

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa dekade terakhir, kualitas udara telah menjadi isu penting di berbagai wilayah di seluruh dunia, termasuk di kota-kota padat penduduk seperti Kota Makassar. Fenomena ini muncul sebagai hasil dari perkembangan industrialisasi, pertumbuhan populasi, dan mobilitas yang tinggi dalam masyarakat modern. Kualitas udara yang buruk dapat memiliki dampak serius terhadap kesehatan manusia dan lingkungan. Udara yang telah tercemar oleh zat-zat polutan bukan saja mempengaruhi kesehatan manusia tetapi seluruh makhluk hidup dan lingkungan juga akan terkena efek dari pencemaran udara tersebut. Paparan jangka panjang pada manusia akan mengakibatkan penyakit berbahaya seperti gangguan pernapasan yang bisa mengakibatkan kematian, (Abidin & Artauli Hasibuan, 2019).

Berdasarkan data dari Badan Pusat Statistik (BPS), jumlah penduduk di Kota Makassar pada tahun 2023 tercatat sebanyak 1.436.626 jiwa. Pertumbuhan populasi yang signifikan ini memberikan dampak besar terhadap lingkungan dan kualitas udara di kota tersebut. Dengan meningkatnya jumlah penduduk, Kota Makassar menghadapi tantangan serius dalam menjaga kualitas udara yang baik. Isu kependudukan dan lingkungan ini adalah dua hal yang saling terkait erat, dimana keberlanjutan ekosistem memiliki pengaruh langsung terhadap kualitas hidup manusia, seperti yang dijelaskan oleh (Ayu dkk, 2021). Kepadatan penduduk yang tinggi, bersama dengan aktivitas manusia yang intensif seperti operasi industri dan transportasi, menyebabkan peningkatan emisi polutan udara. Sehingga, aktivitas sehari-hari seperti penggunaan kendaraan bermotor, konsumsi energi, dan limbah industri, semakin meningkatkan volume emisi gas buang dan partikel polutan ke atmosfer, yang langsung mempengaruhi kualitas udara di Makassar.

Dampak negatif dari pencemaran udara sangat signifikan terhadap kesehatan manusia. Partikel polutan yang terhirup dapat merusak sistem pernapasan dan mengiritasi saluran pernapasan, menyebabkan berbagai masalah kesehatan seperti asma, bronkitis, penyakit paru obstruktif kronis (PPOK), dan bahkan kanker paru-paru. Paparan jangka panjang terhadap polusi udara juga dapat meningkatkan risiko penyakit kardiovaskular dan mempengaruhi sistem kekebalan tubuh. Selain itu, pencemaran udara memiliki dampak negatif terhadap lingkungan; peningkatan emisi gas rumah kaca dan partikel polutan ke atmosfer berkontribusi pada perubahan iklim global. Gas rumah kaca, seperti karbon dioksida (CO₂) dan metana (CH₄), menyebabkan pemanasan global dan perubahan iklim yang tidak terkendali. Perubahan iklim tersebut dapat mengganggu ekosistem, merusak keanekaragaman hayati, dan mempengaruhi pola cuaca yang stabil. Kualitas udara memiliki peranan penting bagi lingkungan dan meningkatkan kualitas hidup masyarakat, seperti yang ditekankan oleh (Delgado dkk, 2020). Dalam beberapa dekade terakhir, algoritma machine learning telah menjadi alat

yang efektif dalam mengklasifikasikan kualitas udara, berkat kemampuan mereka untuk menganalisis dan mengekstraksi pola dari data yang beragam dan kompleks. Dengan menggunakan algoritma ini, dapat dikembangkan model yang memanfaatkan berbagai faktor seperti data cuaca, pola lalu lintas, emisi industri, dan data kualitas udara historis untuk mengklasifikasikan kondisi udara secara efisien.

Penelitian terkait yang telah dilakukan oleh (Kök dkk, 2017) dengan judul "*A Deep Learning Model for Air Quality Prediction in Smart Cities*" didapatkan hasil bahwa LSTM mampu mengklasifikasikan kualitas udara dengan akurasi sebesar 95%. Kemudian penelitian yang serupa juga dilakukan oleh (Delgado dkk, 2020) dengan judul "*Air Quality Prediction (PM2.5 and PM10) at the Upper Hunter Town – Muswellbrook using the Long Short Term Memory Method*" menjelaskan bahwa model klasifikasi menggunakan algoritma LSTM mampu memklasifikasi kualitas udara dengan memanfaatkan fitur urutan waktu. Selanjutnya penelitian serupa juga dilakukan oleh (Wang, J dkk, 2023) yang berjudul "*An air quality index prediction model based on CNN-ILSTM*" Hasil evaluasi klasifikasi AQI (Air Quality Index) menggunakan model CNN-ILSTM lebih baik dibandingkan dengan LSTM dan GRU, hasil MAE dari CNN-ILSTM mengalami penurunan sebesar 0,28479 dan R2 meningkat sebesar 0,013951.

Dari beberapa paper penelitian yang meneliti terkait klasifikasi kualitas udara, terlihat bahwa algoritma dan *Long Short Term Memory (LSTM)* sering kali menghasilkan performa terbaik. Algoritma ini memiliki keunggulan masing-masing yang membuatnya sangat relevan untuk menangani masalah klasifikasi kualitas udara. Penelitian ini bertujuan untuk membangun sebuah sistem klasifikasi kualitas udara di Kota Makassar dengan menggunakan *Long Short Term Memory (LSTM)*. Sistem yang dikembangkan ini secara khusus dirancang untuk mengatasi tantangan dalam mengklasifikasikan kualitas udara yang bersumber dari volume data mentah yang semakin besar.

Data yang digunakan dalam sistem klasifikasi kualitas udara ini bersumber dari Indeks Standar Pencemaran Udara (ISPU), yang merupakan standar pengukuran kualitas udara yang digunakan di Indonesia (Kurniawan, 2018). ISPU mengkategorikan kualitas udara berdasarkan konsentrasi polutan tertentu seperti partikulat dan gas berbahaya yang terdapat dalam udara. Dengan memanfaatkan data historis dari ISPU, sistem klasifikasi yang dikembangkan mampu mengidentifikasi pola dan tren pencemaran udara dalam periode waktu tertentu (Feng dkk, 2024). Sistem ini menggunakan algoritma machine learning canggih untuk mengolah data ISPU, memungkinkan klasifikasi kualitas udara yang akurat dan real-time. Dengan ini, sistem tidak hanya berperan dalam menginformasikan status terkini kualitas udara kepada masyarakat, tetapi juga mendukung pembuat kebijakan dalam merespon masalah pencemaran udara secara lebih efektif (Wang dkk, 2022). Implementasi sistem klasifikasi berbasis data ISPU ini diharapkan dapat membantu meningkatkan kesadaran masyarakat terhadap masalah pencemaran udara dan mengurangi dampak negatifnya terhadap kesehatan dan lingkungan.

1.2 Teori

1.3 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah yang timbul adalah berikut :

1. Bagaimana membangun model klasifikasi kualitas udara di Kota Makassar dengan menggunakan algoritma LSTM?
2. Bagaimana kinerja algoritma LSTM dalam mengklasifikasikan tingkat kualitas udara di Kota Makassar?

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk:

1. Membangun model klasifikasi kualitas udara di Kota Makassar dengan menggunakan algoritma LSTM.
2. Menganalisis kinerja algoritma LSTM dalam mengklasifikasikan tingkat kualitas udara di Kota Makassar.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Pemerintah, penelitian ini diharapkan dapat digunakan sebagai referensi bagi pemerintah untuk membuat sistem klasifikasi kualitas udara di Kota Makassar.
2. Masyarakat, penelitian ini diharapkan mampu memberikan klasifikasi kualitas udara di Kota Makassar.
3. Peneliti, Penelitian ini diharapkan dapat menjadi literatur yang bermanfaat bagi peneliti untuk melakukan penelitian serupa.

1.6 Ruang Lingkup

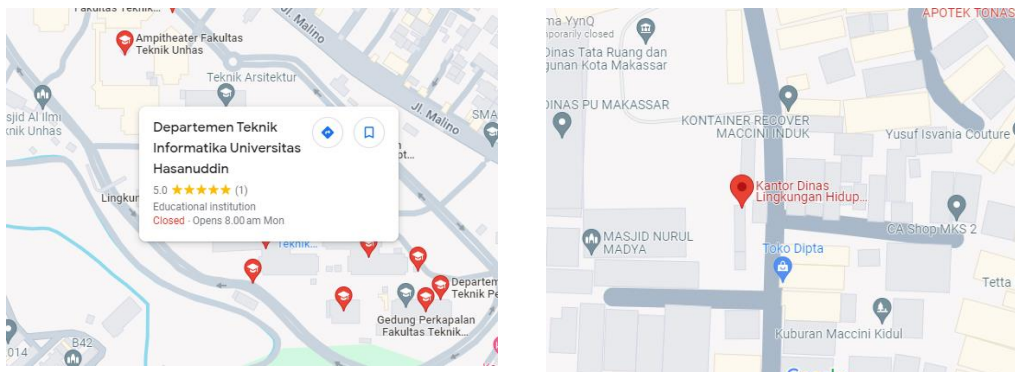
Penelitian ini memiliki beberapa batasan, yaitu:

1. Algoritma yang digunakan adalah algoritma *Long Short Term Memory*
2. Dataset diperoleh dari Dinas Lingkungan Hidup Kota Makassar.

BAB II METODE PENELITIAN

2.1 Lokasi dan Waktu Penelitian

Penelitian ini dimulai sejak proposal penelitian ini disetujui pada bulan juli 2023 hingga penelitian selesai. Pelaksanaan penelitian berlangsung di laboratorium *Cloud Computing* Teknik Informatika Universitas Hasanuddin, sementara pengumpulan data dilakukan di Kantor Dinas Lingkungan Hidup Kota Makassar.



(A) Departemen Teknik Informatika
Universitas Hasanuddin

(B) DLH Kota Makassar

Gambar 10 Lokasi Penelitian

2.2 Instrumen Penelitian

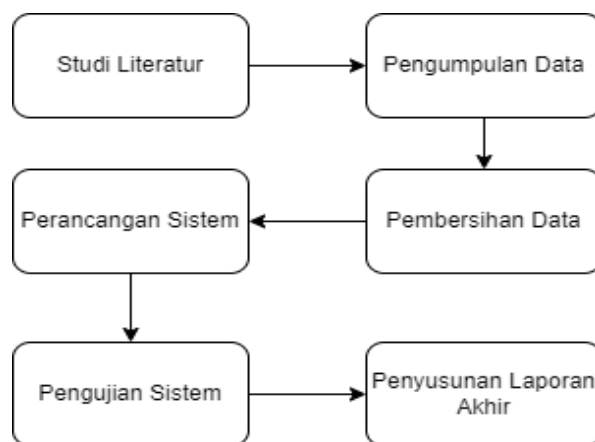
Pada penelitian ini digunakan dua jenis instrumen yaitu instrumen perangkat lunak dan instrumen perangkat keras. Berikut rincian mengenai setiap instrumen penelitian yang digunakan.

2.2.1 Instrumen perangkat lunak

Perangkat lunak yang digunakan pada penelitian ini adalah sebagai berikut :

1. Visual Studio Code
2. Google Chrome
3. Google Collaboratory
4. Google Drive

2.3 Tahapan Penelitian



Gambar 11 Tahapan Penelitian

1. **Studi Literatur**, Pada tahap studi literatur, peneliti akan melakukan pencarian dan analisis literatur terkait klasifikasi kualitas udara menggunakan *Long Short Term Memory* (LSTM) di Kota Makassar. Evaluasi dilakukan terhadap literatur terkini yang relevan dengan lingkungan urban, dan hasil literatur membentuk dasar teoritis untuk merancang model klasifikasi yang efisien dan sesuai dengan kondisi Kota Makassar.
2. **Pengumpulan Data**, Tahap berikutnya setelah studi literatur adalah pengumpulan data dalam penelitian klasifikasi kualitas udara di Kota Makassar menggunakan algoritma *Long Short-Term Memory* (LSTM). Proses ini dimulai dengan penentuan variable fokus, seperti tingkat polutan seperti PM2.5 atau SO₂, dan parameter tambahan seperti data cuaca, lokasi pengukuran, dan waktu pengambilan data. Data dikumpulkan dari tahun 2020 hingga tahun 2023 dan data tahun 2024 digunakan sebagai data uji.
3. **Pembersihan Data (*Preprocessing*)**, Tahap pembersihan data menjadi kunci dalam rangkaian penelitian ini setelah proses pengumpulan data. Langkah pertama yang dilakukan adalah identifikasi anomaly dan kesalahan dalam dataset kualitas udara Kota Makassar. Pemeriksaan teliti dilakukan untuk mengidentifikasi potensi outlier atau nilai yang tidak masuk akal yang mungkin selama proses pengumpulan data. Langkah ini penting untuk memastikan integritas dan konsentrasi data yang akan digunakan dalam analisis. Selain itu, penanganan missing values juga menjadi fokus dalam pembersihan data. Identifikasi dan pengelolaan nilai yang

hilang dari dataset menjadi Langkah kritis untuk mencegah distorsi dalam hasil analisis. Keseluruhan, tahap pembersihan data ini bertujuan untuk menciptakan dataset yang bersih, konsisten, dan siap digunakan dalam pemodelan klasifikasi kualitas udara menggunakan *Long Short Term Memory (LSTM)*.

4. Perancangan Sistem, Setelah tahap pembersihan data, fokus penelitian beralih ke perancangan dan implementasi model klasifikasi kualitas udara menggunakan *Long Term-Short Memory (LSTM)*. Pertama, dilakukan perancangan arsitektur model, yang melibatkan penentuan struktur model termasuk jumlah layer, neuron, dan fungsi aktivasi yang sesuai dengan karakteristik dan kualitas udara di Kota Makassar. Selanjutnya, model *Long Term Short Memory (LSTM)* dilatih menggunakan dataset yang telah dibersihkan untuk memahami pola dan tren dalam data kualitas udara. Setelah model dianggap memadai, dilakukan pengujian menggunakan data uji yang tidak digunakan selama pelatihan dan matrik evaluasi seperti akurasi, presisi, *recall*, dan *F1 Score* digunakan untuk mengevaluasi kinerja model.

5. Pengujian Sistem, Pada tahap ini, model yang telah dirancang dan dilatih akan diuji untuk mengevaluasi kinerjanya dalam memklasifikasi kualitas udara.

6. Penyusunan Laporan Akhir, Pada tahap ini, penelitian mengumpulkan semua hasil penelitian, temuan, dan kesimpulan untuk disusun menjadi suatu laporan yang lengkap dan komprehensif. Laporan akhir ini mencakup deskripsi secara rinci tentang metodologi yang digunakan, mulai dari tahap studi literatur, pengumpulan data, pembersihan data, hingga perancangan dan implementasi model klasifikasi kualitas udara menggunakan *Long Short-Term Memory (LSTM)*.

2.4 Teknik Pengambilan Data

Proses pengumpulan data dilakukan secara langsung di Dinas Lingkungan Hidup Kota Makassar. Langkah ini melibatkan pengambilan data kualitas udara dari sumber yang terpercaya dan langsung terkait dengan lingkungan setempat. Data yang dikumpulkan mencakup berbagai parameter yang relevan dengan kualitas udara, seperti data yang dikumpulkan mencakup parameter-parameter penting seperti:

1. PM10,
2. PM2.5,
3. SO₂,
4. CO,

5. O₃,
6. NO₂,
7. HC

Setiap paramater masing-masing memiliki peran signifikan dalam menentukan kualitas udara. Setelah pengumpulan data, tahap berikutnya adalah pelabelan data. Data dibagi menjadi 3 kelas yaitu 'SEDANG : 0', 'BAIK : 1', dan 'TIDAK SEHAT : 2'.



Waktu	PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	HC	Critical Component
00:00:00	12	38	11	29	18	4	0	PM _{2.5}
01:00:00	13	40	12	29	18	4	0	PM _{2.5}
02:00:00	13	41	12	29	18	4	0	PM _{2.5}
03:00:00	13	41	12	29	18	4	0	PM _{2.5}
04:00:00	13	41	12	29	18	4	0	PM _{2.5}
05:00:00	13	42	12	29	17	4	0	PM _{2.5}
06:00:00	13	41	12	29	17	4	0	PM _{2.5}
07:00:00	13	42	12	29	17	4	0	PM _{2.5}
08:00:00	13	42	12	29	17	4	0	PM _{2.5}
09:00:00	13	42	12	29	17	4	0	PM _{2.5}
10:00:00	14	42	13	29	17	4	0	PM _{2.5}
11:00:00	14	42	13	29	16	4	0	PM _{2.5}
12:00:00	13	42	13	29	16	4	0	PM _{2.5}
13:00:00	13	42	13	28	16	4	0	PM _{2.5}
14:00:00	13	42	13	28	16	4	0	PM _{2.5}
15:00:00	13	41	13	27	16	4	0	PM _{2.5}
16:00:00	13	40	13	26	15	4	0	PM _{2.5}
17:00:00	13	39	14	25	15	4	0	PM _{2.5}
18:00:00	12	38	13	24	15	4	0	PM _{2.5}
19:00:00	12	37	13	24	15	4	0	PM _{2.5}
20:00:00	12	37	13	23	14	4	0	PM _{2.5}
21:00:00	12	37	13	23	14	3	0	PM _{2.5}
22:00:00	12	38	13	23	14	3	0	PM _{2.5}
23:00:00	12	38	13	23	14	3	0	PM _{2.5}

Gambar 12 Contoh Data

2.5 Preprocessing Data

Tahap preprocessing menjadi sangat penting karena data yang tidak lengkap atau tidak konsisten dapat menyebabkan bias atau kesalahan dalam analisis dan model klasifikasi. Tujuan dari tahap *preprocessing* adalah untuk mempersiapkan dataset agar lebih sesuai untuk analisis dan pemodelan. Proses ini melibatkan pembersihan data untuk memastikan konsistensi dan keakuratan. Gambar 13 menampilkan source code preprocessing yang digunakan pada penelitian ini.

```
# preprocessing
df_cleaned = df_cleaned.replace('---', np.nan)
df_cleaned = df_cleaned.replace('nan', np.nan)
df_cleaned = df_cleaned.replace('TIDAK ADA DATA', np.nan)

df_cleaned.isnull().sum()
```

Gambar 13 Source Code Preprocessing

Pada penelitian ini, terdapat berbagai cara penandaan untuk data yang hilang, termasuk '---', 'nan', dan 'TIDAK ADA DATA'. Mengganti semua representasi ini dengan np.nan, standar Python untuk nilai yang hilang (NaN – Not a Number). Berikut adalah contoh data yang akan dihilangkan.

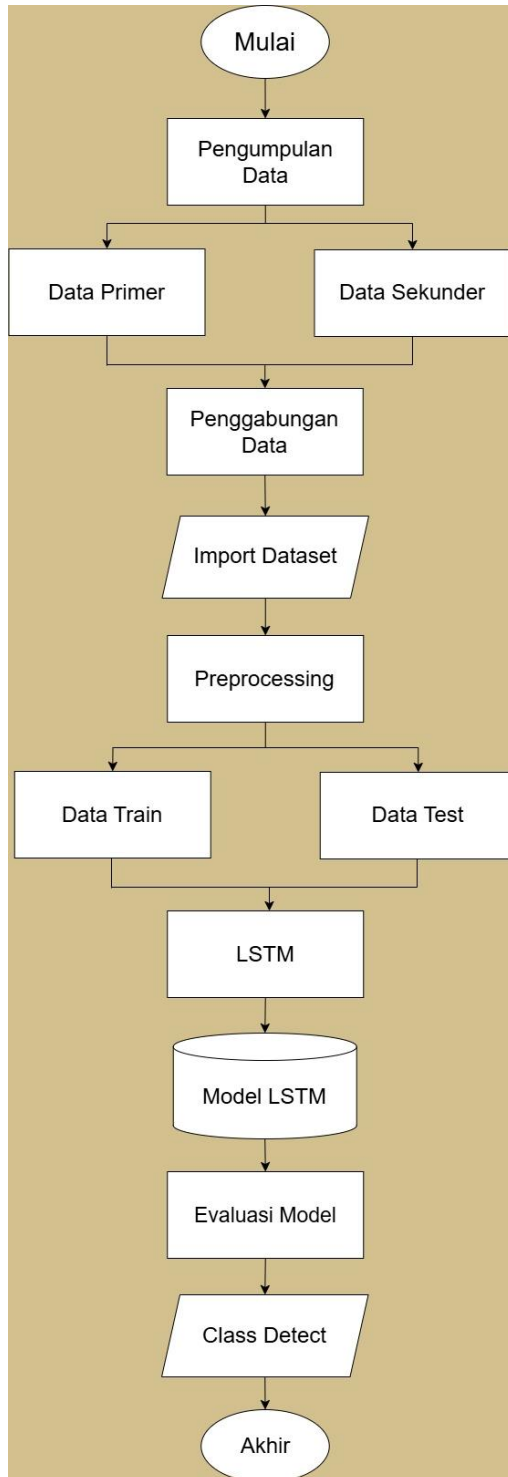
PM ₁₀	PM _{2.5}	SO ₂	CO	O ₃	NO ₂	HC
15	52	11	2	18	0	5
14	50	11	2	18	0	5
12	44	10	2	18	0	5
11	40	10	2	18	0	5
10	39		1	18	0	5
10	38	9	1	18	0	5
10	38	9	1	18	0	5
10	37	9	1	18	0	5
10	37	9	1	18		5
10	39	9	2	18	0	5
10	39	9	2	18	0	5

Gambar 14 Contoh Data yang Dihilangkan

Gambar 14 menunjukkan contoh data yang kosong (berwarna putih), data seperti inilah yang akan dihapus untuk memastikan konsistensi data.

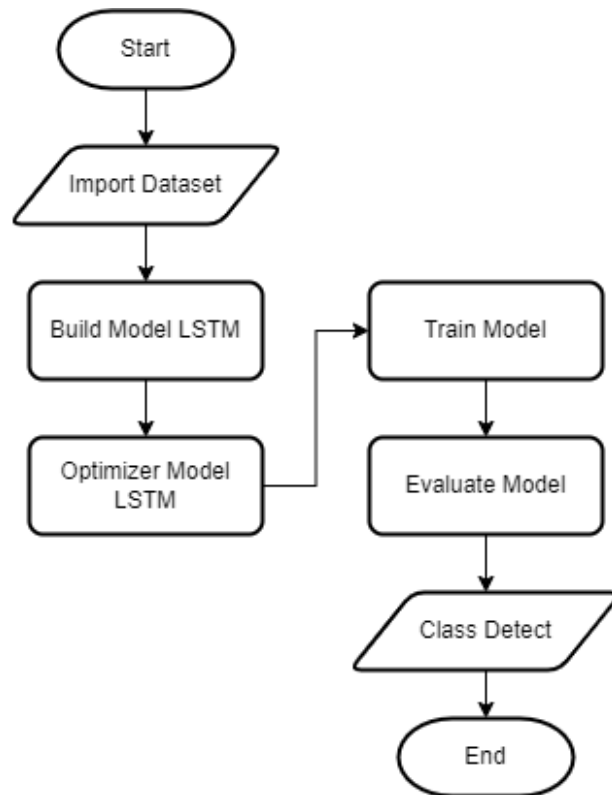
2.6 Perancangan Sistem

Gambar 15 menunjukkan rancangan sistem yang digunakan dalam penelitian ini. Setelah data dikumpulkan dan digabungkan dari data primer dan data sekunder, dataset tersebut diimpor ke dalam sistem untuk dilakukan preprocessing. Proses preprocessing mencakup pembersihan dan transformasi data agar dapat digunakan pada tahap selanjutnya. Setelah preprocessing selesai, dataset kemudian dibagi menjadi dua bagian utama, yaitu data latih dan data uji. Data latih digunakan untuk melatih model LSTM agar mampu mempelajari pola dari dataset. Sementara itu, data uji digunakan untuk mengevaluasi performa model yang telah dilatih dalam melakukan prediksi atau klasifikasi pada data baru. Tahap evaluasi ini memastikan model bekerja dengan akurasi yang diinginkan sebelum digunakan untuk mendeteksi kelas data.



Gambar 15 Rancangan Sistem

LSTM adalah model yang akan digunakan untuk klasifikasi kualitas udara.



Gambar 16 Flowchart model LSTM

Gambar 16 menunjukkan flowchart model LSTM, dimulai dari tahap input dataset, kemudian tahap “*Build Model LSTM*”. Gambar 17 menunjukkan source code yang digunakan untuk membangun model LSTM.

```

# Build LSTM model
model = Sequential()
model.add(LSTM(128, input_shape=(X_train.shape[1], X_train.shape[2])))
model.add(Dense(len(label_encoder.classes_), activation='softmax'))
  
```

Gambar 17 Source code LSTM

Dalam proses membangun model *Long Short-Term Memory* (LSTM), model membuat menggunakan pendekatan Sequential, yang merupakan metode umum dalam pembuatan model neural network dimana lapisan dibangun secara bertahap. Pertama, sebuah lapisan LSTM ditambahkan dengan 128 unit. LSTM adalah jenis *Recurrent Neural Network* (RNN) yang efektif untuk menganalisis data sekuen karena kemampuannya dalam menangkap ketergantungan jangka