



TEMA :
GENERATED CONTENT

PROPOSAL DISERTASI

Qomariyah
99220707

PROGRAM DOKTOR ILMU KOMPUTER DAN TEKNOLOGI
INFORMASI
UNIVERSITAS GUNADARMA
DESEMBER 2021

DAFTAR ISI

DAFTAR ISI.....	ii
DAFTAR GAMBAR	iii
DAFTAR TABEL.....	iv
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Tujuan Penelitian.....	4
1.3 Batasan	4
1.4 Kontribusi Penelitian.....	5
TINJAUAN PUSTAKA	6
2.1 Media Sosial	6
2.2 Review	6
2.3 E-Commerce.....	8
2.4 Text Mining.....	8
2.5 Pembobotan dengan TF-IDF	9
2.6 Klasifikasi.....	10
2.7 Support Vector Machine	10
2.8 Penelitian Terkait	10
METODOLOGI PENELITIAN.....	15
3.1. Tahapan Penelitian	15
3.2. Rencana Kegiatan.....	17
DAFTAR PUSTAKA	18

DAFTAR GAMBAR

Gambar 1. 1. Tren internet dan media sosial tahun 2020 di Indonesia.....	2
Gambar 1. 2. Jenis konten Internet yang banyak di akses tahun 2016 Menurut APJII	3
Gambar 2. 1. Gambaran bagaimana review di buat	7
Gambar 3. 1. Metodologi Penelitian	15

DAFTAR TABEL

Tabel 2. 1. Penelitian Terdahulu	13
--	----

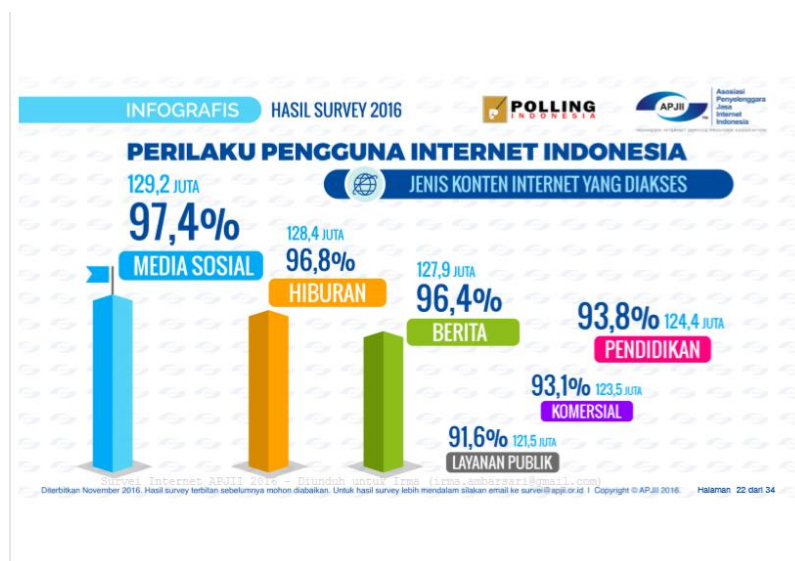
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi saat ini banyak dirasakan dan dapat dimanfaatkan khalayak ramai. Dalam berkomunikasi serta berinteraksi saat ini sudah tidak dibatasi oleh ruang maupun waktu. Komunikasi dan interaksi dapat dilakukan secara online berbasis internet atau melalui operator seluler dengan smartphone (Rahmati, 2009). Internet menjadi sesuatu hal yang tidak terpisahkan dalam melakukan kegiatan sehari-hari, terutama di tengah pandemi Covid-19 seperti yang terjadi saat ini. Dengan adanya internet banyak sekali kegiatan atau pekerjaan yang dapat dilakukan secara cepat dan fleksible tanpa harus keluar rumah. Penggunaan internet yang ada dimaksudkan baik untuk memenuhi kebutuhan primer sehari-hari maupun kebutuhan sekunder. Kebutuhan primer seperti kegiatan berbelanja (contohnya berbelanja makanan, minuman dan pakaian) yang sebelumnya dilakukan secara tradisional melalui pasar atau swalayan dimana pembeli dan penjual harus berinteraksi secara langsung untuk melakukan proses jual-beli, saat ini proses tersebut di gantikan secara online dengan pemanfaatan internet. Salah satu media pemenuhan kebutuhan sehari-hari secara online adalah melalui situs-situs perbelanjaan *E-Commerce*. Situs *E-Commerce* yang sering diakses oleh masyarakat seperti, tokopedia, lazada, shoope dan lain sebagainya. Dengan adanya situs-situs perbelanjaan online banyak masyarakat merasa terbantu dikarenakan kemudahan dalam hal pengaksesan melalui perangkat dan media yang dapat terhubung melalui internet yang mana situs-situs tersebut sudah dapat diakses melalui perangkat-perangkat yang dimiliki oleh kebanyakan masyarakat seperti mobile phone. Saat ini total populasi di Indonesia adalah (jumlah penduduk): 272,1 juta jiwa. Sedangkan pengguna *mobile* unik (mobile phone): 338,2 juta, dimana setiap orang yang ada memungkinkan untuk memiliki lebih dari 1 perangkat *mobile* unik (mobile phone). Pengguna internet di Indonesia sendiri saat ini sebanyak 175,4 juta Pengguna, Dan pengguna media sosial aktif sebanyak: 160 juta. Hal tersebut sesuai dengan data tren internet dan media sosial 2020 di Indonesia menurut tootsuite (*We are social*) dengan banyaknya pengguna perangkat mobile unik (mobile phone) dan sosial media aktif menjadikan banyaknya masyarakat dapat melakukan pengaksesan situs-situs e-commerce bahkan dapat melakukan pertukaran informasi seputar e-commerce dan melakukan review secara online terhadap suatu produk bahkan merek :

oleh konsumen, salah satunya dengan menarik minat pembeli dengan menampilkan review positif atau testimoni pengalaman dari para pembeli sebelumnya sehingga pembeli yang ada dapat percaya dan membeli produk tersebut berdasarkan pengalaman konsumen yang ada, dimana review atau testimoni yang di tampilkan akan menggambarkan kesan dan pesan yang di sampaikan konsumen setelah konsumen tersebut sudah membeli produk sehingga review yang di buat oleh konsumen dibuat berlandaskan pengalaman dan bersifat jujur terhadap suatu produk atau perusahaan sehingga dapat menaikkan citra merek tersebut. Pemanfaatan testimoni dari para pembeli sebelumnya dapat di sebut dengan konten buatan pengguna atau *user-generated content*. Konten buatan pengguna adalah konten yang tersedia untuk umum di Internet yang mencerminkan sejumlah upaya kreatif dan dibuat di luar professional rutinitas dan praktik (OECD, 2007).

Pada 2016, media sosial menjadi konten internet yang paling banyak diakses di Indonesia dengan 129,2 juta orang atau 97,4% dari total pengguna internet di Indonesia dari 132,7 juta orang atau 51,8% dari jumlah penduduk (APJII, 2016). Jenis data konten internet berikut sering kali diakses di Indonesia:



Gambar 1. 2. Jenis konten Internet yang banyak di akses tahun 2016 Menurut APJII
(Hermaren and Achyar, 2018)

Banyaknya konten yang di tampilkan dan di berikan oleh user membuat perusahaan atau penyedia barang harus lebih selektif lagi dalam memilih konten yang dibuat pengguna agar testimoni dan review yang dipilih dapat mewakili dan menarik minat para konsumen yang ada. Selain itu perusahaan juga harus dapat melihat tren di kalangan konsumen agar konten

pengguna yang di tampilkan dapat memberikan kesan tersendiri dan mewakili merek perusahaan tersebut terhadap konsumen serta dapat menarik calon konsumen yang baru. Ulasan atau review produk dari konsumen merupakan salah satu faktor yang penting dalam penjualan suatu produk. Ulasan produk sering digunakan produsen suatu produk untuk menunjang pemasaran suatu produk, terlihat pada penelitian oleh *Dimensional Search* pada 1046 orang koresponden menyatakan dalam menentukan keputusan dalam pembelian suatu produk 90% calon konsumen terpengaruh oleh ulasan produk positif, sedangkan di sisi lain 86% keputusan dipengaruhi oleh ulasan negatif suatu produk (Gesenhues, 2013). Ulasan suatu produk juga dapat menjadi alat yang sangat berguna untuk menarik pelanggan baru, tidak hanya melalui efek psikologis yang didapatkan oleh para pembaca, tetapi dengan meningkatkan peringkat lokal sehingga bisnis lebih mungkin muncul untuk mencari konsumen baru (Murphy, 2020). Oleh karena ulasan, review atau konten-konten yang dibuat pengguna diharapkan dapat dijadikan ringkasan serta tolak ukur untuk dapat membangkitkan citra baru suatu merek agar dapat menjadi salah satu teknik pemasaran dan menarik minat para pelanggan sesuai dengan data yang berasal dari review atau ulasan yang didapatkan dari pengguna. Studi mengenai branding produk atau perusahaan masih jarang dilakukan. Berdasarkan uraian di atas, peneliti akan melakukan penelitian *generated content* terhadap testimoni- testimoni serta review dari suatu produk atau merek yang digunakan konsumen sebagai salah satu teknik pembangkitan citra baru menggunakan teknik klasifikasi.

1.2 Tujuan Penelitian

Berdasarkan latar belakang yang sudah disebut diatas, maka tujuan yang ingin dicapai yaitu mengembangkan satu metode dan algoritma yang dapat:

1. Mengklasifikasi review positive dari pengguna suatu merek.
2. Memberikan persepsi baru terhadap suatu merek untuk menaikkan citra.

1.3 Batasan

Batasan dalam penelitian ini adalah bagaimana caranya meng-*generated content* berdasarkan review atau kesan pengguna yang disebarakan pada sosial media serta mengandung makna positif agar dapat membentuk citra baru terhadap suatu merek.

1.4 Kontribusi Penelitian

Kontribusi dari sisi pengembangan ilmu pengetahuan menghasilkan metoda dan algoritma yang dapat membangkitkan citra merek suatu produk berdasarkan konten yang dibuat pengguna.

Kontribusi dari sisi pengembangan teknologi menghasilkan aplikasi untuk membangkitkan citra merek yang baik terhadap suatu perusahaan dengan mengambil dan menyaring review-review *positive* dari pengguna sosial media terhadap suatu produk yang sedang diminati.

TINJAUAN PUSTAKA

2.1 Media Sosial

Media sosial adalah interaksi yang dilakukan secara online; seperti melakukan diskusi, berbagi pemikiran, pengalaman, dan informasi untuk membuat pilihan yang lebih baik atau lebih terinformasi (Mustonen, 2009). Media sosial dikenal sebagai komunikasi internet partisipatif yang mewakili sekelompok aplikasi berbasis internet untuk menciptakan dan bertukar konten buatan pengguna (video, foto, file grafis, dll) (Osborne-Gowey, 2014).

Menurut (Evans, 2008) media sosial adalah keterbukaan dan kebebasan informasi, mengubah orang dari pembaca konten menjadi penerbit konten. Ini adalah perubahan dari keterbatasan informasi secara personal menjadi publik, berakar pada percakapan antara pembaca berita, orang dan rekan-rekan. Media sosial menggunakan "kebijaksanaan orang banyak" untuk menghubungkan informasi secara kolaboratif. Media sosial dapat berupa platform apa saja, termasuk forum Internet, papan pesan, weblog, wiki, podcast, gambar dan video.

Platform media sosial merupakan produk dari Web 2.0, dengan tawaran interaktif tanpa batasan, yang juga disertai dengan kemudahan untuk berkontribusi seperti memberikan kesempatan bagi para pengguna untuk bersosialisasi dengan bereaksi, mengunggah, memberikan komentar dan berdiskusi pada platform tersebut. (Mustonen, 2009)

Maraknya pemanfaatan berbagai *social media* (media sosial) sebagai sekumpulan aplikasi berbasis internet yang dibangun berdasarkan landasan ideologis dan teknologis dari Web 2.0 (Kaplan and Haenlein, 2010) memungkinkan para konsumen yang juga pengguna sosial media dari seluruh penjuru dunia untuk berinteraksi, berkolaborasi, dan membagi konten yang berupa kata-kata, gambar, video, atau suara yang memaparkan ide, opini, maupun testimonial, mengenai pengalaman mereka dalam menggunakan suatu produk atau jasa yang dikenal dengan konten buatan pengguna secara teks atau yang biasa dikenal sebagai review .

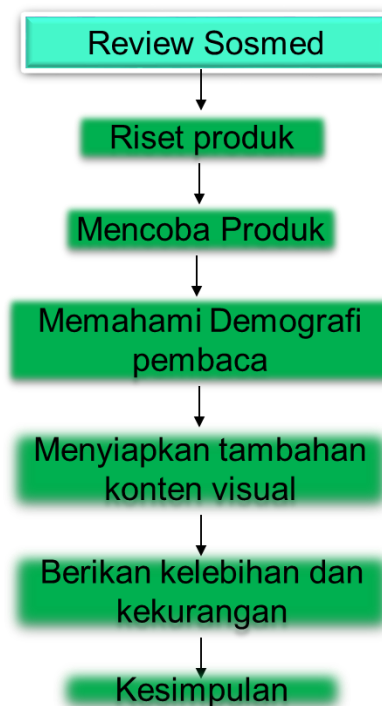
2.2 Review

Online customer reviews (OCR) atau ulasan online pengguna adalah sebuah bentuk dari *electronic word of mouth* (e-WOM) yang mengacu pada sebuah konten yang dibuat oleh pengguna (*user-content*) serta diposting pada situs belanja online atau e-commerce bahkan

situs web pihak ketiga (Mudambi, Schuff and Schuff, 2016). OCR selama ini telah menjadi sebuah sumber utama informasi produk di dunia online (Huang, 2008).

Menurut (Hennig-Thurau *et al.*, 2004) review yang dibuat oleh pengguna adalah “setiap pernyataan positif atau negatif dibuat oleh pelanggan potensial, aktual, atau pelanggan yang pernah merasakan produk atau perusahaan tersebut yang kemudian disebarluaskan melalui media internet yang ditujukan untuk orang banyak”.

Menurut *Organisation for Economic Co-Operation and Development* (OECD, konten buatan pengguna didefinisikan menjadi beberapa bagian ; 1) konten yang dipublikasi dan dapat diakses melalui internet, 2) yang merefleksikan usaha kreatif pengunggahnya, dan 3) diciptakan diluar rutinitas dan konteks profesional. Pada dasarnya konten yang di buat oleh pengguna dimaksudkan untuk berbagi informasi dan menunjukan suatu produk yang di miliki pengguna tersebut melalui halaman sosial media pribadi mereka. Dan perusahaan atau penyedia barang yang mempunyai merek tersebut dapat menampilkannya pada halaman website perusahaan atau sosial media perusahaan dengan terlebih dahulu meminta ijin atau menyertakan pemilik konten dalam publikasi tersebut dengan menyebutkan id pengguna di media sosial atau meminta ijin melalui private message ke pemilik konten tersebut.



Gambar 2. 1. Gambaran bagaimana review di buat

2.3 E-Commerce

Pemanfaatan internet dalam dunia bisnis lebih dikenal dengan istilah '*electronic commerce*'. Pengertian *electronic commerce* (*e-commerce*) Menurut (Efraim Turban, David King, Dennis Viehland, 2006) *E-commerce* adalah pendekatan baru dalam dunia bisnis secara elektronik dan menggunakan jaringan dan *internet*. Dengan cara ini proses jual beli atau pertukaran produk, jasa dan informasi melalui komputer pada jaringan *internet*.

Suatu proses pembelian serta penjualan produk-produk yang dilakukan secara elektronik oleh konsumen, dari perusahaan ke perusahaan dengan komputer sebagai perantara transaksi bisnis. *E-commerce* berguna dalam mengurangi biaya administrasi dan waktu siklus proses bisnis, dan meningkatkan hubungan dengan kedua mitra bisnis dan pelanggan (Laudon and Laudon, 2013). Serta dapat melakukan pemasaran dan penciptaan merek dengan adanya interaksi secara tidak langsung terhadap pelanggan melalui review dan *branding* suatu merek.

2.4 Text Mining

Analisis studi teks dapat dilakukan dengan berbagai cara, yaitu : analisis isi, analisis wacana, analisis semiotika, analisis hermeunetika, analisis gaya teks, analisis struktur teks, analisis leksikal dan banyak analisis lainnya yang dapat dilakukan pada teks dengan banyak metodologi untuk menemukan dan mengevaluasi pengetahuan di dalamnya salah satunya dengan menggunakan metodologi *text mining*.

Text mining memberikan sebuah set metodologi dan tool untuk menemukan (*discovering*), memvisualisasikan (*presenting*), mengevaluasi pengetahuan dari kumpulan besar dari teks dokumen (Purbo, 2019). *Text mining* dapat disebut juga sebagai data *text mining*, adalah proses untuk memperoleh informasi berkualitas tinggi biasanya diperoleh karena memperhatikan pola dan ekstraksi informasi. Menurut (Ronen Feldman, 2007). "*text mining* dapat didefinisikan sebagai sebuah proses pengetahuan intensif yang memungkinkan pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis". Langkah-langkah yang dilakukan dalam *text mining* adalah sebagai berikut:

1. *Tokenizing* adalah proses memecah dokumen menjadi kumpulan kata. *Tokenization* dapat dilakukan dengan menghilangkan tanda baca dan memisahkannya per spasi. Tahapan ini juga menghilangkan karakter-karakter tertentu seperti tanda baca dan mengubah semua *token* ke bentuk huruf kecil (*lower case*).
2. *Filtering* adalah tahap mengambil kata-kata penting dari hasil token dan menghapus karakter angka serta link yang terhubung ke internet. *Filtering* dilakukan dengan

menentukan *term* mana yang akan digunakan untuk merepresentasikan dokumen sehingga dapat mendeskripsikan isi dokumen dan membedakan dokumen tersebut dari dokumen lain di dalam koleksi.

3. *Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*) dimana kata yang ada akan dikembalikan menjadi kata dasarnya. Hal ini bisa dilakukan dengan cara menghilangkan akhiran atau awalan dari sebuah kata.
4. *Tagging* adalah tahapan menemukan bentuk awal atau *root* dari tiap kata lampau atau kata hasil *stemming*.

2.5 Pembobotan dengan TF-IDF

Tahap selanjutnya yaitu pembobotan TF-IDF yang digunakan untuk mengevaluasi seberapa pentingnya sebuah kata di dalam sebuah dokumen. *Term Frequency* (TF) yaitu semakin tinggi frekuensi kemunculan term pada sebuah dokumen maka akan semakin tinggi juga nilai bobot untuk term itu sendiri. Sementara itu, proses *Inverse Document Frequency* (IDF) merupakan kebalikan dari proses TF. Pada IDF, semakin tinggi frekuensi kemunculan term maka nilai bobot term itu sendiri akan semakin kecil (Zakiyuddin, 2021).

Metode TF-IDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Metode ini menggabungkan dua konsep untuk perhitungan bobot, yaitu frekuensi kemunculan sebuah kata di dalam sebuah dokumen tertentu dan *inverse* frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata di dalam dokumen yang diberikan menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut. Sehingga bobot hubungan antara sebuah kata dan sebuah dokumen akan tinggi apabila frekuensi kata tersebut tinggi di dalam dokumen dan frekuensi keseluruhan dokumen yang mengandung kata tersebut yang rendah pada kumpulan dokumen. (Nurjannah and Fitri Astuti, 2013)

Term Frequency Inverse Document Frequency (TF-IDF) merupakan metode pembobotan yang dilakukan untuk ekstraksi data teks yang digunakan untuk menganalisis hubungan antara frase / kalimat dengan sekumpulan dokumen. Algoritma *Term Frequency Inverse-Document Frequency* adalah suatu algoritma yang menggabungkan antara *Term frequency* dengan *Inverse Document Frequency*. *Term frequency* yaitu jumlah kemunculan sebuah *term* pada sebuah dokumen. *Inverse Document Frequency* yaitu pengurangan dominasi *term* yang

sering muncul diberbagai dokumen, dengan memperhitungkan kebalikan frekuensi dokumen yang mengandung suatu kata.

Nilai TF dihitung dengan rumus $TF = \text{jumlah frekuensi kata terpilih} / \text{jumlah kata}$, dan nilai IDF dihitung dengan rumus $IDF = \log (\text{jumlah dokumen} / \text{jumlah frekuensi kata terpilih})$. Selanjutnya adalah melakukan perkalian antara nilai TF dan IDF untuk mendapatkan jawaban akhir.

2.6 Klasifikasi

Klasifikasi merupakan cara pengelompokkan data sesuai dengan ciri-ciri yang dimiliki oleh objek klasifikasi atau karakteristik data. Dalam prosesnya, klasifikasi dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algoritma komputer cerdas . Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya Naïve Bayes, Support Vector Machine, Decission Tree, Fuzzy dan Jaringan Saraf Tiruan. (Wibawa *et al.*, 2018).

2.7 Support Vector Machine

Support Vector Machine (SVM) adalah metode yang mempelajari area yang memisahkan antar kategori dalam sebuah observasi. Dalam terminologi SVM, kita membahas jarak atau margin antar kategori. Setiap kategori memiliki observasi dimana nilai variabel targetnya sama (Williams, 2011). SVM juga dikenal sebagai sistem pembelajaran yang menggunakan hipotesis fungsi linear dalam ruang dimensi tinggi dan dilatih dengan algoritma berdasarkan teori optimasi dengan menerapkan learning bias yang berasal dari teori statistik. Tujuan dari metode ini adalah membangun pemisah optimum yang disebut OSH (*Optimal Separating Hyperplane*) sehingga dapat digunakan untuk klasifikasi.

2.8 Penelitian Terkait

Beberapa penelitian mengenai ini sudah dilakukan antara lain oleh (Roring, Mekel and Soegoto, 2014) dengan tujuan untuk mengetahui tingkat loyalitas konsumen pada PT. Hasjrat Abadi Cabang Manado dalam pembelian toyota avanza berdasarkan citra merek, harga dan inovasi. Sampel penelitian yang digunakan sebanyak 95 orang. Metode penelitian yang digunakan adalah jenis penelitian asosiatif dengan teknik analisa data menggunakan regresi linear berganda. Hasil penelitian yang telah dilakukan menunjukkan bahwa Harga berpengaruh

positif dan signifikan terhadap Loyalitas konsumen pada Mobil Toyota Avanzadi PT. Hasjrat Abadi Cabang Manado, dengan demikian H^a dapat di terima. Penelitian ini sejalan dengan penelitian yang dilakukan sebelumnya tentang: Kualitas Pelayanan, Kualitas produk dan Harga Pengaruhnya terhadap Loyalitas pelanggan kartu AS Telkomsel di kota Manado, yang menyimpulkan bahwa secara parsial Harga berpengaruh signifikan terhadap Loyalitas konsumen pengguna kartu AS Telkomsel di kota Manado. Serta Citra Merek, Harga, dan Inovasi Produk berpengaruh positif dan signifikan terhadap Loyalitas konsumen pada Mobil Toyota Avanza di PT. Hasjrat Abadi Cabang Manado.

(Nurjannah and Fitri Astuti, 2013) melakukan penelitian dengan tujuan agar perusahaan dapat membuat strategi komunikasi di media sosial yang dapat memengaruhi persepsi konsumen dan niat beli, terutama melalui akun resmi merek di LINE. Responden dalam penelitian ini adalah 493 pengguna LINE yang berteman dengan akun resmi sebuah merek. Penyebaran kuesioner dilakukan secara online, dalam kurun waktu 2 bulan. Pengukuran kuesioner menggunakan skala Likert lima poin dan dianalisis menggunakan smart-PLS. Hasil menunjukkan bahwa komunikasi media sosial yang diciptakan oleh perusahaan dan yang dibuat oleh pengguna tidak berpengaruh secara signifikan terhadap ekuitas merek, akan tetapi sikap merek berpengaruh secara positif terhadap ekuitas merek, komunikasi media sosial yang diciptakan oleh perusahaan dan yang dibuat oleh pengguna berpengaruh secara positif terhadap sikap merek, serta ekuitas merek dan sikap merek berpengaruh secara positif terhadap niat beli.

(Herwijayanti, Ratnawati and Muflikhah, 2018) melakukan penelitian dengan tujuan Meminimalkan *human error* serta mengurangi terjadinya kesalahan pengkategorian pada saat memposting berita. Karena banyaknya berita yang dimasukkan ke dalam *website*, sehingga terkadang berita tersebut terposting tidak sesuai dengan kategorinya. *Human error* akan menjadi masalah jika terdapat berita yang salah *posting*. Selain kesalahan *posting* pengelompokan berita online juga penting untuk kenyamanan *user* untuk mencari berita sesuai dengan kategorinya. Agar tidak terjadi kesalahan dalam pengkategorian terhadap berita *online* yang ada, akan diterapkan klasifikasi suatu pengkategorian berita *online* dengan menggunakan pembobotan *tf-idf* dan *cosine similarity*. Adapun teknik yang digunakan dalam ekstraksi fitur yaitu melakukan pembersihan data mulai dari *tokenizing*, *stop words removal* dan *stemming*. Selanjutnya yaitu melakukan *transform* data dengan pembobotan terhadap *term* yang telah dibersihkan. Kemudian dilanjutkan dengan reduksi data. Tahap terakhir yaitu melakukan analisis terhadap proses klasifikasi untuk merepresentasikan hasil informasi yang ditemukan.

(Triyansyah and Fitrianah, 2018) dengan tujuan mengelompokkan produk yang dijual pada online shop menjadi beberapa kluster untuk mengetahui produk mana yang paling diminati sehingga jumlah stok harus banyak untuk produk tersebut. Masalah yang diangkat pada penelitian ini adalah menumpuknya stok sepatu yang belum terjual dan kurang diperhatikannya reseller yang memiliki potensial dalam penjualan. penelitian ini menggunakan metode *K-Means*. *K-Means* adalah metode yang cocok digunakan dengan satu tipe data dan ditunjang dengan validitas kluster yang menggunakan Davies-Bouldin Index untuk mengetahui seberapa baik proses kinerja pengklusteran. Hasil pengklusteran data untuk mendapatkan informasi reseller yang potensial, dengan menggunakan dua kluster, sedangkan untuk mendapatkan informasi tipe sepatu yang laris adalah dengan menggunakan tiga cluster. Dan pengklusteran ini menghasilkan 1 reseller yang potensial yaitu reseller Hoy, sedangkan proses mencari tipe sepatu yang laris mendapat 3 tipe sepatu yang terlaris yaitu Neo Advantage, Cloudfoam Swiftt, dan Bennasi.

Penelitian selanjutnya dilakukan oleh (Maddinsyah and Nabila, 2018) dengan tujuan guna mengetahui apakah citra merek, word of mouth dan makna merek merupakan variabel pembentuk ekuitas merek. Ke-empat variable tersebut dilakukan uji hipotesis dimana setiap variabel diukur dengan menggunakan instrument variabel tersebut menggunakan bantuan SPSS dengan menyebarkan kuisisioner pada 30 responden. Dan untuk mengetahui keterhubungan antara variable yang satu dengan variable yang lainnya digunakan model penuh persamaan SEM dengan menggunakan program LISREL 8.80 diperoleh dua model diagram lintasan, yaitu model *standardized* dan model *t-values*. Hasil perhitungan dengan menggunakan SPSS dan analisis SEM kepada variabel citra merek, *Word of mouth*, dan *Brand meaning* (makna merek) secara bersama-sama didapatkan bahwa ketiga variable tersebut berpengaruh secara signifikan terhadap *Brand equity* (ekuitas merek) pada J.CO *Donuts & Coffee*, dengan t hitung sebesar 10.47 dan koefisien regresi sebesar 0,69.

Tabel 2. 1. Penelitian Terdahulu

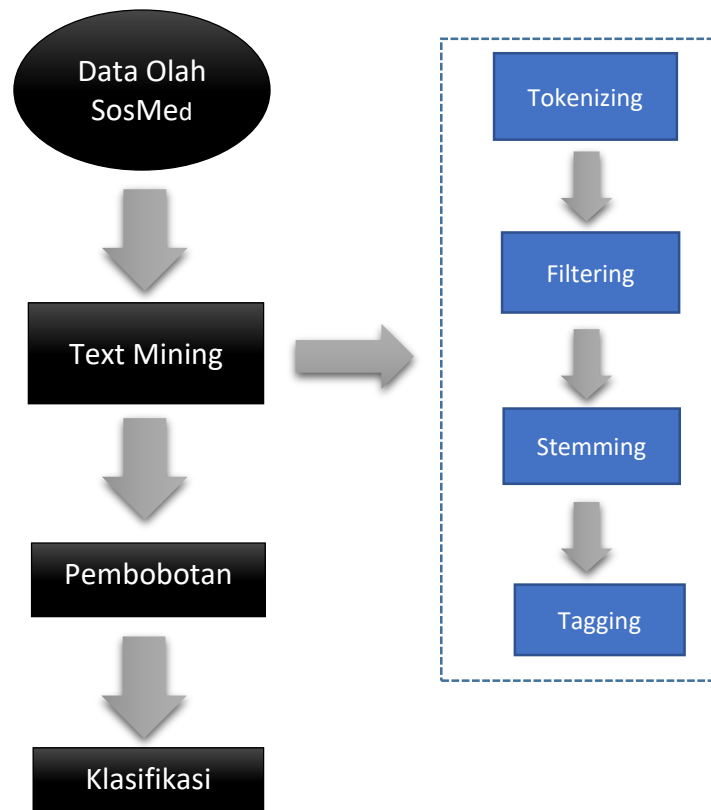
No	Judul Penelitian	Tujuan	Metode	Hasil	Keterbatasan Penelitian
1.	Pencitraan Merek, Penetapan Harga dan Inovasi Produk Pengaruhnya Terhadap Loyalitas Konsumen dalam Membeli Mobil Toyota Avanza di Kota Manado.	Untuk mengetahui tingkat loyalitas konsumen pada PT. Hasjrat Abadi Cabang Manado. Dalam pembelian toyota avanza berdasarkan citra merek, harga dan inovasi.		Citra Merek, Harga, dan Inovasi Produk berpengaruh positif dan signifikan terhadap Loyalitas konsumen pada Mobil Toyota Avanza di PT. Hasjrat Abadi Cabang Manado.	
2.	Penerapan algoritma <i>Term frequency-inverse document</i> (TF-IDF) untuk text mining.	Penelitian ini dilakukan agar perusahaan dapat membuat strategi komunikasi di media sosial yang dapat memengaruhi persepsi konsumen dan niat beli, terutama melalui akun resmi merek di LINE.		Hasil dari penelitian ini memiliki implikasi bagi perusahaan yang berinvestasi di media sosial yaitu meningkatkan pengetahuan bagi manajer pemasaran ketika merencanakan kampanye dan mengembangkan strategi untuk beriklan di media sosial yang memengaruhi persepsi konsumen dan niat beli.	Penelitian ini didominasi oleh responden yang tidak mewakili semua generasi usia, mayoritas responden ini berasal dari generasi millennial dengan usia 17-23 tahun, sehingga untuk penelitian lebih lanjut peneliti dapat menyebarkan kuesioner lebih merata secara usia agar mendapatkan hasil yang lebih beragam dan mewakili semua generasi usia.
3.	Klasifikasi Berita <i>Online</i> dengan menggunakan pembobotan <i>TF-IDF</i> dan <i>Cosine Similarity</i> .	Meminimalkan <i>human error</i> serta mengurangi terjadinya kesalahan pengkategorian pada saat memposting berita.	Metode tf-idf dan cosine similarity.	Dari pengujian I, II, III, dan IV diketahui bahwa semakin banyak jumlah data uji maka semakin tinggi tingkat akurasi. hasil rata-rata pengujian 91.25% dan mempunyai tingkat akurasi paling tinggi pada pengujian ke 3 dengan akurasi 100 %. Hal tersebut dapat terjadi karena hasil pada kategori data hasil sistem sesuai dengan data asli.	Hasil yang didapatkan kurang sesuai dengan penjabaran yang ada. Dimana dari hasil pengujian 4 dengan pembagian data latih 60% dan data uji 40% Didapatkan hasil akurasi sebesar 95% lebih kecil dibandingkan dengan pengujian 3 dengan pembagian data uji sebanyak 30% dan hasil akurasi 100%.

No	Judul Penelitian	Tujuan	Metode	Hasil	Keterbatasan Penelitian
4.	Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing	Mengelompokkan produk yang dijual pada online shop menjadi beberapa kluster untuk mengetahui produk mana yang paling diminati sehingga jumlah stok harus banyak untuk produk tersebut.	Algoritma K-Means.	Dapat mengelompokkan paket penjualan dan diskon dimana jika membeli produk yang banyak diminati dengan yang tidak diminati bersamaan akan mendapatkan diskon. didapatkan informasi reseller potensial dengan proses klustering.	Masih terbatasnya data untuk dilakukan pengelompokkan seperti pada data reseller hanya terdapat 1 reseller potensial untuk melakukan promosi agar penjualan pada hoyweapstore mengalami kenaikan.
5	Analisis pembentukan ekuitas merek berdasarkan citra merek, <i>word of mouth</i> dan makna merek (studi empiris pada J.Co donuts & coffee)	Mengetahui apakah citra merek, <i>word of mouth</i> dan makna merek merupakan variabel pembentuk pada ekuitas merek.	Metode yang digunakan adalah analisis faktor, dan SEM (<i>Structural Equation Modeling</i>).	Citra merek, <i>word of mouth</i> , dan brand meaning secara bersama sama berpengaruh signifikan terhadap brand equity.	Perlu adanya penambahan variable baru seperti konten pengguna secara online, dan penambahan penelitian mengenai ekuitas merek berdasarkan produk yang ada.

METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Metodologi pada penelitian ini terdiri dari beberapa tahapan yang dilakukan berupa: Pengolahan data, text mining dengan beberapa proses : tokenizing, filtering, stemming dan tagging, kemudian dilakukan pembobotan dan yang terakhir adalah klasifikasi. Seperti terlihat pada gambar berikut:



Gambar 3. 1. Metodologi Penelitian

Tahapan pertama adalah melakukan proses pengolahan data di sosial media untuk pengumpulan data secara otomatis terhadap semua kata yang berhubungan atau mengandung kata sebuah “merek”. Pengumpulan kata yang dilakukan pada tahapan olah data sosmed dapat berupa kalimat dengan karakter yang mengandung huruf besar atau huruf kecil, proses perubahan semua karakter pada kalimat tersebut menjadi huruf kecil dan tidak ada huruf besar sama sekali dilakukan pada tahap kedua yaitu tahapan *tokenizing*.

Tahapan selanjutnya adalah proses penyaringan atau *filtering* pada kata-kata yang dirubah menjadi huruf kecil. Langkah ini dilakukan dengan mengambil kata-kata penting pada

token dan melakukan penghapusan karakter angka, link, tanda baca berupa titik dan kata-kata penghubung yang tidak di perlukan sehingga di dapat kata-kata yang memiliki makna namun memiliki imbuhan yang berbeda.

Proses perubahan terhadap kata-kata bermakna dengan imbuhan yang berbeda dilakukan pada tahapan *stemming* agar didapatkan kata dasar dengan menghapus kata penghubung atau imbuhan yang tidak di perlukan yang nantinya akan dikelompokkan kedalam *stoplist*.

Tahapan selanjutnya adalah proses *tagging*. Pada proses ini dilakukan pencarian bentuk awal dari kata-kata yang sudah dilakukan proses *stemming*.

Setelah bentuk awal pada kata dasar di dapatkan perlu dilakukan proses pembobotan menggunakan TF-IDF agar didapatkan informasi pada kumpulan dokumen dan kata apa yang sering muncul dalam kelompok kata dasar dengan format file txt berdasarkan kata kunci yang dimasukan.

Tahapan terakhir adalah tahapan pengklasifikasian data menggunakan metode *support vector machine* yang sebelumnya sudah dilakukan pembobotan terhadap kata yang sering muncul sebagai acuan untuk pembangkitan citra baru dengan kata-kata yang bermakna *positive* dan mengandung merek tertentu untuk pembangkitan citra baru.

3.2. Rencana Kegiatan

	Des-21	Jan-22	Feb-22	Mar-22	Apr-22	Mei-22	Juni-22	Juli-22	Agust-22	Sep-22	Okt-22	Nov-22	Des-22	Jan-23	Feb-23
Seminar Bidang Kajian															
Pemilihan Judul															
Penulisan Bab 1															
Critical Review															
Penulisan Bab 2															
Pengambilan Data															
Pengolahan Data															
Penulisan Bab 3															
Penulisan Bab 4															
Penulisan Bab 5															
Kualifikasi															
Penulisan Paper															
Progress Pertama															
Progress Kedua															
Sidang Tertutup															
Paper 1															
Paper 2															
Paper 3															

DAFTAR PUSTAKA

Damian Ryan, C. J. (2009) *Understanding Digital Marketing: Marketing Strategies for Engaging The Digital Generation*. Edited by Prodi MBTI. London: Kogan Page.

datareportal.com (2020) 'Indonesia Digital report 2020', *Global Digital Insights*, p. 43. Available at: <https://datareportal.com/reports/digital-2020-indonesia?rq=digital-2020-indonesia>.

Efraim Turban, David King, Dennis Viehland, J. L. (2006) *Elrctronik commerce: A managerial perspective*. 180th–183rd edn.

Evans, D. (2008) *Social media marketing: An hour a day*. Indianapolis: Wiley publishing.

Fournier, Susan; Avery, J. (2015) 'Chapter 14 Managing Brands by Managing Brand Relationships', (JANUARY 2011), pp. 1–23.

Gesenhues, A. (2013) *Survey: 90% Of Customers Say Buying Decisions Are Influenced By Online Reviews*. Available at: <https://martech.org/survey-customers-more-frustrated-by-how-long-it-takes-to-resolve-a-customer-service-issue-than-the-resolution/>.

Hennig-Thurau, T. *et al.* (2004) 'Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet?', *Journal of Interactive Marketing*, 18(1), pp. 38–52. doi: 10.1002/dir.10073.

Hermaren, V. and Achyar, A. (2018) 'The effect of firm created content and user generated content evaluation on customer-based brand equity', *INOBIIS: Jurnal Inovasi Bisnis dan Manajemen Indonesia*, 2(1), pp. 86–100. doi: 10.31842/jurnal-inobis.v2i1.63.

Herwijayanti, B., Ratnawati, D. E. and Muflikhah, L. (2018) 'Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity', *Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(1), pp. 306–312. Available at: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/796>.

Huang, A. (2008) 'Similarity measures for text document clustering', *New Zealand Computer Science Research Student Conference, NZCSRSC 2008 - Proceedings*, (April), pp. 49–56.

Kaplan, A. M. and Haenlein, M. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media', *Business Horizons*, 53(1), pp. 59–68. doi: 10.1016/j.bushor.2009.09.003.

Laudon, K. and Laudon, J. (2013) 'Management Information Systems: International Edition, 12/E', *Pearson*, pp. 1–677. Available at: [papers2://publication/uuid/ACC7D75A-6055-4BEA-A385-F48A96400319](https://publication/uuid/ACC7D75A-6055-4BEA-A385-F48A96400319).

Maddinsyah, A. and Nabila, N. I. (2018) 'Analisis Pembentukan Ekuitas Merek Berdasarkan Citra Merek, Word of Mouth Dan Makna Merek (Studi Empiris Pada J.Co Donuts & Coffee)', *Inovasi*, 5(2), p. 1. doi: 10.32493/inovasi.v5i2.y2018.p1-18.

Mudambi, S. M., Schuff, D. and Schuff, D. (2016) 'Quarterly What Makes a Helpful Online Reviews A Study of Customer Review ? on Amazon . com1', 34(1), pp. 185–200.

Murphy, R. (2020) 'Local Consumer Review Survey 2020'. Available at:

<https://www.brightlocal.com/research/local-consumer-review-survey/>.

Mustonen, P. (2009) *Social Media a New Way To Success*. Available at: http://info.tse.fi/julkaisut/kr/Kre1_2009.pdf.

Nurjannah, M. and Fitri Astuti, I. (2013) 'PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING Mahasiswa S1 Program Studi Ilmu Komputer FMIPA Universitas Mulawarman Dosen Program Studi Ilmu Komputer FMIPA Universitas Mulawarman', *Jurnal Informatika Mulawarman*, 8(3), pp. 110–113.

Osborne-Gowey, J. (2014) 'What is Social Media?', *Fisheries*, 39(2), p. 55. doi: 10.1080/03632415.2014.876883.

Purbo, O. W. (2019) *Text Mining - Analisis MedSos, Kekuatan Brand & Intelejen di Internet*. Yogyakarta: Andi Publisher.

Rahmati (2009) *Pemanfaatan E-commerce Dalam Bisnis di Indonesia, Citozcome.Blogspot.Com*. Available at: <http://citozcome.blogspot.com/2009/05/pemanfaatan-e-commerce-dalambisnis-di.html>.

Ronen Feldman, J. S. (2007) *The text mining handbook: advanced approaches in analyzing unstructured data*, *Choice Reviews Online*. doi: 10.5860/choice.44-5684.

Roring, B. C., Mekel, P. A. and Soegoto, A. S. (2014) 'Pencitraan Merek, Penetapan Harga dan Inovasi Produk Pengaruhnya Terhadap Loyalitas Konsumen dalam Membeli Mobil Toyota Avanza di Kota Manado', *Jurnal EMBA*, 2(3), pp. 1225–1236.

Triyansyah, D. and Fitriana, D. (2018) 'Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing', *Jurnal Telekomunikasi dan Komputer*, 8(3), p. 163. doi: 10.22441/incomtech.v8i3.4174.

Wibawa, A. P. *et al.* (2018) 'Metode-metode Klasifikasi', *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, 3(1), pp. 134–138.

Zakiyuddin, H. (2021) 'Penerapan Algoritma Cosine Similarity Dan Pembobotan TF-IDF Pada System Penerimaan Mahasiswa Baru Universitas Bumigoram Mataram', pp. 1–7.