

BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini didasarkan pada hasil studi-studi terdahulu yang telah dilaksanakan sebagai referensi dan analisis kompartif. Setiap penelitian menggunakan metode berbeda, yang mana mempengaruhi tingkat akurasi dan efektivitas analisis sentimen. Tabel di bawah ini memaparkan hasil penelitian terdahulu yang dijadikan foundation teoretis dalam studi ini.

Tabel 2.1 Penelitian Terdahulu

No	Judul	Penulis / Tahun	Dataset	Metode	Hasil
1	Intelligent classification and performance prediction of multi-text assessment with recurrent neural network-long short-term memory[6]	Tukino Paryono; Eko Sedyono; Hendry; Baenil Huda; April Lia Hananto; Aviv Yuniar Rahman; 2024	1.500 Dataset, 70% <i>train</i> dan 30% <i>test</i>	LSTM dengan 5 Algoritma Optimasi	Akurasi terbaik 94,24% dengan optimasi Adamax
2	Analisis sentimen kebijakan MBKM berdasarkan opini	Sio Jurnalis Pipin; Heru Kurniawan; 2022	658 Data Tweet	LSTM	Akurasi terbaik 80,42%

	masyarakat di twitter menggunakan LSTM[7]				
3	Real-time twitter data analytics of mental illness in COVID-19: Sentiment analysis using deep neural network[8]	Poonkuzhali Sugumaran; Anu Barathi Bhagavathi Kannu Uma; 2022	Data dikumpulkan secara <i>real-time</i> selama satu tahun sejak desember 2019 menggunakan <i>Application Programming Interface</i> (API)	RNN dengan algoritma LSTM dan GRU	99,4%
4	Prediksi belanja pemerintah Indonesia menggunakan <i>Long short-term memory</i> (LSTM)[11]	Sabar Sauntomo; Hilman Ferdinandus Pardede; 2021	Dataset berjumlah 721 row data, yang merupakan data harian belanja pemerintah selama	LSTM dan ARIMA (Auto Regressive Integrated Moving Average)	Metode LSTM menghasilkan performa yang lebih baik dengan

			<p>kurun waktu dari 2018 hingga akhir tahun 2020</p>		<p>tiga <i>hidden layer</i> mencapai nilai MSE 0.2325, RMSE 0.4820, MAE 0.3292, dan MAPE 0.4212. LSTM memiliki performa yang lebih baik dengan pengurangan tingkat error prediksi sebesar 84% hingga 87%.</p>
--	--	--	--	--	---

5	Forecasting the stock price of PT Astra International using the LSTM method[12]	Edwin Setiawan Nugraha; Zalfani Alike; Dadang Amir Hamzah; 2024	1.506 dataset harga saham	LSTM	Tingkat <i>accuracy</i> model LSTM dalam meramal kan harga saham PT Astra Internati onal menunju kan hasil sebagai berikut, MSE 151.910, MAE 118.128, MAPE 2.3%
---	---	---	---------------------------	------	---

2.2 Landasan Teori

2.2.1 Social Media X

Perkembangan *Social Media* yang sangat pesat menciptakan suatu platform interaksi sosial secara digital dalam ruang virtual, dimana setiap individu dapat berinteraksi, berkomunikasi, berkirim pesan, berbagi pesan, serta membangun jaringan sosial. Setiap individu dapat menciptakan sebuah situs jejaring sosial dan terhubung dengan individu lainnya[13]. *Social Media* memegang

peranan krusial dan terdapat berbagai alasan yang mendukung penggunaannya dalam kegiatan penelitian. Selama dua dekade terakhir, jaringan *Social Media* telah berkembang secara konsisten menjadi bidang penelitian lintas disiplin ilmu[14].

Twitter atau X merupakan sebuah *platform Social Media* yang sering berguna dalam menyampaikan informasi dengan begitu cepat tentang suatu peristiwa terbaru[15]. Twitter telah digunakan oleh para peneliti di berbagai bidang akademis antara lain, *crisis management, political studies, urban planning, industrial domain, financial institution, tourism industry, media sector, retail commerce, manufacturing sector, educational institutions, dan healthcare service*. Twitter memberikan kemudahan mengakses data melalui taggar atau kata kunci dengan bantuan API[16].

2.2.2 Kasus Kopi Sianida

Dalam peristiwa yang terjadi pada Rabu, 6 Januari 2016, Mirna salihin atau yang dikenal sebagai Mirna bersama teman-temannya yaitu Jessica, dan Hani bertemu di Olivier Cafe, Grand Indonesia. Jessica yang saat itu datang lebih awal memesan minuman untuk mereka termasuk *Ice Coffee Vietnamese* untuk Mirna. Setelah meminum kopi tersebut, Mirna mengalami kejang dan kemudian meninggal dunia di tempat. Hasil otopsi menunjukkan adanya sianida pada tubuh Mirna. Kejadian tersebut menjadi topik pembicaraan di kalangan masyarakat Indonesia sepanjang tahun 2016.

Pada tanggal 30 Januari 2016, Jessica ditangkap lalu pada tanggal 27 Oktober 2016, Jessica menerima sanksi pidana berupa penjara 20 Tahun berdasarkan keputusan Pengadilan Negeri Jakarta Pusat setelah melakukan 46 kali sidang selama 8 bulan. Putusan yang disebutkan memperoleh konfirmasi dari Pengadilan Tinggi DKI Jakarta dan Mahkamah Agung[4].

2.2.3 Analisis Sentiment

Dalam *text mining*, analisis sentimen merupakan sebuah teknik yang digunakan dalam mengidentifikasi serta mengategorikan pendapat publik secara otomatis menjadi positif, negatif ataupun netral[7]. Analisis sentimen menempati peran krusial dalam bidang *Natural Language Processing* (NLP). Melakukan analisis terhadap konten *Social Media* melibatkan tantangan analisis data yang cukup besar[17].

Konten *Social Media* seperti *text review*, forum, tweet, atau *blogger* dapat menjadi objek dalam proses analisis sentimen. Penggunaan analisis sentimen memungkinkan pengolahan informasi *Social Media* menjadi sebuah data yang terstruktur[1].

2.2.4 Text Mining

Text mining sebagai salah satu metode untuk mengekstraksi informasi berkualitas tinggi dari sebuah data teks. Informasi tersebut diperoleh dengan mengamati pola dan tren melalui analisis statistik. Dalam proses *text mining*, dilakukan pembobotan kata untuk memberikan nilai pada setiap istilah dalam sebuah dokumen. Nilai bobot tersebut bergantung pada metode yang diterapkan[18]. Proses berjalannya *text mining* melibatkan beberapa tahapan, antara lain :

a. Text Data Collection

Pengumpulan data atau *data collection* merupakan tahapan mengumpulkan berbagai data mentah dari beragam sumber. Data tersebut dapat berbentuk tidak terstruktur (*unstructured*) ataupun semi-terstruktur (*semi-structured*)[19].

b. Pre-Processing

Tahap *pre-processing* merupakan proses transformasi data teks mentah ke dalam bentuk yang lebih terstruktur dan bersih. Dalam pelaksanaannya, *pre-processing* meliputi beberapa tahapan pengerjaan seperti *Tokenizing*, *Case Folding*, *Stopwords Removal*, *Steaming* atau *Lemmatization*, serta *Normalization*[20].

c. Feature Extraction using Word2vec

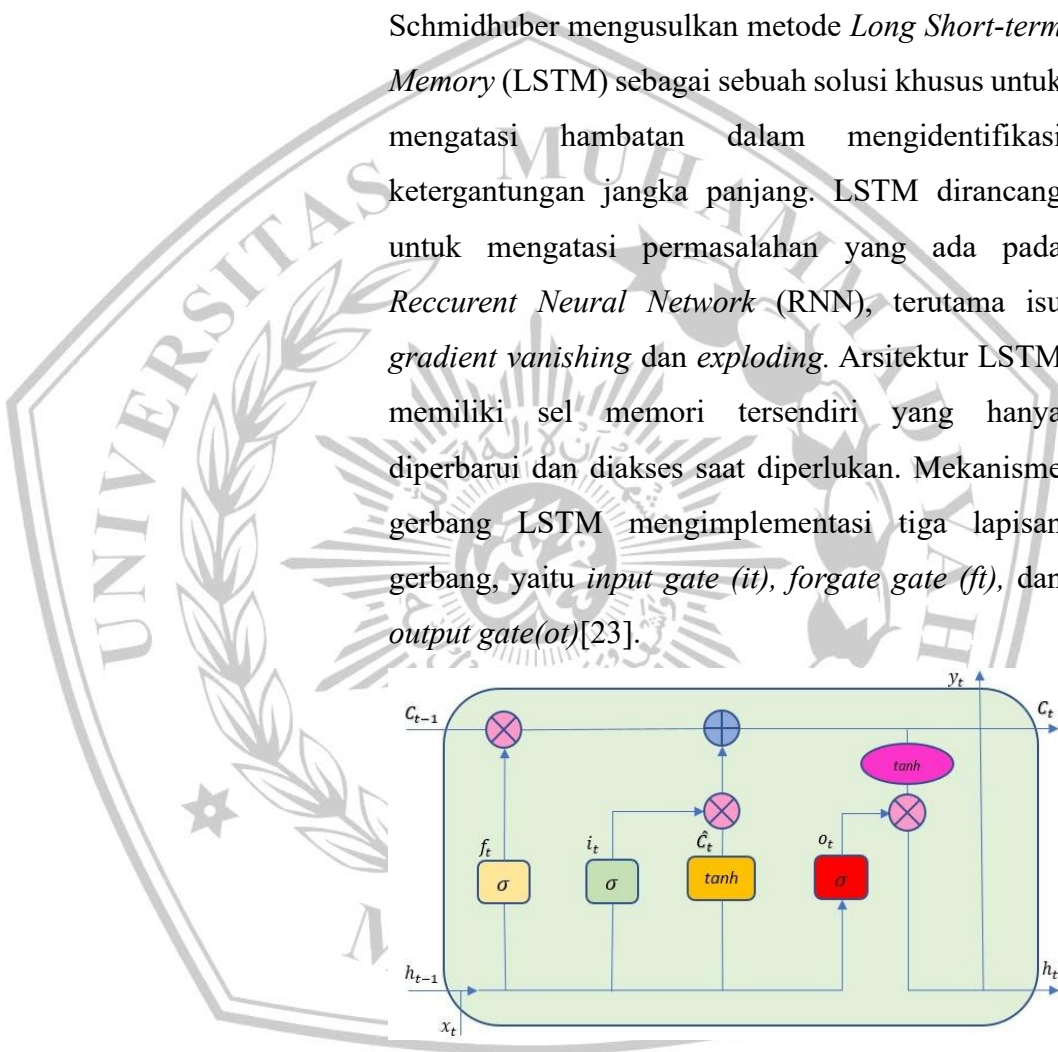
Feature extraction merupakan komponen esensial dalam *machine learning* khususnya untuk pemrosesan data teks. Pada tahun 2013, Mikolov mengusulkan metode *Word2Vec* dengan mempertimbangkan *Corpus* sebagai *input* dan *output* berupa *vector*. Algoritma *Word2Vec* merepresentasikan *vector* kata yang mampu mencapai kinerja terbaik dalam *Natural Language Program* (NLP) dengan tahap mengelompokkan kata yang serupa memiliki *vector* yang sama[21].

Dalam pengembangan model *Word2Vec*, *gensim* digunakan sebagai *library python* yang berfokus pada indeks dokumen. Kemudian pembuatan kamus kata dan kamus *vector* menggunakan algoritma *word embedding*, yaitu *Word2Vec*. Algoritma ini digunakan untuk mengubah kata menjadi *vector* numerik sehingga dapat diproses

oleh komputer dan memungkinkan analisis teks yang lebih efektif[22].

d. Modeling using LSTM

Pada tahun 1997, Hochreiter dan Schmidhuber mengusulkan metode *Long Short-term Memory* (LSTM) sebagai sebuah solusi khusus untuk mengatasi hambatan dalam mengidentifikasi ketergantungan jangka panjang. LSTM dirancang untuk mengatasi permasalahan yang ada pada *Recurrent Neural Network* (RNN), terutama isu *gradient vanishing* dan *exploding*. Arsitektur LSTM memiliki sel memori tersendiri yang hanya diperbarui dan diakses saat diperlukan. Mekanisme gerbang LSTM mengimplementasi tiga lapisan gerbang, yaitu *input gate* (i_t), *forget gate* (f_t), dan *output gate* (o_t)[23].



Gambar 2.1 Arsitektur LSTM

Forget gate (f_t) akan menerima masukan berupa *hidden state* dari sel sebelumnya dan data baru yang berasal dari input kontemporer. Informasi tersebut akan diintegrasikan dan dioleh menggunakan fungsi sigmoid yang menghasilkan

nilai antara 0 hingga 1. Apabila hasil yang diperoleh mendekati nilai 0, maka informasi akan didisposisi, sedangkan jika hasil mendekati nilai 1, maka informasi akan dipertahankan dalam memori.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Input gate (it) akan mengevaluasi parameter mana yang akan diperbaharui melalui konversi nilai ke dalam range biner 0-1. *Input gate (it)* memiliki 2 lapisan, yaitu arsitektur sigmoid yang menentukan nilai mana yang harus di perbarui dan fungsi *tanh* yang menghasilkan representasi *vector* kandidat baru c_t .

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\check{c}_t = \tanh W_c [h_{t-1}, x_t] + b_c \quad (3)$$

Output gate (og) akan menentukan tindakan yang harus diimplementasikan pada *hidden state* berikutnya, karena *hidden state* mempertahankan informasi masukan terdahulu.

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t \times \tanh(c_t) \quad (5)$$