

## BAB II

### KAJIAN LITERATUR

#### 2.1. Tinjauan Pustaka

Bagian ini akan berisi landasan teori-teori yang digunakan dan pekerjaan yang sudah dilakukan peneliti sebelumnya untuk mendukung penyelesaian penelitian yang akan dilakukan.

##### 2.1.1. Analisis Sentimen

Analisis sentimen adalah cabang dari pemrosesan bahasa alami atau *natural language processing* (NLP) yang melibatkan analisis data teks untuk menentukan sentimen yang diekspresikan oleh penulis. Proses ini melibatkan pengklasifikasian teks ke dalam kategori sentimen positif, negatif, atau netral berdasarkan penggunaan bahasa dan fitur kontekstual lainnya. Analisis sentimen telah digunakan dalam berbagai aplikasi, termasuk riset pasar, pemantauan media sosial, dan analisis umpan balik pelanggan [1].

Ada beberapa pendekatan analisis sentimen yang dapat digunakan. Pendekatan analisis sentimen secara umum dibagi menjadi dua, yaitu [22]:

##### 1. *Lexicon-Based*

*Lexicon-Based* atau Pendekatan berbasis leksikon adalah metode *unsupervised learning* yang mengklasifikasikan opini pengguna menjadi positif, negatif, atau netral. Ada tiga tipe pendekatan: manual, berbasis kamus, dan berbasis korpus. Pendekatan manual melibatkan pengkodean leksikon secara manual, sedangkan berbasis kamus menggunakan basis data leksikal. Berbasis korpus mengidentifikasi kata sentimen dalam konteks dan bisa berupa pendekatan statistik atau semantik. Meskipun tidak perlu pelatihan sebelumnya dan berguna untuk menambang data, pendekatan ini bisa membutuhkan upaya manual dan biaya ekstra.

##### 2. *Machine Learning*.

Pendekatan *Machine Learning* dalam analisis sentimen menggunakan algoritme klasifikasi untuk melatih dokumen berlabel pada korpus guna mengklasifikasikan sentimen. Ada tiga jenis pendekatan: *supervised*, *semi-supervised*, dan *unsupervised*. *Supervised* paling sesuai untuk analisis sentimen karena tugasnya adalah klasifikasi. Pendekatan *Machine Learning* memerlukan ekstraksi fitur dan pelatihan pengklasifikasi.

Dua set data berbeda diperlukan untuk pelatihan dan pengujian algoritma. Konstruksi fitur dan pengklasifikasi penting untuk klasifikasi sentimen yang akurat. Proses pembelajaran mesin melibatkan langkah-langkah seperti pemilihan data pelatihan, pembersihan data, teknik pelabelan, pemilihan vektor fitur, dan algoritme klasifikasi.

### 2.1.2. Media Sosial dalam Interaksi Konsumen dengan Merek

Media Sosial merupakan tempat dimana penggunanya dimungkinkan untuk melakukan kegiatan dalam bertukar informasi [23], yang membuat terjadinya penciptaan dan pertukaran konten buatan pengguna atau dalam istilah *user generated content*. Media sosial yang paling banyak mendapat perhatian dari para akademisi dan manajer komunikasi adalah Facebook, Twitter, dan YouTube. Media sosial mempercepat komunikasi antara konsumen dan memungkinkan perilaku konsumen yang lebih proaktif dalam mencari informasi produk dan merek yang diinginkan [24]. Salah satu dari beberapa kegiatan bertukar informasi antar pengguna adalah menemukan informasi yang diinginkan mengenai produk dan juga menyampaikan umpan balik mengenai evaluasi suatu produk. Umpan balik tersebut berisi skenario penggunaan, *best practice*, kualitas produk, hingga penggunaan harian [25]. Ini merupakan data dan pengetahuan yang memiliki nilai yang tinggi bagi perusahaan yang dapat digunakan untuk meningkatkan kualitas produk.

*Brand* atau merek adalah identitas yang melekat pada produk atau jasa suatu perusahaan yang membedakannya dari produk atau jasa sejenis yang ditawarkan oleh perusahaan lain. Merek dapat didefinisikan sebagai kumpulan persepsi yang terbentuk dalam pikiran konsumen mengenai suatu produk atau layanan. Menurut definisi ini, merek mencakup semua hal yang diketahui atau diharapkan oleh konsumen dari produk atau layanan tersebut. [26]. Meskipun pelanggan dapat membentuk hubungan emosional yang kuat dengan merek tertentu yang dapat meningkatkan penjualan, menurunkan sensitivitas harga, dan mengurangi biaya pemasaran, kenyataannya adalah bahwa dalam dunia bisnis yang sangat kompetitif, banyak pelanggan menganggap merek sebagai sesuatu yang sama dan biasa saja [27]. Dalam hal ini dapat diketahui bahwa keterlibatan konsumen terhadap merek sangat jelas, di mana interaksi, promosi dari mulut ke mulut secara elektronik atau eWOM (*electronic Word Of Mouth*) [28], dan tren merupakan elemen kunci yang secara langsung memengaruhi keterlibatan merek konsumen, serta memperkuat kesadaran dan pengetahuan merek [29].

### 2.1.3. Industri Elektronik dan Pengembangan Produk baru

Industri elektronik menjadi salah satu industri terbesar di dunia yang mengalami perkembangan pesat. Persaingan di dalam industri ini semakin ketat dan membutuhkan strategi perusahaan yang tepat untuk memenangkan persaingan dan menarik perhatian konsumen. Dalam persaingan industri elektronik, perusahaan-perusahaan berlomba-lomba mengembangkan produk baru yang inovatif dan unik. NPD menjadi strategi penting dalam memenangkan persaingan dan mempertahankan posisi di pasar. Dalam NPD, perusahaan harus memperhatikan kebutuhan dan keinginan konsumen, serta menyesuaikan dengan perkembangan teknologi yang terus berkembang [30].

Perusahaan harus mempunyai strategi yang tepat dalam memenangkan persaingan. Salah satu strategi yang dapat diterapkan adalah mengembangkan atau menambah inovasi baru yang berbeda dengan produk lainnya. Keterlibatan *supplier* dan integrasi dengan konsumen merupakan salah satu kunci dalam kesuksesan terhadap proses NPD [31]. Analisis lebih lanjut tentang NPD menunjukkan bahwa dengan sumber daya yang berharga dan langka, perusahaan elektronik Taiwan mampu mencapai kesuksesan dengan meningkatkan atribut produk yang ada, dengan menambahkan atribut baru, dengan mengganti komponen baru, atau dengan mengubah konfigurasinya. Dengan demikian, mereka dapat meminimalkan tingkat kegagalan produk baru mereka, mengurangi biaya pengembangan model baru, mengurangi waktu tunggu dalam membawanya ke pasar, mengurangi ketidakpastian produk baru bagi pelanggan, dan bahkan memperkuat strategi mereka terkait para pesaingnya [32]. Dengan demikian, perusahaan dapat memenangkan persaingan dan memperoleh keuntungan yang optimal di pasar.

#### 2.1.4. *New Product Development* (NPD)

*New Development Product* (NPD) atau pengembangan produk baru merupakan salah satu strategi bisnis yang umum dilakukan oleh perusahaan dalam rangka meningkatkan penjualan dan mempertahankan pangsa pasar yang ada [13]. Agar NPD menjadi efektif, diperlukan adanya kerjasama dengan departemen manufaktur, *engineering*, penelitian dan pengembangan (R&D), pemasaran, keuangan, dan pembelian. Departemen pemasaran terlebih dahulu harus mengevaluasi produk baru, dan kemudian tim lintas fungsi dibentuk untuk produk baru terjun langsung ke lapangan [33].

Proses NPD memiliki sebanyak delapan tahap, delapan tahap yang ditunjukkan pada Gambar 2.1 adalah [33]:

### 1. *Idea Generation*

Tahap pertama dalam NPD adalah menghasilkan ide untuk produk baru. Ide dapat berasal dari berbagai sumber seperti departemen Litbang, pelanggan, pesaing, seminar, universitas, investor, dll.

### 2. *Idea Evaluation*

Pada tahap ini, semua ide yang dihasilkan disaring dan dievaluasi untuk memilih ide yang paling berguna. Ide disaring untuk membedakan ide yang lebih berguna dari ide yang kurang berguna.

### 3. *Concept Development*

Ide produk yang paling berguna yang dipilih pada Tahap 2 dikembangkan menjadi konsep produk. Konsep produk ini kemudian diuji dengan kelompok pelanggan terpilih untuk mendapatkan umpan balik. Konsep produk dengan nilai terbaik dipilih sebagai produk baru yang akan dikembangkan.

### 4. *Marketing Strategy*

Strategi pemasaran dibuat untuk konsep yang terpilih. Strategi pemasaran dibuat dalam tiga langkah: mengidentifikasi target pasar, menentukan harga, dan mengembangkan strategi bauran pemasaran.

### 5. *Business Analysis*

Analisis bisnis dibuat dalam dua langkah: memproyeksikan penjualan produk baru dan memproyeksikan biaya dan keuntungan. Penjualan diproyeksikan melalui riset pasar dan menganalisis penjualan produk serupa di masa lalu. Semua biaya yang terlibat dalam NPD diperkirakan, termasuk biaya investasi, operasi, pemasaran, dan R&D. Jika proyeksi tersebut sesuai dengan tujuan bisnis, proses pengembangan dilanjutkan ke tahap berikutnya.

### 6. *Product Development*

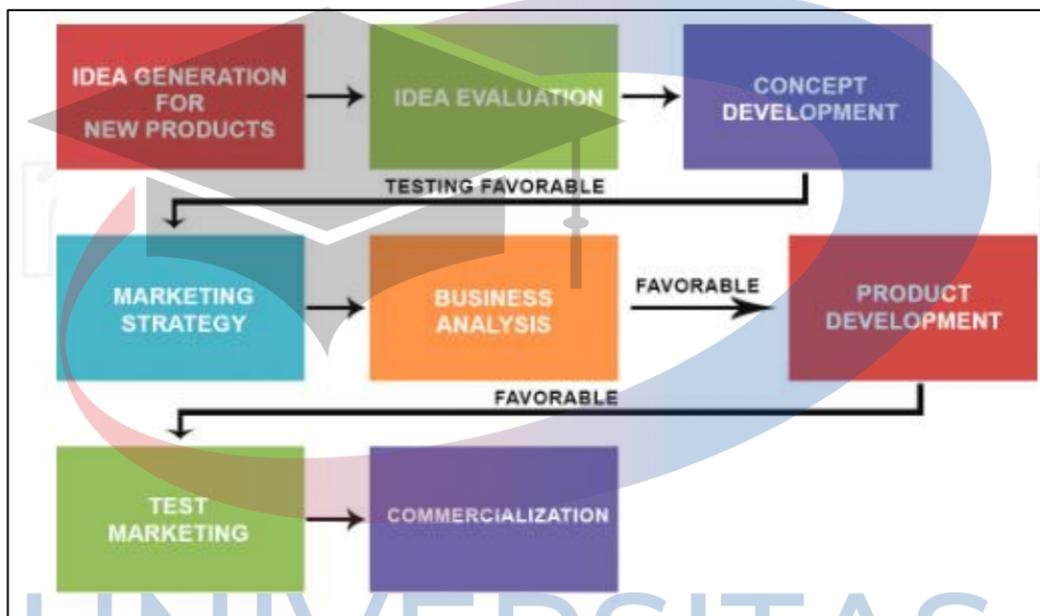
Sampel produk baru dibuat dan diuji oleh departemen Litbang. Beberapa pengujian dilakukan untuk memastikan keamanan, daya tarik, dan efektivitas konsep produk baru.

### 7. *Test Marketing*

Uji pemasaran dilakukan untuk mengidentifikasi strategi pemasaran terbaik untuk konsep produk baru. Semua elemen pemasaran seperti target pasar, iklan, distribusi, dan pengemasan diuji. Lulus uji pemasaran memastikan strategi pemasaran yang sesuai untuk konsep produk baru yang akan dikomersialkan pada tahap berikutnya.

## 8. Commercialization

Tahap terakhir adalah memperkenalkan produk baru ke pasar. Waktu dan skala peluncuran produk ditentukan. Sebagian besar bisnis lebih memilih untuk memperkenalkan produk baru ke pasar dalam skala kecil atau menengah dan memperluas pasar secara bertahap.



Gambar 2. 1 Proses NPD

Keterlibatan konsumen dalam NPD juga menjadi salah satu aspek penting yang harus dipertimbangkan oleh perusahaan. Melibatkan konsumen terjadi pada saat proses yang pertama, yaitu *idea generation for new product* atau menghasilkan ide baru [14]. Konsumen dapat memberikan informasi tentang kebutuhan mereka yang mencerminkan keinginan dan harapan mereka, serta informasi solusi yang diwujudkan dalam saran perbaikan konkret atau ide solusi[15].

### 2.1.5. Keterlibatan Konsumen dalam NPD

Dalam proses NPD khususnya pada fase menghasilkan ide produk baru (*idea generation*), pengumpulan masukan dan umpan balik konsumen sangat penting untuk mendapatkan informasi yang akurat dan relevan terkait preferensi konsumen terhadap produk yang akan dikembangkan. Melibatkan konsumen pada tahap awal pengembangan

produk dapat membantu perusahaan mengidentifikasi desain utama dan fitur fungsional suatu produk dari sudut pandang konsumen [34].

Dalam konteks pengembangan produk baru (NPD), terdapat beberapa model yang mendukung inovasi pada tahap awal yang disebut dengan *Idea Generation*. Penggunaan model-model ini memiliki nilai penting dalam pengambilan keputusan yang efisien dan ekonomis [33]:

#### a. *Product-Service System (PSS)*

*Product-service system* adalah sebuah model bisnis yang menggabungkan pengembangan produk baru dengan layanan terkait, yang melibatkan berbagai pemain, infrastruktur, dan produk untuk memenuhi kebutuhan pelanggan, bersaing di pasar, dan mengurangi dampak lingkungan dibandingkan dengan model bisnis tradisional [33].

#### b. *The Kano Model*

Model Kano adalah kerangka kerja untuk memahami kepuasan pelanggan terhadap suatu produk atau layanan. Model ini menyediakan lima jenis atribut produk [35]:

1. *Must-be*: Ini adalah persyaratan dasar yang diharapkan oleh pelanggan dari suatu produk atau layanan. Jika atribut *must-be* ini tidak terpenuhi, pelanggan akan merasa tidak puas.
2. *One-dimensional*: Atribut-atribut ini berbanding lurus dengan kepuasan pelanggan. Semakin banyak atribut yang dimiliki suatu produk, pelanggan akan semakin puas.
3. *Attractive*: Atribut ini tidak diperlukan untuk kepuasan dasar, tetapi jika ada, mereka dapat menyenangkan pelanggan dan meningkatkan loyalitas mereka.
4. *Indifferent*: Atribut ini tidak mempengaruhi kepuasan pelanggan, baik secara positif maupun negatif. Pelanggan cenderung netral terhadap atribut-atribut ini.
5. *Reverse*: Atribut-atribut ini sebenarnya dapat menurunkan kepuasan pelanggan jika ada. Jika atribut-atribut ini dimiliki oleh produk, pelanggan akan merasa tidak puas.

Model Kano membantu bisnis dalam memprioritaskan atribut-atribut mana yang harus difokuskan untuk meningkatkan kepuasan pelanggan dan menghadirkan produk atau layanan yang lebih unggul. Dengan memahami kebutuhan dan preferensi pelanggan dalam setiap jenis atribut, bisnis dapat mengoptimalkan strategi pemasaran dan pengembangan produk mereka.

#### c. *Conjoint Analysis*

Analisis Konjoin (*Conjoint Analysis/CA*) adalah teknik yang membantu para pemasar dalam menciptakan produk baru berdasarkan preferensi dan kebutuhan pelanggan. Teknik ini menggunakan survei untuk mengumpulkan data dari pelanggan mengenai atribut atau fitur-fitur suatu produk, seperti berat, jenis laser, atau kapasitas tinta untuk

printer. Selanjutnya, data ini diolah untuk menghasilkan ide-ide produk baru yang dapat bersaing di pasar. Analisis Konjoint membantu produsen dalam mendapatkan gambaran lengkap tentang suatu ide dengan pemahaman yang mendalam [36].

#### d. *The Product Value Matrix*

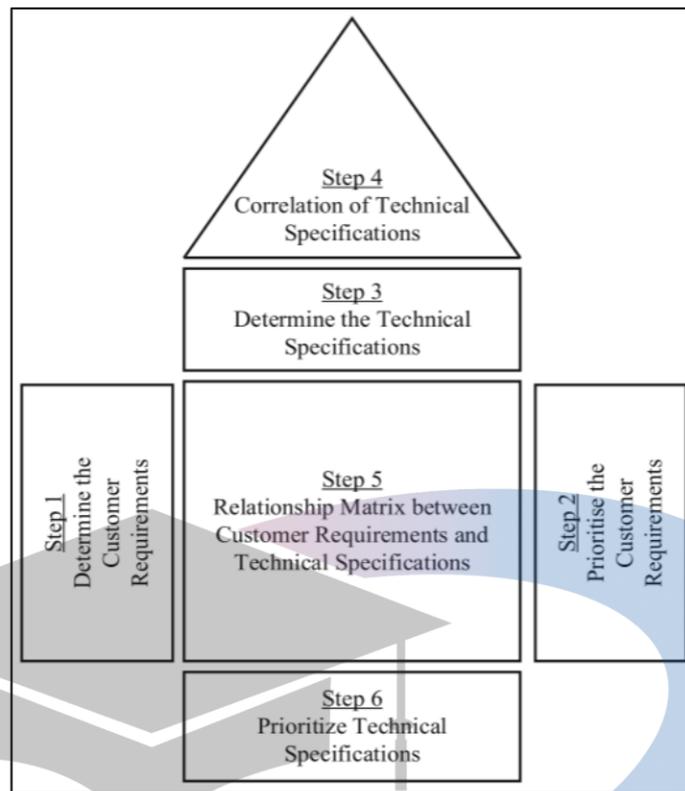
Product value matrix adalah sebuah model yang digunakan untuk menghasilkan spesifikasi kebutuhan pasar atau *market requirement specification* (MRS) untuk produk baru dan mengatur prioritas tugas pengembangan produk. Hal ini memastikan bahwa semua pihak yang terlibat dalam rantai pasok produk baru, seperti pelanggan, pemasok, pengecer, dan lainnya, memiliki kebutuhan yang terpenuhi. Dalam model ini, diasumsikan bahwa tim penjualan dan pemasaran memiliki pemahaman mendalam mengenai semua pihak yang terlibat dan kebutuhan mereka dalam rantai pasok produk tersebut [33]

#### e. *Quality Function Deployment (QFD)*

QFD (*Quality Function Deployment*) adalah pendekatan terstruktur untuk pengembangan produk yang berfokus pada kebutuhan atau persyaratan pelanggan dan menggunakannya untuk memandu seluruh proses pengembangan produk. QFD merupakan pendekatan berbasis pelanggan yang membantu bisnis untuk merancang dan mengembangkan produk yang memenuhi atau melebihi harapan pelanggan [37]. QFD melibatkan langkah-langkah berikut [37]:

1. *Identifying Customer Needs* (Identifikasi Kebutuhan Pelanggan)
2. *Determination of Customer Needs Priorities* (Menentukan Prioritas Kebutuhan Pelanggan).
3. *Determination of Technical Specifications* (Menentukan Spesifikasi Teknis)
4. *Correlation of Technical Specifications* (Korelasi Spesifikasi Teknis)
5. *Relationship Matrix Between Customer Needs and Technical Specifications* (Matriks Hubungan antara Kebutuhan Pelanggan dan Spesifikasi Teknis)
6. *Technical Specifications Priorities* (Prioritas Spesifikasi Teknis).

Tahapan diatas akan digabung ke dalam alat yang dinamakan dengan HOQ atau singkatan dari *House Of Quality* yang dapat dilihat pada Gambar 2.2.



Gambar 2. 2 *Quality Function Deployment House of Quality*

### 2.1.6. Integrasi Analisis Sentimen dengan NPD

Manfaat penerapan analisis sentimen yang digunakan untuk kebutuhan NPD telah dibuktikan dari beberapa penelitian sebelumnya. Dengan memahami pola emosi konsumen yang diperoleh dari platform media sosial Twitter, penelitian yang dilakukan oleh Giannakis dapat membantu perusahaan untuk mengambil keputusan yang berhubungan dengan desain produk mobil [38]. Gambar 2.3 merupakan hasil dari penelitian tersebut yang berfokus pada empat aspek berbeda yang terkait dengan *Style/Appearance* (gaya penampilan), *Performance* (kinerja), *Configuration* (konfigurasi), dan *Service* (layanan produk) sebagai pendekatan *sensemaking-pattern* yang ditandai dengan frekuensi tinggi dalam konten tweet yang dikumpulkan.

First-level indicator	Second-level indicator	Frequency	Second-level score	First-level score	First-level proportion	First-level weight
B1 Style/appearance	C1 Colour	1086	0.78	0.31	0.28	0.08
	C2 Size	928	0.02			
	C3 Shape	876	0.14			
B2 Performance	C4 Speed	2026	-0.36	-0.04	0.33	-0.01
	C5 Engine	563	-0.23			
	C6 Displacement	731	0.46			
B3 Configuration	C7 Sunroof	602	-0.24	0.32	0.18	0.06
	C8 Gps	448	0.85			
	C9 Seat comfort	798	0.35			
B4 Service	C10 Price	1356	-0.54	-0.33	0.21	-0.07
	C11 Maintenance	374	-0.11			
	C12 Cleanliness	422	-0.34			

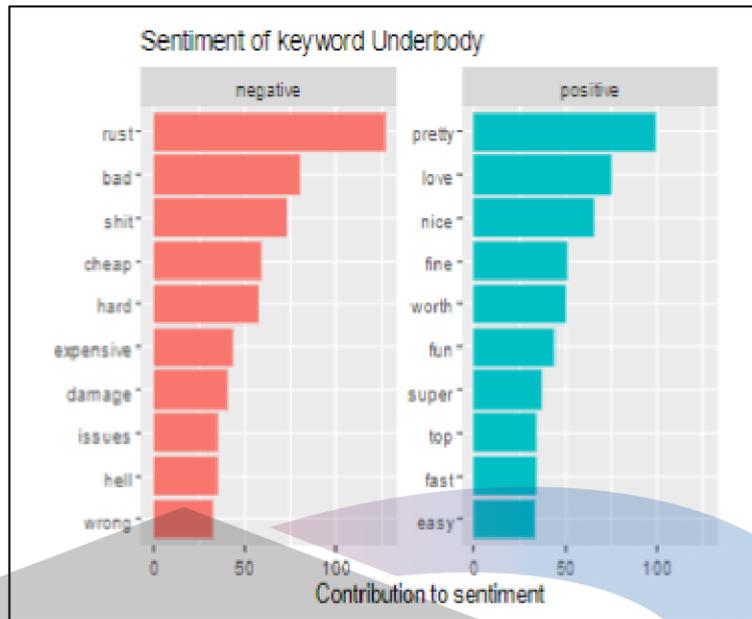
Gambar 2. 3 Tabel Skor Indikator Primer dan Sekunder di Tesla Model X [38]

*User-generated content* dari Twitter dapat digunakan sebagai data alternatif yang diperoleh dari proses survei untuk kebutuhan NPD, seperti yang dilakukan dalam penelitian oleh Rathore [39]. Kemudian, analisis sentimen dilakukan terhadap produk pada waktu sebelum peluncuran dan setelah peluncuran, menghasilkan informasi ekspektasi dan pengalaman konsumen secara *real-time*. Penelitian tersebut menghasilkan tabel ringkasan yang dapat dilihat pada Gambar 2.4 mengenai perbandingan skor sentimen terhadap 3 produk pada waktu sebelum rilis dan setelah rilis.

Domino's Exotic Italian Pizza		Hyundai Creta		Samsung Galaxy S6	
Pre-launch	Post-launch	Pre-launch	Post-launch	Pre-launch	Post-launch
Expectations for new pizza in terms of customization and delivery parameters. Trust and joy are dominant emotions.	Product is not able to fulfil the expectations. Still have trust on Domino's brand and looking for better product.	Overall positive emotions because of existing product performance. Joy is dominant emotion.	Joy and trust are dominant emotions. Positive emotion for design and appearance	Anticipating creative features such as wireless charging, and appearance. Highly negative about poor battery life	Negative due to unavailability of product accessories such as cases. Sad about less creative features as promised before launch.
Unhappy with franchise distribution and non response to queries from Domino's itself.	Anger and disgust are dominant emotions as there is no unique taste.	Anticipating new creative product with better features.	Few technical concepts are still not properly implemented such as beam light, etc.	High sad emotion due to not matching with competitive products (ex. front camera). Trust is higher.	Surprise emotion with review for good quality design.
Looking health awareness topics from Domino's.		Concerns about few technical features such as ergonomics of interior.	Looking for latest technological update in product parts such as brakes		Negative emotions towards offers and discounts.  Joyful emotion for rear camera and look.

Gambar 2. 4 Ringkasan Perbandingan dari Ketiga Produk pada saat Pra-peluncuran dan Pasca-peluncuran. [39]

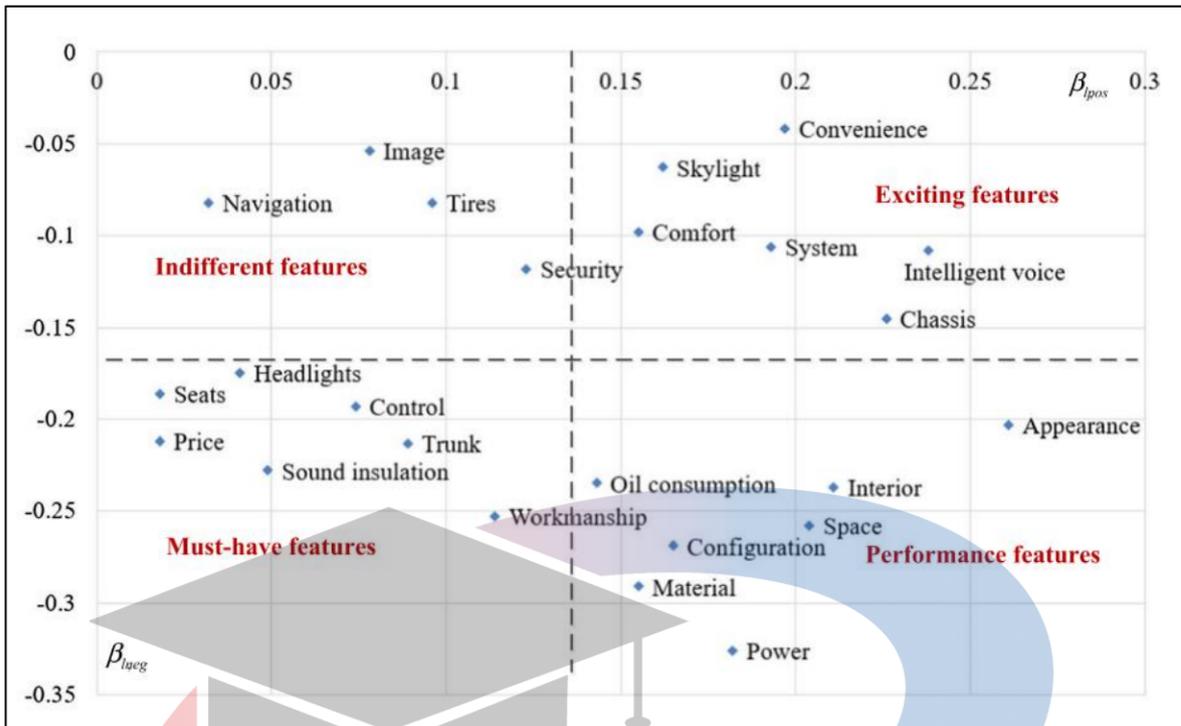
Penelitian oleh Sutrilastyo membahas mengenai peran dari *text mining* dan analisis sentimen pada pengembangan dan desain secara khusus untuk tahap penyaringan ide, kemudian menghasilkan kata sentimen negatif dan sentimen positif yang diurutkan mulai dari yang terbesar hingga terkecil secara khusus untuk bagian produk tertentu yaitu bagian bawah *body* mobil. Dapat dilihat pada Gambar 2.5 bahwa kata sentimen negatif yang paling banyak adalah *rust* (karat) menunjukkan kekhawatiran mengenai daya tahan bagian bawah *body* mobil terhadap karat dan kerusakan [16].



Gambar 2. 5 Sentimen dari Kata Kunci "underbody" [16]

Penelitian yang dilakukan oleh Wang mengusulkan pendekatan IOCM (*inconsistent ordered choice model*) Model, yaitu model yang bertujuan untuk mengukur preferensi pelanggan sehubungan dengan fitur produk dengan mempertimbangkan inkonsistensi antara peringkat produk numerik dan ulasan produk tekstual. *Kano Model* digunakan untuk mengklasifikasikan fitur produk ke dalam empat kategori berdasarkan evaluasi parameter, yaitu *Indifferent Features*, *Exciting Features*, *Must-Have Features*, dan *Performance Features*. Evaluasi parameter membantu mengalokasikan fitur produk ke dalam kategori ini, yang dapat memandu pengembangan produk dengan mengidentifikasi fitur mana yang penting, mana yang perlu ditingkatkan, dan mana yang dapat ditambahkan untuk meningkatkan kepuasan pelanggan [40].

Dapat dilihat pada Gambar 2.6 menunjukkan bahwa pelanggan cenderung mentolerir desain *Exciting Features* saat ini bahkan jika belum sepenuhnya sesuai dengan kebutuhan mereka. Hal ini terjadi karena fitur yang menarik memberikan peluang beragam untuk strategi diferensiasi. Bagi produsen yang mengambil pendekatan diferensiasi, perlu diberikan perhatian khusus pada peningkatan *Exciting Features* karena memiliki potensi untuk memberikan kepuasan kepada pelanggan dan meningkatkan kepuasan mereka.



Gambar 2. 6 Klasifikasi Prioritas Fitur Produk dengan Model IOCM [40]

### 2.1.7. Text Mining

*Text mining* merupakan sebuah proses ekstraksi informasi berkualitas tinggi dari teks yang dapat berupa data terstruktur seperti RDBMS, semi-terstruktur seperti XML dan JSON, dan teks tidak terstruktur seperti dokumen kata, video, dan gambar. *Text mining* melibatkan banyak topik dan algoritma untuk menganalisis teks, seperti pencarian informasi, pemrosesan bahasa alami, pemrosesan data, dan pembelajaran mesin. Teknologi ini memiliki berbagai aplikasi di berbagai domain web dan ilmu biomedis [41].

Dalam *text mining*, data yang dikumpulkan adalah teks atau kata-kata yang diungkapkan oleh konsumen melalui media sosial, forum online, atau platform komunikasi lainnya. Metode ini memungkinkan perusahaan untuk menganalisis dan menggali informasi dari jumlah data yang besar secara efisien dan lebih akurat tanpa terpengaruh oleh bias pengambilan sampel [42].

Teknik yang digunakan dalam *text mining* adalah [42] :

1. **Ekstraksi Informasi:** Identifikasi frasa kunci dan hubungan dalam teks melalui pencocokan pola dan analisis semantik, dan dimasukkan ke dalam database. Cocok untuk teks dalam bentuk dokumen berbahasa alami yang tidak terstruktur.

2. **Kategorisasi:** Menetapkan satu atau lebih kategori untuk dokumen teks dengan menggunakan teknik klasifikasi statistik, seperti Naïve Bayesian, Nearest Neighbour, Decision Tree, dan Support Vector Machines.
3. **Clustering:** Menemukan kelompok dokumen dengan konten yang sama melalui teknik clustering seperti K-means untuk membuat vektor topik untuk setiap dokumen dan mengukur bobot seberapa baik dokumen cocok dengan setiap cluster.
4. **Visualisasi:** Mempermudah penambangan teks melalui metode visualisasi, seperti menggunakan bendera teks untuk menunjukkan kategori dokumen dan warna kepadatan untuk menggambarkan kelompok dokumen.
5. **Summarization:** Mengurangi panjang dan detail dokumen sambil mempertahankan poin-poin terpenting dan makna umum. Membantu menentukan apakah dokumen pantas dibaca secara menyeluruh atau tidak dan layak digantikan dengan sekumpulan dokumen yang diringkas.

#### 2.1.8. *Web Scraping*

*Web scraping* adalah proses mengumpulkan data dalam jumlah besar dari web dan menyimpannya dalam basis data untuk analisis di masa depan dan digunakan di kemudian hari. Teknik ini dapat dianggap sebagai metode untuk mengambil atau mengekstrak konten dari situs web untuk digunakan pada tujuan di luar kendali pemilik situs web [43].

Penelitian ini memilih *Web Scraping* dikarenakan penulis belum menemukan alat yang ketersediaannya sama dengan pengambilan data dari Twitter yang sudah banyak dipakai oleh beberapa peneliti sebelumnya seperti penelitian yang dilakukan oleh Rathore dengan Twitter API [39] atau dengan library python tweepy oleh Giannakis [44]. Proses *Web Scraping* memanfaatkan *crawler* yang melakukan permintaan AJAX YouTube untuk mengambil komentar. *Crawler* membuat sesi dengan agen pengguna, mengirimkan permintaan AJAX ke server menggunakan ID video tertentu. Semua komentar diunduh dan disimpan dalam format JSON, diidentifikasi dengan ID unik [45].

#### 2.1.9. *Text Preprocessing*

*Text Preprocessing* atau preproses data merupakan tahap yang dilakukan terhadap data yang awalnya tidak terstruktur menjadi data yang terstruktur sehingga proses pembobotan dapat dilakukan [41]. Mengurangi dimensi model ruang vektor dapat mempercepat proses klasifikasi. Preproses bertujuan untuk menghapus kata yang tidak relevan dengan penelitian, memperbarui bentuk kata, dan mengurangi jumlah kata [46]. Penelitian ini melibatkan beberapa tahap preproses, yaitu membersihkan data,

menyesuaikan huruf, membagi kata menjadi token, menghilangkan kata stop, dan menghilangkan awalan atau akhiran kata. Tahapan dalam preproses data adalah sebagai berikut [47]:

1. *Data Cleaning*: Pembersihan data pada pengambilan data dari komentar YouTube dimulai dengan menghapus tanda baca, simbol, angka, URL, dan tag HTML. Proses pembersihan dimulai dari pengambilan data mentah dan dilanjutkan dengan menghilangkan segala hal yang tidak dibutuhkan seperti tanda baca, simbol, dan angka agar data dapat diolah dengan lebih baik.
2. *Case Folding*: *Case folding* dilakukan dengan cara membuat semua huruf besar pada dataset menjadi huruf kecil. Dengan membuat semua kata menjadi huruf kecil maka akan sangat membantu untuk melakukan generalisasi.
3. Tokenisasi: Tokenisasi adalah proses yang dilakukan untuk memecah kalimat menjadi potongan-potongan kata, tanda baca, dan ekspresi lain yang bermakna sesuai dengan ketentuan bahasa yang digunakan. Peneliti menggunakan fungsi *word\_tokenize* yang disediakan oleh *library* NLTK pada proses ini.
4. Penghapusan *Stopwords*: Proses penghapusan *Stopwords* dilakukan untuk menghilangkan kata-kata yang tidak bermakna dan terdapat pada daftar *Stopword*, sehingga mengurangi dimensi data pada proses klasifikasi.
5. *Stemming*: *Stemming* adalah proses yang dilakukan untuk mengubah kata-kata dengan imbuhan ke dalam bentuk dasarnya dengan cara menghilangkan imbuhan seperti awalan, akhiran, dan konfiks. Pada tahap ini, stemming dilakukan menggunakan *library* NLTK. *Stemming* sangat penting dalam teknik pemrosesan bahasa alami untuk pencarian informasi yang efektif dan efisien, serta diterima secara luas oleh pengguna. *Stemming* juga dapat digunakan untuk mendukung proses kategorisasi/klasifikasi dan clustering.

#### **2.1.10. Pemilihan Algoritma Klasifikasi untuk Pelabelan Sentimen**

Bagian ini akan membahas mengenai penetapan pendekatan untuk proses pelabelan sentimen terhadap data teks komentar YouTube. Dalam konteks pelabelan sentimen, terdapat beberapa pilihan metode klasifikasi yang dapat diambil, masing-masing memiliki kelebihan dan pertimbangan yang perlu dipertimbangkan. Terdapat dua pendekatan yang dapat digunakan untuk melakukan klasifikasi sentimen, yaitu *Lexicon Based* dan *Machine Learning* menurut Yusof [22].

Pada penelitian yang dilakukan oleh Nguyen yang membandingkan pendekatan *Lexicon Based* dan *Machine Learning*, menemukan bahwa klasifikasi dengan pendekatan *Machine Learning* mengungguli klasifikasi berbasis leksikon di semua metrik saat mengklasifikasikan ulasan konsumen Amazon [48]. Dalam hal ini, *Machine Learning* diandalkan untuk melakukan klasifikasi sentimen pada penelitian ini. Namun, *Machine Learning* memiliki 4 jenis *Machine Learning*, antara lain *Supervised Learning*, *Unsupervised Learning*, *Reinforcement Learning*, dan *Semi-Supervised Learning* [49]. Masing-masing dijelaskan pada Tabel 2.1 berikut:

Tabel 2. 1 Perbedaan Jenis Pembelajaran Mesin[49]

Jenis Pembelajaran	Definisi	Tujuan	Penerapan	Contoh Algoritma
<i>Supervised Learning</i>	Model menggunakan data yang dilabeli untuk belajar dan membuat prediksi baru.	Digunakan untuk masalah klasifikasi (mengelompokkan data ke dalam kategori tertentu) atau regresi (memprediksi nilai kontinu dari data)	Klasifikasi (misalnya, spam email), Regresi	<i>Decision Trees, Naive Bayes, SVM, Random Forest, KNN</i>
<i>Unsupervised Learning</i>	Model mengidentifikasi pola dan struktur dalam data yang tidak memiliki label.	Digunakan untuk masalah klustering (membuat kelompok-kelompok dari data yang mirip) atau reduksi dimensi (mengurangi kompleksitas data dengan menghilangkan fitur-fitur yang kurang penting)	Clustering, Reduksi Dimensi	<i>K-Means, PCA, Apriori Algorithm</i>
<i>Reinforcement Learning</i>	Model pembelajaran yang mengambil tindakan dalam lingkungan untuk mencapai tujuan tertentu.	Digunakan untuk masalah yang melibatkan interaksi dengan lingkungan dinamis dan tujuan jangka panjang	Pengendalian robot, Permainan video	<i>Q-Learning, Deep Q Networks</i>

<i>Semi-Supervised Learning</i>	Model yang bekerja dengan campuran data dilabeli dan tidak dilabeli.	Mengetahui penggabungan data berlabel dan tidak berlabel yang kemungkinan dapat mengurangi kinerja pembelajaran dan merancang algoritma yang memanfaatkan kombinasi tersebut	Pengenalan wajah, Klasifikasi dokumen	<i>Self-Training, Expectation-Maximization</i>
---------------------------------	--	--	---------------------------------------	--

Berkaitan dengan tujuan penelitian ini, maka jenis pembelajaran yang dipilih adalah *Supervised Learning*. Terdapat beberapa contoh algoritma untuk jenis pembelajaran *Supervised Learning*, yaitu *Decision Trees*, *Naive Bayes*, *SVM*, *Random Forest*, dan *KNN* [49]. Hasil perbandingan yang dilakukan oleh Abdualgalil terhadap kelima algoritma yang melakukan deteksi diabetes, ditampilkan pada Tabel 2.2 berikut :

Tabel 2. 2 Perbandingan Antara Decision Tree, Gaussian NB, SVM, RF[49]

Algoritma	Accuracy	F1-Score	Precision	Training time(sec.)	Prediction Time(Sec.)
KNN	80.52	66.66	69.76	0.0009	0.0030
DT	79.87	56.47	63.15	0.0030	0.0010
SVM	81.818	67.56	74.34	2.0845	0.0009
RF	80.52	66.66	69.76	0.0210	0.0020
NB	79.22	64.44	67.44	0.0030	0.0010

Dalam tabel di atas, terlihat bahwa Algoritma SVM menunjukkan kinerja terbaik berdasarkan kolom *Accuracy*. Kolom ini mencerminkan persentase contoh yang diprediksi dengan benar dibandingkan dengan total contoh yang diprediksi, memberikan indikasi sejauh mana prediksi yang diberikan adalah benar. Hal ini menunjukkan hasil yang sama dengan penelitian yang dilakukan oleh Kaur yang membandingkan algoritma *Decision Trees*, *Naive Bayes*, *SVM*, dan *Random Forest* yang dianggap sebagai algoritma yang paling populer dalam klasifikasi sentimen [50] dan merupakan algoritma jenis *Supervised Learning* yang menyimpulkan bahwa pengklasifikasi SVM sebagai pengklasifikasi yang paling akurat dan paling lengkap dalam melakukan klasifikasi sentimen dari ulasan produk di Amazon, dengan nilai yang tinggi untuk metrik: presisi, akurasi, dan skor F1 yang dapat dilihat pada tabel berikut :

Tabel 2. 3 Perbandingan Algoritma Supervised Learning oleh Kaur [50]

	<b>F1-Score</b>	<b>Precision</b>	<b>Re call</b>	<b>Accuracy</b>
<b>SVM</b>	0.82	0.86	0.81	0.82
<b>RF</b>	0.79	0.8	0.79	0.8
<b>NB</b>	0.81	0.85	0.83	0.81
<b>DT</b>	0.8	0.81	0.79	0.8

Pada bagian ini, untuk pemilihan algoritma klasifikasi sentimen pada penelitian diputuskan menggunakan *Support Vector Machine* (SVM). Pertimbangan ini didasarkan pada penilaian dari tingkat akurasi yang dihasilkan oleh masing-masing algoritma. Yang dimana algoritma SVM memiliki akurasi yang paling tinggi diantara algoritma yang lainnya.

## 2.2. Penelitian Sebelumnya

Penelitian ini mengacu pada beberapa penelitian sebelumnya yang membahas analisis sentiment yang digunakan untuk NDP.

Tabel 2. 4 Penelitian Analisis Sentimen untuk NDP terdahulu

No	Nama Peneliti, Judul, Tahun	Metode	Dataset yang digunakan	Temuan terhadap NDP
1.	Giannakis M, <i>Social media and sensemaking patterns in new product development: demystifying the customer sentiment</i> , 2022	1. Pengumpulan data dengan <i>Tweepy</i> 2. Analisis sentimen berbasis leksikon dibantu dengan teori <i>fuzzy</i> 3. Identifikasi data dengan pendekatan <i>sensemaking pattern</i> . 4. Penerapan hasil dengan Skor Indikator	Data komentar mengenai karakteristik Tesla Model X dari platform Twitter	Menggunakan <i>Indicator Score</i> untuk membantu mengukur dan mengevaluasi sentimen pelanggan terhadap fitur spesifik suatu produk selama proses pengembangan produk baru (NPD).

2.	Rathore A, <i>Pre- and post-launch emotions in new product development: Insights from twitter analytics of three products, 2020</i>	<ol style="list-style-type: none"> <li>1. Pengambilan data dengan <i>R script</i></li> <li>2. Analisis sentimen berbasis Leksikon</li> <li>3. Menghitung frekuensi kata-kata yang termasuk sentimen</li> </ol>	Data sebanyak 302.623 <i>tweet</i> dikumpulkan sebelum dan sesudah peluncuran tiga produk.	Tabel perbandingan umpan balik konsumen terhadap tiga produk dari sebelum rilis ke setelah rilis
3.	Sutrilastyo S, <i>Lexicon-based Sentiment Analysis for Product Design and Development, 2021</i>	<ol style="list-style-type: none"> <li>1. Pengambilan data dengan <i>web-scraping</i></li> <li>2. Analisis sentimen dengan <i>Naïve Bayes</i> diteruskan pada kombinasi <i>N-gram</i> dan Leksikon</li> <li>3. Meringkas hasil dengan perbandingan sebelum rilis dan setelah rilis</li> </ol>	Data teks yang sudah digabung dari berbagai sumber ( Facebook, Twitter, Kaskus, Forum My Car, dan Reddit)	Statistik data yang menampilkan jumlah frekuensi kata yang termasuk dalam sentimen negatif dan positif sebagai kriteria desain

UNIVERSITAS  
MIKROSKIL

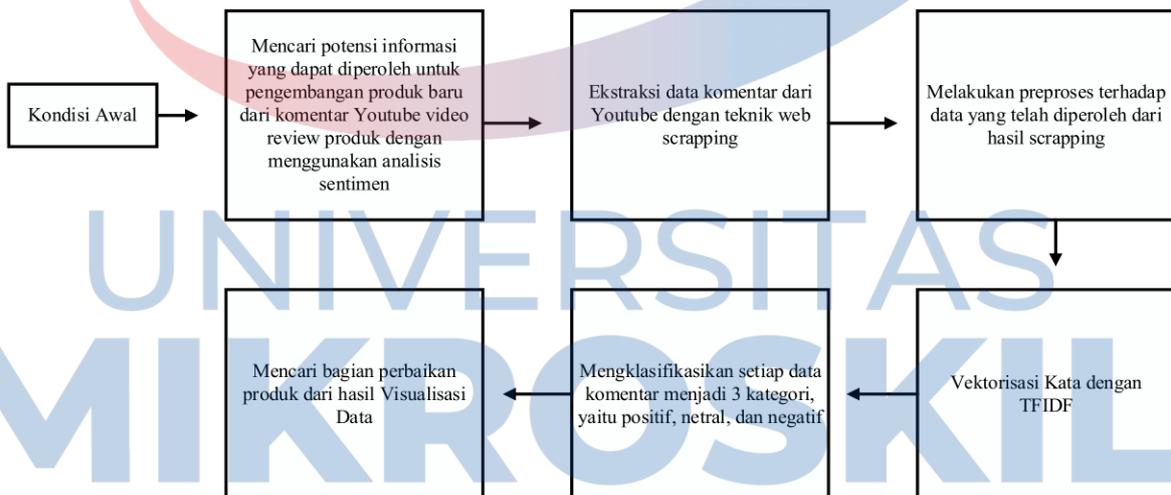
4.	Ng C, <i>Assessing public opinions of products through sentiment analysis: Product satisfaction assessment by sentiment analysis</i> , 2021	1. Pengambilan data dengan teknik <i>crawling</i> . 2. Klasifikasi sentimen dengan <i>k-means</i> . 3. Menentukan prioritas fitur berdasarkan jumlah skor sentimen.	Facebook Fanpage	Pendekatan yang dibuat menawarkan desainer produk sebuah tool yang memungkinkan untuk memutuskan mana yang perlu diperbaiki
5.	Ireland R, <i>Application of data analytics for product design: Sentiment analysis of online product reviews</i> , 2018	<i>Multiple key NLP techniques</i>	Data ulasan produk dari Amazon	Pendekatan yang dibuat menghasilkan informasi bagi desainer mengenai tingkat kepentingan suatu fitur produk.

Dari penelitian diatas, dapat dikemukakan bahwa para peneliti telah berusaha untuk mencoba mengolah umpan balik pelanggan terhadap produk menjadi masukan bagi perusahaan yang membuat produk. Peneliti telah berhasil mengumpulkan dan menganalisis data teks dari berbagai sumber dan kemudian mengolahnya hasil analisis sentimennya menjadi informasi masukan sebagai kebutuhan pelanggan. Namun, salah satu aspek yang kurang dibahas adalah bentuk interpretasi hasil analisis sentimen yang dapat digunakan untuk tahap *idea generation* dalam pengembangan produk baru [33]. Tahap ini merupakan tahap yang penting untuk menghasilkan masukan yang dapat dimanfaatkan oleh pembuat produk. Namun, peneliti hanya melakukan analisis sentimen secara umum tanpa memberikan penjelasan atau rekomendasi tentang bagaimana hasil analisis sentimen dapat diolah menjadi informasi yang berguna yang dapat digunakan oleh pembuat produk. Oleh karena itu, dalam konteks ini, penulis mengusulkan eksplorasi lebih lanjut tentang cara-cara interpretatif yang sesuai untuk tahap *Idea Generation*, serta penerapan hasil analisis

sentimen dalam salah satu kerangka pendekatan yang digunakan dalam tahap *Idea Generation*[33].

### 2.3. Kerangka Konsep

Penelitian ini bertujuan untuk menganalisis sentimen pelanggan terhadap produk *smartphone* berdasarkan fitur-fitur yang diungkapkan dalam komentar video ulasan produk di platform media sosial YouTube menggunakan bantuan pelabelan sentimen berbasis pembelajaran mesin Support Vector Machine (SVM). Kerangka konsep penelitian terdiri dari tiga tahap utama. Tahap pertama adalah pengumpulan data komentar pengguna dengan teknik *web scrapping*. Tahap kedua adalah penerapan teknik TF-IDF untuk vektorisasi kata-kata setiap kalimat agar dapat diproses oleh SVM. Tahap ketiga adalah penggunaan SVM untuk mengklasifikasikan setiap ulasan menjadi tiga kategori sentimen, yaitu positif, negatif, atau netral. Hasil analisis sentimen ini dapat digunakan sebagai masukan bagi perusahaan untuk memperbaiki fitur-fitur produk *smartphone* dan meningkatkan kepuasan pelanggan.



Gambar 2. 7 Kerangka Konsep