebesar 73.27% untuk keberhasilan merekomendasikan produk. Hasil penelitian ini menunjukkan bahwa analisis TF-IDF dan Document Weighting dapat dimanfaatkan untuk pengembangan sistem baru guna mendukung kegiatan promosi pengelola kelompok UMK agar lebih tepat sasaran.

Kata kunci: media sosial; *TF-IDF*; *document weighting*; UMK.

ABSTRACT

The purposes of the research are developing analysis application of status on social media account to support Micro, Small Enterprises (MSE) promotion especially on snack product. The research used techniques in text mining, Term Frequency-Invers Document Frequency analysis and Document Weighting to recommendation the product. The object of research is Facebook status updates as well as products of MSE who registered at the Department of Trade and Cooperatives in Demak District. The final result of this research is prototype social media status analysis system to recommend products from results of system analysis. From the analysis results obtained an average predictive accuracy rate of 73.27% for recommend the product. The results of this study it can be concluded that the analysis of TF - IDF and Document Weighting can be utilized for the development of new systems to support the promotion activities of MSE group managers to be more effective.

Keywords: *social media; TF-IDF; document weighting; MSE.*

1. PENDAHULUAN

Setiap orang memiliki motif berbeda-beda dalam menggunakan media sosial. Motif tertinggi orang menggunakan media sosial adalah ingin menyatakan eksistensi dirinya dengan memperbaharui status serta mendapatkan informasi dari media sosial [1].

Media sosial dapat digunakan untuk menunjang aktivitas keseharian pengguna. Seperti perusahaan atau individu menggunakan media sosial untuk menunjang aktifitas bisnis, saling berbagi karya dan sebagai tempat *sharing* untuk yang memiliki kesamaan hobi. Dengan adanya media sosial di aplikasi

mobile membuat setiap orang bisa menggunakan media sosial dimana saja sepanjang terkoneksi jaringan internet [2].

Media sosial dapat dijadikan sebagai *Integrated Marketing Communication* guna meningkatkan pendapatan dan bertahan dalam persaingan bisnis. UKM dapat menarik calon pelanggan baru atau melakukan komunikasi kepada pelanggan yang telah ada [3].

Penelitian lain tentang promosi menggunakan media sosial dilakukan oleh Priambada yang menghasilkan bahwa web blog, facebook dan E-mail merupakan media yang paling banyak digunakan. Manfaat yang paling dirasakan adalah komunikasi yang efektif sehingga dapat meningkatkan pangsa pasar serta membantu keputusan bisnis. Selain itu, manfaat lainnya untuk mengadakan iklan/promosi, kontak pribadi dengan konsumen, mendata kebutuhan serta memberikan tanggapan kepada konsumen [4].

Penelitian mengenai analisis media sosial pernah dilakukan Muhammad menggunakan teori *five factor model*, hasil penelitian menyimpulkan bahwa kepribadian *Openness to New Experiences* sebanyak 11.68%, kepribadian *Conscientiousness* 5.34%, *Extraversion* 5.5%, *Agreeableness* 8.54%, *Neuroticism* 4.82%, serta 64.12% tidak dapat diklasifikasikan [5].

Untuk penelitian efektifitas iklan pada jejaring sosial yang dilakukan oleh Hasanah mengemukakan hasil dari perhitungan *EPIC rate* yaitu iklan Facebook Huma Harati dinyatakan efektif sebagai media promosi. Namun pada skala efektif menunjukkan bahwa promosi yang dilakukan melalui iklan di Facebook belum mencapai hasil yang maksimal, sehingga perlu ditingkatkan lagi [6]. Rivki melakukan penelitian menggunakan *text mining*, analisis *TF-IDF* serta Algoritma *K-Nearest Neighbor* melakukan klasifikasi ketertarikan *follower* berdasarkan status di Twitter. Hasil nilai akurasi tertinggi sebesar 68% yang didapat pada proses klasifikasi selama empat kali pengujian [7].

Ardi melakukan penelitian dengan membandingkan performa *TF-IDF* dengan *n-gram* menggunakan *Support Vector Machine* (SVM) untuk mendapatkan hasil klasifikasi terbaik untuk menganalisis sentiment masyarakat terhadap *marketplace*. Peneliti menemukan bahwa *Unigram* memiliki nilai akurasi tertinggi, dengan nilai akurasi 80,87% [8]. Sedangkan Chang menggunakan metode *TF-IDF* dan *Social Network Analysis* dalam sebuah studi kasus menyajikan bagaimana sistem rekomendasi bekerja berdasarkan data yang dikumpulkan dari Plurk. Dari hasil analisis, peneliti menemukan perbedaan pengukuran SNA antara berbagai produk yang signifikan, hal itu menunjukkan bahwa sistem rekomendasi dapat dikerjakan untuk merekomendasikan produk yang berbeda untuk menargetkan pelanggan [9].

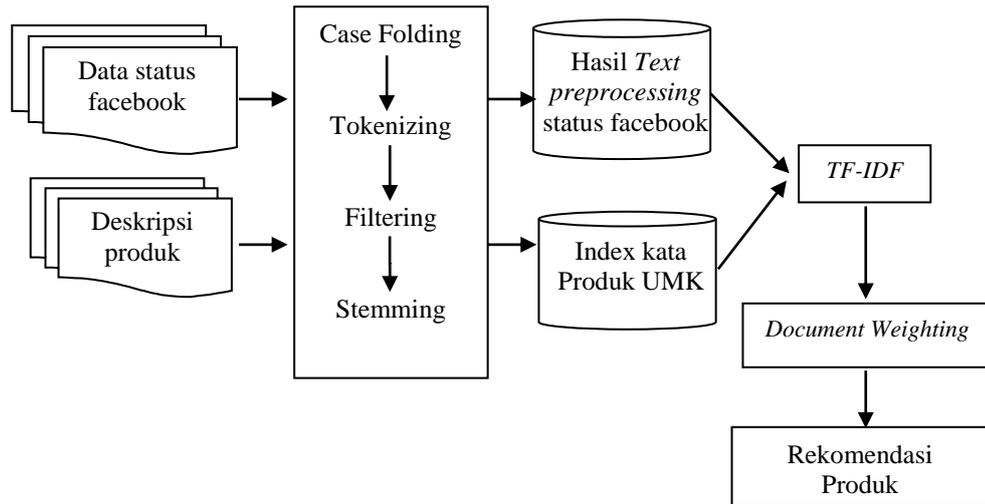
Untuk mengenal calon konsumen secara lebih personal, penelitian ini berusaha menggabungkan data hasil *text mining* status di media sosial seseorang yang diintegrasikan dengan data UKM. Penelitian ini akan menghasilkan sebuah *prototype* sistem rekomendasi produk UMK sesuai dengan hasil analisa status pada *timeline* media sosial masing-masing akun menggunakan analisa *TF-IDF* dan *Document Weighting*.

2. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan dalam tiga tahap yaitu: (1) pengumpulan data penelitian (2) analisis data (3) pengujian hasil analisis produk. Pemrosesan awal data penelitian mencakup aktivitas pengumpulan data produk UMK serta koleksi status akun media sosial. Untuk analisis data terbagi menjadi 2 proses penelitian yaitu *text mining processing*, dan *data comparison*. *Text mining processing* mengolah data input berupa status media sosial serta deskripsi produk UMK. Proses *text mining* dilakukan menggunakan metode yang secara umum dilakukan pada *text mining* yaitu tahap *tokenizing*, *case folding*, *filtering*, serta *stemming*. Proses *text mining* dilakukan pada deskripsi produk makanan ringan yang akan menghasilkan *index* kata pada setiap produk. Proses *text mining* data status media sosial dimulai dengan mengumpulkan semua status facebook calon responden dari awal mendaftar facebook hingga terakhir, kemudian dipilih status yang mengandung kata-kata yang berhubungan makanan ringan seperti makan, lezat, enak, gurih, dan lain-lain.

Proses yang kedua adalah *data comparison*, yaitu membandingkan antara data hasil *text mining* akun media sosial dengan *index* kata pada produk UMK. Proses perbandingan menggunakan *TF-IDF*. Hasil dari *TF-IDF* digunakan untuk memberikan peringkat produk yang terdapat dalam koleksi produk menggunakan kaidah *Document Weighting*. Dari seluruh produk yang terdaftar diambil 5 produk tertinggi untuk di promosikan kepada calon konsumen.

Peneliti melakukan uji coba pertama dengan menggunakan kuesioner untuk menentukan tingkat efektifitas penerapan *TF-IDF* dan *Document Weighting*. Berikut ini adalah diagram tahapan penelitian yang dilakukan (gambar 1).



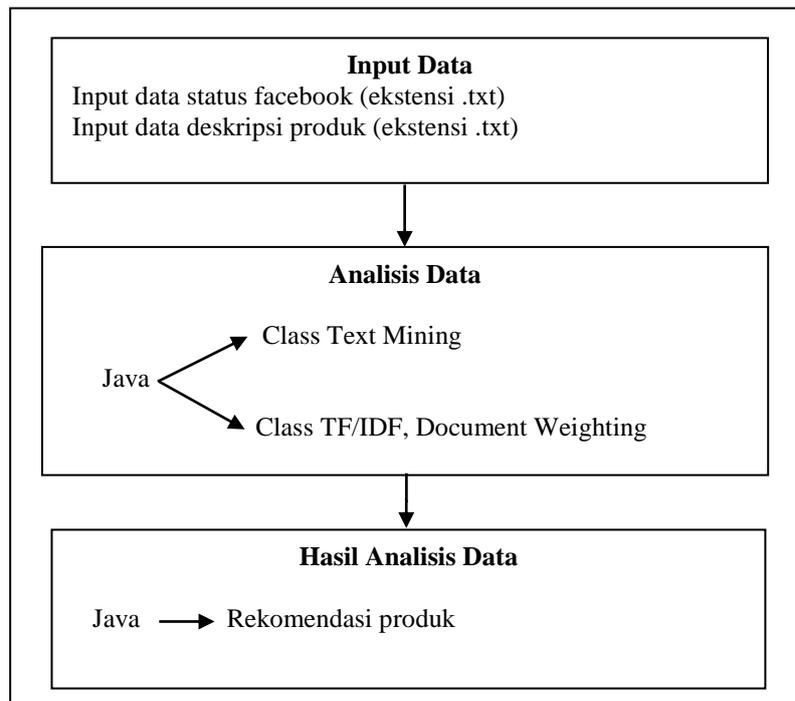
Gambar 1. Diagram Tahap Penelitian

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Desain arsitektural sistem analisis ini dapat dilihat pada Gambar 2. Desain arsitektural sistem terdapat 3 kelompok komponen pada aplikasi yaitu : (1) *input data*, (2) analisis data (3) *output* rekomendasi produk.

Pada gambar 2 memperlihatkan bahwa program memerlukan input data status facebook yang di-copy secara manual dan data deskripsi produk UMK yang disimpan format .txt kemudian dianalisis oleh sistem dengan tahapan yang sudah ditentukan dalam *text mining*. Setelah proses *text mining* selesai, hasil data status facebook dibandingkan dengan data produk UMK menggunakan proses pencairan *TF / IDF* dan *Document Weighting*. Terakhir sistem akan menghasilkan teks rekomendasi produk yang dapat di unduh untuk disimpan.



Gambar 2. Desain Arsitektur

Tampilan dibuat sederhana agar pengguna dapat dengan mudah menggunakan sistem ini. Pada aplikasi juga terdapat halaman utama untuk menganalisis status facebook (Gambar 3). Proses analisis sistem dilakukan dengan tahapan yang telah ditentukan. Proses dilakukan pada *background* sehingga tidak terlihat oleh pengguna. Hasil analisis status facebook akan memberikan rekomendasi produk dengan *sorting* skor tertinggi. Terdapat juga tombol untuk pengaturan produk (Gambar 4) yang berfungsi untuk menambah atau mengurangi produk yang akan dianalisa.

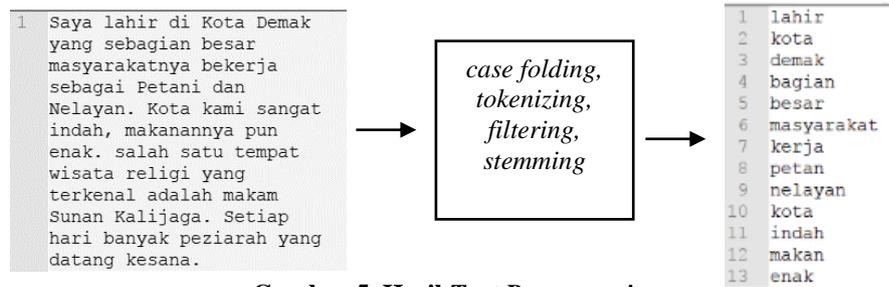
Gambar 3. Form Analisis Data

Gambar 4. Form Pengaturan Produk

3.2 Perhitungan TF-IDF dan Document Weighting

Terlebih dahulu dilakukan *text preprocessing* pada masing-masing deskripsi produk untuk menghasilkan *index* kata tiap produk kemudian dimasukkan ke dalam *database*. Setelah semua data produk tersimpan, selanjutnya data status facebook diproses dengan tahapan yang sama dengan data produk. Data yang status facebook yang sudah dilakukan proses *text preprocessing* diatas kemudian dilakukan perhitungan *TF – IDF* dan *document weighting*. Langkah analisis dilakukan dengan tahapan sebagai berikut :

- a. Langkah pertama dilakukan *text preprocessing* pada produk UMK untuk menghasilkan *index* kata pada masing-masing produk. *Text preprocessing* juga dilakukan pada data status facebook. Tujuan utama dari *text preprocessing* adalah menghilangkan kata-kata tidak penting serta mengubah kata menjadi kata dasar agar mudah digunakan untuk pencarian kata yang sama di masing-masing deskripsi produk. Pada Gambar 5 di bawah ini menunjukkan proses *text preprocessing* pada status facebook.



Gambar 5. Hasil Text Preprocessing

- b. Hitung jumlah kata (*term frequency*) dengan mencari kata yang sama antara data status facebook dengan masing-masing deskripsi produk (*D*). Hasil *term frequency* akan menunjukkan bahwa semakin banyak kata yang sama antara status facebook dengan deskripsi produk maka semakin besar kemungkinan pemilik akun memilih produk tersebut. Pemberian bobot kata pada dokumen menggunakan kode sebagai berikut :

```
private Double tf(List<String> doc, String term) {
    double tf = 0;
    for (String word : doc) {
        if (term.equalsIgnoreCase(word)) {
            tf++;
        }
    }
    System.out.print(" Raw TF : " + tf);
    return tf;
}
```

- c. Tentukan jumlah kata yang memiliki nilai di setiap produk (*dokumen frequency/df*). Jika di dalam produk memiliki nilai *tf* berapapun nilainya maka akan diberi nilai 1, jika tidak memiliki nilai *tf* maka diberi nilai 0.

$$df_{i3} = d1_{i3} + d2_{i3} + d3_{i3} + \dots + d15_{i3} \tag{1}$$

$$= 1 + 0 + 1 + 1 + 1 + 1 + 0 + 0 + 0 + 0 + 0 + 0 + 1 + 1 + 1$$

$$= 10$$

Pencarian nilai *df* bertujuan untuk mengetahui sebaran suatu kata, semakin banyak tersebar di deskripsi produk yang lain maka kata tersebut dianggap kata yang umum, semakin sedikit sebaran suatu kata di produk lain, maka kata tersebut dapat menunjukkan kekhasan suatu produk.

- d. Hitung nilai *IDF* dari tiap-tiap *term*, fungsi rumus menghitung *IDF* sebagai berikut :

$$idf_t = \log \left(\frac{N}{df_t} \right) \tag{2}$$

Dimana :

- N* = total semua dokumen yang diuji
- df* = banyak jumlah dokumen yang mengandung kata (*term*) yang dicari
- t* = kata yang dicari (*term*)

$$idf_{i3} = \text{LOG} \left(\frac{d}{df_{t13}} \right)$$

$$= \text{LOG} \left(\frac{15}{10} \right)$$

$$= \text{LOG}(1,5)$$

$$= 0.176091259$$

Fungsi *IDF* diatas guna mengurangi bobot kata-kata yang bersifat umum dan meningkatkan bobot kata unik. Semakin tinggi bobot kata, di suatu produk semakin besar kemungkinan produk akan dipilih. Berikut *listing* program menghitung *Document Frequency* dan *Invers Document Frequency*.

```

private double idf(List<List<String>> docs, String term) {
    double f = 0;
    for (List<String> doc : docs) {
        for (String word : doc) {
            if (term.equalsIgnoreCase(word)) {
                f++;
                break; }
        }
    }
    System.out.print(" DF : "+f);
    if(f==0){
        System.out.print(" IDF : 0");
        return 0;
    }else{
        System.out.print(" IDF : "+Math.log10(docs.size() / f));
        return Math.log10(docs.size() / f);
    }
}

```

- e. Lakukan pembobotan tiap kata pada deskripsi produk (d) yang telah disimpan. Pemberian bobot kata pada dokumen menggunakan rumus sebagai berikut :

$$W_{dt} = tf_{dt} * IDF_t \quad (4)$$

Keterangan :

d = dokumen ke-d

t = kata ke-t dari kata kunci

W = bobot dokumen ke-d terhadap kata ke-t

tf = jumlah kata yang dicari pada sebuah dokumen

$$\begin{aligned}
 W_{d13} &= tf_{13} \times idf_{13} \\
 &= 1 \times 0.176091259 \\
 &= 0.176091259
 \end{aligned}$$

Berikut *listing* kode untuk menghitung bobot tiap kata.

```

private double Wdt(List<String> doc, List<List<String>> docs, String
term) {
    return tf(doc, term) * idf(docs, term);
}

```

- f. Jumlahkan tiap Bobot W_{dt} untuk mengetahui nilai bobot tiap produk (W_d) semakin besar bobot dokumen, kemungkinan semakin tinggi dipilih oleh konsumen.

$$W_d = W_{dt1} + W_{dt2} + W_{dt3} + \dots + W_{dtn} \quad (5)$$

Dimana :

W_d = Bobot total dokumen

W_{t1-tn} = Bobot tiap dokumen terhadap kata

$$\begin{aligned}
 W_{d1} &= W_{d11} + W_{d12} + W_{d13} + \dots + W_{d125} \\
 &= 0 + 0 + 0.176091259 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0.574031268 + 0 + 0 + 0 + 0 + \\
 &\quad 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 \\
 &= 0.750122527
 \end{aligned}$$

Berikut *listing* kode menghitung bobot setiap produk (W_d).

```

Double[][] hasilWeight = new Double[documents.size()][documents.size()];
TFIDFWeight calculator = new TFIDFWeight();
Double[] docWeight = new Double[documents.size()];
for (int a = 0; a < documents.size(); a++) {

```

```

System.out.println("dokumen ke "+(a+1));
docWeight[a] = 0.0;
System.out.println("dokumen ke "+(a+1));
for (int i = 0; i < medsos.size(); i++) {
    System.out.print(medsos.get(i));
    double Wdt = calculator.Wdt(documents.get(a), documents,
    medsos.get(i));

    //tambahkan bobot produk tiap ada term
    docWeight[a] = docWeight[a] + Wdt;
}

hasilWeight[a][0] = Double.valueOf(a);
hasilWeight[a][1] = docWeight[a];
System.out.println(" Dokumen Weighting = " + docWeight[a]);
}

```

Baris kode pada program diatas menunjukkan perhitungan bobot tiap kata terdapat dokumen akan dijumlahkan hingga selesai untuk satu produk. Jika produk pertama selesai dianalisis akan dilanjutkan ke produk kedua dan seterusnya hingga semua produk selesai di analisis.

Hasil perhitungan pada bobot produk pertama (W_{di}) memiliki bobot 0.750122527. Hasil perhitungan bobot dari setiap produk disajikan pada tabel 1 di bawah ini :

Tabel 1. Hasil pemeringkatan bobot produk

| <i>Kode</i> | <i>Nama Produk</i> | <i>Bobot</i> | <i>Ranking</i> |
|-------------|------------------------|--------------|----------------|
| D2 | bintang 5 | 7.1077 | 1 |
| D14 | rangin super | 2.801275 | 2 |
| D9 | keripik pisang raja | 1.528274 | 3 |
| D8 | keripik pisang enggano | 1.449093 | 4 |
| D12 | kerupuk ikan laut | 1.176091 | 5 |
| D15 | wingko witela | 1.051153 | 6 |
| D1 | aneka cemal cemil | 0.750123 | 7 |
| D5 | carang madu | 0.750123 | 7 |
| D3 | bis bollen | 0.176091 | 9 |
| D4 | brayo | 0.176091 | 9 |
| D6 | ikan berkah | 0.176091 | 9 |
| D13 | kue satelit | 0.176091 | 9 |
| D7 | keripik jamur kancing | 0 | 13 |
| D10 | keripik singkong | 0 | 13 |
| D11 | keripik udang | 0 | 13 |

Dari hasil pemeringkatan diatas maka dipilih 5 produk tertinggi untuk direkomendasikan kepada pelanggan untuk ditawarkan.

3.3 Analisa Hasil

Peneliti melakukan uji coba dengan menggunakan kuesioner untuk menentukan tingkat efektifitas sistem yang dihasilkan. Populasi objek penelitian adalah anggota grup facebook Warga Demak yang telah mempunyai akun facebook lebih dari 5 tahun dan berusia lebih dari 17 tahun. Sedangkan jumlah produk yang digunakan untuk bahan ujicoba dibagi menjadi 3 tahap pengujian yaitu tahap 1 sebanyak 15 produk, tahap 2 sebanyak 30 produk, serta tahap 3 sebanyak 50 produk. Masing-masing tahap sebanyak 50 responden. Pada tabel 2 menunjukkan hasil uji coba sistem yang dihasilkan :

Tabel 2. Hasil analisa

| <i>Jumlah Produk</i> | <i>Memilih</i> | <i>Tidak memilih</i> |
|----------------------|----------------|----------------------|
| 15 | 67% | 30.3% |
| 30 | 75.6% | 24.4% |
| 45 | 77.2% | 22.8% |
| Rata-rata | 73.27% | 26.73% |

4. KESIMPULAN

4.1 Kesimpulan

Dari hasil pertama menunjukkan bahwa tingkat akurasi yang dihasilkan oleh algoritma *TF-IDF* dan *Document Weighting* dalam merekomendasikan produk sebesar 73.27%. Semakin banyak produk yang digunakan dalam penelitian semakin besar tingkat keberhasilan sistem. Sehingga dapat disimpulkan bahwa metode *TF-IDF*, *Document Weighting* dapat digunakan sebagai algoritma untuk merekomendasikan produk pada pelanggan sesuai dengan yang diminati melalui analisis status akun facebook.

4.2 Saran

Saran dalam pengembangan sistem kedepan sebagai berikut :

- a. Penambahan berbagai jenis produk yang lebih beragam tidak hanya makanan ringan.
- b. Kedepan perlu dikembangkan pula meneliti jaringan grup yang diikuti agar lebih maksimal dalam merekomendasikan produk.
- c. Penambahan metode baru untuk mengelompokkan pengguna berdasarkan keseragaman produk.

DAFTAR PUSTAKA

- [1] Utomo DA. Motif Pengguna Jejaring Sosial Google+ Di Indonesia. *J E-KOMUNIKASI*. 2013;1(3).
- [2] Luik JE. Media Sosial dan Presentasi Diri. Bab Buku Komunikasi 2.0: Teoritisasi dan Implikasi. Yogyakarta: AspiKom; 2011.
- [3] Siswanto T. Optimalisasi Sosial Media Sebagai Media Pemasaran Usaha Kecil Menengah. *J Liq*. 2013;2(1):80–6.
- [4] Priambada S. Manfaat Penggunaan Media sosial Pada Usaha Kecil Menengah (UKM). In: Seminar Nasional Sistem Informasi Indonesia. SESINDO; 2015.
- [5] Muhammad HSRRR. Penerapan Data Mining Dalam Menganalisis Kepribadian Pengguna Media Sosial Dengan Naive Bayes Classifier : Studi Kasus Media Sosial Instagram. *J Inform Univ Padjadjaran*. 2016;01(01):5258.
- [6] Hasanah N, Nugroho LE, Nugroho E. Analisis Efektivitas Iklan Jejaring Sosial sebagai Media Promosi Menggunakan EPIC Model. *Sci J Informatics [Internet]*. 2016 Feb 16;2(2):99. Available from: <https://journal.unnes.ac.id/nju/index.php/sji/article/view/5075>
- [7] Rivki M, Bachtiar AM. Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *J Inf Syst*. 2017;13(1).
- [8] Ardi HL, Sedyono E, Ningrum RK. Support Vector Machine Classifier For Sentiment Analysis Of Feedback Marketplace For Sentiment Analysis Of Feedback Marketplace With A Comparison Features At Aspect Level. *Int J Innov Res Adv Eng*. 2017;4(11):7–12.
- [9] Chang P, Ting I-H, Wang S. Towards Social Recommendation System Based on the Data from Microblogs. In: International Conference on Advances in Social Networks Analysis and Mining [Internet]. IEEE; 2011. p. 672–7. Available from: <http://ieeexplore.ieee.org/document/5992655/>