

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian ini merupakan data review pesanan pengguna pada aplikasi Alfacgift. Objek ini dipilih untuk melihat bagaimana pendapat pengguna mengenai pengalaman mereka terhadap pengiriman pada Alfacgift. Data yang diperoleh terlebih dahulu diberikan label untuk mengategorikan review menjadi positif dan negatif. Analisis sentiment dilakukan pada objek menggunakan algoritma *Deep Learning* yakni *bidirectional LSTM* yang akan dibandingkan dengan *simple RNN*. Penelitian ini akan melihat hasil review sejak 1 Oktober 2023 hingga 31 Desember 2023 pada daerah Tangerang untuk melihat pengalaman pengguna terkini dan bisa menjadikannya sebagai bahan evaluasi.



Gambar 3. 1 Logo Alfacgift

Alfacgift merupakan sebuah aplikasi yang bisa dimanfaatkan untuk berbelanja secara online dari Alfamart dengan logo pada Gambar 3.1. Manfaat keanggotaan dari Alfamart juga dapat diakses melalui aplikasi Alfacgift. Alfacgift sendiri diluncurkan pada tahun 2019 dan pada tahun yang sama Alfamart melakukan akuisisi terhadap Ponta. Ponta sendiri merupakan program membership yang digunakan oleh alfamart untuk memberikan benefit pada para member-nya. Alfacgift sendiri memiliki kelebihan yakni pengiriman langsung dari toko Alfamart terdekat yang melayani SAPA (Siap Antar Pesanan Anda) dan terintegrasi dengan lebih dari 15.000 toko Alfamart di seluruh penjuru Indonesia.

Alfagift memiliki pengaruh baik bagi Alfamart terlebih pada saat terjadi pandemi Covid-19 [7]. Alfagift sebagai aplikasi memiliki fitur dan juga manfaat yang membantu masyarakat seperti:

1. *Same Day Delivery* dan Gratis Ongkir:

Produk yang dibeli melalui Alfagift dan memiliki keterangan produk toko dapat dikirimkan secara langsung melalui toko Alfamart SAPA terdekat dari lokasi alamat yang dipilih pembeli dan akan dikirim pada hari yang sama untuk order yang terbuat sebelum jam 9 malam). Alfagift juga menerapkan gratis ongkir tanpa syarat untuk setiap pengiriman.

2. *Membership*

Membership terintegrasi baik pada aplikasi maupun secara langsung pada toko. Pemilik member hanya perlu menyebutkan nomor member atau nomor handphone yang sudah terdaftar untuk mendapatkan A-Poin dari setiap pembelian. Pada aplikasi Alfagift pengguna dapat login menggunakan nomor member. Pengguna juga bisa menunjukkan *barcode* member kepada kasir untuk memperoleh poin.

3. *Store Locator*

Aplikasi Alfagift yang memiliki akses lokasi pada pengaturan dapat menunjukkan toko Alfamart terdekat dari lokasi. Pencarian toko ini juga dapat di filter seperti buka 24 jam, memiliki toilet, memiliki wifi, dan lainnya. Umumnya *store locator* akan mencari toko dengan radius maksimum 2 km dan diurutkan dari toko terdekat.

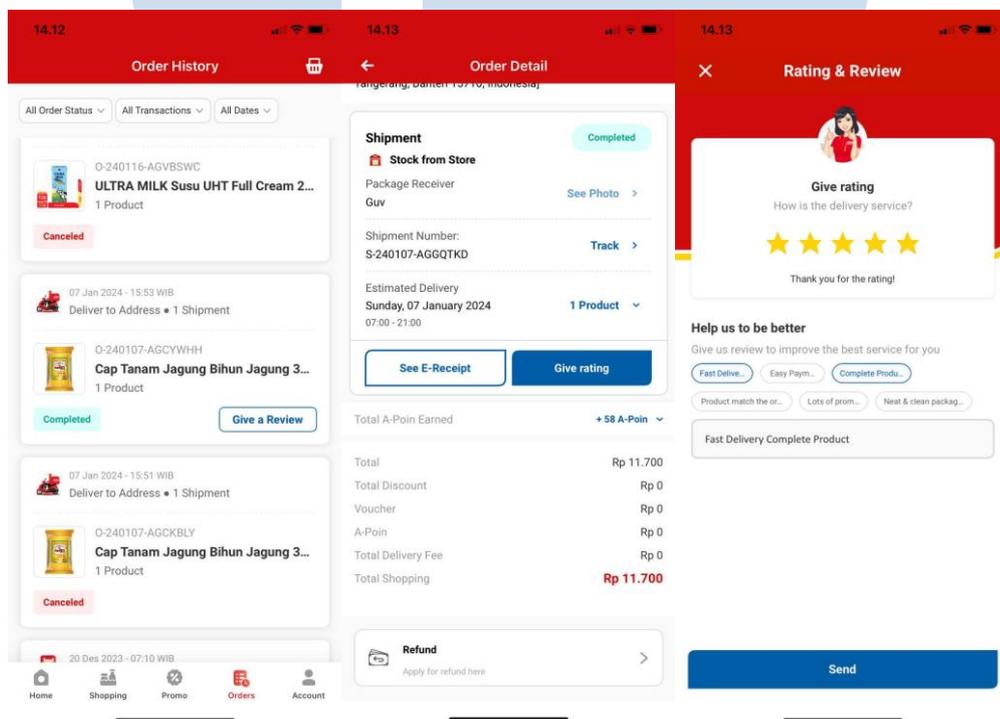
4. *Recommendation Engine*

Ketika pengguna sudah beberapa kali menggunakan aplikasi Alfagift, aplikasi dapat memberikan saran untuk produk yang sering dibeli dan juga *quantity* rutin dari setiap produk.

5. Metode Pembayaran Lengkap

Aplikasi Alfagift memiliki metode pembayaran yang lengkap mulai dari bank transfer menggunakan *Virtual Account*, kartu kredit, COD, menggunakan A-Poin, dan berbagai metode pembayaran lainnya.

Data yang digunakan sebagai objek penelitian ini diambil berdasarkan ulasan terhadap pengiriman pada aplikasi Alfagift. Ulasan dapat dilakukan setelah pengguna selesai melakukan pemesanan pada aplikasi Alfagift dan telah menerima pesanan tersebut. Pesanan akan berganti status menjadi *Completed* dan pengguna dapat memberikan rating dan juga *review*. Alur lengkap dapat dilihat pada Gambar 3.2 yang menunjukkan halaman *Rating & Review*.

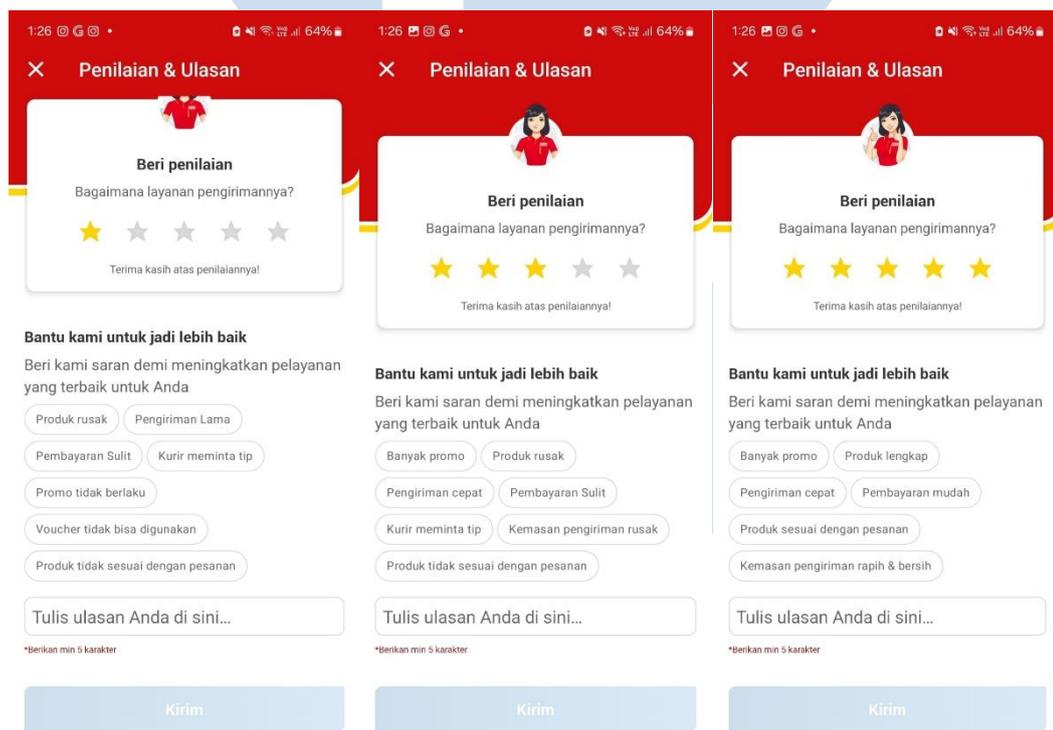


Gambar 3.2 Alur ulasan Alfagift

Pemberian ulasan dapat dilakukan pada menu Orders yang menyajikan *button Give a Review*. Selain itu ketika pengguna melakukan klik pada pesanan maka pengguna akan masuk ke halaman *Order Detail* yang jika di scroll akan menampilkan *button See E-Receipt* dan *Give rating*. *Button Give rating* akan membawa pengguna pada halaman yang sama yakni halaman *Rating & Review*.

Pada halaman ini pengguna dapat memberikan rating berupa bintang satu sampai lima dan juga dapat memasukan ulasan dibawah tulisan *Help us to be better*. Pengguna dapat langsung mengetik ulasan atau juga dapat memilih beberapa *default comment* berdasarkan bintang /rating yang diberikan.

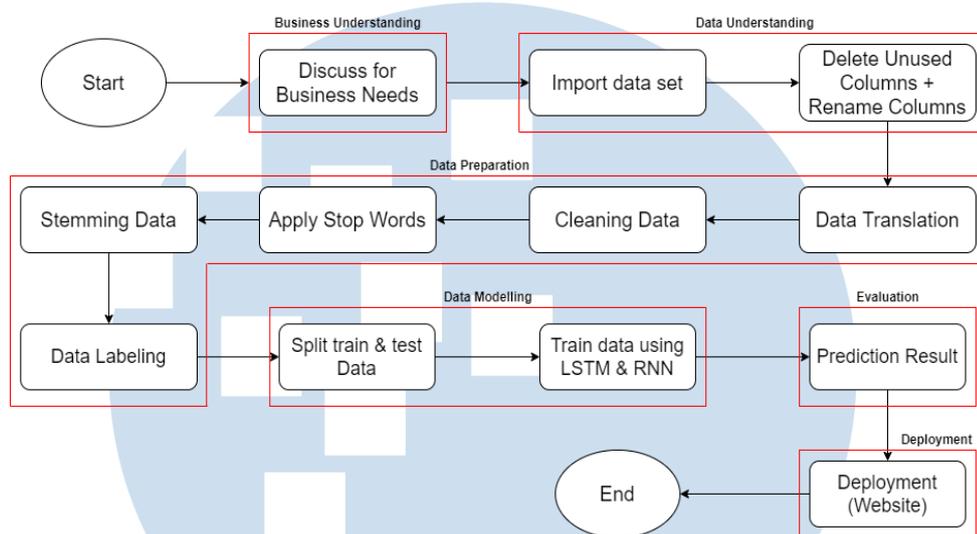
Rating atau bintang satu dan dua merupakan rating negatif sehingga *default* ulasan yang dapat dipilih sama dan menggambarkan ulasan negatif seperti “Produk rusak”, “Pengiriman lama”, “Pembayaran sulit”, dan lain-lain. Untuk rating bintang tiga memiliki sifat lebih netral sehingga memiliki default ulasan yang bersifat kombinasi antara ulasan negatif dan positif. Bintang atau rating empat dan lima menunjukkan ulasan positif dan memiliki default ulasan positif seperti “Banyak promo”, “Produk lengkap”, “Pengiriman cepat” dan lain-lain. Gambaran lebih jelas dapat dilihat pada Gambar 3.3.



Gambar 3.3 Ulasan untuk Review bintang 1&2, bintang 3, dan bintang 4&5

3.2 Metode Penelitian

3.2.1 Alur Penelitian



Gambar 3. 4 Alur Penelitian

Gambar 3.4 menunjukkan alur yang dijalankan penelitian untuk mencapai tujuan penelitian. Penelitian ini dimulai dengan tahapan *business understanding* dengan mendiskusikan permasalahan bisnis yang dapat diselesaikan. Hasil diskusi yang dilakukan menjabarkan perlunya dilakukan pengolahan data ulasan pengguna terhadap pengiriman aplikasi Alfacift untuk mendapatkan gambaran sentimen masyarakat terhadap Alfacift. Langkah selanjutnya adalah melakukan *import* dataset hasil *review* pengguna terhadap pengiriman pada aplikasi Alfacift. *Dataset* yang berhasil di import kemudian akan dipilih kolom yang akan digunakan untuk penelitian, kolom yang tidak memiliki pengaruh atau tidak digunakan akan di-*drop*. Data kemudian akan ditranslasi ke Bahasa Indonesia menggunakan *deep_translator*, dan akan disimpan pada sebuah kolom baru. Langkah selanjutnya adalah melakukan *cleaning* data yang meliputi *case folding* (penyelarasan huruf kapital menjadi huruf kecil), penghapusan angka, spasi kosong, dan juga emotikon. Ketika data sudah selesai melalui data *cleaning* akan dilanjutkan pada proses data *preprocessing*. *Preprocessing* akan dimulai dengan *stop words* untuk menghilangkan kata-kata yang sering muncul tetapi

tidak memiliki pengaruh yang signifikan seperti “dan”, “yang”, dan lain-lain. Proses dilanjutkan dengan melakukan *stemming* untuk merubah kata ke bentuk akar/dasar kata tersebut.

Data labeling dilakukan menggunakan Indonesia Sentiment (InSet) Lexicon yang akan melakukan pembobotan pada kalimat. InSet Lexicon sendiri merupakan *library* yang berisikan kata dengan bobot yang berbeda pada setiap kata. Bobot ini lah yang nantinya akan dijumlahkan untuk menentukan label positif atau negatif yang akan diberikan pada sebuah ulasan. Bobot dari setiap kata adalah -5 sampai 5, nantinya kata pada tiap kalimat akan ditambahkan bobot positif dan negatifnya jika hasil akhir lebih besar dari nol maka data ulasan akan mendapatkan label positif. Bobot akhir negatif pada sebuah kalimat akan mengakibatkan kalimat tersebut mendapatkan label negatif.

Langkah selanjutnya adalah melakukan *split data* menjadi *training* dan *testing* sebesar 80:20. Data yang telah dibagi kemudian akan dilakukan pemodelan menggunakan *bidirectional LSTM* dan akan dibandingkan dengan *simple RNN*. Hasil akhir pemodelan akan dibandingkan menggunakan akurasi, *confusion matrix*, dan juga *classification report*.

3.2.2 Metode Pengembangan Sistem / Metode Data Mining

Terdapat berbagai jenis metodologi framework yang bisa digunakan untuk melakukan proses data mining. CRISP-DM, SEMMA dan KDD menjadi 3 metodologi yang biasa digunakan. Berikut dapat dilihat tabel untuk membandingkan proses antara ketiga metode *data mining*.

Tabel 3.1 Perbandingan proses metode data mining

KDD	SEMMA	CRISP-DM
<i>Pre KDD</i>	-----	<i>Business Understanding</i>
<i>Selection</i>	<i>Sample</i>	<i>Data Understanding</i>
<i>Pre-processing</i>	<i>Explore</i>	
<i>Transformation</i>	<i>Modify</i>	<i>Data Preparation</i>
<i>Data Mining</i>	<i>Model</i>	<i>Modeling</i>
<i>Interpretation/Evaluation</i>	<i>Assesment</i>	<i>Evaluation</i>

KDD	SEMMA	CRISP-DM
<i>Post KDD</i>	-----	<i>Deployment</i>

Berdasarkan proses yang disajikan pada Tabel 3.1 dapat diambil kesimpulan bahwa CRISP-DM dan SEMMA menjadi hasil adaptasi dari proses KDD. CRISP-DM dan SEMMA menjadi metode yang dapat membantu untuk mengetahui bagaimana penerapan data mining secara nyata. Berdasarkan prosesnya CRISP-DM merupakan metode yang lengkap dan terdokumentasi dengan baik untuk setiap tahapannya dimana setiap tahap sudah jelas proses yang akan dilakukan. Selain prosesnya berikut merupakan perbandingan CRISP-DM, KDD, dan SEMMA.

Tabel 3.2 Tabel kelebihan dan kekurangan metode data mining

Metode	Kelebihan	Kekurangan
CRISP-DM	<ol style="list-style-type: none"> Memiliki tahapan dan struktur yang jelas Fokus pada kebutuhan bisnis Banyak digunakan dan dokumentasi lengkap 	<ol style="list-style-type: none"> Memerlukan waktu yang relatif lebih panjang Tidak ada proses untuk mengatasi permasalahan <i>outlier</i> pada data
KDD	<ol style="list-style-type: none"> Mencakup seluruh proses <i>data mining</i> Fokus pada <i>Knowledge discovery process</i> Dapat dilakukan secara <i>repetitive</i> untuk mendapatkan hasil yang lebih baik 	<ol style="list-style-type: none"> Untuk mencapai hasil maksimal perlu dilakukan secara <i>repetitive</i> sehingga memakan waktu lebih lama Memiliki tingkat kompleksitas yang tinggi
SEMMA	<ol style="list-style-type: none"> Memiliki tahapan yang dapat dilakukan secara <i>iterative</i> dengan melakukan trial dan error Fokus pada fleksibilitas pada proses <i>data exploration</i> 	<ol style="list-style-type: none"> Tidak memiliki fase <i>deployment</i> dan evaluasi hanya dilakukan melalui <i>assessment</i> Tidak memiliki tahapan <i>business understanding</i> sehingga tidak fokus pada kebutuhan bisnis

Tabel 3.2 menunjukkan kelebihan dan kekurangan metode data mining, dan berdasarkan tabel tersebut dipilih CRISP-DM sebagai metode *data mining*. Pemilihan ini didasari oleh CRISP-DM yang memiliki fokus pada kebutuhan

bisnis yang pada kasus ini adalah membantu PT. XYZ untuk melakukan analisis sentimen. Penelitian ini menggunakan *metode data mining Cross-Industry Standard Process for Data Mining* atau biasa dikenal dengan CRISP-DM. Tahapan pada penelitian ini terbagi menjadi 6 tahapan yakni:

1. *Business Understanding*

Merupakan tahap pertama yang memiliki tujuan untuk mendapatkan alasan awal dilakukan penelitian. Tahap ini akan mencari maksud dilakukan penelitian dan juga permasalahan apa yang ingin diselesaikan. Dalam penelitian ini memiliki alasan untuk mengetahui sentiment pengiriman Alfagift sebagai salah satu aplikasi yang memiliki transaksi yang besar.

2. *Data Understanding*

Pemahaman data yang dilakukan mulai dari pengumpulan data. Mendeskripsikan data dan juga melakukan eksplorasi terhadap data yang diterima. Pada tahap akan dipilih kolom-kolom yang akan digunakan untuk penelitian. Seperti rating (bintang 1-5) dan juga ulasan yang diberikan oleh pengguna sebagai acuan sentiment berupa positif maupun negatif.

3. *Data Preparation*

Merupakan persiapan pada data untuk dapat digunakan secara maksimal. Tahapan ini meliputi data cleaning untuk menghapus hal-hal yang tidak diperlukan untuk melakukan sentiment analysis. Sebelum melakukan *data cleaning* data akan terlebih dahulu diberikan label apakah ulasan memiliki sentimen positif atau negatif. Penentuan ini dilakukan menggunakan InSet Lexicon yang akan melakukan pembobotan pada tiap kalimat ulasan. Pembobotan dilakukan berdasarkan library yang ada dimana setiap kata memiliki bobot -5 sampai 5. Hasil akhir kalkulasi bobot negatif akan membuat kalimat ulasan memiliki label negatif dan sebaliknya hasil akhir bobot berupa

positif maka akan mendapatkan label positif. *Data cleaning* akan menyelaraskan ulasan menjadi huruf kecil, menghapus data yang memiliki ulasan kosong, menghapus angka, menghapus tanda baca, menghapus emotikon, dan juga menghapus whitespace yang ada. Tahapan selanjutnya adalah melakukan *stemming* untuk merubah kata ke bentuk akar/dasar kata tersebut. Langkah terakhir pada *preprocessing data* adalah *stop words* untuk menghilangkan kata-kata yang sering muncul tetapi tidak memiliki pengaruh yang signifikan seperti “dan”, “yang”, dan lain-lain.

4. *Data Modeling*

Untuk melakukan *data modeling* akan dimulai dengan pemisahan data menjadi *training* dan *testing*. Ketika sudah diterapkan maka langkah selanjutnya adalah menerapkan algoritma *bidirectional LSTM* dan *simple RNN*.

5. *Evaluation*

Hasil modeling yang dilakukan akan dievaluasi menggunakan perhitungan *confusion matrix* mulai dari *accuracy*, *precision*, *recall*, dan *F1-Score*. Kemudian hasil akan dibandingkan antara *bidirectional LSTM* dan *simple RNN*.

6. *Deployment*

Tahapan terakhir ini akan dilakukan *deployment* berdasarkan hasil modeling terbaik. Pada penelitian ini akan dibuat *website* untuk melakukan upload data sentimen mengenai hasil pengiriman dan akan menampilkan hasil analisis terhadap data tersebut.

3.3 Teknik Pengumpulan Data

3.3.1 Populasi dan Sampel

Populasi merupakan hal yang sangat penting dalam penelitian. Populasi merupakan subjek dalam penelitian yang menjadi elemen terpenting dalam suatu penelitian [40]. Pada penelitian ini populasi merupakan data hasil review

orderan Alfagift pada tahun 2023. Sampel merupakan bagian jumlah yang dimiliki dari populasi [41]. Penelitian ini akan menggunakan *Non-probability Sampling* dimana teknik ini tidak memungkinkan peluang yang sama bagi semua bagian dari populasi untuk dipilih menjadi anggota sampel. *Purposive sampling* merupakan salah satu contoh dari *Non-probability sampling* dan akan digunakan pada penelitian ini. *Purposive Sampling* dipilih karena tidak mengambil data secara acak melainkan berdasarkan kriteria data yang diperlukan untuk penelitian [42].

3.3.2 Periode Pengambilan Data

Penelitian ini akan mengambil data ulasan pengiriman Alfagift dengan rentang waktu 1 Oktober 2023 hingga 31 Desember 2023. Alasan memilih periode ini adalah untuk melihat penilaian terkini pengguna terhadap pengiriman Alfagift. Nantinya dapat dimanfaatkan pihak Alfagift untuk melihat sentimen pengguna terhadap aplikasi yang dimiliki serta modul yang paling banyak memiliki sentiment negatif. Total ulasan pada periode ini berjumlah 22,402 ulasan.

3.4 Variabel Penelitian

Terdapat dua variabel yang ada pada penelitian ini yakni variabel dependen dan variabel independen. Kedua variabel digunakan untuk mengidentifikasi hubungan antar variabel yang ada. Variabel terikat (Dependen) akan dipengaruhi oleh variabel bebas (Independen) [43].

1. Variabel Dependen

Variabel dependen atau variabel terikat merupakan variabel yang disebabkan oleh variabel yang lain. Dapat juga diartikan sebagai variabel yang perubahannya dipengaruhi dan bergantung pada variabel lain (variabel independen). Pada penelitian ini label sentimen (positif dan negatif) menjadi variabel dependen.

2. Variabel Independen

Variabel independen atau variabel bebas merupakan variabel yang mempengaruhi variabel dependen. Ketika variabel independen muncul maka akan mengubah kondisi variabel lain. Pada penelitian ini ulasan pengguna Alfacita terhadap pesanan akan menjadi variabel independen.

3.5 Teknik Analisis Data

Data penelitian ini adalah ulasan dan label yang memiliki nilai positif dan negatif sehingga penelitian ini bersifat kualitatif. Penelitian ini menggunakan Google Colaboratory sebagai tempat untuk menjalankan program berbasis Python. Dengan menggunakan Google Colaboratory memiliki keuntungan karena tidak perlu melakukan setup environment dan terkoneksi dengan Google Drive untuk menyimpan data. Data akan melalui langkah pre-processing dan selanjutnya akan dilakukan pengujian menggunakan model. Model yang digunakan adalah bidirectional LSTM yang akan dibandingkan dengan simple RNN. Perbedaan antara kedua Algoritma ini adalah pada kemampuan *bidirectional* LSTM untuk mengatasi masalah long-term dependencies yang dimiliki oleh *simple* RNN yang disebabkan oleh *vanishing gradient*. *Bidirectional* LSTM juga sangat baik untuk data yang kompleks, memerlukan long-term dependencies, dan contextual understanding dari kedua arah sequence. Dalam kata lain bidirectional LSTM merupakan hasil pengembangan dari RNN yang disempurnakan untuk mengatasi keterbatasan yang dimiliki oleh RNN. Perbandingan kedua algoritma dapat dilihat pada Tabel 3.3. Hasil akurasi dari algoritma yang digunakan akan dilakukan pengujian melalui *confusion matrix*.

Tabel 3.3 Perbandingan algoritma bidirectional LSTM dan simple RNN

Fitur	<i>Bidirectional</i> LSTM	<i>Simple</i> RNN
<i>Architecture</i>	Terdiri dari dua <i>layer</i> LSTM, satu untuk memproses <i>input forward</i> dan yang lain secara <i>backward</i>	Hanya memiliki satu <i>layer Recurrent Neural Network</i>
<i>performance</i>	Lebih unggul jika dibandingkan dengan <i>simple</i> RNN dalam tugas	Kurang baik untuk <i>long-range dependencies</i> namun dapat

Fitur	<i>Bidirectional LSTM</i>	<i>Simple RNN</i>
	yang memerlukan <i>long-range dependencies</i>	memiliki performa lebih baik dalam tugas yang sederhana
<i>Training Speed</i>	Lebih lambat karena arsitektur dan tingkat komputasi yang lebih rumit	Lebih cepat karena memiliki arsitektur yang lebih sederhana
<i>Use Cases</i>	<i>Natural Language Processing, Speech Recognition, Time Series Prediction</i>	<i>Simple sequence prediction tasks</i>

