**PREDIKSI PASAR SAHAM MENGGUNAKAN ALGORITMA DEEP LEARNING DENGAN MODEL BILSTM**

****

TESIS

Diajukan sebagai salah satu syarat untuk memperoleh gelar

Magister Ilmu Komputer (M.Kom)

Okki Setyawan

14002429

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER**

**NUSA MANDIRI**

**2021**

# BAB 1

**PENDAHULUAN**

## Latar Belakang Penulisan

Saham merupakan satuan nilai dalam berbagai *instrument* finansial atau keuangan yang mengacu pada kepemilikan sebuah perusahaan. Saat ini saham menjadi salah satu metode utama untuk meningkatkan modal bisnis karena memungkinkan perusahaan yang membutuhkan pendanaan jangka panjang untuk menjualnya kepada investor dengan bentuk imbalan berupa uang tunai. Saham dapat dijual melalui pasar primer maupun pasar sekunder. Pada era ini investasi saham menjadi pilihan investasi yang diminati karena memberikan keuntungan yang cepat dan besar [1]. Salah satu pasar keuangan yang memiliki fungsi perantara adalah pasar modal yang merupakan pasar instrumen keuangan jangka Panjang, yang memperjual belikan hutang maupun modal sendiri yang diterbitkan pemerintah maupun perusahaan swasta [2]. Pasar modal atau saham memiliki peranan yang penting terhadap perekonomian suatu negara dikarenakan pasar modal memiliki dua fungsi, yaitu fungsi ekonomi dan fungsi finansial atau keuangan. Dalam fungsi ekonomi, pasar modal menyediakan fasilitas untuk mempertemukan dua kepentingan, yaitu pihak yang memiliki kelebihan dana atau disebut investor dan pihak yang memerlukan dana. Dengan adanya pasar modal, pihak yang memiliki kelebihan dana dapat menginvestasikan dana tersebut dengan harapan memperoleh keuntungan, sedangkan perusahaan dapat memanfaatkan dana tersebut untuk kepentingan investasi tanpa menunggu tersedianya dana operasional perusahaan. Dalam fungsi finansial atau keuangan, pasar modal memberikan kemungkinan dan kesempatan memperoleh keuntungan bagi investor, sesuai dengan karakteristik investasi yang di pilih [3].

investor yang baru mengambil bagian didalam pasar modal tentunya harus memiliki beberapa pengetahuan tentang pasar modal dengan teknik analisis fundamental maupun teknikal, agar tidak mengalami kerugian saat berinvestasi dipasar modal. Untuk mendapatkan keuntungan yang maksimal di pasar modal investor harus selalu memantau pasar saham dikarenakan hal ini perlu dilakukan untuk menghindari resiko perubahan harga saham yang begitu berfluktuasi dan cepat. Investor dapat menggunakan suatu wadah investasi bernama reksadana untuk mempermudah investor dalam melakukan investasi dan memperoleh keuntungan yang lebih baik [4]. Sebagian besar investor lebih memilih untuk melihat berita-berita mengenai pasar modal dan saham dari media elektronik yang telah tersedia di internet. Berita-berita tersebut mengandung rekomendasi pembelian saham, prediksi saham dalam waktu dekat, nilai IHSG, tips-tips investasi maupun posisi saham yang cenderung menguat atau melemah (naik atau turun) [5].

Pada penelitian sebelumnya telah dilakukan penelitian dengan menggunakan dataset public dari Kaggle (*daily news* DJIA) sebagai dataset yang diujikan dengan metode LSTM serta klasifier Word2vec dan N-gram serta hyperparamter tuning dengan 4 layer dense dan LSTM *deep learning* LSTM yang menghasilkan akurasi sebesar 0.5212 % dan F1 score sebesar 0.6762 % , serta AUC 0.6211 %.

Pada penilitan ini, peneliti menggunakan dataset dari Kaggle menggunakan model deep neural network dengan variasi dari network layer dan hyperparameter tunning 5 layer (ditulis kontribusi disini) Dense & LSTM serta activation function RELU oleh karena itu penulis akan melakukan penelitian kembali dengan *improvement* terbaru dengan model Deep Learning BiLSTM untuk meningkatkan akurasi pengklasifikasian pasar saham dari penelitian sebelumnya.

## Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, terlihat dari hasil penelitian sebelumnya , saat ini penulis akan meneliti bagaimana impementasi, kinerja serta hasil model BiLSTM dengan python untuk mengklasifikasikan Stock Market.

## Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai syarat kelulusan Program Magister Strata Dua (S2) di STMIK Nusa Mandiri.

Selain itu, penelitian ini bertujuan untuk meningkatkan akurasi dari model sebelumnya serta untuk mengetahui proses dan hasil yang didapatkan dari model BiLSTM.

## Ruang Lingkup Penelitian

Agar penelitian ini lebih terarah, maka dilakukan pembatasan ruang lingkup permasalahan dengan menerapkan algoritma *Deep Learning*. Selanjutnya melakukan evaluasi model untuk mengukur tingkat keakuratan dengan matriks accuracy, *F1 Score dan AUC*  menggunakan Python .dan klasifikasi stock market hanya ada pada 2 kelas yaitu naik atau bertahan dan turun.

## Hipotesis

H0. Algoritma Deep Learning diduga dapat digunakan untuk mengklasifikasikan posisi stock market.

H1. Aalgoritma deep learning dapat menghasilkan akurasi terbaik untuk mengklasifikasikan posisi stock market.

## Kontribusi Penelitian

Prediksi stock market yang menjadi penelitian penulis saat ini telah mengalami perubahan dengan hyperparameter dan tuning sehingga menghasilkan kenaikan sebesar 2% dari penelitian sebelumnya yang menggunakan model LSTM dengan konfigurasi 4 hidden layer, layer Dense dan LSTM , activation RELU yang menghasilkan Accuracy 0.5211, F1-Score 0.6762 serta AUC 0.6211. Model yang peneliti gunakan saat ini sebagai kontribusi pada penelitian ini dengan model BILSTM dengan konfigurasi 5 hidden layer, Layer Dense dan BIILSTM, Kernel Uniform, Activation RELU, Learning Rate 0.00005, Epoch 20, Batch Size 16 yang menghasilkan Accuracy 0.54232, F1-Score 0.543625, serta AUC 0.54322.

## Sistematika Penulisan

Pada bab ini berisi tentang kesimpulan dari pembahasan pada bab- bab sebelumnya dan saran untuk penelitian selanjutnya

**BAB I PENDAHULUAN**

Bab ini menjelaskan tentang latar belakang, identifikasi masalah, tujuan masalah, ruang lingkup penelitian, hipotesis, dan sistematika penulisan yangdidasarkan pada fakta-fakta terkait permasalahan yang akan dibahas dalam penelitian tesis ini.

**BAB II LANDASAN/KERANGKA PEMIKIRAN**

Bab ini membahas tentang landasan-landasan teoritis yang digunakan untuk melaksanakan penelitian yang mencakup tinjauan pustaka, tinjauan studi dan tinjauan organisasi.

**BAB III METODOLOGI PENELITIAN**

Bab ini membahas tentang metode penelitian yang akan digunakan dalam penelitian di dalam tesis ini mencakup perencanaan metode penelitian dan konsep dari penelitian yang nantinya akan digunakan sebagai solusi penyelesaian masalah yang diangkat dalam penelitian ini.

**BAB IV HASIL PENELITIAN DAN PEMBAHASAN**

Pada bab ini berisi tentang analisis dan pembahasan yang dilakukan yaitu dimulai dari pengambilan data, pengolahan data dengan model BiLSTM dan penggunaan preprocessing data menggunakan TF – IDF dan Word2Vec.

**BAB V PENUTUP**

Pada bab ini berisi tentang kesimpulan dari pembahasan pada bab- bab sebelumnya dan saran untuk penelitian selanjutnya.

# BAB 2

**LANDASAN/KERANGKA PEMIKIRAN**

### Tinjauan Pustaka

Untuk mendapatkan kesamaan persepsi mengenai pembahasan pada penelitian ini maka perlu penjelasan secara teoritis berkaitan dengan definisi atau pengertian kanker payudara (objek kajian), algoritma *deep learning* sebagai metode analisis, dan metode evaluasi model menggunakan akurasi, *presisi, recall* dan *f1 score.* Referensi diambil dari jurnal nasional dan internasional yang berhubungan dengan judul penelitian.

* + 1. **Stock Market**

Pasar saham atau disebut pasar modal merupakan wadah untuk mencari dana bagi perusahaan dan alternatif sarana investasi bagi masyarakat (investor) dimana di dalamnya terdapat transaksi penawaran umum dan perdagangan efek dari perusahaan publik (emiten) kepada masyarakat investor. Pasar Modal atau sebutan lain seperti Capital Market/Stock Exchange/Stock Market/ dalam pengertian klasik diartikan sebagai suatu bidang usaha perdagangan surat-surat berharga seperti saham, sertifikat saham, dan obligasi atau efek-efek pada umumnya [6].

* + 1. **Machine Learning**

Pembelajaran mesin (ML) adalah studi ilmiah tentang algoritma dan model statistik yang digunakan sistem komputer untuk melakukan tugas tertentu tanpa diprogram secara eksplisit [7].

* + 1. **Deep Learning**

Deep Learning (DL) adalah teknik dalam NN yang menggunakan teknik tertentu seperti Restricted Boltzmann Machine (RBM) untuk mempercepat proses pembelajaran dalam NN yang menggunakan lapis yang banyak atau lebih dari 7 lapis. Dengan adanya DL, waktu yang dibutuhkan untuk training akan semakin sedikit karena masalah hilangnya gradien pada propagasi balik akan semakin rendah [8].

* + 1. **Metode Preprocessing**
    2. **Word2vec**

Word2Vec adalah model untuk merepresentasikan kata-kata sebagai vektor. Kemudian, rumus kesamaan cosinus dapat digunakan untuk menghasilkan nilai kesamaan dari nilai vektor kata yang dihasilkan oleh model Word2Vec. Dalam proses membangun model Word2Vec yang disebut proses *training*, terdapat beberapa karakteristik dari model Word2Vec yang dihasilkan, antara lain konfigurasi ukuran jendela dan dimensi vektor. Beberapa penelitian sebelumnya sebagian besar menggunakan konfigurasi ukuran Windows dan dimensi vektor untuk menghasilkan model Word2Vec. Dalam penelitian ini, beberapa konfigurasi ukuran Windows dan dimensi vektor digunakan untuk membandingkan nilai kesamaan dari setiap konfigurasi model Word2Vec yang dihasilkan. Konfigurasi model Word2Vec yang menghasilkan nilai kemiripan terbaik akan menjadi hasil dari penelitian ini. Penting untuk memahami konfigurasi terbaik dari model Word2Vec untuk menemukan nilai kemiripan kata terbaik [9].

* + 1. **Arsitektur Deep Learning**
       1. Neural Network

Neural Network (NN) pertama kali diusulkan oleh McCulloch dan Pitts (1943). Jaringan saraf tiruan dikembangkan sebagai model matematika, mirip dengan cara berpikir jaringan saraf manusia atau biologis, dan mengasumsikan:

1. Pemrosesan informasi terjadi di banyak elemen sederhana yang disebut neuron.

2. Sinyal yang ditransmisikan antar neuron menggunakan hubungan tertentu.

3. Setiap link antar neuron memiliki bobot yang sesuai dengan sinyal yang dikirimkan. Beban ini dapat diperkuat atau

Kurangi sinyal.

4. Setiap neuron menggunakan fungsi aktivasi untuk input (jumlah sinyal input Weighted) untuk menentukan sinyal keluaran.

Seperti otak manusia, jaringan saraf juga terdiri dari beberapa neuron

Hubungan antara neuron-neuron ini. Neuron ini mengubah informasi yang diterima melalui koneksi keluar ke neuron lain. Dalam jaringan saraf, hubungan ini disebut bobot [10].

* + - 1. Backpropagation

Algoritma backpropagation merupakan salah satu metode jaringan syaraf tiruan yang sangat handal dalam memecahkan masalah. Algoritma tersebut merupakan algoritma pembelajaran terawasi, Biasanya digunakan oleh perceptron dengan layar berlapis-lapis untuk mengubah bobot di lapisan tersembunyinya. Backpropagation adalah algoritma iteratif yang sederhana dan mudah diimplementasikan, bahkan ketika berhadapan dengan data yang kompleks, biasanya dapat bekerja dengan baik. Algoritma ini sangat berguna dalam berbagai aplikasi, seperti pengenalan, pemilihan lokasi, dan evaluasi akhir *training* metode rambat mundur meliputi 3 tahap, yaitu rambat maju, rambat mundur, dan modifikasi bobot. Karakteristik backpropagation melibatkan tiga lapisan: lapisan input, yang memasukkan data ke dalam jaringan; lapisan tersembunyi, yang memproses data; dan lapisan output,Di mana hasil input yang diberikan oleh lapisan input [11].

* + - 1. Loss functions

Fungsi *loss function* digunakan untuk penggunaan *machine learning* seperti regresi dan klasifikasi [12]. Loss Function atau Cost Function merupakan fungsi dari neural network yang menggambarkan kerugian yang terkait dengan semua kemungkinan yang dihasilkan oleh model. Loss Function bekerja ketika model machine learning memberikan kesalahan yang harus diperhatikan. Loss Function yang baik adalah fungsi yang menghasilkan error yang fading rendah. Ketika suatu model memiliki kelas yang cukup banyak, perlu adanya cara untuk mengukur perbedaan antara . probabilitas hasil hipotesis dan probabilitas kebenaran yang asli, dan selama proses training banyak algoritma yang dapat menyesuaikan parameter sehingga perbedaan ini diminimalkan [13].



Keterangan:

H = loss function

p = Hasil akhir

q = hasil prediksi

* + 1. **Hyperparameter Deep Learning**
       1. Neuron

Neuron merupakan bagian pengolah informasi yang merupakan dasar dari proses sebuah Jaringan Saraf Tiruan [14]. Dijelaskan juga ada tiga elemen dasar dari model saraf yaitu:

1.Satu set dari sinapsis, atau penghubung yang masing-masing digolongkan oleh bobot atau kekuatannya.

2.Sebuah penambah untuk menjumlahkan sinyal-sinyal input.

Ditimbang dari kekuatan sinaptik masing-masing neuron.

3.Sebuah fungsi aktivasi untuk membatasi amplitudo output dari neuron.

Fungsi ini bertujuan membatasi jarak amplitude yang diperbolehkan oleh sinyal output menjadi sebuah angka yang terbatas.

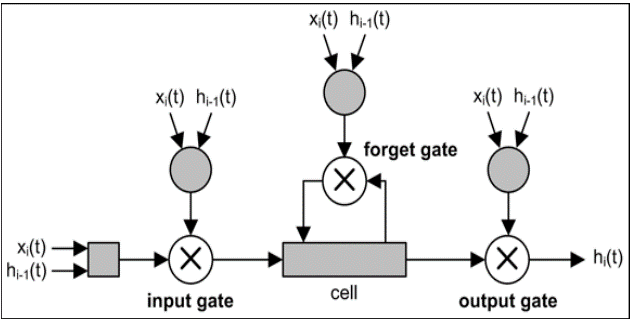
* + - 1. Layer

1. Dense Layer

Dense layer adalah layer yang terkoneksi seluruhnya dari input layer sebelumnya [15]. Dense Layer digunakan untuk menjalankan Fully Connected Layer [16].

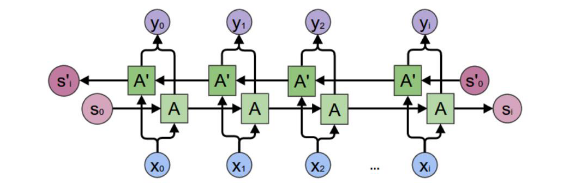
1. LSTM Layer

LSTM adalah arsitektur jaringan saraf berulang spesifik yang dirancang untuk memodelkan urutan temporal. LSTM memiliki dependensi jarak jauh yang membuat LSTM lebih akurat daripada RNN konvensional. Tidak seperti RNN, LSTM berisi unit khusus yang disebut blok memori di lapisan tersembunyi berulang. Blok memori berisi sel memori dengan koneksi sendiri yang menyimpan status temporal jaringan di samping unit perkalian khusus yang disebut gerbang untuk mengontrol aliran informasi. Selain itu, arsitektur LSTM modern berisi koneksi lubang intip dari sel internalnya ke gerbang di sel yang sama untuk mempelajari waktu keluaran yang tepat [17].



1. BiLSTM Layer

BiLSTM adalah perpanjangan dari model LSTM Ada dua lapisan yang prosesnya saling berlawanan.Model ini sangat cocok untuk mengidentifikasi pola dalam kalimat, karena setiap kata dalam dokumen diproses secara berurutan, karena jika Anda mempelajari setiap kata secara berurutan, Anda dapat memahami tweet. Lapisan di bawahnya bersifat maju (forward), yaitu pemahaman dan pengolahan dari kata pertama hingga kata terakhir, sedangkan lapisan di atasnya bersifat mundur (mundur), yaitu dari kata terakhir hingga kata pertama Memahami dan menangani dengan. Dengan dua lapisan yang berlawanan arah ini, model dapat memahami dan melihat melalui kata sebelumnya dan kata pemandu, sehingga membuat proses pembelajaran lebih dalam, yang berdampak pada pemahaman model yang lebih baik tentang konteks dalam tweet gambar. [18]



1. Deep LSTM Layer
2. Deep BiLSTM Layer
   * + 1. Activation Function
3. ReLu
4. Sigmoid
   * + 1. Dropout
       2. Epoch
       3. Batch Size
     1. **Parameter Optimasi**
        1. Optimizer
        2. Learning Rate
     2. **Metrics Performance**
        1. Accuracy Score
        2. F1-Score
        3. AUC Score
   1. **Tinjauan Studi**
   2. **Tinjauan Organisasi**

# BAB 3

**METODOLOGI PENELITIAN**

##### Dataset

Penelitian ini menggunakan dataset bersifat publik yang tersedia pada website *Kaggle* yaitu *Daily News for Stock Market* [https://www.kaggle.com/aaron7sun/stocknews]. Dataset berisi teks data berita yang dipertimbangkan untuk setiap tanggal hanya 25 peringkat berita utama dan teks data saham *Dow Jones Insdustrial Average* (DJIA) yang digunakan untuk membuktikan konsep apakah data berita dapat menentukan saham tersebut *up* ataupun *down*. Data berita dengan rentang waktu yang dimiliki dari 08-06-2008 sampai dengan 07-01-2016 dan data saham dalam rentang waktu 08-08-2008 sampai dengan 07-01-2016.

Dataset terdiri dari 27 *features* yaitu *feature* pertama *date*, *feature* kedua label, dan ketiga sampai dengan *features* 27 merupakan *headlines* dari peringkat Top1 hingga Top25. Karakteristik data terdiri dari satu atribut *categorical numeric* dan 26 atribut *text*. Penjelasan dari setiap karakteristik *features* tersebut dijabarkan pada Tabel 3.1 dibawah ini.

Table 3.1. Deskripsi Atribut Dataset

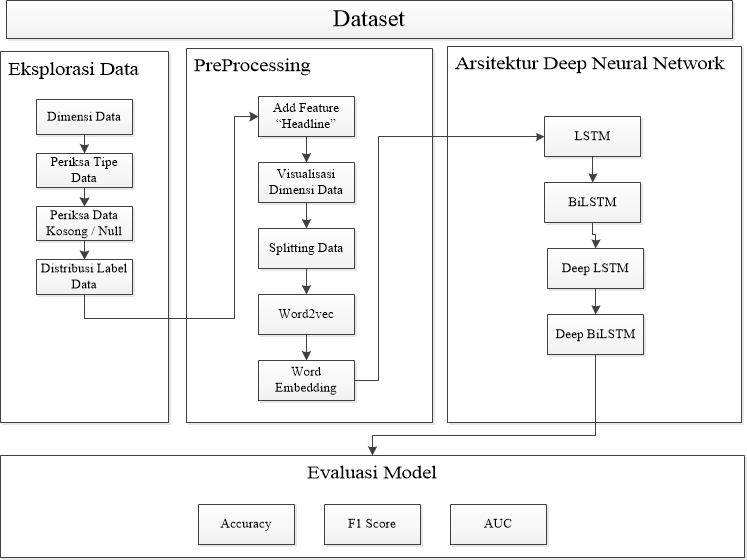
|  |  |
| --- | --- |
| *Attribute* | *Description* |
| Date | Tanggal untuk data berita tersebut terbit |
| Label | Targel kelas untuk teks klasifikasi data teks   * 1 : Ketika nilai DJIA close naik atau tetap sama * 0 : Ketika nilai DJIA close menurun |
| Top1 s.d Top25 | Data teks headline berita yang telah diberi peringkat 1 sampai dengan 25 dalam setiap harinya |

Sepuluh isi data pertama dari deskripsi dataset yang dijelaskan tersebut disajikan pada Table 3.2.

Tabel 3.2. Isi Dataset

##### Eksperimen yang Dilakukan

Tahapan Penelitian pada penelitian ini akan menguji dataset publik yang berisi headline berita Redditnews.com yang diambil dari portal dataset kaggle.com Tujuan digunakannya model tersebut yaitu untuk memprediksi jenis saham berdasarkan headline berita, serta berdasarkan kumpulan-kumpulan data dan variabel yang sudah ditetapkan menggunakan algoritma *Deep Learning*  dengan model BILSTM*.* dengan langkah-langkah yang disajikan dalam metode penelitian (Gambar 3.1) sebagai berikut :



Gambar 3.1 Metode Penelitian

1. Text Preprocesing

1. Data Preprocesing
2. Pembuatan Model Arsitektur DNN
3. Eksperimen dan Validasi
4. Metrics Performance Evaluasi Model

##### DAFTAR PUSTAKA

[1] A. Wulandari, “Perbandingan Klasifikasi Pergerakan Harga Saham Pt. Astra Internasional Tbk Menggunakan Vector Auto Regressive (var) Stasioner Dan Logistic …,” *eProceedings Eng.*, vol. 7, no. 1, pp. 2614–2626, 2020. [1] A. Wulandari, “Perbandingan Klasifikasi Pergerakan Harga Saham Pt. Astra Internasional Tbk Menggunakan Vector Auto Regressive (var) Stasioner Dan Logistic …,” *eProceedings Eng.*, vol. 7, no. 1, pp. 2614–2626, 2020.

[2] L. G. S. Artini, N. T. Aryati, P. V. Lestari, N. P. A. Darmayanti, and G. M. Sudiartha, “Analisis Fundamental Makro Dan Integrasi Pasar Saham Dunia Dengan Bursa Efek Indonesia,” *Matrik J. Manajemen, Strateg. Bisnis dan Kewirausahaan*, vol. 11, no. 2, p. 128, 2017, doi: 10.24843/matrik:jmbk.2017.v11.i02.p03.

[3] F. Muklis, “PERKEMBANGAN DAN TANTANGAN PASAR MODAL.”

[4] Romadhon Febryan Prayuanto, “Pengaruh Inflation, Fund Size Dan Fund Age Terhadap Kinerja Reksadana Saham Di Indonesia,” *J. Akunt.*, pp. 1–17, 2021.

[5] N. C. C. A. Phitaloka, “Web Content Mining Di Sektor Perbankan Pada Lq45 Untuk Pendukung Keputusan Investasi Saham,” *Telematika*, vol. 16, no. 1, p. 18, 2019, doi: 10.31315/telematika.v16i1.2989.

[6] R. Ahli, “No Title.”

[7] B. Mahesh, “Machine Learning Algorithms - A Review,” vol. 9, no. 1, pp. 381–386, 2020, doi: 10.21275/ART20203995.

[8] A. Ahmad, S. Em, and B. Pengetahuan, “Mengenal Artificial Intelligence , Machine Learning , Neural Network , dan Deep Learning Mengenal Artificial Intelligence , Machine.”

[9] D. Jatnika, M. A. Bijaksana, and A. A. Suryani, “Word2vec model analysis for semantic similarities in English words,” *Procedia Comput. Sci.*, vol. 157, pp. 160–167, 2019, doi: 10.1016/j.procs.2019.08.153.

[10] R. A. Fauzannissa, H. Yasin, and D. Ispriyanti, “PERAMALAN HARGA MINYAK MENTAH DUNIA MENGGUNAKAN METODE RADIAL BASIS FUNCTION NEURAL NETWORK,” *J. Gaussian*, vol. 5, no. 1, pp. 193–202, Jan. 2016, doi: 10.14710/J.GAUSS.V5I1.11049.

[11] I. S. Purba and A. Wanto, “Prediksi Jumlah Nilai Impor Sumatera Utara Menurut Negara Asal Menggunakan Algoritma Backpropagation,” *Techno.Com*, vol. 17, no. 3, pp. 302–311, 2018, doi: 10.33633/tc.v17i3.1769.

[12] M. Tanveer, A. Sharma, and P. N. Suganthan, “General twin support vector machine with pinball loss function,” *Inf. Sci. (Ny).*, vol. 494, pp. 311–327, 2019, doi: 10.1016/j.ins.2019.04.032.

[13] U. B. Jaya, A. Fathoni, and U. B. Jaya, “Idetifikasi Ukuran Pakaian,” vol. 5, no. 1, 2021.

[14] S. Solikhun, M. Safii, and A. Trisno, “Jaringan Saraf Tiruan Untuk Memprediksi Tingkat Pemahaman Sisiwa Terhadap Matapelajaran Dengan Menggunakan Algoritma Backpropagation,” *J-SAKTI (Jurnal Sains Komput. dan Inform.*, vol. 1, no. 1, p. 24, 2017, doi: 10.30645/j-sakti.v1i1.26.

[15] I. Akil and I. Chaidir, “Deteksi Karakter Huruf Arab Dengan Menggunakan Convolutional Neural Network,” *Inti Nusa Mandiri*, vol. 15, no. 2, pp. 183–188, 2021.

[16] M. Zulfikar *et al.*, “Prediksi Akor Musik menggunakan Deep Learning berbasis Notasi Angka,” vol. 10, no. 1, 2021.

[17] A. G. Salman, Y. Heryadi, E. Abdurahman, and W. Suparta, “Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting,” *Procedia Comput. Sci.*, vol. 135, pp. 89–98, 2018, doi: 10.1016/j.procs.2018.08.153.

[18] H. F. Fadli, “Identifikasi Cyberbullying pada Media Sosial Twitter Menggunakan Metode LSTM dan BiLSTM,” 2019.