BAB II
TINJAUAN PUSTAKA

2.1 Penelitian Sebelumnya

Penelitian yang bertujuan untuk mengklasifikasi suara menggunakan *Deep Learning* sudah dilakukan dan banyak diterapkan di berbagai bidang teknologi di dunia. Dalam penelitian yang sudah dilakukan sebelumnya menunjukkan bahwa *Deep Learning* adalah teknologi yang memanfaatkan konsep berpikir seperti layaknya otak manusia untuk menyelesaikan sebuah masalah, pendekatan ini meningkatkan secara signifikan kinerja sebuah aplikasi termasuk didalamnya pengklasifikasian suara.

Penelitian pada tahun 2019 yang dilakukan oleh Chih-Yuan Koh, Jaw-Yuan Chang, Chiang-Lin Tai, Da-Yo Huang, Han-Hsing Hsieh, dan Yi-Wen Liu yang berjudul "*Bird Sound Classification using Convolutional Neural Networks*" memaparkan tentang algoritma *Convolutional Neural Networks* yang digunakan untuk mengenali 659 spesies burung dari 50.000 rekaman suara. Mereka menggunakan 2 model yaitu *ResNet* dan *Inception*. Mereka mengubah klasifikasi suara burung menjadi klasifikasi gambar yang mana suara burung tersebut dibuah menjadi spectrogram menggunakan *MEL-scale*. Hasilnya model inception mendapat 0.23 *classification mean average precision* (c-mAP), lebih bagus daripada model *ResNet18* yang mendapat 0.13 dan *ResNet34* yang mendapat 0.11.

Penelitian lain yang berjudul "*Speaker identification based on combination of MFCC and UMRT based features*" yang dilakukan oleh Anett Antony dan R. Gopikakumari pada tahun 2018 tentang mengidentifikasi pembicara baik yang bergantung pada teks maupun yang bukan pada kata-kata berbahasa Inggris seperti "*down*", "*up*", "*left*", "*right*", "*start*", "*stop*", dan "*pause*". Hasil dari penelitian tersebut adalah Ketika kombinasi antara MFCC dan UMRT digunakan, rata-rata akurasi naik 3% baik berbicara bergantung pada teks dan tidak dibandingkan etika hanya menggunakan MFCC saja yaitu 87.5 untuk kategori bergantung dan 86.8 untuk kategori tidak bergantung.

Penelitian selanjutnya berjudul "*EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network*" yang dilakukan pada tahun 2019 oleh Jinru Yang, Xiaofan Huang, Hongkai Wu, dan Xingtong Yang. Penelitian ini mengklasifikasikan 4 kategori perasaan seseorang yaitu senang, sedih, takut, dan netral melalui sinyal *Electroencephalogram* (EEG) menggunakan metode *Bidirectional Long Short-Term Memory Network* (BiLSTM) dengan hasil akurasi sebesar 84.21%.

Penelitian lainnya yaitu berjudul "*Heart sound segmentation via Duration Long–Short Term Memory Neural Network*" yang dilakukan oleh Yao Chen, Jiancheng Lv, Yanan Sun, dan Bijue Jia yang dilakukan pada tahun 2020. Penelitian ini mesegmentasi suara jantung yang bertujuan menganalisa penyakit katup jantung dengan cara mendeteksi suara jantung pertama dan kedua pada phonocardiogram menggunakan metode Duration-LSTM (BiLSTM). Hasil dari penelitian ini yaitu rata-rata nilai F1-Score adalah 96.11±0.27%.

Penelitian selanjutnya yang dilakukan oleh Vitor Guedes, Arnaldo Junior, Joana Fernandes, Felipe Teixeira, dan João Paulo Teixeira pada tahun 2018 berjudul "*Long Short Term Memory on Chronic Laryngitis Classification*" mengklasifikasikan penyakit langritis kronis menggunakan dataset berupa suara manusia. Penelitian ini membandingkan antara metode *Long Short Term Memory (LSTM)* dengan metode *Artificial Neural Network (ANN)*. Hasil dari perbandingan ini adalah metode LSTM lebih baik dalam hal akurasi, sensitifitas dan spesifikasi pada *test set* dibandikan dengan metode ANN.

Penelitian berikutnya berjudul "*Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting*" yang dilakukan oleh Xuan-Hien Le , Hung Viet Ho, Giha Lee , dan Sungho Jung pada tahun 2018. Penelitian ini bertujuan untuk memprediksi banjir dengan debit harian dan curah hujan pada bendungan sungai Da di Vietnam digunaka sebagai data input. Dataset dari tahun 1979-1983 digunakan untuk validasi dan dataset pada tahun 1984 untuk tujuan pengujian. Hasil dari penelitian ini adalah Eror Value sekitar 5% dan 14% untuk satu hari dan dua hari kedepan dimana nilai toleransi dapat diterima di prakiraan hidrologi untuk kasus peramalan debit maksimum.

Penelitian pada tahun 2018 yang berjudul "*Abnormal Heart Sound Detection Using Temporal Quasi-periodic Features and Long Short-Term Memory without Segmentation*" yang dilakukan oleh Wenjie Zhanga, Jiqing Hana, dan Shiwen Deng bertujuan untuk diagnosis awal penyakit jantung menggunankan metode *quasi-periodic features* dan *long short-term memory* (LSTM) dengan cara mendeteksi suara jantung yang abnormal. Hasil dari penelitian ini adalah *overall scores* ketika menggunakan metode LSTM network, seperti A^sLW, A^sLT, dan A^sLC, lebih tinggi disbandingkan tanpa LSTM network, seperti AMDF dan $AMDF^s$.

Pada penelitian yang dilakukan oleh Yagya Raj Pandeya, Dongwhoon Kim dan Joonwhoan Lee pada tahun 2018 yang berjudul "*Domestic Cat Sound Classification Using Learned Features from Deep Neural Nets"* memaparkan hasil penelitian mereka berupa pengklasifikasian suara kucing menurut suasana yang sedang dialami kucing. Ada 10 kategori suara kucing yaitu mood normal ("meow-meow"), pertahanan ("*hissing*"),anak kucing memanggil ibunya ("*pilling*"), kucing kesakitan ("miyoou") , sedang istirahat ("*purring*"), peringatan ("*growling*"), perkawinan ("gay-gay-gay"), bertengkar ("nyaaan"), marah ("momo-mooh"), dan ingin berburu ("trilling or chatting"). Mereka membandingkan 2 buah metode yaitu *convolutional deep belief network* (CDBN) dengan keakuratan 0.995 dan *convolutional neural network* (CNN) dengan keakuratan 0.994.

Dari penjelasan diatas, ringkasan penelitian yang relevan ditunjukkan pada Tabel 2.1 di bawah ini :

Tabel 2. 1 Penelitian terdahulu

No	Judul	Penulis, Tahun	Masalah	Algoritma	Hasil
1.	*Bird Sound Classification using Convolutional Neural Networks*	Chih-Yuan Koh, Jaw-Yuan Chang, Chiang-Lin Tai, Da-Yo Huang, Han-Hsing Hsieh, dan Yi-Wen Liu Tahun : 2019	Memprediksi suara burung untuk mengklasifikasi spesies burung yang bertujuan yang dapat bermanfaat untuk konservasi	*Convolutional Neural Networks*	Dari dua model pada validasi data, hasil dari *inception* lebih baik daripada ResNet
2.	*Speaker identification based on combination of MFCC and UMRT based features*	Annet Antony and R. Gopikakumari Tahun : 2018	Mengidentifikasi pembicara yang bergantung pada teks dan tidak bergantung pada teks pada kata-kata berbahasa Inggris	*Mel Frequency Cepstrum Coefficients* (MFCC) dan *Unique Mapped Real Transform* (UMRT)	Hasil akurasi rata-rata dari kombinasi MFCC dan UMRT naik 3% baik pada kategori bergantung pada text dan tidak dibandingkan dengan koefisien MFCC sendiri
3.	*EEG-based emotion classification based on Bidirectional Long Short-Term Memory Network*	Jinru Yang, Xiaofan Huang, Hongkai Wu, Xingtong Yang Tahun : 2019	Mendeteksi perasaan manusia dengan benar melalui sinyal EEG (*Electroencephalogram*)	*Bidirectional Long Short-Term Memory Network*	Dapat mengklasifikasikan 4 kategori perasaan manusia (senang, sedih, takut, dan netral) dengan akurasi 84.21%

No	Judul	Penulis, Tahun	Masalah	Algoritma	Hasil
4.	*Heart sound segmentation via Duration Long–Short Term Memory Neural Network*	Yao Chen, Jiancheng Lv, Yanan Sun, dan Bijue Jia Tahun : 2020	Deteksi suara jantung yang pertama dan kedua pada fonokardiogram untuk analisis penyakit katup jantung	*Duration Long–Short Term Memory network (Duration LSTM)*	F1 Score dari Duration-LSTM mencapai rata-rata 96.11 ±0.27%
5	*Long Short Term Memory on Chronic Laryngitis Classification*	Vitor Guedes, Arnaldo Junior, Joana Fernandes, Felipe Teixeira, dan João Paulo Teixeira Tahun : 2018	Klasifikasi laringitis kronis menggunakan suara manusia. Dataset suara dibagi menjadi 2 kategori yaitu orang sehat dan sakit langritis kronis.	*Long Short Term Memory (LSTM)* dan *Artificial Neural Network (ANN)*	LSTM lebih baik dalam hal akurasi, sensitifitas dan spesifikasi pada *test set* dibandingkan dengan ANN.
6	*Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting*	Xuan-Hien Le , Hung Viet Ho, Giha Lee , dan Sungho Jung Tahun : 2019	Memprediksi Banjir pada daerah sekitar Bendungan Da Vietnam dengan menggunakan data input dari debit harian dan curah hujan.	*Artificial Neural Network (ANN), Recurrent Neural Network (RNN),*dan *Long Short-Term Memory (LSTM) Neural Network*	*Eror Value* sekitar 5% dan 14% untuk satu hari dan dua hari kedepan dimana nilai toleransi dapat diterima di prakiraan hidrologi untuk kasus peramalan debit maksimum.

No	Judul	Penulis, Tahun	Masalah	Algoritma	Hasil
7	*Abnormal Heart Sound Detection Using Temporal Quasi-periodic Features and Long Short-Term Memory without Segmentation*	Wenjie Zhanga, Jiqing Hana, dan Shiwen Deng Tahun : 2018	Mendeteksi suara jantung yang abnormal untuk diagnosis awal penyakit jantung	*quasi-periodic features* dan *long short-term memory*	*Overall scores* ketika menggunakan metode LSTM network, seperti $A^{s}LW$, $A^{s}LT$, dan $A^{s}LC$, lebih tinggi disbandingkan tanpa LSTM network, seperti AMDF dan $AMDF^{s}$
8.	*Domestic Cat Sound Classification Using Learned Features from Deep Neural Nets*	Yagya Raj Pandeya, Dongwhoon Kim , dan Joonwhoan Lee Tahun : 2018	Klasifikasi suara kucing domestik yang terdiri dari 10 kategori yaitu kesakitan, istirahat, peringatan, marah, pertahanan, bertengkar, senang, berburu, panggilan kawin, dan memanggil induknya.	*Convolutional Neural Network* (CNN) dan *Unsupervised Convolutional Deep Belief Network* (CDBN)	Akurasi CNN = 90.80% Akurasi CDBN = 91.13% F1-Score CNN = 0.91 F1-Score CDBN = 0.91 AUC Score CNN = 0.994 AUC Score CDBN = 0.995

Berdasarkan penelitian terdahulu dapat disimpulkan bahwa penelitian menggunakan metode CNN dan LSTM masih relevan digunakan untuk mengklasifikasikan suara kucing dikarenakan metode CNN dan Metode LSTM dapat memberikan hasil prediksi yang akurat dengan membutuhkan parameter

yang tepat [7]. Penelitian sebelumnya yang menjadi literatur utama adalah penelitian nomor 8 pada tabel karena memiliki tujuan yang sama. Perbedaan terletak pada dataset studi kasus dan juga algoritma yang digunakan.

2.2 Dasar Teori

2.1.1 Suara Kucing

Kucing berkomunikasi melalui suara dengan sesama kucing ataupun manusia. Suara hewan pada saat komunikasi bergantung terhadap lingkup pada saat suara itu dikeluarkan. Suara kucing dikategorikan menjadi 3[3], yaitu :

1. Suara ketika mulut tertutup (*purring, trilling*)
   a. *The Purr* (mendengkur)

      Kucing melakukan *purr* ketika dalam konteks lapar, stress, dan kesakitan baik itu pada saat melahirkan atau ketika akan meninggal. Ibu kucing dan bayinya juga sering berkomunikasi dengan purr dikarenakan sulitnya dideteksi oleh predator.

   b. *The Trill* (getar)

      Trill digunakan ketika menyapa dan mendekat secara ramah, ketika bermain, dan juga terkadang sebagai sebuah konfirmasi seperti “ya”.
2. Suara ketika mulut terbuka kemudian menutup secara perlahan (*meowing, howling, yowling*)
   a. *The Meow* (meong)

      The Meow paling sering digunakan ketika dengan manusia yang memiliki beberapa arti yang berbeda seperti meminta perhatian, stress, atau meminta sesuatu.

   b. *The Howl* (melolong)

      The Howl berdurasi sangat panjang dan sering diulang-ulang. Digunakan sebagai sinyal peringatan pada saat situasi defensive dan juga aggressive.

c. *The Mating Call* (panggilan kawin)

The Mating Call suaranya terkadang mirip seperti anak kecil yang sedang menangis.

3. Suara ketika mulut terbuka dengan posisi yang sama (*growling, snarling, hissing, spitting, chattering, chirping*)

a. *The growl* (menggeram)

The growl berdurasi panjang yang biasanya digunakan sebagai tanda bahaya atau memperingati atau menakuti musuh.

b. *The Hiss* (mendesis) dan *The Spit* (meludah)

The hiss sering dipakai sebagai peringatan atau terkadang sebagai reaksi terkejut ketika ada kemunculan musuh.

c. *The Snarl* (menggeram) - *Scream, Cry, Pain Shriek*

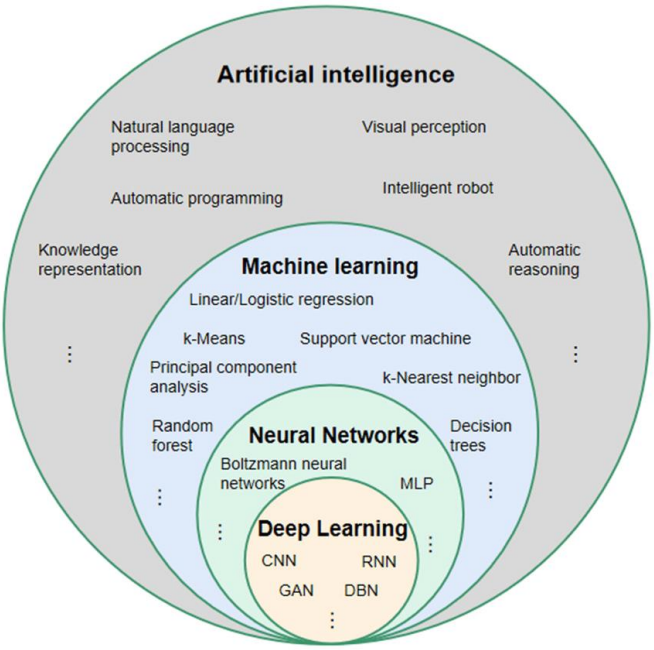
The snarl biasanya sangat keras, terkadang digunakan sebagai peringatan terakhir, tetapi kucing yang sedang terluka atau sakit juga dapat menangis (*cry*) ketika mereka sedang merasakan kesakitan.

d. *The Chirp* (kicauan) dan *The Chatter* (obrolan)

The Chirp dan *The Chatter* merupakan suara yang biasa dihasilkan disekitar mangsa. Kucing mencoba meniru suara dari mangsa.

2.1.2 Deep Learning

Deep learning adalah salah satu cabang Machine Learning (ML) [9] [10] yang menggunakan Deep Neural Network untuk menyelesaikan permasalahan pada bidang Machine Learning. Neural Network adalah model yang terinspirasi oleh bagaimana neuron dalam otak bekerja. Deep learning memanfaatkan dataset yang besar untuk menyelesaikan permasalahan dengan menggunakan jaringan syaraf tiruan yang tersusun oleh beberapa hidden layer, lapisan tersebut merupakan suatu algoritma yang dapat mengklasifikasikan perintah yang diinputkan sehingga menghasilkan keluaran yang diharapkan[11].



Gambar 2. 1 Deep Learning dalam Machine Learning
(Sumber : researchgate.net)

Letak deep learning dalam machine learning dan artificial intelligence dapat dilihat pada Gambar 2.1. Metode CNN dan LSTM yang digunakan dalam penelitian ini masuk ke dalam deep learning.

2.1.3 Metode CNN

Convolutional Neural Network (CNN) merupakan salah satu jenis neural network yang biasa digunakan dalam memproses data[16] yang terdiri dari beberapa lapisan [17]

Lapisan yang ada di CNN adalah sebagai berikut :

- *Convolutional Layer*

Proses konvolusi memanfaatkan apa yang disebut sebagai filter. Seperti layaknya gambar, filter memiliki ukuran tinggi, lebar, dan tebal tertentu. Filter ini diinisialisasi dengan nilai tertentu, dan nilai dari filter inilah yang menjadi parameter yang akan di-updat edalam proses learning[19].

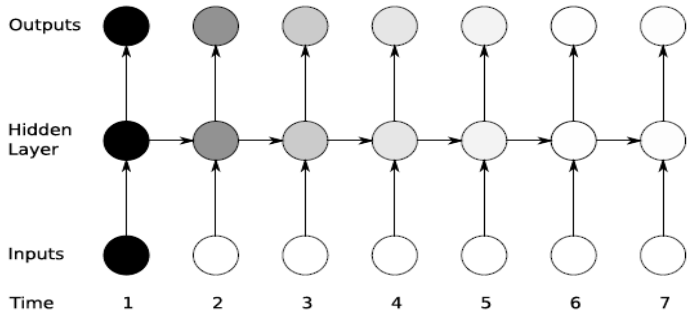
- *Dense*

Pada fase ini, fitur-fitur yang sudah memanjang setelah keluar dari fase LSTM, maka dilakukanlah pengerucutan fitur hingga menjadi beberapa kelas saja sebagai output penentuan klasifikasi gambar[18][19].

2.1.4 Metode LSTM

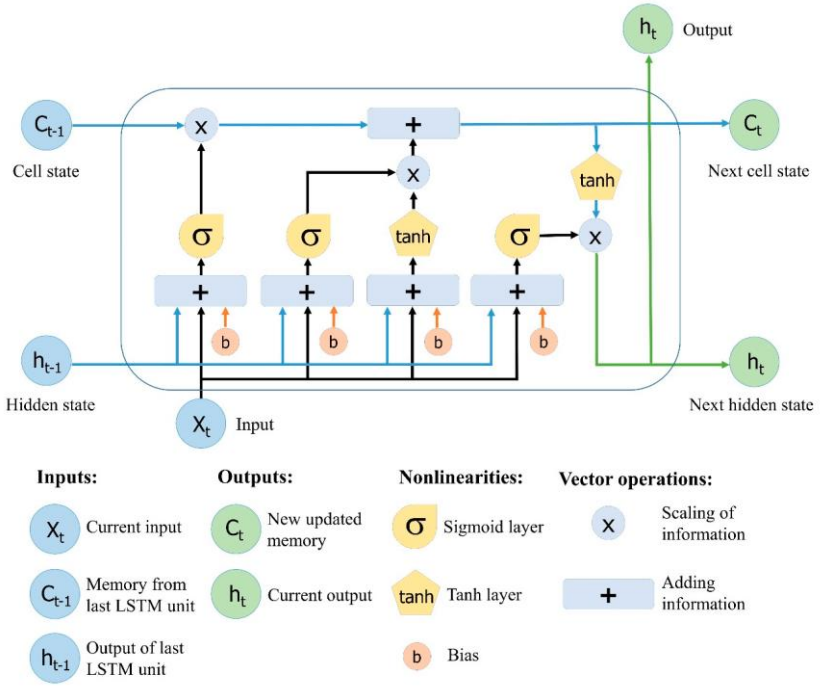
LSTM (Long Short Term Memory) adalah jenis model pemrosesan yang merupakan evolusi dari RNN[12]. LSTM diciptakan oleh Hochreiter & Schmidhuber (1997) yang bertujuan untuk memecahkan masalah gradien yang hilang dari backpropagation melalui algoritma waktu pada RNN[5][6].

Informasi yang relevan untuk kejadian yang akan datang di ekstrak oleh RNN yang disimbolkan oleh lingkaran hitam[8]. Seiring berjalannya waktu dan kejadian baru diproses, informasi tersebut hilang yang seharusnya dipakai untuk jangka panjang[15].



Gambar 2. 2 Gradien pada RNN
(Sumber : researchgate.net)

LSTM mempunyai kemampuan untuk belajar dalam jangka Panjang dan mengingat informasi untuk waktu yang lama[6] [13]. LSTM berbentuk seperti struktur rantai, walaupun modulnya berulang namun memiliki struktur yang berbeda[6] LSTM memilki 3 unit, yaitu gerbang input, gerbang lupa, dan gerbang output.



Gambar 2. 3 Stuktur LSTM

(Sumber : researchgate.net)

Jaringan LSTM tersusun dari blok memori yang disebut sel. *State* ditransfer ke sel berikutnya, *cell state* dan *hidden state*. *State cell* adalah rantai utama pada aliran data, yang memungkinkan data mengalir ke depan dengan data yang tidak berubah, tetapi beberapa transformasi linier bisa terjadi. Data dapat ditambahkan atau dihapus dari cell state melalui gerbang sigmoid. Sebuah gerbang mirip dengan serangkaian operasi matriks, yang berisi bobot individu yang berbeda. LSTM dirancang untuk menghindari masalah ketergantungan jangka panjang karena menggunakan gerbang untuk mengontrol proses menghafal [6].

Langkah pertama dalam membangun jaringan LSTM adalah mengidentifikasi informasi yang tidak diperlukan dan akan dihilangkan dari sel pada langkah itu. Proses mengidentifikasi dan mengecualikan data ini ditentukan oleh fungsi sigmoid, yang mengambil keluaran dari unit LSTM terakhir (h_{t-1}) pada waktu t-1 dan arus input (X_t) pada waktu t. Fungsi sigmoid juga menentukan bagian mana dari keluaran lama

yang harus dihilangkan. Gerbang ini disebut gerbang lupa (atau kaki) di mana f_t adalah vektor dengan nilai yang berkisar dari 0 hingga 1, sesuai dengan setiap angka yang berada dalam status sel, C_{t-1}. [6]

$$f_t = \sigma\,(W_f\,[h_{t-1}, X_t] + b_f\,) \qquad (2.1)$$

Ini adalah fungsi sigmoid (σ), dan W_f dan b_f masing-masing adalah matriks bobot dan bias dari gerbang lupa.

Langkah selanjutnya adalah memutuskan dan menyimpan informasi dari input baru (X_t) dalam keadaan sel serta untuk memperbarui status sel. Langkah ini terdiri dari dua bagian, yang pertama lapisan sigmoid dan yang kedua adalah lapisan tanh. Pertama, lapisan sigmoid memutuskan apakah informasi baru harus diperbarui atau diabaikan (0 atau 1), dan kedua, fungsi tanh memberi bobot pada nilai yang dilewati untuk menentukan level penting (-1 sampai 1). Kedua nilai dikalikan agar status dari sel baru dapat diperbarui. Memori baru ini kemudian ditambahkan ke memori lama C_{t-1} menghasilkan C_t. [6]

$$i_t = \sigma\,(Wi\,[h_{t-1}, X_t] + b_i), \qquad (2.2)$$

$$N_t = \tanh(W_n\,[h_{t-1}, X_t] + b_n), \qquad (2.3)$$

$$C_t = C_{t-1}\,f_t + N_t\,i_t \qquad (2.4)$$

C_{t-1} dan C_t adalah keadaan sel ketika waktu t-1 dan t, sedangkan W dan b adalah matriks bobot dan bias masing-masing dari keadaan sel [6].

Langkah terakhir adalah nilai keluaran (h_t) didasarkan pada status sel keluaran (O_t) tetapi merupakan versi yang difilter. Langkah pertama yaitu lapisan sigmoid memutuskan bagian mana dari status sel yang menghasilkan keluaran. Selanjutnya, keluaran dari gerbang sigmoid (O_t) dikalikan dengan nilai baru yang dihasilkan oleh lapisan tanh dari keadaan sel (C_t), dengan nilai berkisar antara -1 dan 1.

$$O_t = \sigma\,(W_o\,[h_{t-1}, X_t] + b_o), \qquad (2.5)$$

$$h_t = Ot\,\tanh(C_t). \qquad (2.6)$$

Di sini, W_o dan b_o masing-masing adalah matriks bobot dan bias dari gerbang keluaran.[6]

2.1.5 Optimisasi

Optimisasi berfungsi untuk mengoptimalkan proses pembelajaran pada sistem [20] menggunakan nilai *learning rate* tertentu untuk menentukan kemampuan sistem belajar secara cepat atau lambat. Penelitian ini menggunakan optimasi *Adaptive Moment Estimation* (ADAM) yang mempunyai nilai *learning rate* 0,001. ADAM menghitung tingkat pembelajaran secara adaptif untuk setiap parameter dengan cara menyimpan rata-rata gradien secara eksponensial [20][21] dari gradien masa lalu yang mewakili momen pertama (*mean*) dan gradien kuadrat masa lalu yang mewakili momen kedua (*varians*)[21]. Rumus perhitungan ADAM dapat dilihat pada persamaan 2.7

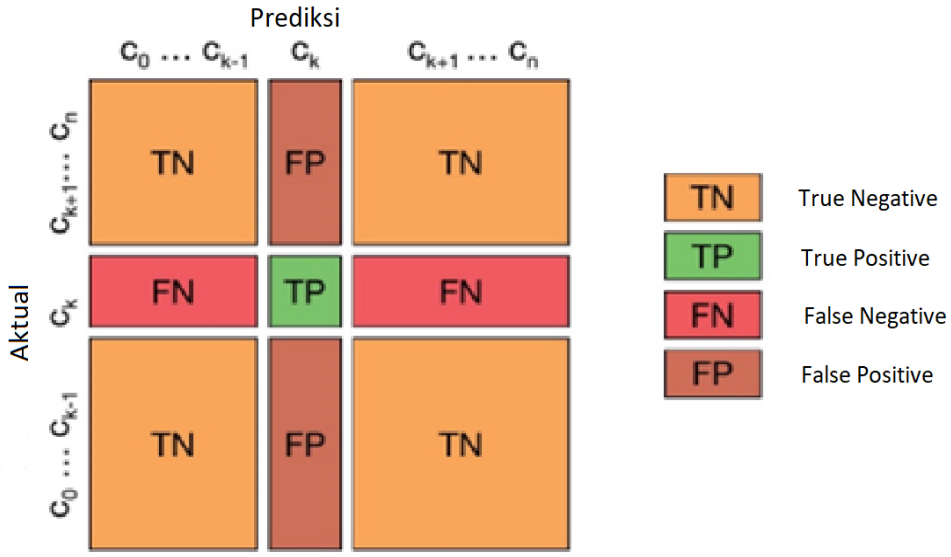
$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t}+\varepsilon} \cdot \hat{m}_t \qquad (2.7)$$

dengan θ_{t+1} adalah parameter hasil pembaruan, θ_t adalah parameter hasil pembaruan sebelumnya, η adalah learning rate, $\hat{m}_t$ adalah gradien kuadrat momen orde pertama, $\hat{v}_t$ gradien kuadrat momen orde kedua, dan ε merupakan scalar kecil untuk mencegah pembagian dengan nol. Persamaan (2.7) menunjukkan perhitungan optimasi Adam dalam memperbarui nilai error dalam proses pelatihan dengan memanfaatkan nilai gradien pada momen orde pertama dan orde kedua [21].

2.1.6 Confusion Matrix

Confusion matrix adalah sebuah visualisasi yang umum yang digunakan dalam mengevaluasi performa dari suatu model klasifikasi

pada algoritma supervised learning. Confusion Matrix berisikan informasi yang aktual dan prediksi pada sistem klasifikasi yang telah dibuat. Berikut merupakan tabel confusion matrix**.**



Gambar 2. 4 Confusion Matrix

(Sumber : researchgate.net)

True Positive (TP) dan True Negative (TN) menandakan jumlah kelas positif dan jumlah kelas negatif yang dikategorikan secara benar, sedangkan False Positive (FP) dan False Negative (FN) menandakan jumlah kelas positif dan jumlah kelas negatif yang tidak dikategorikan secara benar [14]. Berdasarkan confusion matrix tersebut dapat ditetapkan tolak ukur performa seperti *Accuracy, Precision, Recall, Specificity, FMeasure, G-Mea*n dan yang lainnya [14].