AIP Conference Proceedings



Volume 2975

Proceedings of the International Conference on Mathematics, Computational Sciences, and Statistics 2022 Mathematics, Computational Sciences, and Statistics for a Better Future

Surabaya, Indonesia • 2 October 2022 Editors • Elly Pusporani, Nashrul Millah and Eva Hariyanti



ORGANIZING COMMITTEE

GENERAL CHAIR

Prof. Dr. Moh. Yasin, M.Si.

PROGRAM COMMITTEE CHAIR

Dr. M. Fariz Fadillah Mardianto, S.Si., M.Si.

PROGRAM COMMITTEE CO-CHAIR

Dr. Eva Hariyanti, S.Si., M.T

SECRETARY

Nania Nuzulita, S.Kom., M.Kom. Asri Bekti Pratiwi, S.Si., M.Si. Nashrul Millah, M.Si.

EXECUTIVE CHAIR

Dr. Herry SuprajitnoElly Pusporani, S.Si., M.Stat. Dr. Nenik Estuningsih, S.Si, M.Si. Indah Werdiningsih, S.Si., M.Kom . Purbandini, S.Si., M.Kom. Anisa Puspitasari, S.Si. Abdul Latif Yuli Ika Wahyu Utami

TREASURER

Dita Amelia, S.Si., M.Si. Siti Zahidah, S.Si., M.Si.

STEERING COMMITTEE

Cicik Alfiniyah. M.Si., Ph.D. Dr. Fatmawati, M.Si. Dr. Eridani, M.Si. Dr. Herry Suprajitno, M.Si. Dr. Miswanto, M.Si. Dr. Ardi Kurniawan, M.Si.

SCIENTIFIC AND EDITORIAL TEAM

Dr. Nur Chamidah, M.Si. Dr. Windarto, M.Si. Ira Puspitasari, S.T., M.T., Ph.D. Dr. Toha Saifudin, S.Si, M.Si.

EVENT TEAM

Dr. Liliek Susilowati, S.Si, M.Si. Marisa Rifada, S.Si., M.Si. Dr. Yayuk Wahyuni, M.Si. Endah Purwanti, S.Si., M.Kom.

LOGISTIC AND TRANSPORT TEAM

Abdulloh Jaelani, S.Si., M.Si. Muchammad Yusuf Syaifuddin, S.Si., M.Si. Bustomi, S.Si., M.Si. Drs. Suliyanto, M.Si. Muhammadun, S.Si., M.Si. Eko Ardianto, S.T. Moh. Nur Qodim

IT AND DOCUMENTATION

Dr. Rimuljo Hendradi, M.Si. Drs. Edi Winarko, M.Cs Faried Effendy, S.Si., M.Kom. Taufik, S.T., M.Kom. Badrus Zaman, S.Kom., M.Cs Nurhidayat Khoni Iswantoro, S.T. Abdul Aziz, S.T.

COMSUMPTION TEAM

Dr. Inna Kuswandari, M.Si. Auli Damayanti, S.Si., M.Si. Ir. Elly Ana, M.Si.

SPONSORSHIP

Sa'idah Zahrotutl Jannah, S.Si., M.Stat. Drs. H. Sediono, M.Si Drs. Kartono, M. Kom.

SCIENTIFIC COMMITTEE

Prof. Dr. rer.nat. Dedi Rosadi, S.Si., M.Sc.Eng.Math. (Universitas Gadjah Mada, Indonesia) Prof. Dr. Fatmawati, M.Si. (Mathematics, Universitas Airlangga, Indonesia) Prof. Dr. I Nyoman Budiantara, M.Si. (Statistics, Institut Teknologi Sepuluh Nopember, Indonesia) Dr. Achmad Choiruddin, S.Si, M.Sc. (Institut Teknologi Sepuluh Nopember, Indonesia) Dr. Ani Budi Astuti, M.Si. (Universitas Brawijaya, Indonesia) Dr. Anna Islamiyati, M.Si. (Universitas Hasanuddin, Indonesia) Dr. Ardi Kurniawan, M.Si. (Statistics, Universitas Airlangga, Indonesia) Dr. Asmiati, S.Si., M.Si. (Universitas Lampung, Indonesia) Dr. Atina Ahdika, S.Si., M.Si. (Universitas Islam Indonesia, Indonesia) Cicik Alfiniyah, Ph.D (Mathematics, Universitas Airlangga, Indonesia) Dr. Darnah Andi Nohe, M.Si. (Universitas Mulawarman, Indonesia) Dr. Di Asih I Maruddani, M.Si. (Universitas Diponegoro, Indonesia) Dr.Didik Khusnul Arif, S.Si., M.Si. (Institut Teknologi Sepuluh Nopember, Indonesia) Dr. Eridani, M.Si. (Mathematics, Universitas Airlangga, Indonesia) Dr. Eva Hariyati, M.Si. (Universitas Airlangga, Indonesia) Ira Puspitasari, Ph.D. (Universitas Airlangga, Indonesia) Dr. Isniani Rosyida, M.Si. (Universitas Negeri Semarang, Indonesia) Dr. Liliek Susilowati, M.Si. (Mathematics, Universitas Airlangga, Indonesia) Dr. Lyra Yulianti, M.Si. (Universitas Andalas, Indonesia) Dr. M. Fariz Fadillah Mardianto, M.Si. (Statistics, Universitas Airlangga, Indonesia) Dr. Miswanto, M.Si. (Mathematics, Universitas Airlangga, Indonesia) Dr. Moch Idris, M.Si. (Universitas Lambung Mangkurat, Indonesia) Dr. Muhammad Zakir, M.Si. (Universitas Hasanuddin, Indonesia) Dr. Nur Chamidah, M.Si. (Statistics, Universitas Airlangga, Indonesia) Dr. Ridwan Pandiya, S.Si., M.Sc. (Institut Teknologi Telkom Purwokerto) Dr. Sutrima, M.Si. (Universitas Sebelas Maret) Dr. Toha Saifudin, M.Si. (Universitas Airlangga) Dr. Wahyu Wibowo, S.Si., M.Si. (Institut Teknologi Sepuluh Nopember) Dr. Windarto, M.Si. (Universitas Airlangga) Dr. Winita Sulandari, M.Si. (Universitas Sebelas Maret) Dr. Yayuk Wahyuni, M.Si. (Universitas Airlangga) Dr. Yeni Kustiyahningsih, S.Kom., M.Kom. (Universitas Trunojoyo)

Dr. Muhammad Ahsan, S.Si (Institut Teknologi Sepuluh Nopember)

TABLE OF CONTENT

FOREWORD FROM DEAN OF FACULTY OF SCIENCE AND TECHNOLOGY,	
UNIVERSITAS AIRLANGGA	2
FOREWORD FROM CONFERENCE COMMITTEE CHAIR	4
ORGANIZING COMMITTEE	6
SCIENTIFIC COMMITTEE	9
FABLE OF CONTENT	10
CONFERENCE SCHEDULE	14
I ECHNICAL GUIDELINES AND REGULATIONS KEVNOTE SPEAKER RIOCRAPHV	1/
Prof. Paulo Canas Rogdriguez, Ph.D	20
Prof. Zakia Hammouch, Ph.D	21
Assoc. Prof. Mario Rosario Guaracciono, Ph.D.	22
Assoc. Prof. Dr. Kang-Ming Chang	23
KEYNOTE SPEAKER SPEECH	24
The Usefulness of Singular Spectrum Analysis in Hybrid Methodologies for Time Series Forecasting	25
Unsupervised Whole Graph Embedding Methods and Applications	26
Biomedical Signal Processing with Machine Learning- Application on Elder Fall Detection and Meditation.	l 27
INVITED SPEAKER BIOGRAPHY	28
Dr. Lilysuriazna Raya	29
Assoc. Prof. Dr. Haralampos H.	30
Dr. Demi Allen	31
Dr. Eridani	32
Dr. Nur Chamidah	33
Ira Puspitasari, Ph. D	34
INVITED SPEAKER SPEECH	35
The Traveling Salesman Problem in Relation to Industry 4.0 and Sustainability Goals	36
Improving Tumour Predictions: a Bayesian Combination of Dynamic Modelling and Machine Learning Approach	37
Diophantine Approximation for Systems of Linear Forms	38
Integral Operator on Lebesgue and Morrey Spaces	39
Estimating Number of Malaria Parasites Using Nonparametric Negative Binomial Regression Approach Based on Penalized Spline Estimator	40

The Application of Data Analytics, Information Systems Theory, and Interaction Design in Digit Health Initiatives to Promote Well-Being for All Ages	al: 41
ALGEBRA AND COMBINATORIC (AC)	42
AC001 - L _{3,2,1} Labelling of Kalpataru Graph	43
AC002 - General Zeroth-order Randić Index of Zero Divisor Graph for the Ring of Integers	
Modulo n	44
AC003 - Elegant Labeling of Tadpole Graphs $T(n,1)$ and $T(4n,2)$	45
AC005 - On {0, 1}-Distance Labelings of 2-Regular Graphs	46
AC006 - The Dominant Metric Dimension of Generalized Petersen Graph	47
AC007 - Local Multiset Dimension of Amalgamation Cycle	48
AC008 - The Dominant Metric Dimension of the Edge Corona Product Graph	49
AC009 - The Total Non-Zero Divisor Graph of Some Commutative Rings	50
AC010 - The Dominant Metric Dimension On The Vertex Amalgamation Product Graph	51
AC011 - Some Properties of n-Cut Splicing on Labelled Semigraphs	52
AC012 - Multipolar Intuitionistic Fuzzy Ideal in B-Algebras	53
ANALYSIS AND GEOMETRY (AG)	54
AG001 - Local Existence of Regular Solutions in Dynamical Massless Einstein-Skyrme System	55
AG002 - Fixed Point Theorem for A Contractive Mapping on Standard n-Normed Spaces	56
AG003 - Analysis Properties of The Continuous Function in Banach Space	57
APPLIED MATHEMATICS (AM)	58
AM002 - Approximation of the matrix exponential with global symplectic Krylov method	59
AM003 - Dynamic Logistic Models for Stock Price Fluctuation During Covid-19 Pandemic in Indonesia	60
AM004 - Vertex Equitable Labeling of Firecracker and A-Star Graph	61
AM005 - Combination of the Method Based on the Removal Effect Of Criteria (MEREC) and Simple Additive Weighting (SAW) Method in Determining the Best Stock	62
AM007 - A Reduced Two-Layer Non-Hydrostatic Model for Tsunami Generation by Submarine Landslides	63
AM008 - Analysis of the Stability of the Mathematical Model in the Drug Abuse Case Rehabilitation Program in East Java Provinc	64
AM009 - Optimal Control Analysis of Spread Computer Viruses using Linear Quadratic Regulat Method	or 65
AM010 - Rectangular Heatsink Temperature Simulation Based on Two-Dimensional Diffusion Equation	66
AM011 - Stability Analysis and Optimal Control of Social Media Hoax Propagation Mathematic Model with Forgetting Mechanism	al 67

International Conference on Mathematics, Computational Sciences and Statistics	2022
AM012 - Stability Analysis and Optimal Control Strategy of Smoking Behavior Cessation Mathematical Mode	on 68
ACTUARIAL SCIENCES (AS)	69
AS001 - Determination of Premium and Benefit Reserves on Joint Life Insurance Using Prospective Method	7(
COMPUTATIONAL SCIENCES (CS) CS003 - DEVELOPMENT OF AUGMENTED REALITY ON THE SCULPTURE OF UNIVERSITAS AIRLANGGA HISTORY AND CULTURE MUSEUM AS AN EDUC. MEDIA ON HINDU-BUDDHIST HISTORY	71 ATIONAL 72
CS004 - Hybrid Variants of Chaotic Honey Badger Algorithm and Group Teaching Optin Algorithm in Uncapacitated Facility Location Problem	mization 73
CS005 - Reaction Pathways Design Based on the Associative Property of Addition	74
CS008 - Application of the Decision Tree C4.5 Method on the Classification of Diet Typ People with Diabetes Mellitus	es of 75
CS009 - A Scatter Search Algorithm using Biological Evolution Strategies to Solve Mult Uncapacitated Facility Location Problem	ii-level 76
CS010 - Stroke Type Classification Model based on Risk Factors Using Resilient Backpr Neural Networks	opagation 77
CS011 - Implementation For Acute Respiratory Infection Clustering	78
STATISTICS (ST)	79
ST001 - Pairs Trading Strategy Based on Copula-GARCH Model	80
ST002 - Implementation of Agglomerative Hierarchical Clustering Algorithm for Severit COVID-19 in Indonesia	ty Level of 81
ST003 - Forecasting the Composite Stock Price Index in Pandemic Covid-19 in Indonesi ARDL (p,q) Model	a using 82
ST004 - Business Fields Affecting Gross Regional Domestic Product of East Kalimantan using Panel Data Regression Analysis	n Province 83
ST005 - Prediction of the Spread of Infectious Diseases with the Multi-state SIR Model Richards Curve Model in Multi wave Epidemic Cases	and 84
ST006 - Unsupervised Learning of Temporal Data based on Global Prediction Models	85
ST007 - Bootstrap Techniques for Multivariate Time Series. Testing the Equality of Quar Cross-Spectral Densities	ntile 8€
ST008 - A Comparison of Forecasting Method Between Fuzzy Time Series Markov Cha ARIMA in Forecasting Crude Oil Prices Amid the Covid-19 Pandemic	in and 87
ST009 - Zero Inflated Poisson Regression: A Solution of Overdispersion in Stunting Data	a 88
ST011 - Application of Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) on Infrastructure Inequality Mapping in Bantul Regency, Indonesia	89

ST012 - Poverty Gap Index Modeling in Bengkulu Province Using Truncated Spline Regression Longitudinal Data	n for 90
ST013 - Bayesian Survival on Insurance Payments to Predict Consumer Credit Risk	91
ST014 - Modelling Maternal Mortality Based On Health Service Facilities In Indonesia To Supp The Sustainable Development Goals (SDGs) Program In 2030 Using The Geographically Weig Poisson Regression (GWPR) Approach	port hted 92
ST016 - Estimation of Left Truncated Fréchet Distribution Parameters with Maximum Likeliho Method	od 93
ST018 - Forecasting The Volume of Electronic Money Transactions Using ARIMAX-GARCH Model and Support Vector Regression	94
ST019 - Forecasting Marketplace Stock Value in Indonesia Based on The Best Time Series Analysis Model	95
ST020 - Grouping of Provinces in Indonesia Based on Community Welfare Level Indicators Us Hierarchical Cluster Analysis	ing 96
ST021 - Analysis of Factors Influencing Public Perceptions of Smart City Concept of Indonesia New Capital City	ı's 97
ST022 - Predicting International Market Price of Gasoline Using Fourier Series Estimator	98
ST023 - Accounting the Effect of US-China Trade War and COVID-19 in Forecasting Gold Pri- using ARIMAX-GARCH Model	ce 99
ST024 - Factors Affecting the Preeclampsia of Pregnant Women in RSIA Sitti Khadijah 1 Makassar: A Logistic Regression Analysis	100
PARALLEL SESSION SCHEDULES	101

RESEARCH ARTICLE | DECEMBER 22 2023

Comparison method between fuzzy time series Markov chain and ARIMA in forecasting crude oil prices ⊘

S. D. Puspa 🖾; J. Riyono; F. Puspitasari; F. G. Mamarimbing

() Check for updates

AIP Conf. Proc. 2975, 080006 (2023) https://doi.org/10.1063/5.0181088



APL Energy



Latest Articles Online!





Read Now



Comparison Method Between Fuzzy Time Series Markov Chain and ARIMA in Forecasting Crude Oil Prices

S. D. Puspa^{1, a)} J. Riyono¹, F. Puspitasari², F. G. Mamarimbing¹

¹Department of Mechanical Engineering, Faculty of Industrial Technology, Universitas Trisakti, Jakarta 11440, Indonesia ²Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Trisakti, Jakarta 11440, Indonesia

a) Corresponding author: sofia.debi.puspa@trisakti.ac.id

Abstract. Crude oil is a vital natural resource needed worldwide and the most demanded commodity. Fluctuating oil prices can affect a country's economic conditions e.g., economic growth, inflation rate, money supply, exchange rate and interest rates. Consequently, statistical forecasting methods are needed for a more accurate prediction in period t to support decision-making. This study aims to predict crude oil prices during the Covid-19 pandemic and compare the performance of crude oil price forecasting using the Fuzzy Time Series (FTS) Markov Chain method and Autoregressive Integrated Moving Average (ARIMA) method. The data used is daily crude oil prices with West Texas Intermediate (WTI) Standard in US\$/barrel from March 3, 2020, to March 31, 2022. Forecasting with the FTS Markov Chain method resulted in a mean absolute percentage error (MAPE) of 2.76%, and root mean square error (RMSE) of 580.3. The best model for ARIMA is ARIMA (0,1,1) which produces MAPE of 3.85% and RMSE 856.7. Due to the MAPE & RMSE values in the FTS Markov Chain method being smaller than the ARIMA method. Hence, forecasting using the FTS Markov Chain has better performance than the ARIMA method in the forecasting of crude oil prices during the Covid-19 pandemic.

INTRODUCTION

Coronavirus Disease 2019 (Covid-19) is a new type of virus caused by SARS-Cov-2. The new virus variant is known to have originated from Wuhan, China and was first discovered in early December 2019 when a patient was diagnosed with unusual pneumonia [1]. The Covid-19 virus has quickly infected hundreds of countries in the world and has spread to Asia, Europe, the Middle East, America, and other regions. Resulting that in 2020, the World Health Organization (WHO) declared that the Covid-19 pandemic was a global pandemic. The Covid-19 pandemic has not only affected health, but it has also greatly affected the global economy.

Crude oil is a commodity that has an important role in the economy of a country. Crude oil as a vital input is needed in industrial production processes, especially to generate electricity, run production machines, and transport products to the market. In addition, oil is also important for sustainable economic and social development. However, it is clearly seen that during the Covid-19 pandemic, crude oil prices fluctuate. It appeared in Fig 1 that since 2019 the price of WTI (West Texas Intermediate) crude oil has shown a decline and dipped sharply until April 2020 as a result of the Covid-19 pandemic. This is due to the limited space for human activity which has an impact on decreasing demand for crude oil and overproduction [2]. Changes in crude oil prices are not only affected by the Covid-19 pandemic but also influenced by the policies of OPEC (Organization of the Petroleum Exporting Countries) members and conflicts in crude oil producing countries.

Proceedings of the International Conference on Mathematics, Computational Sciences, and Statistics 2022 AIP Conf. Proc. 2975, 080006-1–080006-17; https://doi.org/10.1063/5.0181088 Published by AIP Publishing, 978-0-7354-4773-8/\$30.00



FIGURE 1. The Development of World Oil Supply and Demand (Source: www.worldoil.com)

Along with the decreasing number of Covid-19 cases, industrial activity slowly continued to normal until at the end of 2021 crude oil prices fluctuated and tended to increase in the international market. In the period October - December 2021, the average oil price reached US\$ 77.3 per barrel. The increase in the oil price average in 2021 has escalated by about 42.2% compared to the oil price average in 2020. Fluctuating changes in oil prices have received considerable attention in recent decades because they have a significant political impact. Various attempts were made to explain the behavior of oil prices as well as to see the macroeconomic consequences. In Indonesia, fluctuations in oil prices have an impact on economic growth over a certain period, domestic inflation, money supply, real exchange rates, and interest rates [3]. Due to the resulting impact is being very significant, a statistical forecasting method is needed that can accurately predict crude oil prices during the Covid-19 pandemic.

In this study, oil price forecasting used the Fuzzy Time Series Markov Chain and Autoregressive Integrated Moving Average (ARIMA) methods. The ARIMA method is a suitable time series data forecasting method where the method can handle data that is not stationary in variance and mean such as crude oil prices data that moves fluctuating. ARIMA has been successfully applied at a much larger scale in various fields, mainly due to its easy-to-use concept and utility algorithm [4]. For the advantages of the Fuzzy Time Series (FTS) method, in previous research conducted by [5] that the forecasting results using the FTS Markov Chain method has a low error rate and has a high degree of accuracy.

The fuzzy time series method was first proposed by Song and Chissom by applying fuzzy logic to develop the basis of fuzzy time series. It is a dynamic process of a linguistic variable where the linguistic value is a fuzzy set [6]. However, based on the results of previous research by Tsaur that the FTS method modified with the Markov chain concept obtained a better level of accuracy than the FTS without Markov chain. The implementation of the Markov chain fuzzy time series method was first introduced by Tsaur (2011) in forecasting the Taiwan currency exchange rate against the dollar [7].

The ARIMA method was discovered by Box and Jenskin (1976). The ARIMA method (p, d, q) where p represents the order of the autoregressive process (AR), d represents the difference (differencing) and q represents the order of the moving average (MA) process. Box and Jenkins use ARIMA models for single-variable (univariate) time series. The steps of the ARIMA method are model identification, parameter estimation, parameter significance testing, diagnostic checking and forecasting [8].

Based on the described statement, the purpose of this study is to predict crude oil prices during the Covid-19 pandemic and compare the performance of forecasting crude oil prices using the FTS Markov Chain method and the ARIMA method. Hence, it is hoped that the prediction results can be considered in decision making related to crude oil prices for the government and economic practitioners. This study is limited to only entering historical data values from oil prices without the influence of exogenous factors on the model.

DATA AND METHODS

The secondary data used in this study is the daily data of West Texas Intermediate (WTI) standard crude oil prices during the Covid-19 pandemic, from March 3, 2020, to March 31, 2022. The data is the closing price of crude oil in units of US\$ per barrel and sourced from http://id.investing.com. The total number of data is 547 data with data excluding holidays.

Fuzzy Time Series

Fuzzy times series (FTS) was first developed by Song & Chissom (1993). Fuzzy time series defines a fuzzy relation formed by determining logical relationship of training data. FTS is a forecasting method with the concept of fuzzy logic which can overcome the analysis of data in the form of linguistic value that cannot be handled by classical time series methods.

Suppose U is a set of the universe, $U = \{u_1, u_2, u_3, ..., u_n\}$ then the fuzzy set A of U is defined by Equation (1):

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \dots + \frac{f_A(u_n)}{u_n}$$
(1)

Where f_A is a function member of the fuzzy set A, $f_A: U \to [0,1]$, $f_A(u_i)$ indicates the grade of membership of u_i in the fuzzy set A and $1 \le i \le n$ [9].

Several definitions of fuzzy time series are [10]:

Definition 1

Suppose X(t) ($t = \cdots, 0, 1, 2, \ldots$) is subset of **R**. Let X(t) be a universe of discourse on a set fuzzy $f_i(t)$ ($i = 1, 2, \ldots$). If F(t) is set of $f_i(t)$ then F(t) is referred to as fuzzy time series on X(t)

Definition 2

If F(t) is due to F(t-1) and denoted by $F(t-1) \rightarrow F(t)$, it can be written as follows $F(t) = F(t-1) \circ R(t,t-1)$ (2)

Where " \circ " is max-min composition operator. R(t, t-1) is a fuzzy logical relationship between F(t) and F(t-1), and can be expressed by $R(t, t-1) = \bigcup_{i,j} R_{i,j}(t, t-1)$ where \bigcup is union operator.

Definition 3

Suppose $F(t) = A_i$ is caused by $F(t - 1) = A_j$, then the fuzzy logical relationship is defined as $A_i \rightarrow A_j$

If there are fuzzy logical relationship obtained from state A_2 , then a transition is made to another state A_j where j = 1, 2, ..., n, as $A_2 \rightarrow A_3$, $A_2 \rightarrow A_2$, ..., $A_2 \rightarrow A_1$; hence the fuzzy logical relationship are grouped into a fuzzy logical relationship group as:

$$A_2 \to A_1, A_2, A_3 \tag{3}$$

Fuzzy Time Series Markov Chain

Fuzzy Time Series Markov Chain flowchart shown in Fig 2 and the method steps are as follows [7, 9]: **Step 1:** Define the universe of discourse *U*

Determine minimum value D_{min} and maximum value D_{max} of historical data, then define the universe of discourse U as follows:

$$U = [D_{min} - D_1, D_{max} + D_2]$$
(4)

where D_1 and D_2 are two positive numbers that determined by researcher



FIGURE 2. Flowchart of Fuzzy Time Series Markov Chain

Step 2: Calculate the number of fuzzy intervals into several equal length interval i. Determining the number of class interval using Sturgess Rule in Equation (5):

$$K = 1 + (3,3 \log N)$$
 (5)

where

K : number of intervals

n : amount of data

ii. Determining the length of class interval by Equation (6):

$$l = \frac{[(D_{max} + D_2) - (D_{min} - D_1)]}{K}$$
(6)

where *l* : interval length

Step 3: Defining the fuzzy set A_i in the universe of discourse U. For every fuzzy set A_i (i = 1, 2, ..., n) defined in the number of intervals which have been specified, where $A_1, A_2, ..., A_n$ defined by

$$A_{1} = \{1/u_{1} + 0.5/u_{2} + 0/u_{3} + \dots + 0/u_{n}\}$$

$$A_{2} = \{0, 5/u_{1} + 1/u_{2} + 0, 5/u_{3} + \dots + 0/u_{n}\}$$

$$\vdots$$

$$A_{n} = \{0/u_{1} + \dots + 0, 5/u_{n-1} + 1/u_{n}\}$$
(7)

Step 4: Fuzzification of historical data. Fuzzification is a change in the form of real value (crisp) into the form fuzzy by mapping the real value into fuzzy set that correspond. If a time series data sets are on intervals, u_i then the data is fuzzified into fuzzy set A_i .

Step 5: Determining the fuzzy logical relationship and fuzzy logical relationships group (FLRG). Based on **Definition 3**, fuzzy logical relationship group can be easily obtained.

Step 6: Defining Markov probability transition matrix.

FLRG is utilized to get the probability of the next state. Therefore, we get a Markov transition matrix with dimension $n \times n$. The transition probability formula can be written as:

$$P_{ij} = \frac{M_{ij}}{M_i}; \ i, j = 1, 2, \dots, n$$
(8)

where:

 P_{ij} : Probability of transition from state A_i to state A_j by one step

 M_{ii} : number of transitions from state A_i to state A_j by one step

 M_i : the quantity of data included in the A_i state

The probability matrix state *P* can be written as follows:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

where $\sum_{i=1}^{n} P_{ii} = 1$

For the matrix *P*, several definitions as follows:

1. If $P_{ij} \ge 0$, then state A_j is accessible from state A_i

2. If states A_i and A_j are accessible to each other, then A_i communicates with A_j

Step 7: Define defuzzification forecasting value

To generate forecasting value from the obtained probability matrix then it can be calculated by the rule as following:

1. If FLRG A_i is empty, $(A_i \rightarrow \emptyset)$ then forecasting value is m_i that the middle value of u_i can be written by

$$F(t) = m_i$$

2. If FLRG A_i is one to one (assume $A_i \rightarrow A_k$ where $P_{ij} = 0$ and $P_{ik} = 1, j \neq k$) then the forecasting value is m_k the middle value of u_k

$$F(t) = m_k P_{ik} = m_k$$
(9)
3. If FLRG A_i is one to many (assume $A_j \rightarrow A_1, A_2, \dots, A_n; j = 1, 2, \dots n$). And if $Y(t - 1)$ at time $(t - 1)$ which is on state, A_j then the forecasting value is

$$F(t) = m_1 P_{j1} + m_2 P_{j2} + \dots + m_{j-1} P_{j(j-1)} + Y(t-1) P_{jj} + m_{j+1} P_{j(j+1)} + \dots + m_n P_{j(n)}$$
(10)

where:

 m_1, m_2, \dots, m_n : the midpoint of u_1, u_2, \dots, u_n

Y(t-1) : actual value from state A_i at time t-1

Step 8: Determine the adjustment value to forecasting result. Adjustment forecasting used to review forecasting error. The adjusting rules for forecasting value is explained as follows:

1. If state A_i communicates with A_i , starting in state A_i at time t - 1 as $F(t - 1) = A_i$ and occur an increasing transition into state A_j at time t, (i < j), then the adjusting value is determined as:

$$D_{t1} = \left(\frac{l}{2}\right) \tag{11}$$

where:

- *l* : interval length
- 2. If state A_i communicates with A_i , starting in state A_i at time time t 1 as $F(t 1) = A_i$ and occur a decreasing transition into state A_i at time t, (i > j), then the adjusting value is determined as:

$$D_{t1} = -\left(\frac{l}{2}\right) \tag{12}$$

3. If the current state is in state A_i at time t - 1 as $F(t - 1) = A_i$, and occur a jump-forward transition into state A_{i+s} at time t, $(1 \le s \le n - i)$, then the adjusting value is determined as:

$$D_{t2} = \left(\frac{l}{2}\right)s, (1 \le s \le n - i)$$
(13)

where

- *s* : the number of forward transitions
- 4. If the current state is in state A_i at time t 1 as $F(t 1) = A_i$ and occur a jump-backward transition into state A_{i-v} at time $t, (1 \le v \le i)$, then the adjusting value is determined as:

$$D_{t2} = -\left(\frac{l}{2}\right)v, (1 \le v \le i) \tag{14}$$

where

v: the number of backward transitions

Step 9: Calculate the adjusted forecasting results. In general, for forecasting result
$$F'(t)$$
 can be obtained as:

$$F'(t) = F(t) \pm D_{t1} \pm D_{t2} = F(t) \pm \left(\frac{l}{2}\right) \pm \left(\frac{l}{2}\right) \nu$$
(15)

Stationary in Time Series Forecasting

In a data, it is possible that the data is not stationary because the mean or variance is not constant, hence, to eliminate the non-stationarity to the mean, the data can be made close to stationary by using the method of differencing. The behavior of stationary data includes not having too large variations and tends to approach the mean value, and vice versa for non-stationary data [11].

1. Stationary in Variance

The data that is not stationary in variance can be transformed. Therefore, the data becomes stationary in variance by doing a power transformation calculation. Box and Cox in 1964 introduced power transformation as follows [8]:

$$Z'_{t} = \begin{cases} \frac{Z^{\lambda}_{t} - 1}{\lambda}, \lambda \neq 0\\ \ln(Z_{t}), \lambda = 0 \end{cases}$$
(16)

 Z_t : time series in period t

 λ : transformation parameter

2. Stationary in Mean

The Augmented Dickey-Fuller (ADF) test is one of the tests that can be employed to evaluate the stationary of time series data in mean. It is to see whether the model obtained has or does not have a unit root. The data that is not stationary in the mean can be stationary through the process differencing. This test has a regression model as follows: 1

$$\Delta Z_{t} = \delta Z_{t-1} + \sum_{i=2}^{n} \gamma_{i} \Delta Z_{t-1+1} + a_{t}$$
(17)

Hypothesis testing for ADF model is $H_0: \rho = 0$ (Have unit root) $H_1: \rho \neq 0$ (No unit root)

The null hypothesis is tested by t-statistics which is given by this formula:

$$|\tau| = \frac{\widehat{\delta}}{SE(\widehat{\delta})} \tag{18}$$

where

ŝ : estimated value of parameter δ $SE(\hat{\delta})$: standard error for estimated value δ Reject H_0 if $|\tau| > |\tau_{\alpha;df}|$ or p-value $< \alpha$ [12].

Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average is a forecasting method that can predict time series data stationary in variance and mean. For the data that did not fulfil the stationary, the transformation and differentiation process can be conducted, then ARIMA method can be used if the data has met stationary. In addition, it takes a lot of object data to determine the best ARIMA model on the observed object. The ARIMA (p, d, q) model is a univariate time series that merges the autoregressive (AR) and moving average (MA). The general form of ARIMA (p, d, q) can be written as [13]:

$$\phi_p(B)(1-B)^a y_t = \theta_q(B)\varepsilon_t \tag{19}$$

$$\phi_p(B) = (1 - \phi_1(B^1) - \phi_2(B^2) - \dots - \phi_p(B^p))$$

$$\theta_q(B) = (1 - \theta_1(B^1) - \theta_2(B^2) - \dots - \theta_q(B^q))$$

where

 y_t : observation at time t ϕ_p : AR coefficient on order p θ_q : MA coefficient on order q ε_t : error at time t d: degree of differentiation B: Backshift operator

1

Forecasting steps for ARIMA method is as follows:

Step 1: Model Identification

The identification of ARIMA model is based on the pattern shown in the auto correlation (ACF) and the partial correlation (PACF) plot of the data already stationery. Theoretical behavior of ACF and PACF plot is shown on Table 1: TABLE 1 Theoretical behavior of AP(p) MA(q)

TABLE 1. Theoretical behavior of AR(p), MA(q),						
and ARMA(p, q) models on ACF and PACF plots						
Model ACF PACF						
AR (p)	Tails off	Cuts off after lag p				
MA(q)	Cuts off after lag q	Tails off				
ARMA (p, q)	Tails off and or cuts off	Tails off and or cuts off				

Step 2: Parameter Estimation

There are several methods for parameter estimation such as Moment Method, Maximum Likelihood and Least Square which can be used to estimate the parameters in models [13].

Step 3: Parameter Significance Test

The parameter significance test steps are as follows:

a. The model for parameter significance test:

$$\phi_p(B)(1-B)^d y_t = \theta_q(B)\varepsilon_t \tag{20}$$

with

$$\begin{aligned} \phi_p(B) &= (1 - \phi_1(B^1) - \phi_2(B^2) - \dots - \phi_p(B^p)) \\ \theta_q(B) &= (1 - \theta_1(B^1) - \theta_2(B^2) - \dots - \theta_q(B^q)) \end{aligned}$$

b. Hypothesis testing for AR model

 $H_0: \phi_p = 0$ (Not significant parameter)

 $H_1: \phi_p \neq 0$ (Significant parameter)

Statistical test calculation for AR model is:

$$t_{calc} = \frac{(\hat{\phi}_l - 0)}{SE(\hat{\phi}_l)} \tag{21}$$

Reject H_0 if $|t_{calc}| > t_{\frac{\alpha}{2}(n-1)}$

Hypothesis testing for MA model $H_0: \theta_q = 0$ (Not significant parameter) $H_1: \theta_q \neq 0$ (Significant parameter)

Statistical test calculation for MA model is:

$$t_{calc} = \frac{\hat{\theta}_l - 0}{SE(\hat{\theta}_l)} \tag{22}$$

Reject H_0 if $|t_{calc}| > t_{\frac{\alpha}{2}(n-1)}$

with

If there is one parameter that is not significant then return to Step 1

Step 4: Diagnostic Checking

Diagnostic checking models are carried out to ensure if the remaining models meet the white noise. The following is steps of diagnostic checking [14]:

a. Hypothesis testing Ljung Box

 H_0 : there is no residual autocorrelation

 H_1 : there is residual autocorrelation

b. Determine Ljung Box test statistics:

$$Q = n(n+2)\left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \dots + \frac{\hat{r}_k^2}{n-K}\right)$$
(23)

Where:

Q : Ljung-Box Statistics Test

n: number of observation data

K : number of lag observed

 \hat{r}_i^2 : estimated correlation of residuals on the i-th of the lag with i = 1, 2, ..., K

Reject H_0 if $Q > \chi^2_{(K-p-q)}$ If there is autocorrelation between the residuals, then return to **Step 1**

Step 5: Overfitting

Overfitting is applied to get the best model with addition or subtraction of the order of AR (p) and MA (q) parameters from the tentative model has been obtained. The best model is the model with significant parameter and residual series that does not have autocorrelation.

Step 6: Selection of ARIMA Model

The selection of the best ARIMA model is minimize the information criteria such as Akaike Information Criterion (AIC) value.

$$AIC = n \ln \hat{\sigma}_{\varepsilon}^{2} + 2(p+q+1)$$
(24)
$$\hat{\sigma}_{\varepsilon}^{2} = SSE/n$$
(25)

However, it is known that for the autoregressive model, the AIC criterion does not gives a consistent order of p, hence for comparison using other information criteria such as Schwarzt Bayesian Information Criterion (SBC) $SBC = n \ln \hat{\sigma}_{\varepsilon}^2 + (p + q + 1) \ln n$ (26)

Model Selection Criteria

The selection of the best model is by comparing the error values forecasting of the method used. A method is better compared to other methods if the forecast error value is produced smaller.

If $Z_1, Z_2, ..., Z_n$ state the whole data, and in sample data can be stated as $Z_1, Z_2, ..., Z_m, m < n$. If the adjusted value is $\hat{Z}_1, \hat{Z}_2, ..., \hat{Z}_m, m < n$, the value of MSE, RMSE and MAD for in sample data defined as [15].

$$MSE = \sum_{t=1}^{m} \frac{Z_i - \hat{Z}_i}{n}, \ m < n$$
 (27)

$$RMSE = \sqrt{\sum_{t=1}^{m} \frac{Z_i - \hat{Z}_i}{n}}, \ m < n$$
(28)

$$MAD = \sum_{t=1}^{m} \frac{|Z_i - \hat{Z}_i|}{n}, \ m < n$$
(29)

Furthermore, the accuracy models can be also measured by Mean Absolute Percentage Error (MAPE) with the following formula [13]

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$
(30)

where:

 A_t : Actual data value at period t F_t : Forecast data value at period tn: the number of data

TABLE 2. MAPE Value Interpretation				
MAPE Judgment of Forecast Accurac				
< 10%	Highly Accurate			
$10\% \le MAPE \le 20\%$	Good Forecast			
$20\% \le MAPE \le 50\%$	Reasonable Forecast			
≥ 50%	Inaccurate Forecast			

The smaller the MAPE value, the better the model. Table 2 shows a scale to assess the accuracy of a model based on MAPE value was developed by Lewis (1982) [16]. Root Mean Square Error (RMSE) is the magnitude of the prediction error rate, where the smaller (closer to 0) the RMSE value, the more accurate the prediction results will be. In this study, the model selection criteria used are Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

RESULTS

Data Description

Table 3 shows that the average price of crude oil is US\$58.68 per barrel with spread data of US\$21.92 per barrel from the average. The lowest price was US\$ -37.63 per barrel, while the highest price was US\$123.7 per barrel. Skewness shows a positive value of 0.077 or the values are concentrated on the right side (located to the right of M_0). Thereby, the curve has a tail that extends to the right or the curve skews to the right. Then, with a kurtosis value of less than 3 which is 0.159 then it can be said to be a Platykurtic curve. And Fig 3 shows a time series of crude oil prices for the period March 3, 2020 – March 31, 2022.

TABLE 3 . Descriptive Statistics				
Ν	547			
Mean Value	58.68			
Standard Deviation	21.92			
Minimum Value	-37.63			
Maximum Value	123.70			
Skewness	0.077			
Kurtosis	0.159			



FIGURE 3. Time Series Plot

Fuzzy Time Series Markov Chain Forecasting Analysis

Step 1: Define the universe of discourse U

Based on historical data obtained the minimum value $D_{min} = -37.63$ and maximum value $D_{max} = 123.70$ with $D_1 = 0.37$ and $D_2 = 0.3$ which has been determined by the researcher hence U = [-38, 124]

Step 2: Specifying Class Interval

- a. Determine the number of class intervals The calculation of the number of class intervals is determined by using Sturgess Rule and obtained $K = 1 + (3,3 \log N)$ $K = 1 + (3,3 \log 547)$ $K = 10,035 \approx 10$
- b. Determine the length of the class interval $[(D_{111}, \dots, D_{n+1})^{-1}, (D_{111}, \dots, D_{n+1})^{-1}]$

$$l = \frac{\left[(D_{max} + D_2) - (D_{min} - D_1)\right]}{K}$$
$$l = \frac{\left[(123.7 + 0.3) - (-37.63 - 0.37)\right]}{10}$$
$$l = 16.2 \approx 16$$

Furthermore, divide the universe of discourse U into several partitions according to the number of class intervals, which is 10 and the length of the class interval is 16. Hence the interval and the middle interval can be seen in Table 5.

TABLE 4. Class Interval and Middle Value Interval					
Interval	Middle Value Interval				
$u_1 = [-38, -21.8]$	-29.9				
$u_2 = [-21.8, -5.6]$	-13.7				
$u_3 = [-5.6, 10.6]$	2.5				
$u_4 = [10.6, 26.8]$	18.7				
$u_5 = [26.8, 43]$	34.9				
$u_6 = [43, 59.2]$	51,1				
$u_7 = [59.2, 75.4]$	67.3				
u ₈ = [75.4 , 91.6]	83.5				
$u_9 = [91.6, 107.8]$	99.7				
$u_{10} = [107.8, 124]$	115.9				

Step 3: Fuzzification of actual data

Based on the fuzzy set that has been formed, where the oil price data is converted into the form of linguistic values. The results of fuzzification are notated into linguistic numbers can be seen in Table 5.

TABLE 5. Crude Oil Price Data Fuzzification						
Date	Actual Data (Y _t)	Interval	Fuzzification			
03/03/2020	47.18	$u_6 = [43, 59.2]$	A_6			
04/03/2020	46.78	$u_6 = [43, 59.2]$	A_6			
05/03/2020	45.90	$u_6 = [43, 59.2]$	A_6			
06/03/2020	41.28	$u_5 = [26.8, 43]$	A_5			
07/03/2020	31.13	$u_5 = [26.8, 43]$	A_5			
:	:		÷			
30/03/2022	107.82	$u_{10} = [107.8, 124]$	A_{10}			
31/03/2022	100.28	$u_9 = [91.6, 107.8]$	A_9			

Step 4: Determining the Fuzzy Logical Relationship and Fuzzy Logical Relationships Group (FLRG). In Table 6 is shown Fuzzy Logical Relationship (FLR) which is the relationship between each data to the next data in the form of a fuzzy set based on the determination of fuzzification. After obtaining the FLR, then the FLRG is determined which is the grouping of each state transfer namely the current state and the next state.

TABLE 6. Fuzzy Log	ical Relationship (FLR)
Data Order	FLR
1 - 2	$A_6 \rightarrow A_6$
2 - 3	$A_6 \rightarrow A_6$
3 - 4	$A_6 \rightarrow A_5$
4 - 5	$A_5 \rightarrow A_5$
5 - 6	$A_5 \rightarrow A_5$
:	:
545 - 546	$A_{10} \rightarrow A_9$
546 - 547	$A_9 \rightarrow A_9$

Step 5: Defining Markov probability transition matrix.

The transition probability matrix in this study is 10×10 because, in **Step 1**, the number of class intervals is 10. The following is the transition probability matrix:

			J	-						
	^[0.857]	0.142	0	0	0	0	0	0	0	ך 0
	0.040	0.800	0.160	0	0	0	0	0	0	0
	0	0.267	0.600	0.133	0	0	0	0	0	0
	0	0	0.016	0.812	0.172	0	0	0	0	0
D —	0	0	0.011	0.105	0.863	0.021	0	0	0	0
г –	0	0	0	0	0.029	0.941	0.029	0	0	0
	0	0	0	0	0	0	0.867	0.133	0	0
	0	0	0	0	0	0	0.059	0.871	0.071	0
	0	0	0	0	0	0	0	0.067	0.892	0.041
	LΟ	0	0	0	0	0	0	0	0.086	0.914

Step 6: Define defuzzification and determine the adjustment value to forecasting result. After the transition probability matrix is formed, the next step is the calculation process for forecasting and defuzzification of the previously obtained FLRG. Before determining the forecasting results, the Markov chain fuzzy time series method has a step to adjust the forecasting results for each relationship between the current state and the next state.

Step 7: Determine the adjusted forecasting

After the adjustment value is obtained, then the forecasting results are determined which have been adjusted can be seen in Table 7.

TABLE 7. Adjusted Forecasting Results						
Date	Actual	Initial	Adjustment	Final		
	Data (Y _t)	Forecasting	Value	Forecasting		
		(\boldsymbol{F}_t)		$(\boldsymbol{F'}_t)$		
03/03/2020	47.18	-	-	-		
04/03/2020	46.78	45.921	0	45.921		
05/03/2020	45.90	45.575	0	45.575		
06/03/2020	41.28	46.244	-2	44.816		
07/03/2020	31.13	36.338	-4	32.828		
08/03/2020	34.36	32.233	1	33.177		
:	:	:	:	:		
30/03/2022	107.82	105.598	0	105.598		
31/03/2022	100.28	100.043	0	100.043		



FIGURE 4. Plot Using FTS Markov Chain

The Fuzzy Time Series Markov Chain method has a very good performance in forecasting crude oil prices during the Covid-19 pandemic. This is indicated by the resulting MAPE value of 2.76% (less than 10%) and RMSE 580.3 on daily data for March 2020 - March 2022. This is supported by the similarity plot between the forecast data and the actual data as presented in Fig 4. And the results of forecasting crude oil prices for the next 5 consecutive days period (April 1st -7th, 2022) are US\$99.83, US\$105.12, US\$101.89, US\$98.31, US\$96.93 (per barrel).

ARIMA Forecasting Analysis

In time series analysis, stationarity is required to minimize model errors in order to get the best model. The first step to determine the stationarity of the data is to make a time series plot. From Fig 3 it is obtained that the data is not stationary due to the fluctuating data where the data is not around a constant mean or on average does not form an almost horizontal trend.

In addition, the stationarity of the data can be determined by the unit root test using the Augmented Dickey Fuller (ADF) test. The unit root test hypothesis is:

 $H_0: \rho = 0$ (There is a unit root)

 $H_1: \rho \neq 0$ (No unit root)

Augmented Dickey-Fuller Test data: price Dickey-Fuller = -3.0149, Lag order = 7, p-value = 0.1487 alternative hypothesis: stationary

FIGURE 5. Unit Root Test Result with ADF Test

From Fig 5 above obtained p-value, 0.1487 is greater than $\alpha = 0.05$ then accept H_0 so that it can be said that there is a unit root which means the data is not stationary. Therefore, a differencing process is needed which was previously carried out by the natural logarithm transformation process on 547 crude oil price data during the Covid-19 pandemic hence the data becomes stationary with respect to the mean and variance.

Furthermore, the natural logarithm transformation process and differencing process are carried out. And from Fig 6 it is obtained that the data is stationary at the first difference level, thus indicating the value of d = 1. To strengthen the existence of stationary data, Fig 7 shows the results of the ADF test with the data from the logarithmic differencing process. The p-value obtained from the unit root test results is 0.01 < 0.05 (α) so that H_0 is rejected, which means that there is no unit root so that the data is stationary.



FIGURE 6. The Result of Differencing-Natural Logarithm Plot



FIGURE 7. The Result of Differencing-Natural Logarithm by Using ADF Test

This is also in line with the ACF plot and PACF plot where there are no more than 3 lags that come out of the confidence interval shown in Fig 8(a) & 8(b). Because of the time series data has met the stationary requirements. Therefore, forecasting can be done and there is no need to do further differencing (second order differencing).



FIGURE 8. (a) ACF Plot & (b) PACF Plot After Natural Logarithm Differencing Process

After pre-processing the data, the next step is to identify and estimate the best ARIMA model from the results of the transformation and differencing processes. Based on the analysis of the order correlogram on the Auto Regressive (p) PACF plot in Fig 8 (b), the significant order is lag 18 and from Fig 8 (b) the ACF plot also shows that the significant order in the Moving Average (q) is lag 18. Significant order is obtained by considering the outgoing correlogram. of confidence intervals. So, the models estimated on crude oil price data during the Covid-19 pandemic are ARIMA (1,1,0), ARIMA (0,1,1), and ARIMA (1,1,1).

Furthermore, the parameters selected into the model are if the p-value or significance value for each parameter is less than α (0,05). The hypothesis used is as follows:

 H_0 : Parameters are not significant in the model

 H_1 : Significant parameters in the model

The results of the significance test on the three ARIMA models estimated in Table 8 show that only ARIMA (0,1,1) has significant parameters where p-value MA (1) is 0.023 less than α (0.05) then reject H_0 which means only ARIMA model (0,1,1) has significant parameters in the model. Furthermore, the selection of the best model is also carried out by analyzing the Akaike information criterion (AIC), Schwarz information criterion (SIC) and Hannan-

24 December 2023 10:15:30

Quinn criterion (HQC). Table 8 shows that the AIC, SIC and HQC values in ARIMA (0,1,1) have the lowest values compared to other ARIMA models. In addition, the MSE value on ARIMA (0,1,1) is also the lowest, so it can be said that ARIMA (0,1,1) is the best model to use on crude oil price data during the Covid-19 pandemic.

	TABLE 8.	ARIMA M	Iodel Est	timation		
Model/	Variable	p-value	MSE	AIC	SIC	HQC
Parameter		-				
ARIMA (1,1,0)	AR (1)	0.058	2.83	3.750	3.776	3.760
Constant		0.459				
ARIMA (0,1,1)	MA (1)	0.023	2.47	3.689	3.694	3.690
Constant		0.400				
ARIMA (1,1,1)	AR (1)	0.246	2.78	3.723	3.741	3.753
	MA (1)	0.812				
Constant		0.427				

	TABLE 9. Dia	agnostic Checking
Lag	p-value	Decision
12	0.392	white noise
24	0.299	white noise
36	0.618	white noise
48	0.656	white noise

To obtain the best ARIMA model, diagnostic checking includes white noise and normal distribution of residuals. The white noise test with the null hypothesis is that there is no residual correlation between lags. Table 9 shows the p-value by Ljung Box Statistics for ARIMA (0,1,1). From Table 9 the value of all p-values at lags to 12, 24, 36 and 46 is more than α (0.05) so accept H_0 that the residuals do not contain autocorrelations or white noise. This is reinforced through the residual ACF and PACF correlograms, in Fig 9 it is shown that probability > α (0.05) which resulting the white noise model.

Autocome lation	Partal Canalation	<u>.</u>	AQ.	PAC	0-8tøl	Prop.
10	10 (1)	1.1	0.001	0.000	T.E-05	
30	116	12	-0.015	-0.045	0.1037	0.747
0.0	0.0	3	-0.018	-D.04R	0.2588	0.875
	- 4C	14	0.013	0.013	0.3375	0.953
	100	1.5	0.058	0.057	1.9603	0.743
10	ų ie	6	-0.024	-0.024	2.2439	0.814
10	0.	7	0.053	0.055	3.6042	0.730
10	0.	8	0.038	0.037	4,2200	0.754
0	0	0	0.077	0.077	7.1438	0.521
4	•	10	-0.048	-0.04B	8.2077	0.047
10	10	11	-0.021	-0.016	8.4843	0.582
10	13	12	0.065	0.059	10.567	0.480
101	()	13	-3.025	-D.033	10,938	0.533
10	96	14	0.052	0.055	12.707	0.464
- D	40	15	-0.023	-0.014	12,982	0.525
19		16	0.068	0.051	14,765	0.468
11	di C	17	-0.008	-0.015	\$4,797	0.540
10	10 C	18	0.087	0.084	18.622	0.361
	()	19	-0.058	-0.056	19,886	0.355
10	10 10	20	0.017	0.016	20.030	0.383

FIGURE 9. The Output of ACF and PACF Plot for Residual Autocorrelation Test

Normality testing was carried out using the Kolmogorov-Smirnov test with the null hypothesis is the residuals are normally distributed. The residual is normally distributed if the p-value is more than α with α value is 0.05.

The results of the normality test using the Kolmogorov-Smirnov test showed that the p-value in ARIMA (0,1,1) residual probability plot has a value of more than 0.120. Hence, it can be said that the residuals in the model are normally distributed. Due to the model satisfies the white noise test and has a normal distribution of error, the ARIMA model (0,1,1) is the best model so that it can proceed to the next stage, which is forecasting.

The best model used is ARIMA (0,1,1) with constants written in the form of an equation, the following model is obtained:

$$(1-B)\hat{Y}_t = 0.0587 + 0.0293\varepsilon_{t-1} \tag{31}$$

where $\hat{Y}_t = \ln \hat{Z}_t$, then the forecast value for Z_t is

$$\hat{Z}_t = \exp(\hat{Y}_t)$$
(32)

where \hat{Z}_t : crude oil price prediction

MAPE & RMSE resulting from the ARIMA method are 3.85% and 856.7 which shows that ARIMA performance is highly accurate. Fig 10 shows a plot between the predicted and actual values using the ARIMA method. Table 10 shows the prediction results by using the ARIMA Model (0,1,1) for the next 5 consecutive days period (April 1st -7th, 2022):



FIGURE 10. Plot Using ARIMA Method

	TABLE IV. C	rude Oli Prices (Confidence	ce Interval 95%)
Data		Crude Oil Prices (US\$	per barrel)
Date	Prediction	The Lowest Prediction	The Highest Prediction
04/01/2022	100.831	94.285	108.364
04/04/2022	102.976	96.199	109.752
04/05/2022	100.598	93.287	108.451
04/06/2022	95.711	91.143	105.639
04/07/2022	95.832	92.547	107.433

ABLE 10 Crude Oil Prices (Confidence Interval 05%)

Comparison Forecasting Methods

Comparison of forecasting accuracy can be done visually and analytically. A comparison of forecasting accuracy visually is done by comparing the estimated value and the actual value using a time series plot. While the comparison of the accuracy of forecasting time series data analytically is done by comparing the forecasting error values between methods. The comparison of the forecasting results of the ARIMA and Fuzzy Time Series (FTS) Markov Chain methods visually on the crude oil price data is done by comparing the plots of the actual value and the forecast value of crude oil prices as presented in Fig 11. Based on Fig 11, the forecast value plot of the FTS Markov Chain method is more similar or closer to the actual value pattern when compared to the ARIMA method, which means that forecasting using the FTS Markov Chain method is better than the ARIMA method on the data as many as 547 observations.



FIGURE 11. Actual Data & Forecasted Data Plots Using FTS Markov Chain & ARIMA Methods

Analytically, the comparison of forecasting accuracy in this study is conducted by using MAPE and RMSE values as a measure of forecasting error. The MAPE value obtained by the FTS Markov Chain method is 2.76% and the RMSE is 580.3. While the MAPE value generated by the ARIMA method is 3.85% and RMSE is 856.7. Due to the MAPE and RMSE values in the FTS Markov Chain method are smaller than the ARIMA method, it shows that the FTS Markov Chain method works better than the ARIMA method to forecast crude oil price data during the Covid-19 pandemic for the 2020-2022 period. The results visually and analytically yield the same conclusion that the FTS Markov Chain method has better performance than the ARIMA method. The results of forecasting time series data on crude oil prices during the Covid-19 pandemic using the FTS Markov chain method and the ARIMA method alongside the MAPE and RMSE values can be seen in Table 11.

Data	Crude C	il Prices (US\$ per barre	el)
Date	Actual Data (Y_t)	FTS Markov Chain	ARIMA
03/03/2020	47,18	-	-
04/03/2020	46,78	45.921	47.238
05/03/2020	45,90	45.575	46.852
06/03/2020	41,28	44.816	45.986
07/03/2020	31,13	32.828	41.477
08/03/2020	34,36	33.177	31.492
:	:	:	:
30/03/2022	107.82	105.598	104.182
31/03/2022	100.28	100.043	107.364
	MAPE	2.76%	3.85%
	RMSE	580.3	856.7

TABLE 11. Comparison of Actual Data with Predicted Data for Crude Oil Prices

CONCLUSION

This study aims to predict crude oil prices during the Covid-19 pandemic and compare the performance of crude oil price forecasting by using the Fuzzy Time Series (FTS) Markov Chain method and Autoregressive Integrated Moving Average (ARIMA) method. The data used is daily data on the crude oil prices with West Texas Intermediate (WTI) Standard in US\$/barrel for the period March 3, 2020 – March 31, 2022. The Fuzzy Time Series Markov Chain method has an excellent performance in forecasting crude oil prices during the Covid-19 pandemic. This is indicated by the resulting MAPE value of 2.76% (less than 10%) and RMSE 580.3 on daily data for March 2020 - March 2022. ARIMA model (0,1,1) with constants is the best ARIMA model for modeling actual data on crude oil prices for the period of March 2020 - March 2022 during the Covid-19 pandemic so that this model can be used for prediction of future world crude oil prices Covid-19 pandemic. MAPE of this model is 3.85% and RMSE is 856.7. Based on a visual and analytical comparison, it can be concluded that the Fuzzy Time Series Markov Chain method works better than the ARIMA method in forecasting crude oil prices during the Covid-19 pandemic on period March 2020 - March

2022. These results have several important implications for Indonesia, especially on policy recommendations and economic development due to changes in oil prices that have an impact on several sectors.

ACKNOWLEDGMENTS

This research is supported by PDP Universitas Trisakti 2021 & Faculty of Industrial Technology USAKTI

REFERENCES

- 1. Y.C. Wu, et al, Journal of the Chinese Medical Association, vol. 83, issue 3, pp. 217-220, march 2020.
- 2. N.L. Widyastuti and H. Nugroho, JPP, vol. 4, no. 2, pp. 166-176, June. 2020.
- 3. M. A. Nizar, Buletin Ilmiah Litbang Perdagangan, vol.6, no. 2, 2012.
- 4. O.D. Ilie, et al, Microorganisms, 8(8), 1158.
- 5. Zaenurrohman, S Hariyanto, T.Udjiani, "Fuzzy Time Series Markov Chain and Fuzzy time series Chen & Hsu for Forecasting", *Journal of Physics: Conference Series*, 1943(1): 012128, July 2021.
- 6. Q. Song and B. S. Chissom, Fuzzy Sets and Systems, vol. 54, no.1, pp.1-9, February 1993.
- R. C. Tsaur, International Journal of Innovative Computing, Information and Control, vol 8, no 7(b), pp 4931-4942, July 2012.
- 8. A. Aswi dan Sukarna. Analisis Deret Waktu: Teori dan Aplikasi. Makassar: Andira Publisher, 2006.
- 9. H.T. Jasim, A.G.J. Salim and K.I. Ibraheem, Applications and Applied Mathematics, Vol. 7, Issue 1, pp. 385-397. June 2012.
- 10. Q. Song and B.S. Chissom, Fuzzy Sets and Systems, vol. 54, Issue 3, pp. 269-277, March 1993.
- 11. D. N. Gujarati. Basic Econometric (4th ed). The McGraw-Hill Companies, 2004.
- D. A. Dickey and W.A. Fuller, Journal of The American Statistical Association, vol.74, no.366, pp. 427-431, June 1979.
- 12. D. C. Montgomery, C. L. Jennings, and M. Kulahci. *Introduction to Time Series Analysis and Forecasting*. New Jersey (US): J Wiley, Inc, 2008.
- 13. J. D. Cryer and K. S. Chan. Time Series Analysis with Applications in R (2nd ed.). New York: Springer, 2008.
- 14. D. Chicco, M.J. Warrens, and G. Jurman, PeerJ Computer Science, 7:e623.
- 15. C. D. Lewis. Industrial and business forecasting methods: A practical guide to exponential smoothing and curve *fitting*. London: Butterworth Scientific, 1982

Comparison method between fuzzy time series Markov chain and ARIMA in forecasting crude oil prices

by Sofia Debi Puspa Debi Puspa

Submission date: 11-Jun-2024 10:05AM (UTC+0700) Submission ID: 2400042896 File name: Proceeding_Fuzzy_ARIMA_AIP_Sofia.pdf (1.64M) Word count: 7109 Character count: 32674

AIP Conference Proceedings

RESEARCH ARTICLE | DECEMBER 22 2023

Comparison method between fuzzy time series Markov chain and ARIMA in forecasting crude oil prices @

S. D. Puspe 🛋 J. Riyono; F. Puspitasari; F. G. Mamarimbing

Check for updates
 Conf. Proc. 2975, 080006 (2023)
 IDst.ndstrog/10.10635, 0151034

		CrossMark
View Online	Export	

24 December 2023 10 10:00

と AIP Publishing









Comparison Method Between Fuzzy Time Series Markov Chain and ARIMA in Forecasting Crude Oil Prices

S. D. Puspa^{1, a)} J. Riyono¹, F. Puspitasari², F. G. Mamarimbing¹

¹Department of Mechanical Engineering, Faculty of Industrial Technology, Universitas Trisakti, Jakarta 11440, Indonesia ²Department of Industrial Engineering, Faculty of Industrial Technology, Universitas Trisakti, Jakarta 11440, Indonesia

*1 Corresponding author: so fia.debi.puspa@trisakti.ac.id

Abstract. Crude oil is a vital natural resource needed worldwide and the most demanded commodity. Fluctuating oil prices can affect a country's economic conditions e.g., economic growth, inflation rate, money supply, exchange rate and interest rates. Consequently, statistical fonecasting methods are needed for a more accurate prediction in period 1 to support decision-making. This study rates to predict crude oil prices during the Covid-19 partonic and compare the performance of crude oil price forecasting using the Fuzzy Time Series (FTS) Markov Chain method and Autoregressive Integrated Moving Average (ARIMA) method. The data used is daily crude oil prices with West Texas Intermediate (WTI) ratio and used in USS/barrel from March 3, 2020, to March 31, 2022. Forecasting with the FTS Markov Chain method resulted in a mean absolute percentage error (MAPE) of 2.76%, and noot mean square error (RMSE) of 580.3. The best model for ARIMA is ARIMA (0,1,1) which produces MAPE of 3.85% and RMSE 856.7. Due to the MAPE & RMSE value of the FTS Markov Chain method her ARIMA method in the ARIMA method in the fTS Markov Chain has beller performance than the ARIMA method in the forecasting of crude oil prices during the FTS Markov Chain has beller performance than the ARIMA method in the forecasting of crude oil prices during the FTS Markov Chain has beller performance than the ARIMA method in the forecasting of crude oil prices during the Covid-19 pandemic.

28 INTRODUCTION

Corunavirus Disease 2019 (Covid-19) is a new type of virus caused by SARS-Cov-2. The new virus variant is known to have originated from Wuhan, China and was first discovered in early December 2019 when a patient was diagnosed with unusual pneumonia [1]. The Covid-19 virus has quickly infected hundreds of and has spread to Asia, Europe, the Middle East, America, and other regions. Resulting that in 2020, the World Health Organization (WHO) declared that the Covid-19 pandemic was a global pandemic. The Covid-19 pandemic has not only affected health, but it has also greatly 23 fected the global economy.

Crude oil is a commodity that has an important to generate electricity, run production machines, and transport needed in industrial projection processes, especially to generate electricity, run production machines, and transport products to the market. In addition, oil is also important for sustainable economic and social development. However, it is clearly seen that during the Covid-19 pandemic, crude oil prices fluctuate. It appeared in Fig 1 that since 1019 the price of WTI (West Texas Intermediate) crude oil has shown a decline and dipped sharply until April 2020 as a result of the Covid-19 pandemic. This is due to the limited space for human activity which has an impact on decreasing demand for crude oil and overproduction [2]. Changes in crude oil prices are not only affected by the Covid-19 pandemic but also influenced by the policies of OPEC (Organization of the Petroleum Exporting Countries) members and conflicts in crude oil producing countries.

> Miccodings of the International Conference on Medianan 35 computational Sciences, and Scotteria 2422 NP Conf. Proc. 2975, 080006-1 - 080006-17, https://doi.org/10.10635.0131088 Postedeedby AIP Publishing, 978-07358-4773-8 \$50.00



Along with the decreasing number of Covid-19 cases, industrial activity slowly continued to normal until at the end of 2021 crude oil prices fluctuated and tended to increase in the international market. In the period October - December 2021, the average oil price reached US\$ 17.3 per barrel. The increase in the oil price average in 2021 has escalated by about 42.2% compared to the oil price average in 2020. Fluctuating changes in oil prices have received considerable attention in recent decades because they have a significant political impact. Various attempts were made to explain the behavior of oil prices as well as to see the macroeconomic consequences. In Indonesia, fluctuations in oil prices have an impact on economic growth over a certain period, domestic inflation, money supply, real exchange rates, and interest rates [3]. Due to the resulting impact is being very significant, a statistical forecasting method is needed that can accurately predict crude oil prices during the Covid-19 pandemic.

In this study, oil price forecasting used the Fuzzy Time Series and Autoregressive Integrated Moving Average (200A) methods. The ARIMA method is a saitable time series data forecasting method where the method of analysis and that is not stationary in variance and mean such as crude oil prices data that moves fluctuating. ARIMA has been successfully applied at a much larger scale in various fields, mainly due to its easy-to-use concept and utility algorithm [4]. For the advantages of the Fuzzy Time Series (FTS) method, in previous research conducted by [5] that the forecasting results using the FTS Markov Chain method has a low error rate and has a high degree of accuracy.

The fuzzy time series method was first proposed by Song and Chissom by applying fuzzy logic to develop the basis of fuzzy time series. It is a dynamic process of a linguistic variable where the linguistic value is a fuzzy set [6]. However, based on the results of previous research by Tsaur that the FTS method modified with the first kov chain concept obtained a better level of accuracy than the FTS without Maria chain. The implementation of the Markov chain fuzzy time series method was first introduced by Tsaur (2011) in forecasting the Taiwan currency exchangerate against the dollar [7].

The ARIMA method was discovered by Box and Jenskin (1976). The ARIMA method (p, d, q) where p represents the order of the autoregressive process (AR), d represents the difference (differencing) and q represents the order of the moving average (MA) process. Box and Jenkins use ARIMA models for single-variable (univariate) time series. The steps of the ARIMA method are model identification, parameter estimation, parameter significance testing, diagnosite checking and forecasting [3].

Based on the detabled statement, the purpose of this study is to predict crude oil prices during the Covid-19 pandemic and compare the performance of forecasting crude oil prices using the FTS Markov Chain method and the ARIMA method. Hence, it is hoped that the prediction results can be considered in decision making related to crude oil prices for the government and economic practitioners. This study is limited to only entering historical data values from oil prices without the influence of exogenous factors on the model.

DATA AND METHODS

The secondary data used in this study is the daily data of West Texas Intermediate (WTI) standard crude oil prices during the Covid-19 pandemic, from March 3, 2020, to March 31, 2022. The data is the closing price of crude oil in units of USS per barrel and sourced from http://id.investing.com. The total number of data is 547 data with data excluding holidays.

Fuzzy Time Series

Fuzzy times series (FTS) was first developed by Song & Chissom (1993). Fuzzy time series defines a fuzzy relation formed by determining logical relationship of training data. FTS is a forecasting method with the concept of fuzzy logic which can overcome the analysis of data in the form of linguistic value that cannot be handled by classical time series methods.

Suppose U is a set of the universe, $U = \{u_1, u_2, u_3, ..., u_n\}$ then the fuzzy set A of U is defined by Equation (1): $A = f_A(u_3) + f_A(u_3) + f_A(u_3) + f_A(u_3)$ (1)

$$A = \frac{(n+1)}{3} + \frac{(n+1)}{u_2} + \dots + \frac{(n+1)}{u_n}$$
(1)

Where f_A is a function member of the fuzzy set A, $f_A: U \rightarrow [0,1]$, $f_A(u_i)$ indicates the grade of membership of u_i in the fuzzy set A and $1 \le i \le n$ [9].

everal definitions of fuzzy time series are [10]:

Definition 1

Support X(t) $(t = \dots, 0, 1, 2, \dots)$ is subset of **R**. Let X(t) be a universe of discourse on a set fuzzy $f_i(t)$ $(i = 1, 2, \dots)$. If F(t) is set of $f_i(t)$ then F(t) is referred to as fuzzy time series on X(t)

Definition 2

If F(t) is due to F(t-1) and denoted $\lim_{t \to T} F(t-1) \to F(t)$, it can be written as follows

 $\overline{F(t)} = F(t-1) \circ \overline{R}(t,t-1)$ (2) Where ' * ' is max-min composition operator. R(t,t-1) is a fuzzy logical relationship between F(t) and F(t-1), and can be expressed by $R(t,t-1) = \bigcup_{t,t} R_{t,t}(t,t-1)$ where U is union operator.

Definitio 47

Suppose $F(t) = A_i$ is caused by $F(t-1) = A_i$, then the fuzzy logical relationship is defined as $A_i \rightarrow A_j$. If there are fuzzy logical relationship obtained from state A_2 , then a transition is made to another state A_i where j = 1, 2, ..., n, as $A_2 \rightarrow A_2, A_2 \rightarrow A_2, ..., A_2 \rightarrow A_4$; hence the fuzzy logical relationship are grouped into a fuzzy logical relationship group as:

$$A_2 \rightarrow A_1, A_2, A_3 \tag{3}$$

Fuzzy Time Series Markov Chain

Step 1: Define the universe of discourse U
 Determine minimum value D_{min} and maximum value D_{max} of historical data, then define the universe of discourse U as follows:

$$U = [D_{min} - D_1, D_{max} + D_2]$$
(4)

where D₃ and D₂ are two positive numbers that determined by researcher



Step 5: Determining the fuzzy logical relationship and fuzzy logical relationships group (FLRG). Based on Definition 3, fuzzy logical relationship group can be easily obtained.

Step 6: Defining Markov probability transition matrix.

080006-4

24 Docember 2023 10:15:30

FLRG is utilized to get the probability of the next state. Therefore, we get a Markov transition matrix with dimension $\pi \times \pi$. The transition probability formula can be written as:

$$P_{ij} = \frac{M_{ij}}{M_i}; \ i, j = 1, 2, ..., n$$
 (8)

 P_{ij} : 39 bability of transition from state A_i to state A_j by one step

M₁₁ : number of transitions from state A₁ to state A₁ by one step

 M_t : the quantity of data incluing in the A_i state

The probability matrix state P can be written as follows:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nn} \end{bmatrix}$$

where $\sum_{i=1}^{n} P_{ii} = 1$

when 39

For the matrix P, several definitions as follows:

If P_{ij} ≥ 0, then state A_i is accessible from state A_i

2. If states A₂ and A₃ are accessible to each other, then A₂ communicates with A₃

Step 7: Define defazzification forecasting value

- To generate forecasting value from the obtained probability matrix then it can be calculated by the rule as following:
- 1. If FLRG A_i is empty, $(A_i \rightarrow \emptyset)$ then forecasting value is m_i that the middle value of u_i can be written by

$$F(t) = m_t$$

- If FLRG A_l is one to one (assume A_l → A_k where P_{lj} = 0 and P_{lk} = 1, j ≠ k) then the forecasting value is m_k the middle value of uk
- $F(t) = m_k P_{ik} = m_k$ 3. If FLRG A_i is one to many (assume $A_j \rightarrow A_1, A_2, ..., A_n; j = 1, 2, ..., n$). And if Y(t-1) at time (t-1) which is on state, A, the forecasting value is

$$F(t) = m_1 P_{j1} + m_2 P_{j2} + \dots + m_{j-1} P_{j(j-1)} + Y(t-1) P_{jj} + m_{j+1} P_{j(j+1)} + \dots + m_n P_{j(n)}$$
(10)
here:

wh

 m_1, m_2, \dots, m_n : the midpoint of u_1, u_2, \dots, u_n V(t-1): actual value from state A_i at time t = 1

Step 8: Determine the adjustment value to forecasting result. Adjustment forecasting used to review forecasting error. The adjusting rules for forecasting value is explained as follows:

1. If state A_i communicates with A_i , starting in state A_i at time t - 1 as $F(t - 1) = A_i$ and occur an increasing transition into state A_i at time t, (i < j), then the adjusting value is determined as:

$$D_{l1} = \left(\frac{1}{2}\right) \qquad (11)$$

where:

1 ; interval length

2. If state A_t communicates with A_t , starting in state A_t at time time t - 1 as $F(t - 1) = A_t$ and occur a decreasing transition into state A_i at time t, (i > f), then the adjusting value is determined as:

$$D_{l1} = -\left(\frac{1}{2}\right)$$
 (12)

 If the current state is in state A_i at time t - 1 as F(t - 1) = A_i, and occur a jump-forward transition into state A_{i+s} at time t, $(1 \le s \le n - i)$, then the adjusting value is determined as:

$$D_{t2} = \left(\frac{i}{2}\right)s_{i}(1 \le s \le n - i)$$
 (13)

where

s : the number of forward transitions

 If the current state is in state A_i at time t − 1 as F(t − 1) = A_i and occur a jump-backward transition into state A_{i-v} at time $t_i (1 \le v \le l)$, then the adjusting value is determined as:

$$D_{t2} = - \left(\frac{i}{2}\right) v_i \left(1 \le v \le i\right) \qquad (14)$$

where

Step 9: Calculate the adjusted forecasting results. In general, for forecasting result
$$F'(t)$$
 can be obtained as:
 $F'(t) = F(t) \pm D_{t1} \pm D_{t2} = F(t) \pm {t \choose 2} \pm {t \choose 2} v$
(15)

Stationary in Time Series Forecasting

In a data, it is possible that the data is not stationary because the mean or variance is not constant, hence, to eliminate the non-stationarity to the mean, the data can be made close to stationary by using the method of differencing. The behavior of stationary data includes not having too large variations and tends to approach the mean value, and vice versa for non-stationary data [11].

1. Stationary in Variance

The data that is not stationary in variance can be transformed. Therefore, the data becomes stationary in variance by doing a power transformation calculation. Box and Cox in 1964 introduced power transformation as follows [8]:

$$Z_{\ell}^{t} = \begin{cases} \frac{Z_{\ell}^{1} - 1}{\lambda}, \lambda \neq 0\\ \ln(Z_{\ell}), \lambda = 0 \end{cases}$$
(16)

 Z_t : time series in period t

3 : transformation parameter

2. Star21 ary in Mean

The Augmented Dickey-Fuller (ADF) test is one of the tests that 601 be employed to evaluate the stationary of time series data in mean. It is to see whether the model obtained has or does not have a unit root. The data that is not stationary in the mean can be stationary through the process differencing. This test has a regression model as follows: $\Delta Z_{\varepsilon} = \delta Z_{\varepsilon-1} + \sum_{i=2}^{n} \gamma_i \Delta Z_{\varepsilon-1+1} + a_{\varepsilon}$ (17)

Hypothesis testing for ADF model is $H_0: \rho = 0$ (Have unit root) $H_5: \rho \neq 0$ (No unit root)

The null hypothesis is tested by t-statistics which is given by this formula:

$$|\tau| = \frac{\delta}{SE(\delta)}$$
(18)

where

8 ; estimated value of parameter δ

 $SE(\delta)$: standard error for estimated value δ

Reject H_0 if $|\tau| > |\tau_{\alpha;ef}|$ or p-value $< \alpha$ [12].

Autoregressive Integrated Moving Average (ARIMA)

Autoregressive Integrated Moving Average is a forecasting method that can predict time series data stationary in variance and mean. For the data that did not fulfil the stationary, the transformation and differentiation process can 34 conducted, then ARIMA method can be used if the data has been been added and the set of object data to determine the best ARIMA model on 33 observed object. The ARIMA (p, d, q) model is a univariate time series that merges the autoregressive (AR) and moving average (MA). The general form of ARIMA (p, d, q) can be written as [13]:

$$\phi_p(B)(1-B)^d y_t = \theta_q(B) \varepsilon_t \qquad (19)$$

$$\begin{split} \phi_p(B) &= (1 - \phi_1(B^1) - \phi_2(B^2) - \dots - \phi_p(B^p)) \\ \theta_q(B) &= (1 - \theta_1(B^1) - \theta_2(B^2) - \dots - \theta_q(B^q)) \end{split}$$

where

 y_t : observation at time t ϕ_p : AR coefficient on order p θ_q : MA 56 fficient on order q ε_t : error at time t d: degree of differentiation B: Backshift operator

Forecasting steps for ARIMA method is as follows: Step 1: Model Identification

The identification of ARIMA model is based on the pattern shown in the auto correlation (ACF) and the partial correlation (PACF) plot of the data already stationery. Theoretical behavior of ACF and PACF plot is shown on Table 1:

TABL and A	E 1. Theoretical behaver of RMA(p, q) models on ACF a	AR(p), MA(q), nd PACF plots
Model	ACF	PACF
AR (p)	Tails off	Cuts off after log p
MA.(q)	Cuts off after lag q	Tails off
ARMA (p, q)	Tails off and or cuts off	Tails off and or cuts off

Step 2: Parameter Estimation

There are several methods for parameter estimation such as Moment Method, Maximum Likelihood and Least Square which can be used to estimate the parameters in models [13].

Step 3: Parameter Significance Test

The parameter significance test steps are as follows:

a. The model for parameter significance test:

with

$$\phi_{p}(B)(1-B)^{d}y_{t} = \theta_{q}(B)x_{t}$$
(20)
$$\frac{17}{\phi_{p}(B)} = (1 - \phi_{1}(B^{1}) - \phi_{2}(B^{2}) - \dots - \phi_{p}(B^{p}))$$

$$\theta_{q}(B) = (1 - \theta_{1}(\underline{B}^{1}) - \theta_{2}(\underline{B}^{2}) - \dots - \theta_{q}(B^{q}))$$

Hypothesis testing for AR model H₀: φ_p = 0 (Not significant parameter) H₁: φ_p ≠ 0 (Significant parameter)

if the particular parameter)

Statistical test calculation for AR model is:

$$t_{calc} = \frac{(\tilde{\phi}_l - 0)}{s_E(\tilde{\phi}_l)} \qquad (21)$$

Reject H_0 if $|t_{calc}| > t_{\frac{a}{2}(n-1)}$

Hypothesis testing for MA model $H_0: \partial_q = 0$ (Not significant parameter) $H_1: \theta_q \neq 0$ (Significant parameter)

Statistical test calculation for MA model is:

$$t_{calc} = \frac{\hat{\theta}_{l} - 0}{s_{E}(\hat{\theta}_{l})} \qquad (22)$$

Reject H_0 if $||t_{calc}|| > t_{\frac{\alpha}{2}(n-1)}$

080006-7

24 Docember 2023 10-16:30

with

If there is one parameter that is not significant then return to Step 1

Step 4: Diagnostic Checking

Diagnostic checking models are carried out to ensure if the remaining models meet the white noise. The following is steps of diagnostic checking [14]:

a. Hypothesis testing Ljung Box

H₀ : there is no residual autocorrelation

H1 : there is residual autocorrelation

b. Determine Ljung Box test statistics:

$$Q = n(n + 2) \left(\frac{t_1^2}{n-1} + \frac{t_2^2}{n-2} + \dots + \frac{t_k^2}{n-k} \right) \qquad (23)$$

Where:

Q : Ljung-Box Statistics Test

n : number of observation data

X : number of lag observed

 \hat{r}_i^2 : estimated correlation of residuals on the i-th of the lag with i = 1, 2, ..., K

Reject H_0 if $Q > \chi^2_{(R-p-q)}$ If there is autocorrelation between the residuals, then return to Step 1

Step 5: Overfitting

Overfitting is applied to get the best model 7th addition or subtraction of the order of AR (p) and MA (q) parameters from the tentative model has been obtained. The best model is the model with significant parameter and residual series that does not have autocorrelation.

Step 6: Selection of ARIMA Model

The selection of the best ARIMA model is minimize the information criteria such as Akaike Information Criterion (AIC) value.

$$AIC = \pi \ln \bar{\sigma}_{\delta}^{2} + 2(p + q + 1) \qquad (24)$$

$$\partial_{\delta}^{2} = SSE/n \qquad (25)$$

However, it is known that for the autoregressive model, the AIC criterion does not gives a consistent order of p, hence for comparison using other information criteria such as Schwarzt Bayesian Information Criterion (SBC) $SBC = n \ln \sigma_t^2 + (p + q + 1) \ln n$ (26)

Model Selection Criteria

10 eselection of the best model is by comparing the error values forecasting of the method used. A method is better compared to other methods if the forecast error value is produced smaller.

If $Z_1, Z_2, ..., Z_n$ state the whole data, and in sample data can be stated as $Z_1, Z_2, ..., Z_m, m < n$, the value of MSE, RMSE and MAD for in sample data defined as [15].

$$MSE = \sum_{i=1}^{m} \frac{z_i - \bar{z}_i}{n}, \ m < n$$
 (27)

$$RMSE = \sqrt{\sum_{t=1}^{m} \frac{z_t - \hat{z}_t}{n}}, \ m < n$$
 (28)

$$MAD = \sum_{t=1}^{m} \frac{|z_t - \hat{z}_t|}{n}, \ m < n$$
 (29)

Furthermore, the accuracy models can be also measured by Mean Absolute Percentage Error (MAPE) with the following formula [13]

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \qquad (30)$$

At : Actual data value at period t

 F_{g} : Forecast data value at period t

n : the number of data

whe 44

MAPE	Judgment of Forecast Accuracy
< 10%	Highly Accurate
10% MAPE < 20%	Good Forecast
20%≤ MAPE< 50%	Reasonable Forecast
≥ 50%	Inaccurate Forecast

The smaller the MAPE value, the better the model. Table 2 shows a scale to assess the accuracy of a model based on MAPE value was developed by Lewis (1982) [16]. Root Mean Square Error (RMSE) is the magnitude of the prediction error rate, where the smaller (closer to 0) to 12 (MSE value, the more accurate the prediction results will be. In this study, the model selection criteria used are Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

RESULTS

Data Description

Table 3 shows that the average price of crude oil is US\$58.68 per barrel with spread data of US\$21.92 per barrel from the average. The lowest price was US\$ -37.63 per barrel, while the highest price was US\$ 123.7 per barrel. Skewness shows a positive value of 0.077 or the values are concentrated on the right side (located to the right of M_0). Thereby, the curve has a tail that extends to the right or the curve skews to the right. Then, with a kartosis value of less than 3 which is 0.159 then it can be said to be a Platykurtic curve. And Fig 3 shows a time series of crude oil prices for the period March 3, 2020 – March 31, 2022.

N	547
Mean Value	58.68
Standard Deviation	21.92
Minimum Value	-37.63
Maximum Value	123.70
Skewness	0.077
Kurtesis	0.159



Fuzzy Time Series Markov Chain Forecasting Analysis

9 Step 1: Define the universe of discourse U

Based on historical data obtained the minimum value $D_{min} = -37.63$ and maximum value $D_{max} = 123.70$ with $D_1 = 0.37$ and $D_2 = 0.3$ which has been determined by the researcher hence U = [-38, 124]

Ste 43: Specifying Class Interval

a. Determine the number of class intervals
 46): calculation of the number of class intervals is determined by using Sturgess Rule and obtained
 K = 1 + (3,3 log N)
 K = 1 + (3,3 log 547)
 K = 10,035 = 10

b. Determine the length of the class interval

$$l = \frac{|(D_{max}+\partial_2)-(D_{min}-D_1)|}{K}$$

$$l = \frac{|(123.7+0.3)-(-37.63-0.37)|}{10}$$

$$l = 16.2 \approx 16$$

Furthermore, divide the universe of discourse U into several partitions according to the number of class intervals, which is 10 and the length of the class interval is 16. Hence the interval and the middle interval can be seen in Table 5.

TABLE 4. Class Interval and Middle Value Interval

Interval	Middle Value Interval
$u_1 = [-38, -21.8]$	-29.9
$u_2 = [-21.8, -5.6]$	-13.7
$u_3 = [-5.6, 10.6]$	2.5
$u_4 = [10.6, 26.8]$	18.7
$u_5 = [26.8, 43]$	34.9
$u_6 = [43, 59.2]$	51,1
$u_7 = [59.2, 75.4]$	67.3
$u_0 = [75.4, 91.6]$	83.5
$u_0 = [91.6, 107.8]$	99.7
$u_{10} = [107.8, 124]$	115.9

Step 3: Fuzzification of actual data

Based on the fuzzy set that has been formed, where the oil prigstata is converted into the form of linguistic values. The results of fuzzification are notated into linguistic numbers can be seen in Table 5.

1	TABLE 5. Crude Oil	Price Data Fuzzification	n':
Date	Actual Data (Y ₂)	Interval	Fuzzification
03/03/2020	47.18	$u_6 = [43, 592]$	Aa
04/03/2020	46.78	$u_6 = [43, 592]$	As
05/03/2020	45.90	$u_6 = [43, 592]$	Aa
06/03/2020	41.28	$u_5 = [26.8, 43]$	As
07/03/2020	31.13	$u_5 = [26.8, 43]$	As
1	1	and the second first second	1
30/03/2022	107.82	$u_{10} = [107.8, 124]$	A10
31/03/2022	100.28	$u_9 = [91.6, 107.8]$	A_{2}

Step 4: Determining the Fuzzy Logical Relationship and Fuzzy Logical Relationships Group (FLRG). In Table 6 is shown Fuzzy Logical Relationship (FLR) which is the relationship between each data to the next data in the form of a fuzzy set based on the determination of fuzzification. After obtaining the FLR, then the FLRG is determined which is the grouping of each state transfer namely the current state and the next state.

Data Order	FLR.
1-2	$A_n \rightarrow A_n$
2-3	$A_c \rightarrow A_c$
3 - 4	$A_6 \rightarrow A_5$
4-5	$A_5 \rightarrow A_5$
5 - 6	$A_5 \rightarrow A_5$
- I and	1
545-546	$A_{10} \rightarrow A_{0}$
546 - 547	$A_{\alpha} \rightarrow A_{\alpha}$

Step 5: Defining Markov probability transition matrix.

The transition probability matrix in this study is 10 × 10 because, in Step 1, the number of class intervals is 10. The following is the transition probability matrix:

	0.857	0.142	0	0	0	0	0	0	0	0 1
	0.040	0.800	0.160	0	0	0	0	0	0	0
	0	0.267	0.600	0.133	0	0	0	0	0	0
	0	0	0.016	0.812	0.172	0	0	0	0	0
D	D	0	0.011	0.105	0.863	0.021	0	0	0	0
<i>e</i> =	0	0	0	0	0.029	0.941	0.029	0	0	0
	Ð	0	0	0	0	0	0.867	0.133	.0	0
	0	0	0	0	0	0	0.059	0.871	0.071	0
	0	0	0	0	0	0	0	0.067	0.892	0.041
	L 0	0	0	0	0	0	0	0	0.086	0.914

Step 6: Define defuzzification and determine the adjustment value to forecasting result. A fler the transition probability matrix is formed, the next step is the calculation pn₂₀ s for forecasting and defuzzification of the previously obtained FLRG. Before determining the forecasting results, the Markov chain fuzzy time series method has a step to adjust the forecasting results for each relationship between the current state and the next state.

Step 7: Determine the adjusted forecasting

After the adjustment value is obtained, then the forecasting results are determined which have been adjusted can be seen in Table 7.

TABLE 7. Adjusted Forecasting Results							
Date	Actual Data (Y ₁)	Initial Forecasting (F _t)	Adjustment Value	Final Forecasting (F't)			
03/03/2020	47.18		- ÷	- 19 Mar			
04/03/2020	46.78	45,921	0	45.921			
05/03/2020	45.90	45.575	0	45.575			
06/03/2020	41.28	46.244	-2	44.816			
07/03/2020	31.13	36.338	-4	32.828			
08/03/2020	34.36	32,233	1	33.177			
10 miles	1	1	E o	E.:			
30/03/2022	107.82	105.598	0	105.598			
31/03/2022	100.28	100.043	0	100.043			



FIGURE 4. Plot Using FTS Markov Chain

The Fuzzy Time Series Markov Chain method has a very good performance in forecasting crude oil prices during the Covid-19 pandemic. This is indicated by the resulting MAPE value of 2.76% (less than 10%) and RMSE 580.3 on daily data for March 2020 - March 2022. This is supported by the similarity plot between the forecast data and the actual data as presented in Fig 4. And the results of forecasting crude oil prices for the next 5 consecutive days period (April 1* -7*, 2022) are US\$99.83, US\$105.12, US\$101.89, US\$98.31, US\$96.93 (per barrel).

ARIMA Forecasting Analysis

In time series analysis, stationarity is required to minimize model errors in order to get the best model. The first step to determine the stationarity of the data is to make a time series plot. From Fig 3 it is obtained that the data is not stationary due to the fluctuating data where the data is not around a constant mean or on average does not form an almost horizonta 32 end.

In addition, the stationarity of the data can be determined by the unit root test using the Augmented Dickey Fuller (ADF) test. The unit root test hypothesis is:

 $H_0: \rho = 0$ (There is a unit root) $H_1: \rho \neq 0$ (No unit root)



FIGURE 5. Unit Root Test Result with ADF Test

From Fig S above obtained $p-\sqrt{34}$ c, 0.1487 is greater than $\alpha = 0.05$ then accept H_0 so that it can be said that there is a unit root which means the data is not stationary. Therefore, a differencing process is needed which was previously carried $\sqrt{52}$ by the natural logarithm transformation process on 547 crude oil price data during the Covid-19 pandemic hence the data becomes stationary with respect to the mean and variance.

Furthermore, the natural logarithm transformation process and differencing process are carried out. And from Fig 6 it is obtained that the data is stationary at the first difference level, thus indicating the value of d = 1. To strengthen the existence of stationary data, Fig 7 shows the results of the ADF test with the data from the logarithmic differencing regress. The p-value obtained from the unit root test results is 0.01 < 0.05 (a) so that H_0 is rejected, which means that there is no unit root so that the data is stationary.



FIGURE 7. The Result of Differencing-Natural Logarithm by Using ADF Test

This is also in line with the ACF plot and PACF plot where there are no more than 3 lags that come out of the confidence interval shown in Fig 8(a) & 8(b). Because of the time series data has met the stationary requirements. Therefore, forecasting can be done and there is no need to do further differencing (second order differencing).



After pre-processing the data, the next step is to identify and estimate the best ARIMA model from the results of the transformation and differencing processes. Based on the analysis of the order correlogram on the Auto Regressive (p) PACF plot in Fig 8 (b), the significant order is lag 18 and from Fig 8 (b) the ACF plot also shows that the significant order in the Moving Average (q) is lag 18. Significant order is obtained by considering the outgoing correct gram. of confidence intervals. So, the models estimated on crude oil price data during the Covid-19 pandemic are ARIMA (1,1,0), ARIMA (0,1,1), and ARIMA (1,1,1).

Furthermore, the parameters selected into the model are if the p-value or significance value for each parameter is less than α (0.05). The hypothesis used is as follows:

- H₀: Parameters are not significant in the model
- H1: Significant parameters in the model

The results of the significance test on the three ARIMA models estimated in Table 8 show that only ARIMA (0,1,1) has significant parameters where p-value MA (1) is 0.023 less than α (7)5) then reject H_0 which means only ARIMA model (0,1,1) here H_0 which means only ARIMA model (0,1,1) here H_0 which means only arithmeters in the model. Furthermore, the selection of the best model is also carried out by analyzing the Akaike information criterion (AIC), Schwarz information criterion (SIC) and Hannan-



Quian criterion (HQC). Table 8 shows that the AIC, SIC and HQC values in ARIMA (0,1,1) have the lowest values compared to other ARIMA models. In addition, the MSE value on ARIMA (0,1,1) is also the lowest, so it can be said that ARIMA (0,1,1) is the best model to use on crude oil price data during the Covid-19 pandemic.

TABLE 8. ARIMA Model Estimation						
Model/ 23 Parameter	Variable	p-value	MSE	AIC	SIC	HQC
ARIMA (1,1,0) Constant	AR (1)	0.058 0.459	2.83	3.750	3.776	3.760
ARIMA (0,1,1)	MA (1)	0.023 0.400	2.47	3.689	3.694	3.690
ARIMA (1,1,1) Constant	AR (1) MA (1)	0.246 0.812 0.427	2.78	3.723	3.741	3,753

TABLE 9. Diagnostic Checking			
30	p-value	Decision	
12	0.392	white noise	
24	0.299	white noise	
36	0.618	white noise	
48	0.656	white noise	

To obtain the best ARIMA model, diagnostic checking includes white noise and normal distribution of residuals. The white noise test with the null hypothesis is that there is no residual correlation between lags. Table 9 shows the p-value by Ljung Box Statistics for ARIMA (0,1,1). From Table 9 the value of all p-values at lags to 12, 24, 36 and 46 is more than α (0.05) so accept H₀ that the residuals do not contain autocorrelations or white noise. This is reinforced through the residual ACF and PACF correlograms, in Fig 9 it is shown that probability > α (0.05) which resulting the white noise model.

Autocorrelation	Patial Convision		AC:	PAC	0-934	Prot
10	1 1	1.1	0.008	6.061	76-05	
100		2	-0.015	-4.016	1.1107	0.74
14	-	1.2	-0.018	-8.018	3.2588	0.87
10	30	4	0.013	0.013	4.3075	0.953
10		1.5	0.068	1.087	1.8508	0.74
	346	14	-0.094	+1024	2,2439	0.81
- 10	30	7	0.053	1.055	3.6342	0.738
100		1 A	0.036	0.037	4.2209	0.75
- 0			0.077	\$ \$07	7.1630	0.52
- etc.		110	-0.048	-0-048	8-2577	0.500
40-	10	11	-0.021	0.016	8.4940	:0.565
19	10	12	0.065	0.058	18.567	0.48
	4	13	0.00%	6.603	13,300	0.02
10	30 -	14	0.062	0.055	12.797	0.45
441	16	16	-0.098	-8.014	12,462	0.62
10	10	15	0.065	0.051	14,765	0.46
10		117	-0.008	4015	14,797	0.54
100	di i	112	-0.067	4.084	18.522	0.26
	16	110	-0.058	-8-048	19.885	0.33
- de		128	0.017	8.015	28 930	0.10

FIGURE 9. The Output of ACF and PACF Plot for Residual Autocorrelation Test

16 Normality testing was carried out using the Kolmogorov-Smirnov test with the null hypothesis is the residuals are nor 16 lly distributed. The residual is normally distributed if the p-value is more than a with a value is 0.05. The results of the normality test using the Kolmogorov-Smir 45 test showed that the p-value in ARIMA (0.1,1) residual probability plot has a value of more than 0.120. Hence, it can be said that the residuals in the model are normally distributed. Due to the model satisfies the white noise test and has a normal distribution of error, the ARIMA model (0.1,1) is the best model so that it can proceed to the next stage, which is forecasting.

The best model used is ARIMA (0,1,1) with constants written in the form of an equation, the following model is obtained:

$$(1 - B)Y_t = 0.0587 + 0.0293\varepsilon_{t-1}$$
 (31)

where $\hat{Y}_t = \ln \hat{Z}_t$, then the forecast value for Z_t is $\hat{Z}_r = \exp(\hat{Y}_r)$

where \hat{Z}_t : crude oil price prediction

MAPE & RMSE resulting from the ARIMA method are 3.85% and 856.7 which shows that ARIMA performance is highly accurate. Fig 10 shows a plot between the predicted and actual values using the ARIMA method. Table 10 shows the prediction results by using the ARIMA Model (0,1,1) for the next 5 consecutive days period (April 1st -7st, 2022):



FIGURE 10. Plot Using ARIMA Method

TABLE 10. Crude Oil Prices (Confidence Interval 95%)

Dista	Crude Oil Prices (USS per barrel)				
Date	ase Prediction The Lower		The Highest Prediction		
04/01/2022	100.831	94.285	108.364		
04/04/2022	102,976	96.199	109.752		
04/05/2022	100.598	93.287	108.451		
04/06/2022	95.711	91.143	105.639		
04/07/2022	95.832	92.547	107.433		

Comparison Forecasting Methods

Comparison of forecasting accuracy can be done visually and analytically. A15 imparison of forecasting accuracy visually is done by comparing the estimated value and the actual value using a time series plot. While the comparison of the accuracy of forecasting time series data analytically is done by comparing the forecasting error values between methods. The comparison of the forecasting results of the ARIMA and Fuzzy Time Series (FTS) Markov Chain methods visually on the crude oil price data is done by comparing the plots of the actual value and the forecast value of crude oil prices as presented in Fig 11. Based on Fig 11, the forecast value plot of the FTS Markov Chain method is more similar or closer to the actual value pattern when compared to the ARIMA method, which means that forecasting using the FTS Markov Chain method is better than the ARIMA method on the data as many as 547 observations.

(32)



FIGURE 11. Actual Data & Forecasted Data Plots Using FTS Markov Chain & ARIMA Methods

Analytically, the comparison of forecasting accuracy in this study is conducted by using MAPE and RMSE values as a measure of forecasting error. The MAPE value obtained by the FTS Markov Chain method is 2.76% and the RMSE is 580.3. While the MAPE value generated by the ARIMA method is 3.85% and RMSE is 856.7. Due to the MAPE and RMSE values in the FTS Markov Chain method are smaller than the ARIMA method, it shows that the FTS Markov Chain method works better than the ARIMA method to forecast crude oil price data during the Covid-19 pandemic for the 2012 2022 period. The results visually and analytically yield the same conclusion that the FTS Markov Chain method has better performance than the ARIMA method. The results of forecasting time series data on crude oil prices during the Covid-19 pandemic using the FTS Markov chain method and the ARIMA method alongside the MAPE and RMSE values can be seen in Table 11.

FABLE 11, Comparison	of Actual Data with Predicted	Data for Crude Oil Prices
----------------------	-------------------------------	---------------------------

Date	Crude C	hil Prices (US\$ per harn	el)
Date	Actual Data (1/2)	FTS Markov Chain	ARIMA
03/03/2020	47,18		
04/03/2020	46,78	45.921	47.238
05/03/2020	45,90	45.575	46.852
06/03/2020	41,28	44.816	45,986
07/03/2020	31,13	32.828	41.477
08/03/2020	34,36	33.177	31.492
	the second second	1	÷
30/03/2022	107.82	105.598	104.182
31/03/2022	100.28	100.043	107.364
	MAPE	2.76%	3.85%
	RMSE	580.3	856.7

CONCLUSION

This study aims to reactice crude oil prices during the Covid-19 pandemic and 10 mpare the performance of crude oil price forecasting by using the Fuzzy Time Series (158) Markov Chain method and Autoregressive Integrated Moving Average (ARIMA) method. The data used is daily data on the crude oil prices with West Texas Intermediate (WTI) Standard in USS/barrel for the period March 3, 2020 - March 31, 2022. The Fuzzy Time Series Markov Chain method has an excellent performance in forecasting crude oil prices during the Covid-19 pandemic. This is indicated a the resulting MAPE value of 2.76% (less than 10%) and RMSE 580.3 on daily data for March 2020 - March 20 for ARIMA model (0,1,1) with constants is the best ARIMA model for modeling actual data on crude oil prices for the period of March 2020 - March 2022 during the Covid-19 pandemic so that this model can be used for prediction of future world crude oil prices Covid-19 pandemic. MAPE of this model is 3.85% and RMSE is 856.7. Based on a visual and analytical comparison, it can be concluded that the Fuzzy Time Series Markov Chain method works better than the ARIMA method in forecasting crude oil prices during the Covid-19 pandemic on period March 2020 - March

2022. These results have several important implications for Indonesia, especially on policy recommendations and economic development due to changes in oil prices that have an impact on several sectors.

ACKNOWLEDGMENTS

This research is supported by PDP Universitas Trisakti 2021 & Faculty of Industrial Technology USAKTI

REFERENCES

- 1. Y.C. Wu, et al, Journal of the Chinese Medical Association, vol. 83, issue 3, pp. 217-220, march 2020.
- 2. N.L. Widyastuti and H. Nugroho, JPP, vol. 4, no. 2, pp. 166-176, June. 2020.
- 3. M. A. Nizar, Buletin Ilmiah Litbang Perdagangan, vol.6, no. 2, 2012.
- 4. O.D. Ilie, et al, Microorganisms, 8(8), 1158.
- Zaenurrohman, S Hariyanto, T.Udjiani, "Fuzzy Time Series Markov Chain and Fuzzy time series Chen & Hsu for Forecasting", Journal of Physics: Conference Series, 1943(1): 012128, July 2021.
- 6. Q. Song and B. S. Chissom, Fuzzy Sets and Systems, vol. 54, no.1, pp.1-9, February 1993.
- R. C. Tsaur, International Journal of Innovative Computing, Information and Control, vol 8, no 7(b), pp 4931-4942, July 2012.
- 8. A. Aswi dan Sukarna. Analisis Deret Waktu: Teort dan Aplikasi. Makassar: Andira Publisher, 2006.
- H.T. Jasim, A.G.J. Salim and K.I. Ibraheem, Applications and Applied Mathematics, Vol. 7, Issue 1, pp. 385-397. June 2012.
- Q. Song and B.S. Chissom, Fuzzy Sets and Systems, vol. 54, Issue 3, pp. 269-277, March 1993.
- 11. D. N. Gujarati. Basic Econometric (4th ed). The McGraw-Hill Companies, 2004.
- D. A. Dickey and W.A. Fuller, Journal of The American Statistical Association, vol.74, no.366, pp. 427-431, June 1979.
- D. C. Montgomery, C. L. Jennings, and M. Kulahci. Introduction to Time Series Analysis and Forecasting. New Jersey (US): J Wiley, Inc, 2008.
- 13. J. D. Cryer and K. S. Chan. Time Series Analysis with Applications in R (2nd ed.). New York: Springer, 2008.
- 14. D. Chieco, M.J. Warrens, and G. Jurman, PeerJ Computer Science, 7:e623.
- C. D. Lewis. Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. London: Butterworth Scientific, 1982

Comparison method between fuzzy time series Markov chain and ARIMA in forecasting crude oil prices

ORIGIN	ALITY REPORT				
SIMILA	7% ARITY INDEX	11% INTERNET SOURCES	11% PUBLICATIONS	8% STUDENT PA	APERS
PRIMAR	Y SOURCES				
1	Submitte Bangalo Student Paper	ed to Indian Ins [.] ore	titute of Mana	agement,	1%
2	Ansari S Safty, Sa "SutteAl the Infa Comput Publication	aleh Ahmar, Eva amirah AlZahrar RIMA: A Novel M nt Mortality Rat ers, Materials &	a Boj del Val, N ni, Hamed El-K 4ethod for For e in Indonesia Continua, 202	A. A. El hawaga. recasting i", 22	1 %
3	Shyi-Mir betweer IEEE Tra Cyberne Publication	ng Chen. "Simila n vague sets and insactions on Sy etics, Part B (Cyb	rity measures d between ele stems, Man a pernetics), 199	ments", nd 97	1 %
4	reposito	ory.usd.ac.id			1%
5	Submitt UMT Student Paper	ed to Universiti	Malaysia Tere	ngganu	<1%

Green Arther Sandag, Adinda Marcellina Manueke, Michael Walean. "Sentiment Analysis of COVID-19 Vaccine Tweets in Indonesia Using Recurrent Neural Network (RNN) Approach", 2021 3rd International Conference on Cybernetics and Intelligent System (ICORIS), 2021 Publication

<1%

<1%

<1 %

Wiwik Anggraeni, Abdolatul Abdillah, Pujiadi, Lulus Tjondro Trikoratno et al. "Modelling and Forecasting the Dengue Hemorrhagic Fever Cases Number Using Hybrid Fuzzy-ARIMA", 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH), 2019 Publication

8	biroauk.uinjkt.ac.id Internet Source	<1%
9	Cheng, C.H "Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost", Technological Forecasting & Social Change, 200606	<1 %

Publication

11

7



		<1%
12	repositorio.iscte-iul.pt Internet Source	<1%
13	www.ncbi.nlm.nih.gov Internet Source	<1%
14	Dodi Devianto, Aulia Zuardin, Maiyastri Maiyastri. "TIME SERIES MODELING OF NATURAL GAS FUTURE PRICE WITH FUZZY TIME SERIES CHEN, LEE AND TSAUR", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2022 Publication	<1%
15	F Fitri, R Rahmat. "Comparison of SSA and SARIMA in Forecasting the Rainfall in Sumatera Barat", Journal of Physics: Conference Series, 2021 Publication	<1%
16	hts.org.za Internet Source	<1%
17	Ramalho, Octávio Ferreira. "Implementação integrada de spc e epc na melhoria contínua do processo", Universidade NOVA de Lisboa (Portugal), 2024 Publication	<1%

19	Submitted to Universitas Jenderal Soedirman Student Paper	<1%
20	Submitted to Universitas Trunojoyo Student Paper	<1%
21	Submitted to 於2012-03-15提交至South Bank University ^{Student Paper}	<1%
22	Submitted to National University of Singapore Student Paper	<1%
23	icas2019.icar.gov.in Internet Source	<1%
24	pubs.aip.org Internet Source	<1%
25	rstudio-pubs-static.s3.amazonaws.com	<1%
26	Liam Vaughan, Muyang Zhang, Haoran Gu, Joan B. Rose et al. "An exploration of challenges associated with machine learning for time series forecasting of COVID-19 community spread using wastewater-based epidemiological data", Science of The Total Environment, 2023 Publication	<1%
27	Submitted to Stefan cel Mare University of Suceava	<1%

Student Paper

28	WWW.Scitepress.org	<1%
29	id.123dok.com Internet Source	<1%
30	repository.its.ac.id	<1%
31	repository.ntu.edu.sg	<1%
32	www.rsisinternational.org	<1%
33	Submitted to Asian Institute of Technology Student Paper	<1%
34	Umi Mahmudah. "Autoregressive Integrated Moving Average Model to Predict Graduate Unemployment in Indonesia", Practice and Theory in Systems of Education, 2017 Publication	<1%
35	Vira Shendryk, Yuliia Parfenenko, Sergii Tymchuk, Yevhen Kholiavka, Yaroslava Bielka. "Modeling techniques of electricity consumption forecasting", AIP Publishing, 2022 Publication	<1%



soar.wichita.edu Internet Source <1 %

37	www.investopedia.com	<1%
38	www.omicsdi.org Internet Source	<1%
39	J. Garcia-Frias. "Decoding of Low-Density Parity-Check Codes Over Finite-State Binary Markov Channels", IEEE Transactions on Communications, 2004 Publication	< 1 %
40	Taiyong Li, Yingrui Zhou, Xinsheng Li, Jiang Wu, Ting He. "Forecasting Daily Crude Oil Prices Using Improved CEEMDAN and Ridge Regression-Based Predictors", Energies, 2019 Publication	<1%
41	web2.uconn.edu Internet Source	<1%
42	WWW.eco.uc3m.es Internet Source	<1%
43	Harris Ehrlich, . "Measure Phase", Transactional Six Sigma and Lean Servicing Leveraging Manufacturing Concepts to Achieve World-Class Service, 2002. Publication	< 1 %
44	Rim Anggrainingsih, Gilang Romadhon Aprianto, Sari Widya Sihwi. "Time series	<1%

forecasting using exponential smoothing to

predict the number of website visitor of Sebelas Maret University", 2015 2nd International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), 2015

Publication

45	Submitted to Universiti Sains Malaysia Student Paper	<1%
46	123dok.com Internet Source	<1%
47	Submitted to De Montfort University Student Paper	<1%
48	Submitted to Michigan Technological University Student Paper	<1%
49	Submitted to School of Oriental & African Studies Student Paper	<1%
50	Submitted to UT, Dallas Student Paper	<1%
51	Submitted to University of Melbourne	<1%
52	M Alawiyah, D A Kusuma, B N Ruchjana. "Application of Generalized Space Time Autoregressive Integrated (GSTARI) model in	<1%

the phenomenon of covid-19", Journal of Physics: Conference Series, 2021 Publication

53	Submitted to University of Kent at Canterbury Student Paper	<1%
54	committees.westminster.gov.uk	<1%
55	www.jatit.org	<1%
56	V. Prakash, K. C. Sharma, R. Bhakar, H. P. Tiwari, S. Sreekumar, S. Chawda, F. Teotia. "Primary frequency response with stochastic scheduling under uncertain photovoltaic generation", 2017 IEEE Power & Energy Society General Meeting, 2017 Publication	<1%
57	brage.bibsys.no Internet Source	<1%
58	eprints.whiterose.ac.uk Internet Source	<1%
59	escholarship.org Internet Source	<1%
60	research.unilus.ac.zm Internet Source	<1%
61	sec.report Internet Source	<1%



Exclude quotes On Exclude bibliography On Exclude matches

< 10 words

Comparison method between fuzzy time series Markov chain and ARIMA in forecasting crude oil prices

GRADEMARK REPORT

FINAL GRADE	GENERAL COMMENTS
/100	
PAGE 1	
PAGE 2	
PAGE 3	
PAGE 4	
PAGE 5	
PAGE 6	
PAGE 7	
PAGE 8	
PAGE 9	
PAGE 10	
PAGE 11	
PAGE 12	
PAGE 13	
PAGE 14	
PAGE 15	
PAGE 16	
PAGE 17	
PAGE 18	