IMPLEMENTATION OF LONG SHORT TERM MEMORY (LSTM) ALGORITHM FOR TIME SERIES FORECASTING OF INSTAGRAM ACCOUNT ENGAGEMENT

(Undergraduate Thesis)

By

AHMAD YUSRIL YUSRO

DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITY OF LAMPUNG
BANDAR LAMPUNG

2023

ABSTRACT

IMPLEMENTATION OF LONG SHORT TERM MEMORY (LSTM) ALGORITHM FOR TIME SERIES FORECASTING OF INSTAGRAM ACCOUNT ENGAGEMENT

By

AHMAD YUSRIL YUSRO

The determination of advertising rates driven by advertising service providers on Instagram must be adjusted to the number of interactions obtained by advertising service providers within a certain period. Information on the number of interactions in the future period can be obtained from the application of time series forecasting which in this study uses Long Short Term Memory (LSTM) which includes the stages of data preparation, model building, prediction and forecasting. This research aims to apply the LSTM for forecasting account engagement data of Instagram @lampuung. Collecting account engagement data was conducted by manual observation directly from @lampuung Instagram in 365 day period. This analysis builds the best model using the LSTM method in predicting the subsequent 30 periods based on comparing each parameter's evaluation value. The results of this study indicate that the best model was constructed using time series cross-validation and LSTM units, batch size, and epoch with each value of 25, 16, and 50. The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) obtained are 1024.12 and 3.87%. By the forecasting value of the account engagement data obtained, this research concludes that PT Lampung Geh Helau that centrally organize @lampuung Instagram can increase advertising rates in June 2023 by up to 24%.

Keywords: LSTM, Instagram Account Engagement Data, RMSE, MAPE, Advertising Rate.

ABSTRAK

IMPLEMENTASI ALGORITMA *LONG SHORT TERM MEMORY* (LSTM) UNTUK PERAMALAN DERET WAKTU *ENGAGEMENT* AKUN INSTAGRAM

Oleh

AHMAD YUSRIL YUSRO

Penentuan tarif iklan yang dilakukan oleh penyedia jasa periklanan di Instagram harus disesuaikan dengan jumlah interaksi yang didapatkan oleh penyedia jasa periklanan dalam kurun waktu tertentu. Informasi jumlah interaksi pada masa mendatang dapat diperoleh dari pengaplikasian peramalan deret waktu yang pada penelitian ini menggunakan Long Short Term Memory (LSTM) yang meliputi tahap pemersiapan data, membangun model, prediksi dan peramalan. Penelitian ini bertujuan untuk menerapkan LSTM untuk peramalan data *engagement* akun Instagram @lampuung. Pengumpulan data *engagement* akun dilakukan dengan observasi manual langsung dari Instagram @lampuung dalam kurun waktu 365 hari. Analisis ini membangun model terbaik dengan menggunakan metode LSTM dalam memprediksi 30 periode berikutnya berdasarkan perbandingan nilai evaluasi setiap parameter. Hasil dari penelitian ini menunjukkan bahwa model terbaik dibangun dengan menggunakan *time series cross validation* dan LSTM unit, *batch size*, dan *epoch* dengan masing-masing nilai 25, 16, dan 50. Nilai *Root Mean Square Error* (RMSE) dan *Mean Absolute Percentage Error* (MAPE) yang diperoleh sebesar 1024,12 dan 3,87%. Dengan nilai peramalan dari data *account engagement* yang diperoleh, penelitian ini menyimpulkan bahwa PT Lampung Geh Helau yang mengelola Instagram @lampuung secara terpusat dapat meningkatkan tarif iklan pada bulan Juni 2023 hingga 24%.

Kata Kunci: LSTM, Data *Engagement* Akun Instagram RMSE, MAPE, Tarif Iklan.

IMPLEMENTATION OF LONG SHORT TERM MEMORY (LSTM) ALGORITHM FOR TIME SERIES FORECASTING OF INSTAGRAM ACCOUNT ENGAGEMENT

By

Ahmad Yusril Yusro

Undergraduate Thesis

Submitted in a Partial Fulfilment of
The Requirements for
BACHELOR OF MATHEMATICS

In

Department of Mathematics
Faculty of Mathematics and Natural Sciences

DEPARTMENT OF MATHEMATICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
UNIVERSITY OF LAMPUNG
BANDAR LAMPUNG
2023

Judul Skripsi	: **IMPLEMENTATION OF LONG SHORT TERM MEMORY (LSTM) ALGORITHM FOR TIME SERIES FORECASTING OF INSTAGRAM ACCOUNT ENGAGEMENT**
Nama Mahasiswa	: ***Ahmad Yusril Yusro***
Nomor Pokok Mahasiswa	: **1917031039**
Program Studi	: **Matematika**
Fakultas	: **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Khoirin Nisa, S.Si., M.Si.	**Dr. Aang Nuryaman, S.Si., M.Si.**
NIP. 19740726 200003 2 001	**NIP. 19740316 200501 1 001**

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Khoirin Nisa, S.Si., M.Si.** ....

Sekretaris : **Dr. Aang Nuryaman, S.Si., M.Si.** ....

Penguji
Bukan Pembimbing : **Drs. Nusyirwan, M.Si.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

Dr. Eng Heri Satria, S.Si., M.Si.
NIP. 197110012005011002

Tanggal Lulus Ujian Skripsi : **11 Juli 2023**

PERNYATAAN

Yang bertanda tangan dibawah ini:

Nama : **Ahmad Yusril Yusro**

Nomor Pokok Mahasiswa : **1917031039**

Jurusan : **Matematika**

Judul Skripsi : **IMPLEMENTATION OF LONG SHORT TERM MEMORY (LSTM) ALGORITHM FOR TIME SERIES FORECASTING OF INSTAGRAM ACCOUNT ENGAGEMENT**

Dengan ini menyatakan bahwa penelitian ini adalah hasil pekerjaan saya sendiri dan apabila di kemudian hari terbukti bahwa skripsi ini merupakan hasil salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 11 Juli 2023
Penulis

METERAI TEMPEL
E8AAKX532126807

Ahmad Yusril Yusro
NPM. 1917031039

CURRICULUM VITAE

Ahmad Yusril Yusro was born in Bandar Lampung on February 20, 2002. He is the third of three siblings of Mr. Imanuddin Rowiyan (deceased) and Mrs. Robiah Adawiyah. He has two older siblings, each named Iqlima Amelia and Muhammad Wildan El Kirom.

Ahmad Yusril Yusro started his elementary school education at MIN 5 Bandar Lampung in 2007-2013. He then continued his Junior High School Education at MTsN 2 Bandar Lampung in 2013-2016 and continued his Senior High School education at SMA Kebangsaan Lampung majoring in Mathematics and Natural Sciences in 2016-2019.

In 2019, Ahmad Yusril Yusro continued his undergraduate education at the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung. through the SNMPTN pathway. During his time as a student, he was active in several organizations and communities, namely the English Society Universty of Lampung as a member of the Human Resources and Development department in 2022, Chief Executive Officer of Ruang Pangan in 2022, and Social Media Specialist of Lampung Geh @lampuung by 2022 until 2023. He is also an awardee of scholarships from various programs such as: Smart Scholarship by YBM Bank Rakyat Indonesia, Djarum Scholarship Plus by Djarum Foundation, and BRILiaN Scholarship by Bank Rakyat Indonesia.

In January to February 2022, Ahmad Yusril Yusro carried out Kuliah Praktek (KP) or Internship at the Regional Tax and Retribution Management Agency (BPPRD) of Bandar Lampung as a form of self-development and applying the knowledge gained during lectures. Subsequently, from June to August 2022, the author carried out the Period II Real Work Lecture (KKN) in Beringin Jaya Village, Kemiling District, Bandar Lampung City as a form of community service.

During his time as a student, Ahmad Yusril Yusro also participated in the MBKM program, namely the Kampus Mengajar Program at SD IT Insan Kamil Bandar Lampung from August to December 2021 and the Magang dan Studi Independen Bersertifikat (MSIB) Program at Ruang Guru in the field of Data and Business from January to June 2022. He is also a selected student who is fully funded to participate in the International Exposure program by Djarum Foundation which allows him to attend and be involved in the Harvard MUN International Conference in Paris, France as a delegate from Indonesia in March 2023. Before completing his studies, he already has been accepted as a Management Trainee by PT Bank Rakyat Indonesia, Tbk, in the field of Information and Technology (IT), especially in the area of Data Science.

MOTTO

" … Therefore, surely with hardship comes ease, surely with that hardship comes more ease …"
(Q.S Ash - Sharh: 5 - 6)

"Paris, I believe, is a man in his twenties in love with an older woman."
(John Peter Berger)

"Dream on, knowing that God will embrace those dreams."
(Andrea Hirata - Edensor)

"Everyone keeps telling me how my story is supposed to go. Nah, I am going to do my own thing"
(Miles Morales – Across The Spider Verse)

"Set me on fire and I will blaze through the battle."
(X.borg Firaga Armor- Mobile Legends)

"If you do not fight, you cannot win."
(Eren Jeager- Attack on Titan)

"You are on your own, kid."
(Taylor Swift – You are on your own, kid)

DEDICATIONS

By uttering my deepest praise and gratitude to Allah SWT, who has given me His mercy, guidance, and grace. I proudly present this humble work with full sincerity as a tribute of my love and affection to:

Mom, Dad, Sister, and Brother

Thank you for providing endless encouragement for the completion of every successful step and for providing continuous assistance to the author.

Advisory and Examination Committee

Thank you to the advisory and examination committee who have helped, provided motivation, direction, and valuable knowledge to the author.

Most Beloved Alma Mater University of Lampung

ACKNOWLEDGEMENTS

All praises and gratitudes are due to Allah SWT, who has given His grace, guidance, and gifts to the researcher, therefore the researcher can complete the undergraduate thesis entitled "Implementation of Long Short Term Memory (LSTM) for Time Series Forecasting of Instagram Account Engagement". The completion of this undergraduate thesis is not separated from the support, guidance, advice, and prayers from various parties. With all humility and sincere, the researcher would like to humbly thank to:

1. Dr. Khoirin Nisa, S.Si, M.Si., as the first supervisor for her patience and willingness to provide countless guidance, criticism, and suggestions in the process of completing this thesis.
2. Dr. Aang Nuryaman, S.Si., M.Si., as second supervisor and Head of the Mathematics Department, Faculty of Mathematics and Natural Sciences, University of Lampung, who has provided advice and direction to the researcher and always makes the time for guidance.
3. Drs. Nusyirwan, M.Si., as the thesis reviewer who has provided criticism, suggestions and input that greatly helped the author in improving and completing the undergraduate thesis.
4. Dra. Dorrah Aziz, M.Si, as Academic Supervisor who has provided guidance and support during the study period which helped in completing the lectures well.
5. Dr. Eng Heri Satria, S.Si., M.Si, as Dean of the Faculty of Mathematics and Natural Sciences, University of Lampung.
6. All lecturers, staff, and employees of the Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung who have helped during the lectures.

7. Father, Mother, Sister, and Brother who always wholeheartedly provided the most sincere prayers and support during my studies and the writing of this undergraduate thesis.
8. Deepest appreciation towards to: Matematika Cohort 2019, McDonalds Rangers, Kasmaran Boys, GPS, Dhevandra, CC8 Rumah Inggris, ESoers, Food Heroes of Ruang Pangan, Sekelik Lampung Geh, KKN Beringin Jaya Folks, Kampus Mengajar Troops, Ruang Guru Lads, Parisian of International Exposure, Beswan Djarum Lampung, Sobat BRILiaN, and any unnamed groups and pairs.
9. All parties who have helped and been involved in completing this thesis.

The researcher realizes that there are many shortcomings in the writing of this undergraduate thesis. Therefore, the researcher expects constructive criticism and suggestions from all parties.

Bandar Lampung, July 11th 2023
Researcher

Ahmad Yusril Yusro
NPM. 1917031039

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURE

I. INTRODUCTION

1.1 Research Background

Technological developments result in future developments that are highly dependent on technology and these developments assistance for future forecasting. Forecasting is an exciting and relevant problem that combines the need to make sense of information on a time series with predicting the most likely future using that information (Hahn, et al., 2023).

One of the methods for prediction is the Long Short Term Memory (LSTM) algorithm, a unique Recurrent Neural Network (RNN) architecture that can be utilized to study temporal sequences and long-term dependencies more accurately than conventional deep neural networks and RNNs. According to Qiu, et al., (2020), Long Short Term Memory uses the most common form of RNN. The intended form is a form that is used to avoid problems in processing and predicting time series data, namely in the form of long-term dependency problems.

Instagram account @lampuung is a social media that shares compelling information related to Lampung Province such as news, tourism, culture, lifestyle, culinary, and charity. This instagram account has approximately 557.000 followers and 21.000 posts that centrally managed by PT. Lampung Geh Helau. This instagram account has expanded to paid promotion for any SME (Small Medium Enterprise) and company to get promoted that funds the finance of the PT. Lampung Geh Helau to grow and develop Lampung Province through the positive-image publication. The business development of social media can be measured by monitoring the instagram insights. One of the key indicators of instagram insights is Account Engagement rate.

Instagram account @lampuung engagement rate is the number of accounts that have interacted with @lampuung's content, including in advertisements. Content includes posts, stories, reels, videos, and live videos. Interactions can include actions such as likes, saves, comments, share, or replies. The researcher observes that the account engagement rate can be forecasted by sufficient algorithm method. The result of prediction can be beneficial to preview the business growth impacted by upcoming interaction in instagram account @lampuung.

Some research using LSTM algorithm to predict has been priorly done. The research that predicting the productivity of seismic waves was produced by Shi, et al. (2021), which results RMSE 0,53. Subsequent research was conducted by Sen, et al. (2020) for rice price forecasting, which produces an RMSE accuracy of 0.49 on training data and 0.27 on testing data. Other research was conducted by Pothuganti (2021) regarding predictions on the stock market exchange, which has an accuracy value of 97%.

There are no prior researches that has discussion on the business growth in Social Media account engagement using LSTM. Therefore, This study aims to implement the LSTM algorithm to forecast instagram account @lampuung engagement in the purpose of previewing the business growth of PT. Lampung Geh Helau. This research intends to construct the model of LSTM for time series forecasting.

1.2 Research Objectives

The objectives of composing this research are:

1. Constructing the model of LSTM for time series forecasting of Instagram Account Engagement in period from June 2022 to May 2023.
2. Obtaining the result of forecasting the account engagement in Instagram @lampuung.

1.3 The Use of the Research

The following are the use of composing this research:

1. Comprehending the data analytics process using LSTM algorithm for time series forecasting.
2. Improving the knowledge of analytics thinking for the researcher and the reader of this research.
3. Contributing to scientific publication with research model using LSTM algorithm for time series forecasting of Instagram Account Engagement.

II. LITERATURE REVIEW

2.1 Time Series Forecasting

Time series forecasting is an analysis that considers the influence of time sequentially. At the same time, the time-series data itself is data that is collected based on specific sequences and time intervals, such as in hours, days, weeks, months, quarters, semesters, and years (Ruhiat & Suwanda, 2019). Time series forecasting is a quantitative forecasting method based on a series of data tied to a time period variable. The data used in this method is observational data based on various variations of the time series used (hours, days, weeks, months, quarters, quarters, and years).

Standard stages that have been agreed upon in applying the time forecasting method include identifying forecasting objectives, determining forecasting periods, selecting forecasting methods, preparing data, applying forecasting methods, analyzing forecasting results, and evaluating forecasting results. Several widely used forecasting time series methods include the mean forecast, naïve forecast, linear trend forecast, non-linear forecast, exponential smoothing, and moving average (Auliasari et al., 2020).

2.2 Machine Learning

Machine learning is a science that originates from artificial intelligence and can also be interpreted as computer applications and mathematical algorithms derived from data to obtain predictions. This machine learning relies heavily on data for training and testing. Machine learning is divided into three categories: supervised, unsupervised, and reinforcement (Roihan, et al., 2020). The supervised learning method is based on a collection of labeled data samples. The sample collection is used to summarize the characteristics of the behavior size distribution in each type of application to form a behavioral model from the data.

In this type of unsupervised learning, the system is provided with some sample inputs, but no output is present. Since there is no desired output here, the categorization is done so that the algorithm differentiates properly between the data sets. Reinforcement learning comes from animal learning theory. This learning does not require prior knowledge, can independently obtain optional policies with knowledge gained through trial and error, and continuously interacts with a dynamic environment.

2.3 Deep Learning

According to Supriyadi (2020), Deep Learning network is one of them development of machine learning that utilizes artificial Artificial Neural Network (ANN). Artificial Neural Network is a new computational system and method for machine learning, knowledge demonstration, and applying the acquired knowledge to maximize the response output of complex systems (Chen, et al., 2019). Artificial Neural Network (ANN) is designed like the human brain, with connected nerve nodes like a network. The Artificial Neural Network (ANN) architecture is as follows:

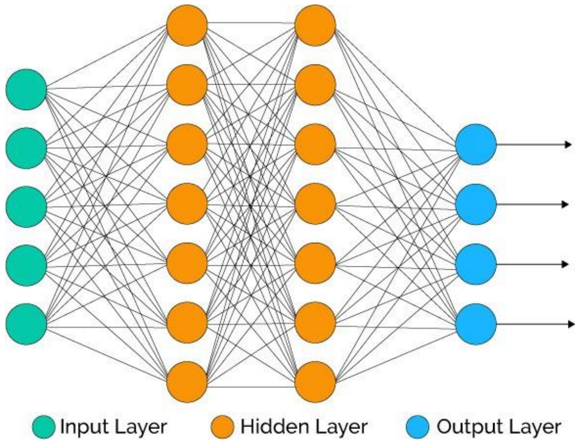

Figure 1. Architecture of ANN

2.4 Recurrent Neural Network (RNN)

Recurrent Neural Network (RNN) is a recurrent neural network structured to process sequential data such as time series data. RNN is said to be a recurrent neural network because the output from the previously hidden layer will be reused as input data for further processing.

This RNN consists of an input layer, a hidden layer, and an output layer. The distinctive feature of the RNN in making a prediction is not only using one input at a time but requires input and previous input. Therefore the inputs are interconnected and can provide information to the hidden layer (Sen, et al., 2020).

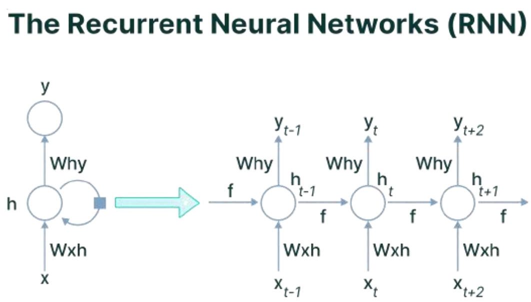

Figure 2. Architecture of Recurrent Neural Network

2.5 Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a neural network development that can be used for modeling time series data. LSTM can learn which data will be stored and which will be discarded because each LSTM neuron has several gates that regulate the memory of each neuron itself. LSTM cells can link previous information with subsequent information, and the effectiveness of storing long information is very much needed in processing time series data.

According to Wiranda and Sadikin (2019), the LSTM architecture consists of an input, hidden, and output layer. Each memory cell in the LSTM has three sigmoid layers and one tanh layer, and one memory cell is composed of three gates, namely the forget gate, input gate, and output gate.

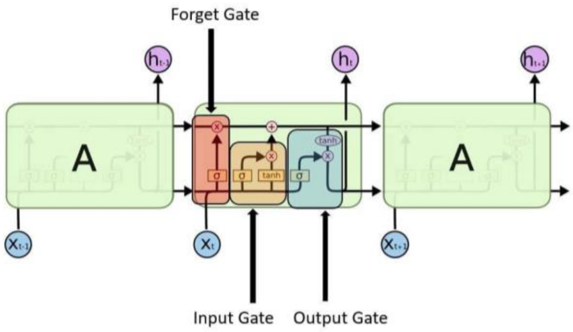

Figure 3. Architecture of LSTM

Two activation functions are used in neural networks, namely the sigmoid activation function and the tanh activation function, where the activation function purposes to activate or not activate neurons (Nwankpa, et al., 2020).

1. The sigmoid activation function transforms values between -1 and 1 into values between 0 and 1. This function has the following equation.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

2. The tanh activation function ranges from -1 to 1 and is an alternative function of the sigmoid layer. This function has the following equation.

$$\mathrm{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2.2}$$

The following are gates in one LSTM cell.

1. Forget Gate, this gate functions to determine the information that is less needed or not meaningful to be removed using the sigmoid function. With the following equation

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f). \tag{2.3}$$

Where:

f_t	: Forget gate
σ	: Sigmoid Function
W_f	: Value of weight for forget gate.
h_{t-1}	: Value of output before ordo -t
x_t	: Value of input for ordo -t
b_f	: Value of bias for forget gate

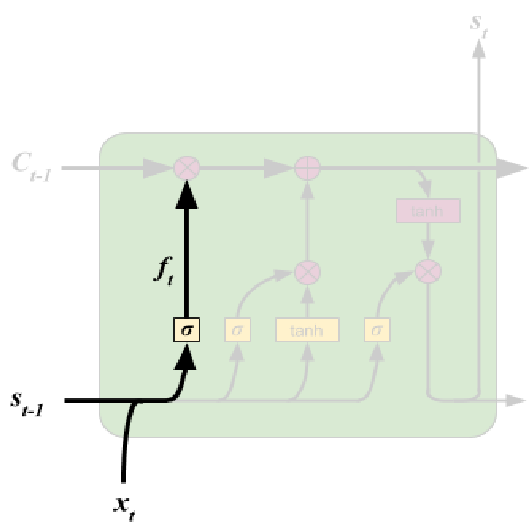

Figure 4. Structure of Forget Gate

2. Input Gate, this gate will sort and determine certain information that will be updated to the cell state using the sigmoid activation function. This step also forms a new candidate vector using the tanh function. With the following equation.

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i). \quad (2.4)$$

Where:

i_t	: Input gate
σ	: Sigmoid Function
W_i	: Value of weight for input *gate*.
h_{t-1}	: Value of output before ordo -t
x_t	: Value of input for ordo -t
b_i	: Value of bias for input *gate*

The equation of new candidate is as follow

$$\tilde{C}_t = tanh(W_c * [h_{t-1}, x_t] + b_c). \quad (2.5)$$

Where:

$\tilde{C}_t$	: New value added to cell state
$tanh$	: Tanh function
W_i	: Value of weight for cell state
h_{t-1}	: Value of output before ordo -t
x_t	: Value of input for ordo -t
b_i	: Value of bias for cell state

After that, the old cell state will be updated into a new cell state with the equation.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t. \quad (2.6)$$

Where:

C_t	: Cell state
f_t	: Forget gate
C_{t-1}	: Cell state before ordo -t
i_t	: Input gate
$\tilde{C}_t$	: New value added to cell state

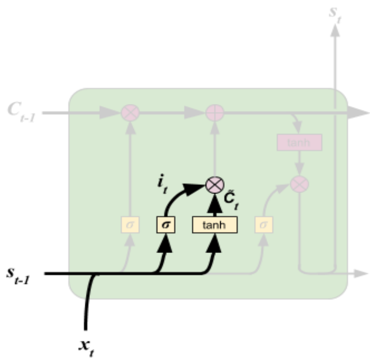

Figure 5. Structure of Input Gate

3. Output Gate, this gate functions to decide the output value results in the hidden state and place the cell state in the tanh based on input and block memory. After generating the sigmoid output value and the second tanh output value, the activation results are multiplied. With the following equation.

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o). \tag{2.7}$$

Where:

o_t	: Output gate
σ	: Sigmoid Function
W_o	: Value of weight for output gate
h_{t-1}	: Value of output before ordo -t
x_t	: Value of input for ordo -t
b_o	: Value of bias for output gate

$$h_t = o_t * \tanh(C_t). \tag{2.8}$$

Where:

h_t	: Value of output for ordo -t
o_t	: Output gate
$tanh$	: Tanh Function
C_t	: Cell state

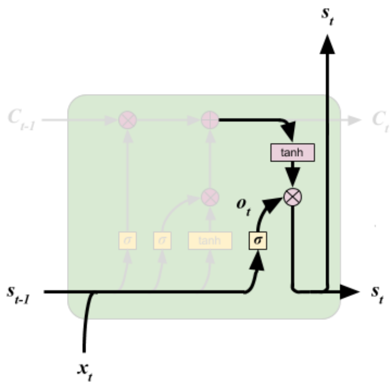

Figure 6. Structure of Output Gate

2.6 Data Normalization

Data normalization or scaling is a data transformation process that changes the original data into another form. The normalization process must be done to produce the smallest possible error value.The normalization technique used account engagement data forecasting is Min-Max Scaler, called Min-Max Scaling. This technique transforms the actual data into values with a range of [0,1]. The Min-Max Scaling technique is formulated into the following equation (De Amorim, et al., 2023):

$$x' = \frac{x - X_{min}}{X_{max} - X_{min}}. \tag{2.9}$$

Where:

x' : Normalized data of account engagement

x : Actual data of account engagement

X_{min} : Minimum Value of account engagement data

X_{max} : Maximum Value of account engagement data

2.7 Data Denormalization

The denormalization process is the process of changing the forecasting results using normalized data into the original data form. Process is carried out to determine the predicted data and make comparisons to actual data to see the performance of the model that has been formed. If previously normalization was carried out in the interval [0,1], then denormalization is expressed in the following equation (Wiranda & Sadikin, 2019):

$$x_t = x'(X_{max} - X_{min}) + X_{min}. \tag{2.10}$$

Where:

x_t : Denormalized data of account engagement

x' : Normalized data of account engagement

X_{min} : Minimum Value of account engagement data

X_{max} : Maximum Value of account engagement data

2.8 Cross Validation

In general, the model performance evaluation process is carried out on test data with model training on training data, where data has been divided into two parts: training data and test data. However, this is prone to overfitting (Mundra, 2020). Overfitting is a fundamental issue in supervised machine learning which prevents us from perfectly generalizing the models to well-fit observed data on training data, as well as unseen data on the testing set (Ying, 2019).

Another way to evaluate model performance and prevent overfitting is by cross-validation. The method that can be used for time series models is rolling cross-validation. Where a small portion of the initial data is used as training, then checking the accuracy of some data is later used as forecasting. The data previously used for forecasting is combined with the previous training data and reused for the next training, then re-check the accuracy of some of the data afterward. These steps are repeated until the last data. The model evaluation results are obtained from the average of each forecasting accuracy calculation.

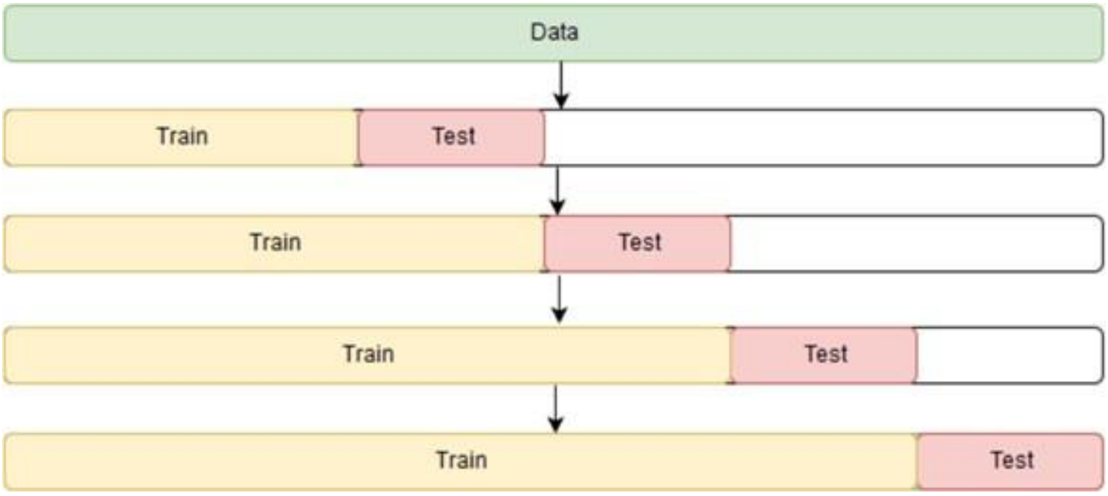

Figure 7. Illustration of Cross Validation

2.9 Model Evaluation

Evaluation is the process of testing the accuracy of the model performance that is formed in predicting data that has been tested before to know the performance of the model that has been trained. According to Wiranda & Sadikin (2019), Evaluating a model can use Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

2.9.1 Root Mean Square Error

Root Mean Square Error is a method used in evaluating a model that is formed to calculate the level of accuracy. The RMSE value is obtained from the actual account engagement data minus the account engagement prediction data squared and then divided by the number of account engagement forecasting periods. The last one is its root. The determination of the RMSE value is formulated in the following equation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n}} \quad (2.11)$$

Where,

Y_i : The actual data in period-*i*

$\hat{Y}_I$: The predicted data in period-*i*

n : Total number of observations

2.9.2 Mean Absolute Percentage Error

To assess the performance of this LSTM model, the researcher also applies the MAPE calculation. MAPE is a proportion of error values that shows data accuracy (Sitepu et al., 2021). The calculation method can be analyzed in the equation below.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \quad (2.12)$$

Where,

y_i = The actual data in period-i

$\hat{y}_i$ = The predicted data in period-i

n = Total number of observations

Table 1. Interpretation range of MAPE

MAPE	Interpretation
$\leq$ 10%	Highly accurate forecasting
10% $-$ 20%	Good forecasting
20% $-$ 50%	Reasonable forecasting
$\geq$ 50%	Inaccurate forecasting

2.10 Analytics Tools

The data analysis process requires tools that assist researchers in completing this study using the Python programming language, the LSTM method and the Google Colaboratory as workflows notebooks.

2.10.1 Python

Python is a popular programming language for data processing. Numerous data scientists have employed Python because it is relatively easy to use (Prasetya et al., 2020). In addition, the library available for processing this data is quite a lot that makes the analytics process more straightforward for the researcher in this study.

2.10.2 Google Colaboratory

Google Colaboratory is a document that allows you to write, run, and share Python code within your browser. It is a version of the popular Jupyter Notebook within the Google suite of tools. Jupyter Notebooks (and therefore Google Colaboratory) serves as a virtual lab notebook to support detailed workflows, code, data, and visualizations (Randles et al., 2017) which is then stored in cloud and shareable to peers and colleagues for editing, commenting, and viewing.

2.11 Instagram Account Engagement

Instagram account engagement rate is a metric standard utilized in social media marketing to measure the performance of the content on social media platforms, particularly on Instagram and facebook (Ariescey and Amriel, 2021). This indicator is significant to be understood by instagram users, influencers, and digital marketers to assess audience engagement with the posts they have published. If content can get much attention from followers, the public knowledge of a brand will automatically increase. Engagement rates can also be used as a research tool to understand the audience's desires based on the number of their interactions with some or specific content (Ariesey and Amriel, 2021).

2.12 OSEMN Framework

OSEMN is data science framework in the form sequences of activities in a data science process that is useful for problem solving in the field of data science analysis (Sitepu et al., 2021). The following are an explanation of the steps contained in the framework:

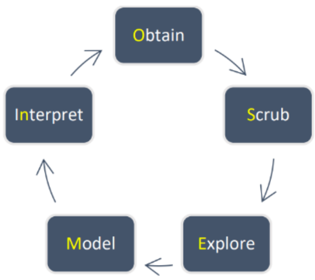

Figure 8. OSEMN Framework

2.12.1 Obtain

To begin with, the first activity in this framework is obtaining data that the research requires from the available and valid resources. There are some formats of data that can be provided which are Google Big Query, Google Spreadsheet, MySQL, Microsoft Excel, and etcetera.

There are other popular techniques to acquire the dataset by using Web API (Application Programming Interface). A number of websites such as Facebook and Twitter allow the users to access their data. Furthermore, there is a conventional technique to acquire the datasets by directly downloading it which formed as CSV (Comma Separated Value) or TSV (Tab Separated Value) from Kaggle or the official websites of company.

2.12.2 Scrub

The next activity in this framework is scrubbing data that preprocess dataset attentively right before it will be operated. The dataset and their variables are mainly to be cleaned, filtered, replaced, combined, eliminated, and extracted (Sitepu et al., 2021)

2.12.3 Explore

The explore activity is a process that inspects the dataset and its properties. The required treatments to the dataset are based on the data types such as numerical data, categorical data, ordinal, and nominal data. The next step is to compute the descriptive statistics to extract features and to utilize data visualization to help the researcher to identify significant patterns and trends in the dataset.

2.12.4 Model

The fourth activity is data modelling in purpose to train model for predicting or forecasting data in the future. The model determination requires to be matched with the desired purpose of data analytics that conducted on the research for the suitable result

2.12.5 Interpret

The interpret is the final activity in purpose to conclude the model and the dataset. Interpreting data and model performance is particularly presented in the basic and fundamental form so that the presentation is comfortably understood by the common and non-technical people. This activity is the vital OSEMN framework process to provide comprehension of the data analytics about the result of the research. The result can be taken advantage for the upcoming policy decision.

III. METHODOLOGY

3.1 Research Setting

The research is conducted in the even semester academic year 2022/2023. It takes a place at Research Laboratory in Mathematics Department, Faculty of Mathematics and Natural Sciences, University of Lampung.

3.2 Research Data

The dataset that operates is a secondary data in which is the historical dataset of account engagement at instagram account @lampuung in period from June 1st 2022 to May 31st 2023. This dataset is amount to 365 daily data obtained by manually operating from insight instagram account @lampuung.

3.3 Research Method

This research is conducted with the use of Long Short Term Memory algorithm under OSEMN framework. This is the following steps which are:

1. Collection (Obtain)

   This step contains the activity of collecting data. The dataset utilized in this research is the historical dataset of account engagement of instagram account @lampuung.

2. Variable Selection and Data Split (Scrub)

   The activities variable selection and data split is executed in this scrubbing step. The variables that will be selected are the date stamp ideally formed YYYY-MM-DD and the measurement of daily total of engaged account in numeric form that the researcher wishes to forecast. The dataset splits into the two parts of data division which are training data and testing data,

3. Perspective Analysis (Explore)

   The explore activity can be executed by three perspective steps which are understanding data and its property, calculating descriptive statistic, and visualizing data (Sitepu et al., 2021)

4. Model Estimation (Model)

   This research uses the model design by applicating with LSTM algorithm. The application of LSTM algorithm to predict the account engagement of instagram account @lampuung will result the model estimation as the base of forecasting.

5. Analysis of Result (Interpret)

   This step contains performing accuracy test to the model by using MAPE, forecasting the case study in this research, and visualizing of the result of forecast of account engagement of instagram.

Here is the flow chart:

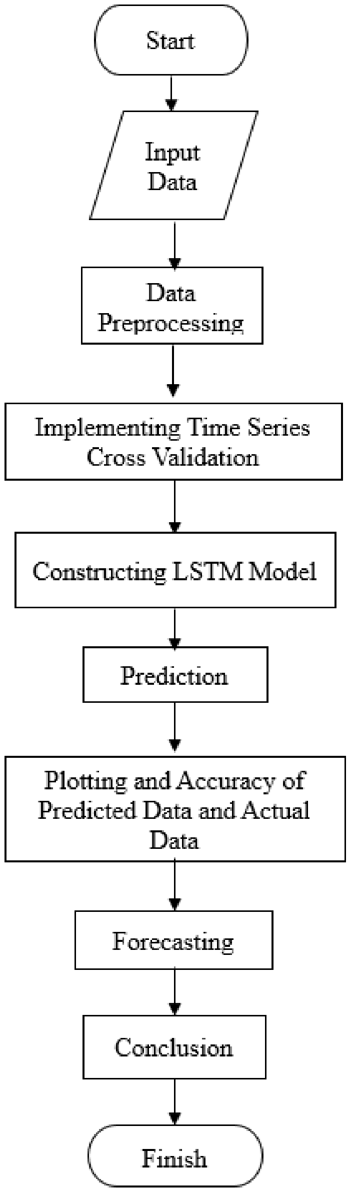

Figure 9. Flow Chart

V. CONCLUSION

Based on the results and discussion, the conclusions of this research are obtained as it follows:

1. The best LSTM model in account engagement forecasting based on dataset division composition testing is formed from the K-Fold Time Series Cross Validation with K=4. The best LSTM model for account engagement forecasting is built from 25 LSTM Units, 16 batch sizes, and 50 epochs. The RMSE and MAPE values of the best LSTM model for account engagement are 1024,12 and 3,87%.
2. The results of forecasting account engagement from June 1st, 2023, to June 30th, 2023, continue to experience an upward trend. It indicates that the amount of Instagram account engaged with @lampuung are provenly increasing. According to the results and discussion, on June 2023, Instagram account @lampuung received a total account engagement of 1.224.591. In comparison, on June 2022, Instagram account @lampuung received a total of account engagement of 922.280. It concludes that PT Lampung Geh Helau can increase promotional and advertising rates as of June up to 24%

REFERENCES

Ariescey, R. R., and Amriel, E. V., 2021. Analisa Engagement Rate di Instagram: Fenomena Like dan Komentar. *Journal of Management Media Service*. **9**: 2502 - 3632.

Auliasari, K., Kertaningtyas, M., and Kriswantono, M. 2020. Penerapan Metode Peramalan untuk Identifikasi Permintaan Konsumen. *INFORMAL: Informatics Journal*, **4**(3): 121

Bengio, Y. 2012. Practical recommendations for gradient-based training of deep architectures. *In Neural networks: Tricks of the trade* (pp. 437-478). Springer

Chen, M., Chalita, U., Saad, W., Yin, C., and Debbah, M. 2019. Artificial Neural Networks-Based Machine Learning for Wireless Networks: A Tutorial. *IEE: Comunication Surveys & Tutorials.* **21**(4): 3039-3071.

Choudhury, N. A., & Soni, B. 2023. An Adaptive Batch Size based-CNN-LSTM Framework for Human Activity Recognition in Uncontrolled Environment. *IEEE Transactions on Industrial Informatics*, **42**(13): 197

De Amorim, L. B., Cavalcanti, G. D., & Cruz, R. M. 2023. The choice of scaling technique matters for classification performance. *Applied Soft Computing*, **133**: 109924.

Hahn, Y., Langer, T., Meyes, R., and Melsen, T. 2023. Time Series Dataset Survey with Deep Learning. *Forecasting*. **5**(1):315-335.

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. 2016. On large-batch training for deep learning: Generalization gap and sharp minima. *International Conference on Learning Representations*.

Mehedi, M. A. A., Khosravi, M., Yazdan, M. M. S., & Shabanian, H. (2022). Exploring temporal dynamics of river discharge using univariate long short-term memory (LSTM) recurrent neural network at East Branch of Delaware River. *Hydrology*, **9**(11): 202.

Mundra, A., Mundra, S., Verma, V. K., & Srivastava, J. S. (2020). A deep learning based hybrid framework for stock price prediction. *Journal of Intelligent & Fuzzy Systems*, **38**(5): 5949-5956.

Nwankpa, C., Winifred, I., Anthony, G., and Stephen, M. 2018. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. *Computer Science*. 124-133.

Prasetya, B. D., Pamungkas, F. S., and Kharisudin, I. 2020. Pemodelan dan Peramalan Data Saham dengan Analisis Time Series menggunakan Python. *Proceeding of National Mathematics Seminar*. **3**: 714 - 718.

Pothuganti. 2021. Long Short-Term Memory (LSTM) Algorithm Based Prediction of Stock Market Exchange. *International Journal of Research Publication and Reviews*. **2**(1): 90-93.

Qiu, J., Wang, B., and Zhou, C. 2020. Forecasting Stock Prices with Long-Short Term Memory Neural Network Based on Attention Mechanism. *PLOS ONE*. (**15**)1: 2-3.

Randles, B. M., Pasquetto, I. V., Golshan, M. S., and Borgman, C. L. 2017. Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*. 17–18.

Roihan, A., Sunarya, P.A., and Rafika, A.S. 2020. Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *Indonesian Journal on Computer and Information Technology*. **5**(1): 75-82.

Ruhiat, D., and Suwanda, C. 2019. Peramalan Data Deret Waktu Berpola Musiman Menggunakan Metode Regresi Spektral (Studi Kasus : Debit Sungai Citarum-Nanjung). *TEOREMA : Teori Dan Riset Matematika*, **4**(1): 1.

Sen, S., Sugiarto, D., and Rochman, A. 2020. Komparasi Metode Multilayer Perceptron (MLP) dan Long-Short Term Memory (LSTM) dalam Peramalan Harga Beras. *Jurnal Teknik Informatika*. **11**(1): 35-41.

Shi, P., Seydoux, L., and Poli, P. 2021. Unsupervised Learning of Seismic Wavefield Features: Clustering Continuous Array Seismic Data During the 2009 L'Aquila earthquake. *Journal of Geophysical Research: Solid Earth*, **126**(1).

Sitepu, F. T. B., Sirait, V. A. P., and Yunis, R. 2021. Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Prophet Facebook. *Journal of Computer and Informatics*. **23**(1): 99–105.

Supriyadi, E. 2020. Prediksi Parameter Cuaca Menggunakan Deep Learning Long-Short Term Memory (LSTM). *Jurnal Meteorologi dan Geofisika*. **21**(2): 55-67

Wiranda, L. and Sadikin, M. 2019. Penerapan Long Short Term Memory Pada Data Time Series Untuk Memprediksi Penjualan Produk PT. Metiska Farma. *Pendidikan Teknik Informatika*. **8**(3): 184-196.

Ying, X. 2019. An Overview of Overfitting and Its Solutions. *Journal of Physics*: Conference Series. **1168**(2019)