

Home
Guide for authors
Submit Article
Editorial Board
Archives
Contact

TEM JOURNAL - Technology, Education, Management, Informatics

Editorial Board:

Editor-in-Chief:

Prof. dr Camil Sukic,
(International University of Novi Pazar, Novi Pazar, Serbia)

Lector:

Irma Halilovic

Members:

Prof. dr Keiichi Kaneko

(Tokyo University of Agriculture and Technology | TUAT · Department of Computer and Information Sciences, Tokyo, Japan)

Prof. dr Klaus Tochtermann

(ZBW - Leibniz Information Centre for Economics, Germany)

Prof. dr Gillich Gilbert-Rainer

(Eftimie Murgu University in Resita, Department of Mechanical Engineering, Romania)

Samir Khatir, PhD

(Technologiepark, Ghent University, Belgium)

Prof. dr Álvaro Rocha

(University of Lisbon - ISEG, Lisbon, Portugal)

Asst. Prof. Enes Sukic

(UIKTEN - Association for Information Communication Technology Education and Science, Serbia)

Erica Magagnini, PhD

(Università Politecnica delle Marche, Italy)

Maria José Sousa, PhD

(ISCTE-Instituto Universitario de Lisboa, Portugal)

Prof. Dr. Khaled Shaalan

(Faculty of Engineering & IT, The British University in Dubai, UAE)

Asst. Prof. Isaac Lera Castro

(University of the Balearic Islands, Department of Mathematic and Computer Sciences, Spain)

Le Thanh Cuong, PhD

(Faculty of civil engineering, Ho Chi Minh City Open University, Vietnam)

Prof. dr Borut Buchmeister

(University of Maribor, Slovenia)

Ing. Darina Duplákova, PhD

(Faculty of Manufacturing Technologies with a seat in Prešov, Technical University of Košice, Slovakia)

Stamatis Papadakis, PhD

(Faculty of Education, University of Crete, Crete, Greece)

Assoc. Prof. dr. Kuaanan Techato

(Faculty of Environmental Management, Prince of Songkla University, HatYai, Songkhla, 90110, Thailand)

[Publication Ethics and
Publication Malpractice Statement](#)

[Open Access and copyright](#)

Scopus®

Clarivate
Web of Science™

DOAJ

EBSCOhost



Central and Eastern European Online Library





Prof. dr Bekim Fetaji
(University Mother Teresa, Faculty of Informatics,
Skopje, North Macedonia)

Asst. Prof. Dominik Zimon
(Rzeszow University of Technology,
Department of Management Systems and Logistics, Poland)

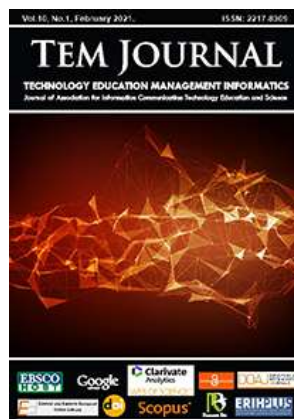
Prof. dr Petre Lucian Ogrutan
(Transilvania University of Brasov,
Faculty of Electrical Engineering and Computer Science, Romania)

Prof. dr Danijela Milosevic
(University of Kragujevac, Serbia)

Prof. dr Marko Hölbl
(University of Maribor, Faculty of Electrical Engineering and Computer
Science (FERI), Slovenia)

Prof. dr Aleksandar Kupusinac
(University of Novi Sad, Faculty of Technical Sciences, Serbia)

Prof. dr Marina Klacmer Calopa
(University of Zagreb, Faculty of Organization and Informatics (FOI),
Croatia)



[Open Access and copyright](#)

[Publication Ethics and](#)

[Publication Malpractice Statement](#)



WEB OF SCIENCE™



Universität Regensburg

[University Library of Regensburg](#)



Title: *The Influence of the Human Factor on the Success of the Localization Project of the Automated Technological Line for Wood Production*

Authors: Radka Vaničková

DOI: 10.18421/TEM101-01

[| abstract |](#) [full text |](#) [how to cite |](#)

Title: Smart University: A University In the Technological Age

Authors: Anne Bakupa Mbombo, Nadire Cavus

DOI: 10.18421/TEM101-02

[| abstract |](#) [full text |](#) [how to cite |](#)

Title: *Reactor Design Pattern*

Authors: Marko Mijač, Antonio García-Cabot, Vjeran Strahonja

DOI: 10.18421/TEM101-03

[| abstract |](#) [full text |](#) [how to cite |](#)

Title: *A Systematic Design Approach in Building Digitalization Services Supporting Infrastructure*

Authors: Pavel Petrov, Mihail Radev, Georgi Dimitrov, Adrian Pasat, Alexander Buevich

DOI: 10.18421/TEM101-04

[| abstract |](#) [full text |](#) [how to cite |](#)

Title: Improved Pohlig - Hellman using Sieve of Eratosthenes for Three-Pass Protocol Security Enhancement

Authors: Robbi Rahim, Nik Adilah Hanin Zahri, Mohd Nazri Bin Mohd Warip

Title: *Forecast Accuracy Evaluation of the Enterprise's Industrial Safety Integral Risk*

Authors: Leyla M. Bogdanova, Sergey Ya. Nagibin, Aleksandr S. Chemakin

DOI: 10.18421/TEM101-06

Title: *Translation Learning Tool for Local Language to Bahasa Indonesia using Knuth-Morris-Pratt Algorithm*

Authors: Harco Leslie Hendric Spits Warnars, Jessica Aurellia, Kendrick Saputra

DOI: 10.18421/TEM101-07



Title: *Effects on Using Tutoring Application in Integration with Self-Directed Learning to Improve Statistical Analysis Skills*

Authors: Sawanan Dangprasert

DOI: 10.18421/TEM101-08



Title: *Mobile Applications to Fight against COVID-19 Pandemic: The Case of Saudi Arabia*

Authors: Imdadullah Hidayat-ur-Rehman, Arshad Ahmad, Manzoor Ahmed, Aftab Alam

DOI: 10.18421/TEM101-09



Title: *Medical Informatics in Indonesia: Importance, Development, and Future Directions*

Authors: Seng Hansun

DOI: 10.18421/TEM101-10

Title: *ABSA: Computational Measurement Analysis Approach for Prognosticated Aspect Extraction System*

Authors: Maganti Syamala, N.J.Nalini

DOI: 10.18421/TEM101-11

Title: *Readiness Model in Adopting E-health: An Indonesian Experience*

Authors: Vera Pujani, Hardisman Dasman, Rima Semiaty, Refdinal Nazir

DOI: 10.18421/TEM101-12

Title: *IoT Telemonitoring System for COVID-19 Quarantine*

Authors: Kobkiat Saraubon, Benchaphon Limthanmaphon

DOI: 10.18421/TEM101-13

[. | abstract | full text | how to cite |](#)

Title: Opportunities of Digital Technologies in Leveling Financial and Socio-economic Exclusion Problems

Authors: Viktoriia Druzhynina, Ganna Likhonosova, Lyudmyla Davidiyuk, Aleksandr Kievich, Galyna Lutsenko

DOI: 10.18421/TEM101-14

[. | abstract | full text | how to cite |](#)

Title: *Economic Feasibility Analysis in Developing 5G Infrastructure and Locations in Indonesia*

Authors: Yudha Aprilianto, Muhammad Asrol, Fergyanto E Gunawan

DOI: 10.18421/TEM101-15

[. | abstract | full text | how to cite |](#)

Title: *Comparing Performance of Machine Learning Algorithms for Default Risk Prediction in Peer to Peer Lending*

Authors: Yanka Aleksandrova

DOI: 10.18421/TEM101-16

[. | abstract | full text | how to cite |](#)

Title: *Estimation of Gourami Supplies Using Gradient Boosting Decision Tree Method of XGBoost*

Authors: I Made Sukarsa, Ngakan Nyoman Pandika Pinata, Ni Kadek Dwi Rusjayanthi, Ni Wayan Wisswani

DOI: 10.18421/TEM101-17

[. | abstract | full text | how to cite |](#)

Title: *Utilization of the Ultrasonic Diagnostic Method in Rail Status on a Defined Railway Section*

Authors: Jakub Kaščák, Jozef Török, Monika Töröková

DOI: 10.18421/TEM101-18

[. | abstract | full text | how to cite |](#)

Title: *Micro Grid Operation Mode With Different MPPT Controller*

Authors: Ihsan Mizher Baht, Farah Fahem Al Sarhan, Petre-Marian Nicolae, Nameer Bahat

DOI: 10.18421/TEM101-19

[. | abstract | full text | how to cite |](#)

Title: *FPGA-based Visible Light Communication System for Real-Time File Transmission*

Authors: Syifaul Fuada, Rosmianto Aji Saputro, Muhammad Luthfi, Trio Adiono

DOI: 10.18421/TEM101-20

[. | abstract | full text | how to cite |](#)

Title: *Express Methods and Procedures for Determination of the Main Egg Quality Indicators*

Authors: Jakhfer Alikhanov, Aidar Moldazhanov, Akmaral Kulmakhambetova, Zhandos Shynybay, Stanislav M. Penchev, Tsvetelina D. Georgieva, Plamen I. Daskalov

DOI: 10.18421/TEM101-21

[. | abstract | full text | how to cite |](#)

Title: *Solution Proposal for Utilization of the Waste Heat in Refrigeration Systems*

Authors: Halima Hadžiahmetović, Rejhana Blažević, Emina Peco

DOI: 10.18421/TEM101-22

[. | abstract | full text | how to cite |](#)

Title: *Diagnosis of Pneumatic Systems on Basis of Time Series and Generalized Feature for Comparison with Standards for Normal Working Condition*

Authors: Rosen Kosturkov, Veselin Nachev, Tanya Titova

DOI: 10.18421/TEM101-23

[. | abstract | full text | how to cite |](#)

Title: *Conceptualization of Smart System Based on RFID Technologies for Controlling Vehicle Speed*

Authors: Evan Asfoura, Mohammed Samir Abdel-Haq, Gamal Kassem

DOI: 10.18421/TEM101-24

[. | abstract | full text | how to cite |](#)

Title: *Proposal of the Reworking Station Model using Plant Simulation*

Authors: Naqib Daneshjo, Peter Malega

DOI: 10.18421/TEM101-25

[. | abstract | full text | how to cite |](#)

Title: *Evaluation of a Strategic Management Program: Context, Input, Process, Product Model as a Prototype for Business Academies*

Authors: Bengi Birgili, Kamil Arif Kırkıç

DOI: 10.18421/TEM101-26

[. | abstract | full text | how to cite |](#)

Title: *The Use of Business Process Management in Hotel Direct Sales Improvement*

Authors: Stepan Chalupa, Martin Petricek, Zdenek Ulrych

DOI: 10.18421/TEM101-27

[. | abstract | full text | how to cite |](#)

Title: *The Reinvesting Impact of Promotional Activity and Store Atmosphere on Impulse Buying Behavior: The Mediating Role of Payment Facility*

Authors: Syed Mir Muhammad Shah, Jan Muhammad Sohu, Sarfraz Ahmed Dakhan, Raja Sikander Ali, Ikramuddin Junejo, Irfan Mir Chouhan

DOI: 10.18421/TEM101-28

[. | abstract | full text | how to cite |](#)

Title: *Assessing the Maturity of Corporate Social Responsibility Practices: An Applied Study on Telecom Companies in KSA*

Authors: Abderrazak Ahmed Laghouag, Faycal Farhi, Faiz Bin Zafrah

DOI: 10.18421/TEM101-29

[. | abstract | full text | how to cite |](#)

Title: *Estimation of Management Effectiveness of Electricity Supply Enterprises in Emerging Economies*

Authors: Skibina Tetiana, Kurbatova Tetiana, Sotnyk Iryna, Telizhenko Oleksandr, Sotnyk Mykola, Hyrchenko Yevgeniy

DOI: 10.18421/TEM101-30

[. | abstract | full text | how to cite |](#)

Title: *Approaches to the Organization of International Accounting in Russian Enterprises*

Authors: Ekaterina V. Surkova, Galina A. Skachko, Larisa K. Nikandrova, Maria M. Starkova, Nina F. Sakharova

DOI: 10.18421/TEM101-31

[. | abstract | full text | how to cite |](#)

Title: *The Influence of Multi - Cultural Higher Education on Economic Development*

Authors: Hüseyin Baykan

DOI: 10.18421/TEM101-32

[. | abstract | full text | how to cite |](#)

Title: *Improving Aviation Insurance System in Russia*

Authors: Svetlana Yu. Sivoplyasova, Sergey M. Voinov

DOI: 10.18421/TEM101-33

[. | abstract | full text | how to cite |](#)

Title: *The Effectiveness of Network Systems in Providing Project Maturity of Public Management*

Authors: Svitlana Bondarenko, Oleksandr Halachenko, Leonid Shmorgun, Iryna Volokhova, Alla Khomutenko, Valerii Krainov

DOI: 10.18421/TEM101-34

[. | abstract | full text | how to cite |](#)

Title: *Tech Start-up Incubation Program: Business Model Evaluation on Government Based Incubator in Indonesia*

Authors: Tifa Noer Amelia, Armanu Thoyib, Gugus Irianto, Ainur Rofiq

DOI: 10.18421/TEM101-35

[. | abstract | full text | how to cite |](#)

Title: *Optimization of Information Analytical Support for Assessing the Reliability of Counterparties from the Standpoint of Economic Security*

Authors: Larysa Hnylytska, Ievgeniia Gavrylychenko, Stepan Melnyk, Olena Illiashenko, Yevhenii Rudnichenko, Nataliia Havlovska

DOI: 10.18421/TEM101-36

[. | abstract | full text | how to cite |](#)

Title: *Development of the Organization's Key Performance Indicators System in Order to Improve the Effectiveness of Its Human Capital and Risk Management*

Authors: Rushan N. Zaripov, Ilnur M. Murakaev, Anatoly V. Ryapukhin

DOI: 10.18421/TEM101-37

[. | abstract | full text | how to cite |](#)

Title: *Integrative Viewpoint for Implementing Sustainable Management Agricultural Business Excellence*

Authors: Mykhailo Sahaidak, Mariia Tepluk, Victoria Zhurylo, Natalia Rudenko, Olesia Samko

DOI: 10.18421/TEM101-38

[. | abstract | full text | how to cite |](#)

Title: *Compliance with Tax Laws in the Russian Federation: Trust or Control*

Authors: Murat Aliev, Ekaterina Milovanova, Irina Moiseenko, Vladimir Molodykh

DOI: 10.18421/TEM101-39

[. | abstract | full text | how to cite |](#)

Title: *Assessment of the Competitive Positions of National Airlines of Ukraine*

Authors: Ovsak Oksana Pavlivna, Liskovych Nazarii Yuriiiovych, Nazarenko Oleksandra Pavlivna

DOI: 10.18421/TEM101-40

[. | abstract | full text | how to cite |](#)

Title: *Review of Decentralized Finance Applications and Their Total Value Locked*

Authors: Viktorija Stepanova, Ingars Eriņš

DOI: 10.18421/TEM101-41

[. | abstract | full text | how to cite |](#)

Title: *Entrepreneurship Motivation in the 21st Century in Terms of Pull and Push Factors*

Authors: Zsuzsanna Gódnány, Renáta Machová, Ladislav Mura, Tibor Zsigmond

DOI: 10.18421/TEM101-42

[. | abstract | full text | how to cite |](#)

Title: *Development of a Set of Marketing Activities in the Construction of an Innovative Energy Efficient Cluster*

Authors: Alexey G. Palamarchuk

DOI: 10.18421/TEM101-43

[. | abstract | full text | how to cite |](#)

Title: *Workflow Based Prototype for Criminal Investigation in Bulgaria*

Authors: George Pashev, Silvia Gaftandzhieva

DOI: 10.18421/TEM101-44

[. | abstract | full text | how to cite |](#)

Title: *Application of Cognitive Modelling for Operation Improvement of Retail Chain Management System*

Authors: Oksana Dankeieva, Nataliia Solomianiuk, Larysa Strashynska,
Nataliia Fiedotova, Yuliia Soloviova, Viktor Koval

DOI: 10.18421/TEM101-45

[. | abstract | full text | how to cite |](#)

Title: *Implementation of Industry 4.0 Using E-learning and M-learning Approaches in Technically-Oriented Education*

Authors: Dusan Mital', Darina Duplákova, Ján Duplák,
Zuzana Mital'ová, Svetlana Radchenko

DOI: 10.18421/TEM101-46

[. | abstract | full text | how to cite |](#)

Title: *Teaching Accompaniment in Colombia's Official Educational Institutions*

Authors: Romel Ramón González-Díaz, Santos Lucio Guanilo Gómez,
Juan Carlos Vegas-Ochoa, Elena Cachicatari Vargas

DOI: 10.18421/TEM101-47

[. | abstract | full text | how to cite |](#)

Title: *Guiding the Students in High School by Using Machine Learning*

Authors: Mustafa Ababneh, Aayat Aljarrah, Damla Karagozlu, Fezile Ozdamli

DOI: 10.18421/TEM101-48

[. | abstract | full text | how to cite |](#)

Title: *Promoting Sustainable Development in School Classrooms: Using Reciprocal Teaching in Mathematics Education*

Authors: Sarfraz Aslam, Atif Saleem, Ali Usman Hali, BaoHui Zhang

DOI: 10.18421/TEM101-49

[. | abstract | full text | how to cite |](#)

Title: *Education during a State of Emergency*

Authors: Radoslava Topalska

DOI: 10.18421/TEM101-50

[. | abstract | full text | how to cite |](#)

Title: *Ecosystem Functions of Individual Style in a Digital Educational Environment*

Authors: Lyudmila Andryukhina, Nadezhda Sadovnikova, Svetlana Semenova,
Tatiana Sumina, Natalya Tserkovnikova

DOI: 10.18421/TEM101-51

[. | abstract | full text | how to cite |](#)

Title: *Methodical Scheme for Introduction of Combinatorial Compounds in Mathematics Education*

Authors: Ivan Georgiev, Ivo Andreev

DOI: 10.18421/TEM101-52

[. | abstract | full text | how to cite |](#)

Title: *Cyberloafing Among University Students*

Authors: Fezile Ozdamli, Erinc Ercag

DOI: 10.18421/TEM101-53

[. | abstract | full text | how to cite |](#)

Title: *A New Teaching Approach in the Study of “Chemical Kinetics” and “Chemical Equilibrium” with the Contribution of Moodle Learning Platform*

Authors: Anna Thysiadou, Ioanna Sakla, Sophoclis Christoforidis, Panagiotis Giannakoudakis

DOI: 10.18421/TEM101-54

[. | abstract | full text | how to cite |](#)

Title: *Comparative Analysis of Perfectionism and Value-Semantic Barriers of the Student's Personality*

Authors: Galina Epanchintseva, Irina Bukhtiyarova, Natalia Panich

DOI: 10.18421/TEM101-55

[. | abstract | full text | how to cite |](#)

Title: *The Mediating Role of Information Flow and Factors for Supplier Selection*

Authors: Manzoor Ali Mirani, Ikramuddin Junejo, Jan Muhammad Sohu, Sikandar Ali Qalati, Hafiz Muhammad Naveed, Amna Shabir

DOI: 10.18421/TEM101-56

[. | abstract | full text | how to cite |](#)

Title: *Using Chess for Identifying and Correcting “Problem Areas” in the School Math Course*

Authors: Svetlana N. Dvoryatkina, Galina A. Simonovskaya

DOI: 10.18421/TEM101-57

[. | abstract | full text | how to cite |](#)

Title: *Formation of Entrepreneurial Culture of Students*

Authors: Aliya Janissenova, Elmira Kaliyeva, Gulzhazira Kosmaganbetova

DOI: 10.18421/TEM101-58

[. | abstract | full text | how to cite |](#)

Title: *Using Case Study Method for Forming Clinical Thinking Ability in Nursing and Midwifery Education*

Authors: Tsveta Hristova, Teodora Todorova, Maya Markova

DOI: 10.18421/TEM101-59

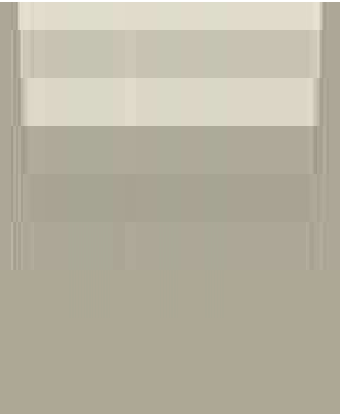

[. | abstract | full text | how to cite |](#)



[Home](#) | [Instructions for the Authors](#) | [Submit paper](#) | [Editorial Board](#) | [Archives](#) | [Contact](#)
[Publication Ethics and Publication Malpractice Statement](#)

Copyright © 2012-2021 UIKTEN, All Rights reserved

Copyright licence: All articles are licenced via Creative Commons CC BY-NC-ND 4.0 licence



Estimation of Gourami Supplies Using Gradient Boosting Decision Tree Method of XGBoost

I Made Sukarsa¹, Ngakan Nyoman Pandika Pinata¹,
Ni Kadek Dwi Rusjyanthi¹, Ni Wayan Wisswani²

¹ *Departement of Information Technology, Udayana University, Bali, Indonesia*

² *Department of Informatic Management, Bali State Polytechnic, Bali, Indonesia*

Abstract – The need for food supplies are very crucial in a food business, therefore it is necessary to estimate the right supplies to maximize profit. One of the methods to determine these is by looking for patterns and forecasting transaction data. The purpose of this research is to estimate the gourami supplies using transaction data to forecast using the gradient boosting decision tree method from XGBoost. The transaction data used comes from Restaurant X with a time period from 2016 to 2019. A measurement error rate of the model is achieved by using MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error). This study tried five XGBoost models with different features such as lag, rolling window, mean encoding, and mix. The results of this study indicate that the mixed feature model produces an accuracy of 97.54% with an MAE of 0.63 and a MAPE of 2.64%.

Keywords – Forecasting, Time Series, GBDT, XGBoost, Gourami Inventory.

1. Introduction

Gourami (*Osphronemus goramy*) is a type of freshwater fish native to Indonesia that has long been

cultivated and consumed by the public because of its delicious meat taste, so it has a high economic value. Therefore, many food businesses sell gourami [1].

Business players in business competition use information technology to be able to analyze transactions carried out by consumers, including forecasting transaction data so that they can provide useful information for the company [2].

Most sales transaction data are only used as archives without being properly utilized, even though these data sets contain information that can be very useful [3]. Cases of data imbalance with the resulting information are usually referred to as the data-rich but information poor phenomenon where organizations have excess data from business processes that are captured and stored still lack of information [4]. Finding new information, patterns or certain rules in a large amount of data can be done using data mining techniques [5].

Data mining also has many methods and categories, in this study, it focuses on forecasting. Forecasting future data based on analysis of past data is an important way to properly explore data for decision making [6], such as sales in the retail industry. Forecasting sales transactions can have a crucial impact on the success and performance of companies. Shortages of inventory and excess inventory that results in losses for the companies usually occurs due to inaccurate forecasts [7].

Researchers made predictions at a restaurant X that sells a special menu of gourami. Restaurant X has a problem in the supplies namely, the supply exceeds demand or is unable to fulfill demand. Uncertainty in sales requires restaurant X to develop a strategy to maximize profits. One of the methods to determine these is by looking for patterns and forecasting data transactions.

Forecasting on sales data is usually performed with a statistical-based methods approach such as research that conducts short-term forecasting of international tourism demand with the F.Y.R. case study Macedonia uses the ARIMA method [8]. Machine learning approaches have also been widely used such

DOI: 10.18421/TEM101-17

<https://doi.org/10.18421/TEM101-17>

Corresponding author: I Made Sukarsa,
Departement of Information Technology, Udayana University, Bali, Indonesia.


Email: sukarsa@unud.ac.id

Received: 18 August 2020.

Revised: 16 January 2021.

Accepted: 21 January 2021.

Published: 27 February 2021.

 © 2021 I Made Sukarsa et al; published by UIK TEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The article is published with Open Access at www.temjournal.com

as research on forecasting outdoor air temperature and humidity using XGBoost which uses data from Shenzhen, China with excellent model results [9], and it was also used to predict hourly PM2.5 (aerosol particles has aerodynamic diameter of less than $2.5 \mu m$) concentration in China, which has a good performance compared with the random forest, decision tree regression, support vector machine, and multiple linear regression [10]. According to [11] doing traffic flow forecasting for one day using the Gradient Boosting Decision Tree (GBDT) method is very effective with a MAPE error rate of 0.097%. XGBoost also produces the best forecast and fast execution times than several other methods such as ARIMA, SARIMA, and Random Forest in load forecasting [12].

This study applies the decision tree gradient boosting method in forecasting because based on previous research, it is found that the use of the gradient boosting method produces very good accuracy. Forecasting is done on transaction data for gourami and uses the XGBoost library. The results of this study are expected to provide recommendations in purchasing raw gourami or determining the supply of raw gourami.

2. Literature Review

2.1. XGBoost

XGBoost is a machine learning ensemble algorithm that is based on gradient boosting. XGBoost has been used in several studies, particularly in data and machine learning competitions. XGBoost has shown great results over the other methods. XGBoost can be used in regression or classification problems, such as prediction of store sales, prediction of customer behavior, prediction of a number of ad clicks, prediction of risk of a case, classification of web text, and classification of malware. [13].

XGBoost requires a number of parameters to choose from. Some of the parameters in XGBoost are as follows.

- 1) **eta**, the learning rate which serves to prevent the model from overfitting.
- 2) **gamma**, to determine tree pruning when the resulting split experiences a positive reduction in the loss function.
- 3) **max depth**, maximum depth of tree.
- 4) **min child weight**, the minimum weight a node should have. If the minimum is not met, splits in a tree will not take place.
- 5) **subsample**, to use the entire row of data (default 1) while 0.5 means randomly using half of the rows of data.

- 6) **colsample bytree**, to use the entire column in the data (default 1) while 0.5 means using half of the existing columns.

2.2. Evaluation Metrics

A forecast model can be said to be accurate or not, there are two aspects that have to be considered. First, it is known that the model can follow the training data pattern, and second, how well the model predicts the test data. One way to evaluate a forecast model is to calculate the Mean Absolute Percentage Error (MAPE). MAPE is one of the most used error metric for measuring regression model [14]. Although MAPE is very sensitive to outliers, MAPE is still very widely used because it is so easy to interpret [15].

MAPE is the average percentage of the absolute result of the difference between the actual value and the predicted value divided by the actual value [16]. MAPE can be found using the formula in Equation (1).

$$MAPE = \frac{100}{n} \sum \left| \frac{A_t - y_t}{A_t} \right| \quad (1)$$

The explanation of Equation (1) is as follows:

A_t = actual value on data t

y_t = predicted value on data t

n = the amount of data

A good regression model is a model that has a MAPE smaller than 10% [17]. MAPE according to [18] is divided into 4 of bad, reasonable, good, and excellent. The range of MAPE values and criteria are shown in Table 1.

Table 1. MAPE score criteria

MAPE	Criterion
<10%	Excellent
10% - 20%	Good
20% - 50%	Reasonable
>50%	Bad

Researchers also use another measure, namely MAE (Mean Absolute Error) which is a simpler value than MAPE. Similar to MAPE, MAE is the absolute value of the difference between the actual value and the predicted value divided by the amount of data expressed in Equation (2).

$$MAE = \sum \left| \frac{A_t - y_t}{n} \right| \quad (2)$$

3. Methodology

3.1. Datasets

The data used to forecast are the sales transactions of Restaurant X in the period January 2016 to December 2019. A snapshot of the transaction data can be seen in Table 2.

Table 2. Transaction data of Restaurant X

Date	TrxCode	Qty	ItemName
14/01/2016	KS160100002	1	Soup Gurame
14/01/2016	KS160100002	1	Gurame 400 gr
15/01/2016	KS160100006	1	Orange Juice
15/01/2016	KS160100006	1	Soda Gembira
15/01/2016	KS160100006	2	Udang Panggang
15/01/2016	KS160100009	1	Gurame 350 gr

The data is received from the Restaurant X cashier program. A lot of data obtained is around 182,756 data with various food menus, but the focus here is the menu that uses gourami.

3.2. Research Flow

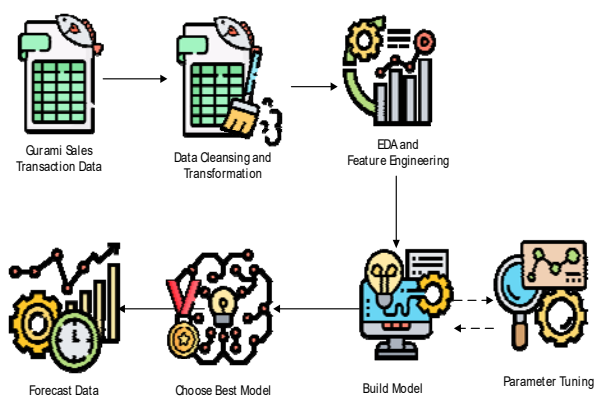


Figure 1. Research Flow

Figure 1 is an overview of this study. The first step is to load gourami sales transaction data in excel file format. Furthermore, data cleaning is carried out, such as deleting unused columns, deleting zero transactions, and transforming data into kilograms. Clean data is used in the EDA (Exploratory Data Analysis) process to see what information the data can provide by visualizing. The next step is performing feature engineering, namely the process of converting data into better features to increase the accuracy of the model.

Model building and parameter configuration are done by trying various models with features that have been made at the feature engineering stage such as lag, rolling window, and mean encoding, then looking for the optimal parameter configuration. All the models that have been built are compared to the MAE and MAPE error values from the predicted results of each model, then a model with a low MAE and MAPE is selected. The last step is if the best model has been found, the model is used to forecast unseen data.

3.3. Data Cleansing

The first step of data cleansing is loading the data then sorting out the data, which ones will be used, and removing those that are not used. The transaction data is then converted into kilograms to make it easier to provide suggestions from the prediction results. The next step is to check the data whether there is a zero sales if it is true that the transaction will be deleted (Figure 2).

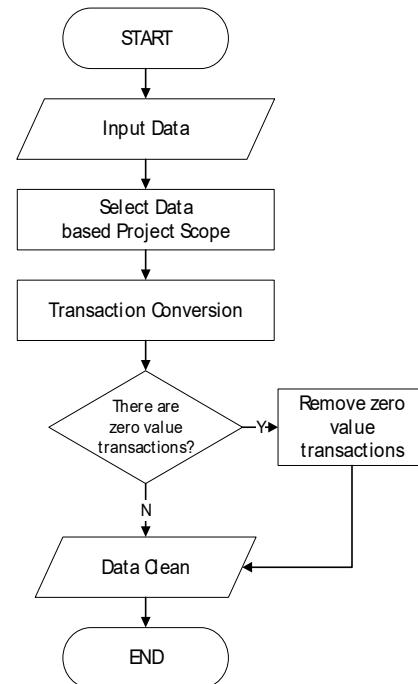


Figure 3. Flowchart Data Cleansing

3.4. Exploratory Data Analysis

Exploratory data analysis (EDA) is a critical first step in analyzing data. EDA is usually carried out by analyzing data descriptively [19]. Exploratory data analysis is used to obtain information from the data, find out the distribution of data, data outliers, the correlation between variables and to obtain the underlying assumptions that occur.

Most EDA techniques are visualizing data. The main reason for visualizing is to explore data and graphics from visualizing give the insight to see data distribution, trends, correlations, and outliers [20].

3.5. Build and Testing Model

The first step in model building process is the initiation of the XGBoost model then inputting training data, test data, and parameters such as max_depth, min_child_weight, eta, subsample, colsample_bytree, and gamma. The next process is

training the XGBoost model. Details of the model training process can be seen in Figure 3. Model evaluation is done by measuring MAE and MAPE. If the error results are still high enough, change the parameters until the best parameter is found. The best model is selected, then visualized to see the trend between the predicted results and the actual data.

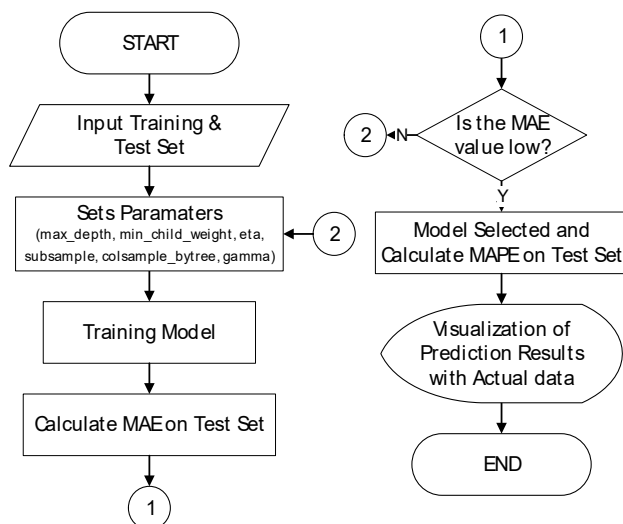


Figure 4. Flowchart Build and Testing Model

4. Implementation and Results

The implementation is carried out in the Kaggle private kernel with the Python version 3. Implementation is done in the Kaggle kernel to speed up computation. The libraries used include the following:

- 1) Numpy, to do math calculations like matrices and algebra;
- 2) Pandas, for data tabulated;
- 3) Plotly, data visualization;
- 4) Sklearn, calculation of MAE and MAPE error metrics;
- 5) XGBoost, machine learning to build model based on gradient boosting decision tree.

The first implementation is data cleansing, which includes selecting data based on project scope, conversion of transactions, and eliminating zero transactions. Second, EDA and finally building and testing models.

4.1. Select Data based Project Scope

The results of the special data selection process for gourami menu can be seen in Figure 4.

Date	qty	Noltem	Namaltem
2016-01-16 13:19:22	1.0	F4	Gurame 400 gr
2016-01-16 13:19:22	1.0	F9	Soup Gurame
2016-01-16 13:29:00	3.0	F3	Gurame 350 gr
2016-01-16 13:56:12	1.0	F4	Gurame 400 gr
2016-01-16 14:04:49	1.0	F4	Gurame 400 gr

Figure 5. Gourami Menu Transactions

The menu with raw gourami ingredients is the number item F1, F2, F3, F4, F5, F6, F7, F8, F9. The total data that has the gourami menu is 37,010 data.

4.2. Transformation

There are two stages in data transformation, namely the transformation of transaction units into kilograms and the transformation of transaction times into daily. The results of the transformation of the gourami transaction data are as follows.

Date	qty	kilo
2016-01-16	37.0	12.85
2016-01-17	30.0	10.10
2016-01-18	35.0	11.40
2016-01-19	26.0	8.50
2016-01-20	20.0	6.65

Figure 6. Result of Transformation Data

The result of the transformation of the kilogram unit contained in the kilo column in Figure 5 is obtained by multiplying each row of the qty column value by the kilogram unit owned by the menu. Daily transformations are obtained by daily resampling using the Pandas library.

4.3. Delete Zero Sales

The zero-value transaction write-off stage is carried out to make the model work better. The results of this stage obtained 79 zero transactions. The 79 transactions were later deleted. So the total rows of data after cleaning this data are 1367 rows of data with daily intervals.

4.4. Exploratory Data Analysis

The Exploratory Data Analysis (EDA) stage was carried out by the researcher by visualizing the transaction data with various types of visualization including line charts, bar charts, and box plots. Visualization of the need for gourami according to transactions that occurred from 2016 to 2019 is shown in Figure 6.

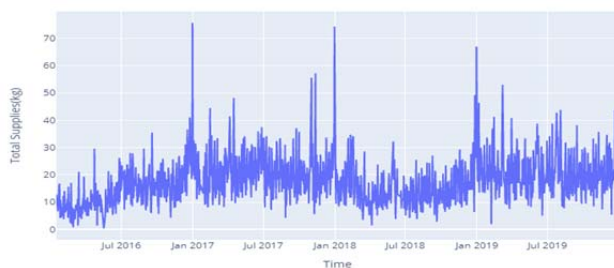


Figure 7. History Transactions

In this visualization, it is known that transactions always experience a spike at the end of each year. The distribution of gourami needs with histogram visualization is as follows.

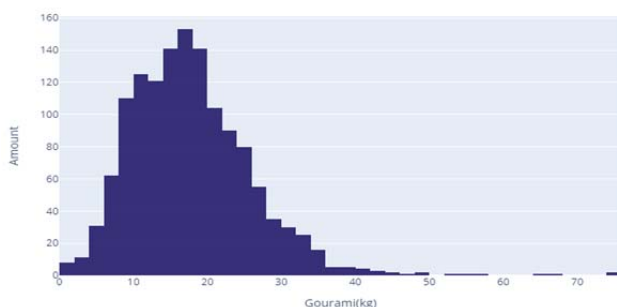


Figure 8. Distribution of Gourami Transaction

Based on the histogram in Figure 7, it is known that there is an abnormal distribution, there are several transactions that are far from the distribution. Transaction distribution top 3 ranges from 14-20kg. The initial assumption of the abnormal distribution is an outlier at the end of the year.

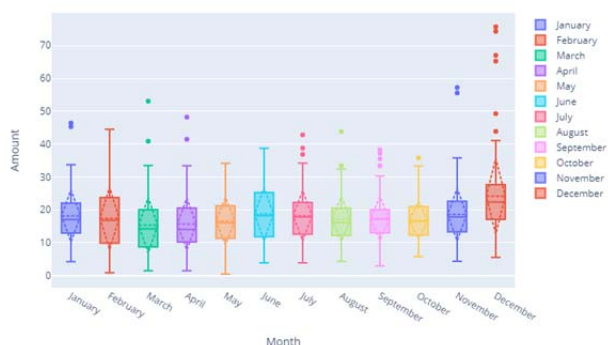


Figure 9. Boxplot of Gourami Transaction Monthly

It turns out that from the boxplot graph (Figure 8) the researcher can see data outliers from each month. Outliers are data that are far from the distribution of other data, therefore in certain months that have high outliers results in a high mean value, in other words, the mean is influenced by outliers. The initial assumption that the abnormal distribution on the histogram is reinforced by the outliers that occurred in December which can be seen in the boxplot visualization. Some of the other visualizations are as follows.

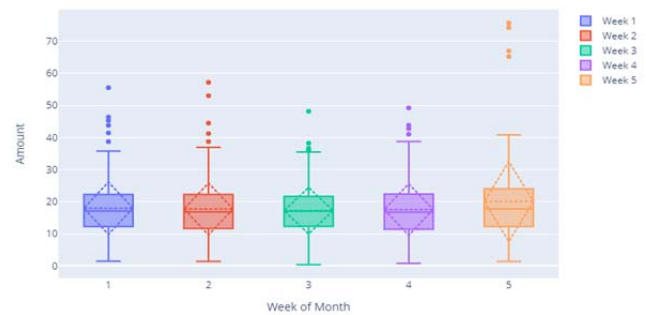


Figure 10. Boxplot of Gourami Transaction Week of Month

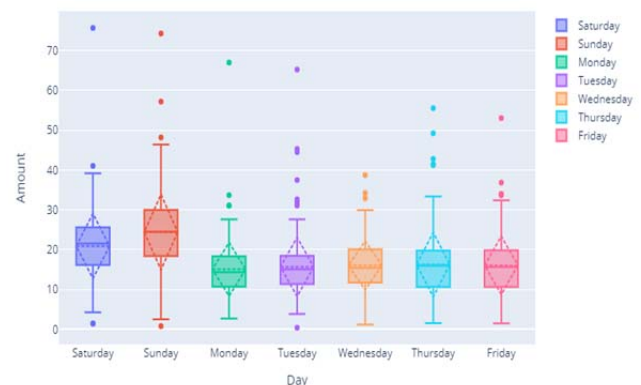


Figure 11. Boxplot of Gourami Transaction Daily

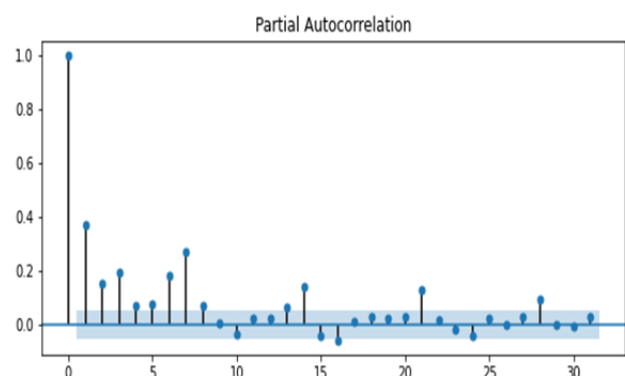


Figure 12. PACF of Gourami Transaction

The information obtained from the exploratory data analysis on the transaction data of gourami at Restaurant X is as follows:

- 1) The use of gourami supplies in 2018 has decreased as shown in Figure 6;
- 2) Every year in December the use of gourami supplies is always high as shown in Figure 6 and Figure 8;
- 3) The use of gourami supplies tends to be high at the end of December;
- 4) Every week in every month tends to be stable as shown in Figure 9;
- 5) Early at the end of the month, the use of gourami supplies is quite stable, except in December;
- 6) Weekends of gourami supplies usage are always high as shown in Figure 10;

- 7) The use of gourami supplies may be influenced by the use of raw materials from yesterday to one week ago as shown in Figure 11.

4.5. Building and Testing Models

The building and testing models are carried out by splitting the data into training sets and test sets as shown in Figure 12. Training set with 1351 data is from 16 January 2016 to 15 December 2019 (blue line Figure 12) and a test set with 16 data from 16 December 2019 to 31 December 2019 (red line Figure 12).

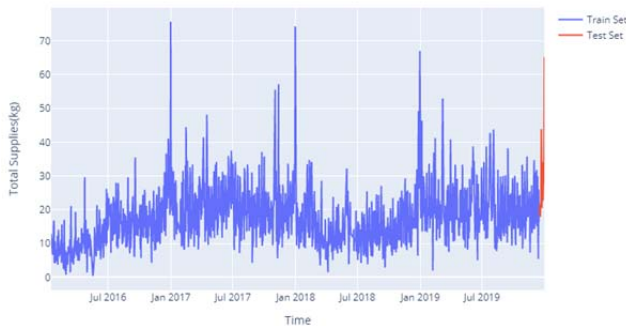


Figure 13. Data Splitting

The models are built with different kinds of features such as date feature, lag feature, mean rolling window feature, mean encoding feature, and combined feature. The description of these features is as follows.

1) Date

Date feature is done by taking components from the time series itself, such as the day component of the year, day of the month, month, and year.

Date	dayofmonth	dayofweek	weekofmonth	month	dayofyear	weekofyear	year
2016-01-16	16	5	3	1	16	2	2016
2016-01-17	17	6	3	1	17	2	2016
2016-01-18	18	0	3	1	18	3	2016
2016-01-19	19	1	3	1	19	3	2016
2016-01-20	20	2	3	1	20	3	2016

Figure 14. Date Features

In Figure 13 January 16, 2016, in the first row data, some components are taken into sections such as *dayofmonth* (day in each month), *dayofweek* (day of the week), *weekofmonth* (week of each month), *month* (month), *dayofyear* (day of the year), *weekofyear* (week of the year), and *year* (year). This date feature can see how the transaction patterns in each time component such as transactions on weekdays and weekends.

2) Lag

Lag Feature is a feature that assigns the value of t to the value of $t-1$. The value at time t is strongly influenced by the value at time $t-n$. The previous value is known as lag. Thus, $t-1$ is lag 1, $t-2$ is lag 2.

Date	kilo	lag_1	lag_2	lag_3
2016-01-16	12.85	NaN	NaN	NaN
2016-01-17	10.10	12.85	NaN	NaN
2016-01-18	11.40	10.10	12.85	NaN
2016-01-19	8.50	11.40	10.10	12.85
2016-01-20	6.65	8.50	11.40	10.10

Figure 15. Lag Features

Based on Figure 14, it is known that on January 16, 2016, the needed gourami is 12.85. Lag 1 takes data on the previous day which is 12.85 placed at $t+1$, lag 2 takes data on the previous 2 days is placed at $t+2$.

3) Rolling Window

Rolling Window feature is a feature that is obtained by taking a summary of several previous values with the aggregate. This research uses the *mean* as the aggregate of the rolling windows obtained.

Date	kilo	roll_mean_2	roll_mean_3	roll_mean_4
2016-01-16	12.85	NaN	NaN	NaN
2016-01-17	10.10	11.475	NaN	NaN
2016-01-18	11.40	10.750	11.45	NaN
2016-01-19	8.50	9.950	10.00	10.7125
2016-01-20	6.65	7.575	8.85	9.1625

Figure 16. Rolling Mean Window Feature

Based on Figure 15, the *roll_mean_2* column takes the average of the two previous observation data, *roll_mean_3* takes the average of the three previous observation data.

4) Mean Encoding

Mean encoding feature is a feature obtained by finding the average of the component time series itself. For example, the average for each month of December, the average for each day of the week, and the average for every 31st.

	kilo	weekofyear	weekofyear_enc
Date			
2016-01-16	12.85	2	19.202273
2016-01-17	10.10	2	19.202273
2016-01-18	11.40	3	16.161111
2016-01-19	8.50	3	16.161111
2016-01-20	6.65	3	16.161111

Figure 17. Mean Encoding Features

Based on Figure 16 the *weekofyear_enc* column takes the average set of values from the week of year time component which has a value of 2. Besides that, this study also uses several time components to be used as a mean encoding feature such as months, days, and days of the year.

5) Combined

Combined features are done by combining pre-made features such as date features, lags, rolling windows and mean encoding.

Model testing is done by searching for the best model of all models with several parameters such as *eta*, *max_depth*, *min_child_weight*, *subsample*, *colsampe_bytree*, and *gamma*. The performance of each model is measured using MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error) on the test data. The results of all models found that the best model is with combined features as it is shown in Table 3 and Figure 17.

Table 3. Result of All Models

Model	MAE	MAPE	Time
Model 1 (Date)	4.03	13,59%	113ms
Model 2 (Lag)	3.39	10,62%	90.6ms
Model 3 (Rol. Window)	2.51	10,34%	231ms
Model 4 (Mean Enc)	3.52	12,56%	139ms
Model 5 (Combined)	0.63	2,46%	348ms



Figure 18. Comparison of the results of all models

4.6. Select Best Model

The best model which has the smallest error value is chosen, namely model 5 which uses combined features. The comparison of the actual data with the predicted results of model 5 on the test data can be seen in Figure 18. The accuracy produced by model 5 is 97.54% with 16-time steps from 16 December 2019 to 31 December 2019.

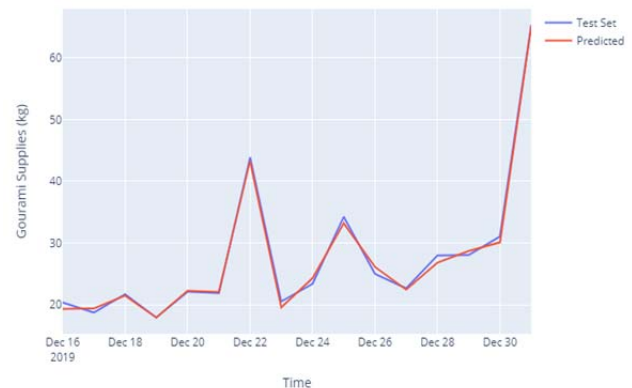


Figure 19. Actual Data vs Predicted

The important features in making model 5 can be seen in Figure 19.

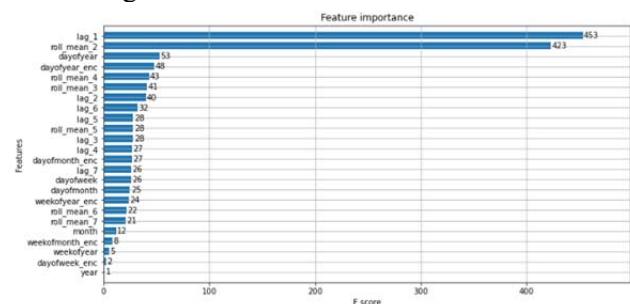


Figure 20. Feature Importance

Based on Figure 19, it is known that two important features in making model 5 are the *lag_1* and *roll_mean_2* features.

5. Conclusion

Based on the results of testing with 5 XGBoost models, the best XGBoost model is obtained with combined features consisting of components of time, lag, rolling window, and mean encoding. The best parameter is the number of *colsampl_bytree* is 1, *eta* 0.175, *min_child_weight* 6, *gamma* 0.6, *max_depth* 11, and *subsample* 0.75 resulting in an accuracy of 97.54% to predict gourami supplies. Suggestions for the further development of this research are to add some external features such as holidays and weather or climate that occurs in the field. Besides that, we can also do a stacking model which makes models with various other methods such as linear SVM models, and neural networks.

References

- [1]. Budiana, B., & Rahardja, B. S. (2018). Teknik Pembenihan Ikan Gurame (*Osphronemus gouramy*) Di Balai Benih Ikan Ngoro, Jombang. *Journal of Aquaculture and Fish Health*, 7(3), 90-97.
- [2]. Anharudin, A. (2015). Perencanaan Strategis Sistem Informasi Untuk Meningkatkan Pelayanan Menggunakan Metode Ward And Peppard (Studi Kasus: PT Pos Indonesia Cilegon-Banten). *Prosisko: Jurnal Pengembangan Riset dan Observasi Sistem Komputer*, 2(2).
- [3]. Elisa, E. (2018). Market Basket Analysis Pada Mini Market Ayu Dengan Algoritma Apriori. *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, 2(2), 472-478.
- [4]. Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.
- [5]. Ngatmari, N., Musthafa, M. B., Rahmad, C., Asmara, R. A., & Rahutomo, F. (2020). Pemanfaatan Data Pddikti Sebagai Pendukung Keputusan Manajemen Perguruan Tinggi. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(3), 555-564.
- [6]. Zhang, Y., Qu, H., Wang, W., & Zhao, J. (2020). A Novel Fuzzy Time Series Forecasting Model Based on Multiple Linear Regression and Time Series Clustering. *Mathematical Problems in Engineering*, 2020, 1-17. <https://doi.org/10.1155/2020/9546792>
- [7]. Beheshti-Kashi, S., Karimi, H. R., Thoben, K. D., Lütjen, M., & Teucke, M. (2015). A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering*, 3(1), 154-161.
- [8]. Petrevska, B. (2017). Predicting tourism demand by Arima models. *Economic research-Ekonomska istraživanja*, 30(1), 939-950.
- [9]. Ma, X., Fang, C., & Ji, J. (2020). Prediction of outdoor air temperature and humidity using Xgboost. In *IOP Conference Series: Earth and Environmental Science* (Vol. 427, No. 1, p. 012013). IOP Publishing. <https://doi.org/10.1088/1755-1315/427/1/012013>.
- [10]. Pan, B. (2018, February). Application of XGBoost algorithm in hourly PM2. 5 concentration prediction. In *IOP Conference Series: Earth and Environmental Science* (Vol. 113, No. 1, p. 012127). IOP publishing. <https://doi.org/10.1088/1755-1315/113/1/012127>.
- [11]. Xia, Y., & Chen, J. (2017, April). Traffic flow forecasting method based on gradient boosting decision tree. In *2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017)* (pp. 413-416). Atlantis Press. <https://doi.org/10.2991/fmsmt-17.2017.87>.
- [12]. Hadri, S., Naitmalek, Y., Najib, M., Bakhouya, M., Fakhri, Y., & Elaroussi, M. (2019). A comparative study of predictive approaches for load forecasting in smart buildings. *Procedia Computer Science*, 160, 173-180.
- [13]. Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). <https://doi.org/10.1145/2939672.2939785>
- [14]. Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669-679.
- [15]. Chen, C., Twycross, J., & Garibaldi, J. M. (2017). A new accuracy measure based on bounded relative error for time series forecasting. *PloS one*, 12(3), e0174202. <https://doi.org/10.1371/journal.pone.0174202>.
- [16]. Putro, B., Furqon, M. T., & Wijoyo, S. H. (2018). Prediksi Jumlah kebutuhan pemakaian air menggunakan metode exponential smoothing (Studi Kasus: PDAM Kota Malang). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. e-ISSN, 2548, 964X.
- [17]. Almazrouee, A. I., Almeshal, A. M., Almutairi, A. S., Alenezi, M. R., & Alhajeri, S. N. (2020). Long-Term Forecasting of Electrical Loads in Kuwait Using Prophet and Holt–Winters Models. *Applied Sciences*, 10(16), 5627. <https://doi.org/10.3390/app10165627>.
- [18]. Chang, P. C., Wang, Y. W., & Liu, C. H. (2007). The development of a weighted evolving fuzzy neural network for PCB sales forecasting. *Expert Systems with Applications*, 32(1), 86-96.
- [19]. R. Peng. (2012). *Exploratory Data Analysis with R*. LeanPub.
- [20]. Komorowski, M., Marshall, D. C., Saliccioli, J. D., & Crutain, Y. (2016). Exploratory data analysis. *Secondary Analysis of Electronic Health Records*, 185-203. https://doi.org/10.1007/978-3-319-43742-2_15.