

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Sebelumnya

Penelitian yang dilakukan untuk menganalisis suatu produk telah banyak dilakukan. Tidak sedikit pula diantaranya mampu memberikan hasil analisis yang akurat. Dari banyaknya penelitian yang sudah dilakukan sebelumnya menunjukkan bahwa dengan menganalisis suatu produk dapat memberikan manfaat dalam membantu perusahaan untuk meningkatkan kualitas layanan yaitu dengan analisis sentimen. Berikut penelitian terdahulu yang menurut penulis mempunyai keterkaitan dengan penelitian yang akan dilakukan.

Pertama, penelitian yang dilakukan oleh Saleh Hasan Badjrie dkk pada tahun 2021 [12]. Penelitian tersebut dilakukan untuk menganalisis *review* produk *IndiHome* dan *First Media*. Melalui model CNN pada penelitian tersebut didapatkan bahwa hasil analisis pada *review* produk *IndiHome* dan *First Media* mendapatkan nilai *validation accuracy* sebesar 98.4% dan 91.1%. Kemudian, untuk hasil dari *confolution matriks* pada tiap-tiap *provider* didapatkan hasil rata-rata nilai *accuracy*, *precission*, *recall*, dan *f1-score* diatas 90%, dimana hasil tersebut membuktikan bahwa algoritma ini bekerja dengan efektif.

Kedua, penelitian yang dilakukan oleh Fany Alifian Irawan dan Dwi Anindyani Rochmah pada tahun 2022 [9]. Pada penelitian ini dilakukan suatu proses analisis untuk memperoleh hasil terhadap kebijakan vaksinasi Covid-19 melalui ulasan di aplikasi twitter. Dalam penelitian itu, didapatkan bahwa evaluasi dengan metode *K-Fold Cross Validation* memperoleh nilai akurasi training terbaik yaitu 100% dan akurasi testing terbaik yaitu 99.61%.

Ketiga, penelitian yang dilakukan oleh Sukma Nindi Listyarini dan Dimas Aryo Anggoro pada tahun 2021 [13]. Penelitian tersebut dilakukan untuk menganalisis. Dalam penelitian ini, penulis menggunakan metode CNN dengan data bersumber dari twitter yang kemudian dilakukan proses *crawling* lalu melakukan pelabelan pada data dan dilakukan tahap preprocessing yang selanjutnya dilakukan tahap klasifikasi dengan CNN. Hasil penelitian didapatkan hasil model terbaik dengan CNN sebesar 90% dengan 4-layer *convolutional* dan 100 *epoch*.

Keempat, penelitian yang dilakukan oleh I Made Adi Susilayasa, AAIN Eka Karyawati, Luh Gede Astuti, Luh Arida Ayu Rahning Putri, I Gede Arta Wibawa, I Komang Ari Mogi pada tahun 2022 [14]. Penelitian tersebut dilakukan untuk mengukur performa terbaik yang dilakukan oleh algoritma CNN dalam mengklasifikasi sentimen berdasarkan kategori yang diteliti yaitu kategori *tops*, kategori *bottoms*, kategori *jackets*, kategori *dresses*, dan *intimates* pada *e-commerce* melalui data yang ada pada kaggle. Penelitian menunjukkan hasil penggunaan algoritma CNN dalam melakukan klasifikasi sentimen pada kategori *bottoms* didapatkan hasil akurasi validasi sebesar 82,36% dengan *hyperparameter* yang digunakan yaitu 16 untuk *batch size* dan 0.001 *learning rate*, serta dengan 10 *epoch*. Pada kategori *dresses* diperoleh hasil terbaik sebesar 81,5% dengan kombinasi *hyperparameter* adalah 16 untuk *batch size* dan 0.001 untuk *learning rate*, serta *epochs* 5. Sedangkan kategori *tops* diperoleh hasil terbaik sebesar 80,7% dengan kombinasi *hyperparameter batch size* adalah 16 dan *learning rate* nya 0.001 serta *epochs* 10.

Kelima, penelitian oleh Erwin Yudi Hidayat dan Devioletta Handayani pada tahun 2022 [15]. Penelitian tersebut dilakukan untuk menganalisis sentimen terhadap ulasan produk *skincare* dengan data diperoleh pada website *Female Daily Review*. Dari penelitian tersebut ditemukan bahwa hasil analisis yang dilakukan berdasarkan 30 eksperimen didapatkan *best* model terdapat pada eksperimen ke-12 dengan akurasi 80.22%.

Tabel 2. 1 Penelitian Sebelumnya

No	Judul	Comparing	Contrasting	Criticize	Synthesize	Summary
1	ANALISIS SENTIMEN <i>REVIEW CUSTOMER</i> TERHADAP PRODUK <i>INDIHOME</i> DAN <i>FIRST MEDIA</i> MENGGUNAKAN ALGORITMA <i>CONVOLUTIONAL NEURAL NETWORK</i>	Penelitian dilakukan untuk menganalisis <i>review</i> customer pengguna <i>indihome</i> dan <i>first media</i> pada data twitter dengan model CNN untuk menganalisisnya.	Penelitian dilakukan dengan algoritma CNN yang dimulai dari tahap pengambilan data, <i>preprocessing</i> data, <i>labelling</i> data, hingga visualisasi data.	Penelitian ini menggunakan model CNN didapatkan visualisasi data untuk tiap-tiap provider memiliki tingkat spesifikasi yang berbeda. Dimana, diantara kedua <i>provider</i> tersebut sentimen terhadap harga dengan data positif sama-sama sedikit.	Pada penelitian ini menggunakan data hasil <i>review</i> pelanggan produk <i>IndiHome</i> dan <i>First Media</i>	Pada penelitian ini, metode algoritma CNN menghasilkan nilai <i>validation accuracy</i> untuk layanan <i>IndiHome</i> sebesar 98.4% dan 91.1% untuk layanan <i>First Media</i> . Dimana pada tiap-tiap <i>provider</i> didapatkan hasil rata-rata nilai <i>accuracy</i> , <i>precision</i> , <i>recall</i> , dan <i>f1-score</i> diatas 90%.
2.	Penerapan Algoritma	Penelitian dilakukan	Penelitian	Penelitian ini	Pada penelitian ini, data	Penelitian dengan

	<p>CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19</p>	<p>guna memperoleh hasil sentiment dari masyarakat mengenai kebijakan vaksinasi covid-19 dengan algoritma CNN</p>	<p>dilakukan dengan algoritma CNN data yang digunakan dilakukan tahap <i>preprocessing</i> terlebih dahulu kemudian dilanjutkan dengan tahap ekstraksi data, dan proses pengujian. Tahap terakhir yaitu evaluasi dengan menggunakan <i>precision, recall, f1-score</i>, dan <i>accuracy</i>.</p>	<p>menghasilkan perbandingan pengujian antara model <i>convolutional neural network</i> dengan Lapisan <i>Global Max Pooling</i> dan pengujian dengan algoritma <i>Naïve Bayes</i>. Perbandingan tersebut menunjukkan bahwa nilai akurasi algoritma CNN lebih besar dibandingkan nilai akurasi <i>Naive Bayes</i>. Berdasarkan hasil tersebut</p>	<p>yang digunakan adalah data hasil <i>crawling</i> pada aplikasi twitter dengan menggunakan API twitter pada rentang waktu bulang Januari 2021 dengan jumlah komentar sebanyak 1424 komentar.</p>	<p>menggunakan model CNN ini didapatkan hasil akurasi sebesar 98.66%, <i>precision</i> 98.33%, <i>recall</i>, 98.33%, dan <i>f1-score</i> 98.66%. Pengujian <i>K- Fold Cross Validation</i> dengan Layer <i>Global Max Pooling</i> juga memperkuat hasil dari penelitian ini.</p>
--	--	---	--	---	--	---

				didapatkan respon positif mengenai kebijakan vaksinasi.		
3.	Analisis Sentimen Pilkada di Tengah Pandemi Covid-19 Menggunakan <i>Convolution Neural Network (CNN)</i>	Penelitian ini dilakukan untuk memperoleh hasil analisis sentimen terhadap pilkada di tengah pandemi covid-19 dengan menggunakan algoritma CNN	Penelitian dilakukan dengan pendekatan <i>deep learning</i> menggunakan CNN. Metode ini diawali dengan <i>crawling</i> data, pemberian label dataset, <i>pre-processing</i> , dan yang terakhir dilakukan klasifikasi.	Penelitian ini menghasilkan akurasi tertinggi yaitu 90% yang didapatkan dari model dengan 4- <i>layer convolutional</i> dan 100 <i>epoch</i> . Empat <i>layer</i> ini mengalami peningkatan akurasi sejalan dengan meningkatnya jumlah <i>epoch</i> yang digunakan. <i>Epoch</i> 50, 75, hingga 100	Pada penelitian ini, data yang digunakan berasal dari twitter yang dilakukan dengan Bahasa pemrograman python dan <i>library tweepy</i> yang dihubungkan dengan twitter API. Data yang diperoleh sebanyak 500 tweet.	Penelitian ini didapatkan hasil terbaik terdapat pada <i>layer</i> ke-4 dengan jumlah <i>epoch</i> 100 dengan nilai akurasi yaitu sebesar 90%

				mengalami peningkatan akurasi sekitar 2-4%.		
4.	Analisis Sentimen Ulasan <i>E-Commerce</i> Pakaian Berdasarkan Kategori dengan Algoritma <i>Convolutional Neural Network</i>	Penelitian ini dilakukan untuk menganalisa sentimen berlandaskan kategori pada dataset menggunakan metode CNN.	Terdapat beberapa tahapan dalam penelitian ini yaitu pengumpulan data berupa data sekunder, tahap <i>pre-processing</i> untuk mengamati bahwa data yang digunakan benar-benar bersih, tahap <i>word embedding</i> , dan yang terakhir adalah pengolahan dataset dengan CNN.	Hasil dari penelitian ini menunjukkan kombinasi <i>hyperparameter</i> terbaik dengan 5 <i>epoch</i> , <i>learning rate</i> 0,001, dan <i>batch size</i> 64, akurasi validasi 88,1%. Dari ketiga jenis tersebut didapatkan hasil akurasi terbaik ada pada kategori bottoms dengan nilai akurasi 82.36%	Pada penelitian ini data yang digunakan adalah data yang diperoleh dari website <i>kaggle</i> dengan jumlah 23.486 data dengan 10 data <i>feature</i> , tetapi yang digunakan hanya 3 data <i>feature</i> saja.	Hasil pengujian model untuk data <i>testing</i> menunjukkan performa yang baik. Hasil akurasi dari model tersebut yaitu 88%. Dengan masing-masing nilai <i>precision</i> , <i>recall</i> , <i>f1-score</i> sebesar 88%. Hasil evaluasi untuk keseluruhan juga berada pada performa yang baik dengan nilai validasi akurasi yaitu 88.1%.

5.	Analisis Sentimen Ulasan Produk Kosmetik Berdasar <i>Female Daily Review</i>	Bertujuan untuk menganalisa sentimen pada produk kosmetik dengan mengacu pada situs <i>Female Daily Review</i> menggunakan ID-CNN.	Dilakukan dengan mengekstrak data menggunakan metode web <i>scarping</i> dan dilanjutkan <i>preprocessing</i> . Tujuan dilakukan <i>preprocessing</i> adalah untuk memperbaiki data teks yang tidak terstruktur. Selanjutnya adalah dilakukan <i>labeling</i> untuk mengetahui ketahanan sentiment pada ulasan produk. <i>Tokenizing</i> perlu dilakukan untuk	Dari setiap tahapan penelitian dihasilkan data ulasan produk sebesar 11119 data, <i>preprocessing</i> 11084 data. <i>Labeling</i> menghasilkan data kelas netral sebanyak 1666, kelas positif sebanyak 8089, dan kelas negatif sebanyak 1329. Hasil <i>tokenizing</i> adalah 13446 kata. Eksperimen ini dilakukan sebanyak 30 kali dengan model terbaik dalam analisa sentimen adalah eksperimen ke-12. Apabila ditinjau dari nilai	Pada kasus ini menggunakan data <i>review</i> produk kosmetik pada website <i>Female Daily Review</i> dan data yang dihasilkan disimpan pada file csv.	Model CNN terbaik ada pada percobaan ke-12 dengan <i>learning rate</i> sebesar 0.008. Dengan model yang digunakan terdiri dari beberapa layer seperti <i>Conv1D</i> (128,3), <i>MaxPooling 1D</i> , <i>Dropout</i> (0.5), <i>Dense</i> (64), <i>Dense</i> (32), <i>Dropout</i> (0.2), <i>Dense</i> (16), <i>GlobalMaxPooling 1D</i> dan 60 epoch menghasilkan nilai akurasi yaitu 80.22%
----	--	--	--	---	--	--

			<p>menghasilkan susunan indicator integer setiap kata dan memiliki daftar <i>vocanulary</i> dataset. Cara terakhir adalah melakukan pengujian dan pengukuran model.</p>	<p>akurasi, hasil penelitian ini tidak sebaik penelitian sebelumnya, yang mungkin diakibatkan oleh <i>imbalanced</i>.</p>		
6.	<p>Opinion Mining Terhadap Pemberitaan Corona di Instagram menggunakan <i>Convolutional Neural Network</i></p>	<p>Penelitian ini bertujuan untuk mengimplementasikan <i>opinion mining</i> pada kasus Corona yang terjadi di Indonesia melalui komentar pada akun Instagram.</p>	<p>Penelitian ini dilakukan dengan <i>crawling data</i>, <i>labeling</i>, <i>preprocessing</i>, dan yang terakhir dilakukan pemodelan dengan algoritma CNN.</p>	<p>Penelitian ini menghasilkan data proses <i>crawling</i> ketiga akun Instagram yaitu @kumparancom 368 komentar dengan 160 positif, komentar negatif dan netral sebanyak 104 dan 55 data. Instragram</p>	<p>Pada penelitian ini, data yang digunakan adalah data dari ketiga akun Instagram, yaitu @kumparan.com, @cnnindonesia, dan @detikcom. Dataset tersebut diambil dari 16 unggahan dari ketiga akun dalam rentang waktu 1 bulan.</p>	<p>Penelitian dengan menggunakan CNN untuk opinion mining terhadap pemberitaan corona pada akun media sosial menunjukkan bahwa sistem berhasil melakukan proses opinion mining dalam dataset. Dataset</p>

				<p>@detikcom memiliki 347 data komentar dengan komentar positif berjumlah 188, komentar negatif berjumlah 104, dan 55 komentar bersifat netral. Instagram @cnnindonesia memiliki data komentar berjumlah 316 dengan komentar positif berjumlah 147 data, komentar negatif 105, dan 64 komentar netral.</p>		<p>yang diperoleh dari proses <i>crawling</i> mencapai 1031 data. Sistem tersebut dapat mengklasifikasikan komentar berdasarkan kategori. Presentase presisi mencapai 96% dengan akurasi hingga 85%.</p>
--	--	--	--	--	--	--

7.	Analisis Sentimen Pada Data Saran Mahasiswa Terhadap Kinerja Departemen di Perguruan Tinggi	Penelitian dilakukan untuk mengklasifikasi sentimen terhadap kinerja atau pelayanan sebuah unit kerja dengan menggunakan algoritma CNN.	Penelitian ini dilakukan dengan 2 model arsitektur yang berbeda untuk mengukur kinerja model. Arsitektur pertama dengan 2 <i>layer</i> dengan masing-masing <i>layer Max-Pooling</i> di masing-masing <i>layer (DoubleMax CNN)</i> . Sedangkan model kedua terdapat 2 lapisan <i>convolutional</i> dengan <i>layer Max-Pooling</i> di <i>layer</i> terakhir.	Hasil pengujian terbaik pada filter 5 menghasilkan <i>DoubleMax CNN</i> Sedangkan pada klasifikasi 3 yang menghasilkan <i>Simple CNN</i> memperoleh hasil terbaiknya yaitu ketika memakai filter ukuran 2 dengan nilai untuk <i>Precision</i> , <i>Recall</i> , dan <i>F1-Score</i> adalah 90%, dan hasil akurasi yaitu 91%.	Pada penelitian ini, data yang digunakan merupakan QA kinerja layanan pada unit kerja dan departemen yang ada pada Politeknik Caltex Riaudi pada tahun ajaran 2018/2019 yang bersumber dari BP3MPCR.	Hasil yang didapatkan dari model tersebut bekerja dengan sangat baik. Penggunaan algoritma CNN untuk arsitektur yang memiliki kekompleksitasan lebih kompleks memiliki performa yang baik dalam analisis.
----	---	---	--	--	--	---

	Menggunakan <i>Convolutional Neural Network</i>					
8.	Analisis Sentimen Berbasis Fitur pada Ulasan Tempat Wisata Menggunakan Metode <i>Convolutional Neural Network</i> (CNN)	Penelitian ini digunakan untuk menerapkan sentimen analisis berbasis aspek pada opini yang ada di TripAdvisor dengan CNN.	Penelitian ini dilakukan dengan cara mengumpulkan data, melakukan preprocessing, pelabelan data, ekstraksi fitur, <i>aspect based sentiment analysis</i> (ABSA), dan melakukan evaluasi.	Penelitian ini menghasilkan model CNN terbaik dengan penambahan 2 model berbeda berupa CNN+LST dan CNN+GRU. Nilai Akurasi sentiment tertinggi adalah 0,9085 dengan nilai presisi 0,9078, dan Recall 0,8915. Akurasi scenario tertinggi dihasilkan nilai 0.9569, presisi sebesar 0,9568, dan recall 0,9567.	Pada penelitian ini data yang digunakan diambil dari situs tripadvisor.co.id dan data disimpan kedalam bentuk file (.csv)	Penelitian ini memiliki akurasi sentiment 0,9085 dan akurasi aspek kategori 0,9569. Dengan arsitektur model yang terdiri dari layer embedding, layer dropout, layer convolution, layer pooling, dan layer FC. Penggunaan fitur ekstraksi pada POS Tag dan negation handling dinyatakan dapat membantu

						model dalam memperoleh sentimen,
9.	<i>Implementation of Convolutional Neural Network Algorithm in Sentiment Analysis on User Reviews of MySAPK Application</i>	Penelitian ini memiliki tujuan untuk mengimplementasikan algoritma CNN guna melakukan analisa sentimen berdasarkan <i>review</i> pengguna aplikasi MySAPK dan untuk mengetahui tingkat akurasi algoritma CNN ketika dilakukan analisis sentimen.	Penelitian ini dilakukan dengan mengumpulkan data <i>review</i> terlebih dahulu. Data tersebut kemudian dilabeli untuk diklasifikasikan menjadi sentimen positif ataupun negatif. Data selanjutnya dilakukan preprocessing. Data yang telah diproses akan dibagi menjadi data	Penelitian ini menghasilkan data <i>review</i> yaitu 2208 <i>review</i> positif dan 2176 data <i>review</i> negatif. Dari hasil training didapatkan tingkat akurasi sebesar 99% dengan 17 epoch. Berdasarkan hasil uji dengan algoritma CNN didapatkan akurasi 90% sentimen positif	Penelitian ini menggunakan data <i>review</i> konsumen pada aplikasi MySAPK dengan jumlah data yaitu 5000 data.	Penelitian mengenai implementasi algoritma CNN untuk analisis sentimen pada komentar pengguna aplikasi MySAPK berhasil dilakukan. Program ini dapat memetakan ulasan menjadi sentimen positif dan negative. Model CNN ini menggunakan 17 epoch untk memperoleh akurasi training sebesar

			pelatihan dan data uji yang akan dilakukan tahap akhir yaitu evaluasi.			99%. Pengujian CNN memperoleh akurasi sebesar 90% dengan sentiment paling positif.
--	--	--	--	--	--	--

2.2 Landasan Teori

2.2.1 *Brand Image*

Menurut Aaker dalam Tingkir dalam [16] menerangkan bahwa citra merek adalah rangkaian asosiasi pada konsumen mengenai suatu merek. Selain itu *Brand image* dapat dikatakan sebagai persepsi konsumen terhadap suatu produk tertentu. Adanya citra sebuah produk yang positif juga berpengaruh pada pemasaran dan memiliki nilai lebih bagi perusahaan dalam menarik konsumen [17]. Sebelum membeli sebuah produk, konsumen tentunya mencari informasi terkait produk yang akan dibeli atau informasi mengenai *brand image* yang bersumber dari *marketplace* melalui media sosial. *Brand image* bisa diperkuat dengan adanya ulasan atau *review* dari pengguna pada *marketplace* [18].

Dengan adanya *brand image* yang kuat mampu membuat yakin pelanggan terhadap kualitas dari produk tersebut tentunya akan memberikan informasi melalui sebuah *review*. Hal tersebut mampu menjadi pendorong kemunculan niat beli dari *customer* lain sehingga mereka memiliki kecenderungan untuk membeli produk tersebut[7].

Citra merek pada dasarnya mengacu pada kesan pertama konsumen ketika merek tersebut ditempatkan. Apabila *brand image* memiliki *influencer* dengan citra yang baik sebagai sarana untuk mempromosikan produknya, itu dapat menjadi suatu keunggulan lebih untuk menarik minat beli konsumen. Hal itu berarti seorang *influencer* dapat dipercaya. Adapun beberapa indikator tentang *brand image* seperti [19]:

1. *Influencer* yang dapat diandalkan
2. *Influencer* merupakan symbol status sosial
3. *Influencer* yang memiliki reputasi baik dan dikenal

2.2.1.1 Faktor-Faktor yang mempengaruhi *Brand Image*

Ada beberapa faktor pendukung terbentuknya *brand image* adalah sebagai berikut [20].

- 1) *Brand Identity* : Identitas yang ditampilkan pada suatu produk
- 2) *Brand Personality* : Karakter yang dihasilkan dari suatu merek sehingga konsumen memiliki kesan tersendiri terhadap produk tersebut
- 3) *Brand Association* : Kesan konsumen terhadap suatu produk yang melekat
- 4) *Brand Attitude & Behaviour* : Perilaku antara konsumen dengan suatu merek produk
- 5) *Brand Benefit & Competence* : Keunggulan yang dihasilkan oleh suatu produk sehingga memberikan manfaat bagi konsumen

2.2.2 Label Harga

Harga ialah segala sesuatu yang diberikan oleh pelanggan untuk mendapatkan keunggulan yang ditawarkan oleh bauran pemasaran perusahaan [21]. Penentuan harga dibatasi adanya permintaan, persaingan, dan biaya. Dimana harga akan bergerak fluktuasi sesuai dengan ruang gerak persaingan yang ada di pasar. Perubahan yang terjadi tidak akan melebihi dari batas harga tertinggi, pun tidak akan lebih rendah dari harga yang ditentukan oleh produsen [17].

Menurut Amilia dalam [22] terdapat beberapa indikator harga yaitu :

1. Harga yang terjangkau. Dalam membeli sebuah produk hal ini penting bagi konsumen karena tentunya mereka akan membeli suatu produk sesuai dengan budget mereka.
2. Keselarasan harga. Sejalan dengan kualitas suatu produk yang baik, konsumen tidak ada masalah dengan harga apabila harga produk tersebut mahal.
3. Perang harga. Dalam memasarkan produknya, perusahaan perlu untuk mengetahui harga dari pesaingnya dengan harapan produk yang ia jual mampu bersaing di pasaran.

4. Keselarasan harga dengan kegunaan. Harga yang mahal sekalipun tidak masalah bagi konsumen apabila produk tersebut memiliki kegunaan kegunaan yang dapat di peroleh.

2.2.3 Skincare

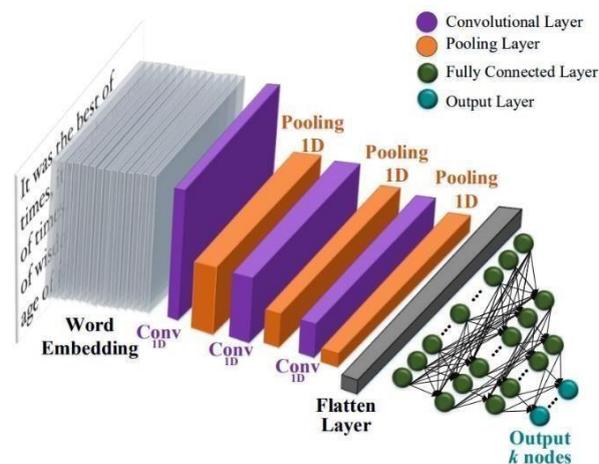
Skincare ialah kegiatan untuk merawat kulit selain badan menggunakan suatu produk tertentu. Umumnya perempuan melakukan kegiatan tersebut, meskipun laki-laki juga bisa melakukannya. Menurut *Oxford dictionary* dalam [23] *skincare* ialah penggunaan krim dan produk khusus yang digunakan untuk merawat kulit. Adapun produk kecantikan yang termasuk kedalam *skincare* seperti, (1) *facial wash*, digunakan untuk membersihkan wajah; (2) Toner, dapat digunakan untuk mengurangi minyak di wajah; (3) *Moisturizer*, digunakan sebagai pelembab dan mencegah kerusakan kulit; (4) *Sunscreen*, sebagai pelindung dari paparan sinar matahari. Beberapa tujuan utama dari penggunaan *skincare* yaitu membantu menutrisi kulit, menynda munculnya garis halus, mencegah munculnya *dark spot* dan mengurangi risiko kulit kusam.

2.2.4 Analisis Sentimen

Analisis sentimen adalah ekstraksi proses yang dilakukan pada sebuah teks. Analisis sentimen dapat menentukan sikap konsumen terhadap sesuatu [13]. Analisis sentimen memiliki inputan teks berupa sentimen positif, negatif, maupun netral [24]. Tujuan utama analisis sentimen adalah untuk menganalisis informasi dalam sebuah teks dengan berbagai metode [25].

2.2.5 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) ialah salah satu metode dari *machine learning* yang dikembangkan dari *Multi Layer Perceptron* (MLP) [26]. Algoritma ini mempunyai beberapa lapis *layer* yang saling berhubungan. CNN termasuk ke dalam jaringan *multi-layer* karena setiap *output* tersebut akan dijadikan masukan pada layer selanjutnya [27]. Algoritma ini pada umumnya digunakan untuk memproses klasifikasi gambar. Akan tetapi, algoritma CNN juga bisa diaplikasikan pada NLP (*Natural Language Processing*) dengan cara mengubah sebuah kalimat menjadi matrix [14]. Berikut disajikan gambar arsitektur CNN pada gambar 2.1.



Gambar 2. 1 Arsitektur CNN for text classification [28]

Pada CNN terdapat tiga jenis *layer* yang digunakan, pertama lapisan *convolutional*, kedua lapisan *pooling*, terakhir adalah lapisan *fully-connected*. Layer *Convolutional* merupakan layer yang memiliki filter dengan ukuran yang tetap yang dapat mengkonvolusi data [15]. Keluaran dari layer ini yaitu *feature maps*. Persamaan ini dapat dilihat pada persamaan 2.1 dibawah.

$$FM_{a,b} = bias + \sum_c^C \sum_d^D Z_{c,d} + X_{a+c-1,b+d-1} \quad (2.1)$$

Layer Pooling merupakan lapisan yang menetapkan bahwa jaringan tersebut berfokus untuk model yang paling berpengaruh, dengan merangkum keberadaan fitur dengan menggeser jendela melintasi peta fitur, selanjutnya diterapkan sebagian operasi *linear* atau *non linear* pada data yang ada. Selain itu, lapisan *pooling* memiliki kegunaan untuk mengecilkan dimensi dari peta fitur yang akan digunakan untuk lapisan berikutnya. Berikut persamaan lapisan pooling dapat dilihat pada 2.2.

$$f_h(0, FM_{a,b}) = \max(0, FM_{a,b}) = \begin{cases} FM_{a,b}, & \text{jika } FM_{a,b} \geq 0, \\ 0, & \text{jika } FM_{a,b} < 0 \end{cases} \quad (2.2)$$

Fully-Connected Layer merupakan *layer* terakhir yang berfungsi untuk memahami pola pada *layer* sebelumnya. Lapisan ini memiliki *neuron* dengan koneksi penuh hingga ke semua aktivasi pada lapisan sebelumnya. Adapun fungsi aktivasi yang digunakan yaitu *ReLU*. Sedangkan aktivasi *Softmax* digunakan untuk fungsi aktivasi output. Aktivasi ini bertujuan untuk mendapatkan hasil klasifikasi dan menghasilkan nilai yang diinterpretasikan dengan probabilitas untuk tiap kelas. Nilai kelas tersebut dihitung dengan rumus sebagai berikut :

$$y_{ijk} = \frac{e^{X_{ijk}}}{\sum_{t=1}^D e^{X_{ijt}}} \quad (2.3)$$

Loss function merupakan fungsi terakhir dengan tujuan menghitung *error value* dengan *categorical cross-entropy* dengan persamaan seperti dibawah ini.

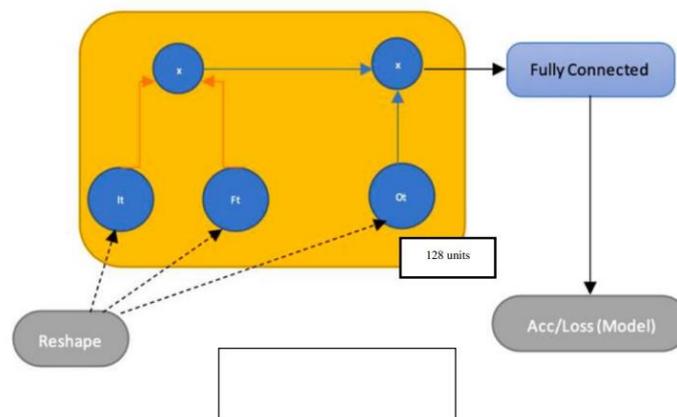
$$L_{log}(Y, Y_{pred}) = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C 1_{y_i \in C_c} \log p_{model} [y_i \in C_c] \quad (2.4)$$

2.2.6 Data Scraper

Data *scraper* merupakan *tools* untuk melakukan proses *scraping* data. *Tools* ini dapat diunduh melalui ekstensi google chrome. Hasil dari *scraping* menggunakan *tools* ini dapat disimpan kedalam file CSV maupun JSON. *Tools* ini juga memudahkan dalam pengambilan web *scraping* tanpa perlu menggunakan coding. Akan tetapi, data yang bisa di *scraping* terbatas.

2.2.7 Long Short Term Memory (LSTM)

Algoritma *Deep Learning* mampu menghasilkan performa lebih baik dan waktu komputasi yang lebih cepat dibandingkan algoritma *machine learning*. Salah satu algoritma dalam *deep learning* yang dapat digunakan untuk klasifikasi kata adalah LSTM. LSTM merupakan model algoritma yang ditingkatkan dari RNN [29]. Metode LSTM terdiri dari 3 bagian yaitu *forget gate*, merupakan proses memilah informasi yang ada pada *cell state*, informasi akan dibuang dari *cell state* jika *forget gate* bernilai 0, sebaliknya informasi akan disimpan *cell state* jika *forget gate* bernilai 1. Berikut disajikan gambar arsitektur LSTM pada gambar 2.2.



Gambar 2.2 Arsitektur LSTM [30]