AIP Conference Proceedings

scitation.org/journal/apc

Volume 2668

The 3rd International Conference on Mathematics and Sciences (The 3rd ICMSc)

A Brighter Future with Tropical Innovation in the Application of Industry 4.0

East Kalimantan, Indonesia • 12–13 October 2021

Editors • Rudy Agung Nugroho, Veliyana Londong Allo, Meiliyani Siringoringo, Surya Prangga, Wahidah, Rahmiati Munir and Irfan Ashari Hiyahara

The 3rd ICMSc 2021

FMIPA UNMUL

International Conference on Mathematics and Sciences

AIP Publishing

Table of Contents

THE 3RD INTERNATIONAL CONFERENCE ON MATHEMATICS AND SCIENCES (THE 3RD ICMSc): A Brighter Future with Tropical Innovation in the Application of Industry 4.0

PREV NEXT

Conference date: 12–13 October 2021
Location: East Kalimantan, Indonesia
ISBN: 978-0-7354-4214-6
Editors: Rudy Agung Nugroho, Veliyana Londong Allo, Meiliyani Siringoringo, Surya Prangga, Wahidah, Rahmiati Munir and Irfan Ashari Hiyahara
Volume number: 2668
Published: Oct 11, 2022

DISPLAY : 20 50 100 all

PRELIMINARY

AIP Conference Proceedings **2668**, 010001 (2022); https://doi.org/10.1063/12.0013094

BIOLOGY

No Access . October 2022

Growth and digestive enzyme activity of *Clarias gariepinus* fed *Myrmecodia pendens* merr. & perry bulb extract

Lani Nurfakhira, Rudy Agung Nugroho, Retno Aryani, Yanti Puspita Sari, Hetty Manurung and Rudianto Rudianto

AIP Conference Proceedings **2668**, 020001 (2022); https://doi.org/10.1063/5.0111704

SHOW ABSTRACT

No Access . October 2022

Effects of ant-nest plant (*Myrmecodia pendens*) bulb extract on histology of intestinal, liver and proximate fillet of Sangkuriang catfish (*Clarias gariepinus Var*)

Nurlaila Ismah Sardi, Retno Aryani, Rudy Agung Nugroho, Yanti Puspita Sari, Hetty Manurung and Rudianto Rudianto

AIP Conference Proceedings **2668**, 020002 (2022); https://doi.org/10.1063/5.0111711

SHOW ABSTRACT

No Access . October 2022

Preliminary study of forensic investigation: Insect arrival pattern on Wistar rat (*Rattus norvegicus*) carrion in Samarinda, East Kalimantan

Feiky Aprilasari, Nova Hariani and Sus Trimurti

AIP Conference Proceedings **2668**, 020003 (2022); https://doi.org/10.1063/5.0114146

SHOW ABSTRACT

No Access . October 2022

Antioxidant activity of single bulb garlic callus (*Allium sativum* L.) extract with *in vitro* method

Yanti Puspita Sari and Merdi Sahara Anwar

AIP Conference Proceedings **2668**, 020004 (2022); https://doi.org/10.1063/5.0112192

SHOW ABSTRACT

No Access . October 2022

Subchronic toxicity of *Ficus deltoidea* jack. leaves on the histology of ventriculus and intestinum tenue male mice

Retno Aryani, Rudy Agung Nugroho, Hetty Manurung, Nur Ainun Oktavia Pusparini, Widha Prahastika and Rudianto Rudianto

AIP Conference Proceedings **2668**, 020005 (2022); https://doi.org/10.1063/5.0111944

SHOW ABSTRACT

CHEMISTRY

No Access . October 2022

Preparation of sodium silicate to synthesize silica with polymerization $Si(OH)_4$ from rice husk (INPARI 64, Deli Serdang)

Supiyani Supiyani, Harry Agusnar, Purwantiningsih Sugita and Irwana Nainggolan

AIP Conference Proceedings **2668**, 030001 (2022); https://doi.org/10.1063/5.0114582

SHOW ABSTRACT

No Access . October 2022

Effect of extraction time and solid-liquid ratio on tannin extraction from guava (*Psidium guajava*) leaves

Helda Niawanti, Fitri Yani, Mu'min Herman and Husnul Rafliansyah

AIP Conference Proceedings **2668**, 030002 (2022); https://doi.org/10.1063/5.0111847

SHOW ABSTRACT

No Access . October 2022

Utilization of Urang-aring (*Eclipta prostrata* L.) leaf as

solar cells

Noor Hindryawati, Irfan Ashari Hiyahara, M. Syaiful Arief and Yunistira Yunistira

AIP Conference Proceedings **2668**, 030003 (2022); https://doi.org/10.1063/5.0111706

SHOW ABSTRACT

No Access . October 2022

Methylene blue degradation using Fe_3O_4-ZnO/WO_3 catalyst

Noor Hindryawati, Teguh Wirawan, Soerja Koesnarpadi, Daniel Daniel, Ainun Rahmi Ramadhannur, Mukhamad Nurhadi, Uripto Trisno Santoso, Triastuti Sulistyaningsih and Abu Masykur

AIP Conference Proceedings **2668**, 030004 (2022); https://doi.org/10.1063/5.0111705

SHOW ABSTRACT

No Access . October 2022

In silico analysis of flavonol compound against Mpro COVID-19

Veliyana Londong Allo, Gabby Encha Farhanah and Rahmat Gunawan

AIP Conference Proceedings **2668**, 030005 (2022); https://doi.org/10.1063/5.0111695

SHOW ABSTRACT

No Access . October 2022

Comparing CO_2 concentrations from ground and space

Waluyo Eko Cahyono, Nani Cholianawati, Indah Susanti and Zahra Syahida Sabila

AIP Conference Proceedings **2668**, 030006 (2022); https://doi.org/10.1063/5.0124116

SHOW ABSTRACT

No Access . October 2022

Method verification and absorption of Cr^{6+} ion by the bacterium *Pseudomonas* sp

Daffa Anugerah, Rudi Kartika and Rahmat Gunawan

AIP Conference Proceedings **2668**, 030007 (2022); https://doi.org/10.1063/5.0112011

SHOW ABSTRACT

No Access . October 2022

Adsoprtion of Cu(II) ion in aqueous solution by *Pseudomonas* sp. biosorbent

Achmad Dwi Setiawan, Rudi Kartika and Rahmat Gunawan

AIP Conference Proceedings **2668**, 030008 (2022); https://doi.org/10.1063/5.0112558

SHOW ABSTRACT

No Access . October 2022

Riki Riki, Rudi Kartika and Rahmat Gunawan

AIP Conference Proceedings **2668**, 030009 (2022); https://doi.org/10.1063/5.0112190

SHOW ABSTRACT

No Access . October 2022

Synthesis of silver nanoparticles (AgNPs) using aggregate mangrove leaf extract (*Sonneratia alba*) for colorimetric analysis of chloramphenicol

Moh. Syaiful Arif, Erika Risdayana, Ika Yekti Liana Sari and Aman Sentosa Panggabean

AIP Conference Proceedings **2668**, 030010 (2022); https://doi.org/10.1063/5.0112721

SHOW ABSTRACT

No Access . October 2022

Fulvic acids coated magnetite nanoparticles for methylene blue adsorption

Jerry Darmawan, Soerja Koesnarpadi and Noor Hindryawati

AIP Conference Proceedings **2668**, 030011 (2022); https://doi.org/10.1063/5.0111738

SHOW ABSTRACT

No Access . October 2022

Pyrolysis of green microalgae by thermogravimetric analysis (TGA) and TG-FTIR

Polonius Dosi Miten, R. R. Dirgarini J. N. Subagyono, Rahmat Gunawan, Alan L. Chaffee, Marc Marshall and Rudy Agung Nugroho

AIP Conference Proceedings **2668**, 030012 (2022); https://doi.org/10.1063/5.0112297

SHOW ABSTRACT

No Access . October 2022

Synthesis of magnetic coffee grounds from cafe as a dye adsorbent

Teguh Wirawan, Noor Hindryawati and Nanang Tri Widodo

AIP Conference Proceedings **2668**, 030013 (2022); https://doi.org/10.1063/5.0111749

SHOW ABSTRACT

GEOPHYSICS

No Access . October 2022

Influences *EL Nino Southern Oscillation (ENSO)* on rainfall in East Kalimantan-Indonesia

Idris Mandang and Rahmiati

AIP Conference Proceedings **2668**, 040001 (2022); https://doi.org/10.1063/5.0113918

SHOW ABSTRACT

No Access . October 2022

Numerical assessment of flushing time in Balikpapan Bay

Mutiara Rachmat Putri, Najwan Alghifari, Lamona Irmudyawati Bernawis, Agus Setiawan and Idris Mandang

AIP Conference Proceedings **2668**, 040002 (2022); https://doi.org/10.1063/5.0111776

SHOW ABSTRACT

MATHEMATICS

No Access . October 2022

Series solutions for the spherically symmetric droplet transient heating problem

Kwassi Anani

AIP Conference Proceedings **2668**, 050001 (2022); https://doi.org/10.1063/5.0112066

SHOW ABSTRACT

No Access . October 2022

Analysis of service provider marketing strategy in achieving customer loyalty using game theory for senior high school

Wasono Wasono, Syaripuddin Syaripuddin and Darnah Andi Nohe

AIP Conference Proceedings **2668**, 050002 (2022); https://doi.org/10.1063/5.0112779

SHOW ABSTRACT

No Access . October 2022

Continuous monotonic decomposition of corona product $C_n \odot \overline{K_m}$

Suaidah Suaidah and Purwanto Purwanto

AIP Conference Proceedings **2668**, 050003 (2022); https://doi.org/10.1063/5.0111839

SHOW ABSTRACT

No Access . October 2022

Calculation of region in more than two different UTM zone using shoelace formula based on the GADM database

Adi Setiawan and Eko Sediyono

AIP Conference Proceedings **2668**, 050004 (2022); https://doi.org/10.1063/5.0111701

SHOW ABSTRACT

No Access . October 2022

Predicting the number of Covid-19 cases in Surabaya using

optimization

Mohamad Handri Tuloli, Syaiful Anam and Nur Shofianah

AIP Conference Proceedings **2668**, 050005 (2022); https://doi.org/10.1063/5.0112007

SHOW ABSTRACT

No Access . October 2022

Hill Cipher algorithm using two keywords and 94 ASCII characters

Fidia Deny Tisna Amijaya, Syaripuddin Syaripuddin, Qonita Qurrota A'yun, Yuki Novia Nasution, Wasono Wasono and Moh. Nurul Huda

AIP Conference Proceedings **2668**, 050006 (2022); https://doi.org/10.1063/5.0111696

SHOW ABSTRACT

No Access . October 2022

A mathematical modelling in determining the portion of Tabarru' fund for sharia life insurance based on wakalah scheme

Galuh Oktavia Siswono, Wawan Hafid Syaifudin and Wisnowan Hendy Saputra

AIP Conference Proceedings **2668**, 050007 (2022); https://doi.org/10.1063/5.0112031

SHOW ABSTRACT

No Access . October 2022

Detecting the tomato leaf disease using clustering based on fuzzy adaptive turbulence particle swarm optimization

Syaiful Anam, Indah Yanti and Zuraidah Fitriah

AIP Conference Proceedings **2668**, 050008 (2022); https://doi.org/10.1063/5.0111988

SHOW ABSTRACT

No Access . October 2022

Regularized quasi semigroups

Risky Yoga Suratman and Christiana Rini Indrati

AIP Conference Proceedings **2668**, 050009 (2022); https://doi.org/10.1063/5.0112850

SHOW ABSTRACT

PHYSICS

No Access . October 2022

SNI standard freight software design with social platform media for cost efficiency and effectiveness for product advantages for cooperative found MSME products

Diena Noviarini and Eko Arif Syaefudin

AIP Conference Proceedings **2668**, 060001 (2022); https://doi.org/10.1063/5.0113102

SHOW ABSTRACT

No Access . October 2022

Effect of width, thickness, and length on static bending of red meranti (*Shorea* sp.)

Irvin Dayadi

AIP Conference Proceedings **2668**, 060002 (2022); https://doi.org/10.1063/5.0111713

SHOW ABSTRACT

No Access . October 2022

Synthesis and characterization of TiO_2 from Lampung's iron sand using acid leaching method with variation of HCl concentration

Roniyus Marjunus, Adi Sucipto, Diah Purwarini and Iqbal Firdaus

AIP Conference Proceedings **2668**, 060003 (2022); https://doi.org/10.1063/5.0111689

SHOW ABSTRACT

No Access . October 2022

Effect of truncating the Taylor series terms on derivative results of the complex conjugate approach and its applications in geophysics and statistics

Rani Chahyani and Abdul Manan

AIP Conference Proceedings **2668**, 060004 (2022); https://doi.org/10.1063/5.0111981

SHOW ABSTRACT

No Access . October 2022

Physic laboratory by video tracker and Visual Basic for Application at home during Covid-19 pandemic: Material elasticity measurement

Rahmawati Munir, Nadya Amalia and Rahmiati Munir

AIP Conference Proceedings **2668**, 060005 (2022); https://doi.org/10.1063/5.0113768

SHOW ABSTRACT

STATISTICS

No Access . October 2022

ASEAN foreign trade analysis using dynamic panel data and simultaneous regression model

Dwi Endah Kusrini, Zakiatul Wildani and Nur Azizah

AIP Conference Proceedings **2668**, 070001 (2022); https://doi.org/10.1063/5.0117377

SHOW ABSTRACT

No Access . October 2022

Negative binomial regression analysis on dengue hemorrhagic fever cases in East Kalimantan Province

M. Fathurahman, Ika Purnamasari and Surya Prangga

AIP Conference Proceedings **2668**, 070002 (2022); https://doi.org/10.1063/5.0111999

SHOW ABSTRACT

No Access . October 2022

Study of features importance level identification of machine learning classification model in sub-populations for food insecurity

Endang Yuliani, Bagus Sartono, Hari Wijayanto, Alfian Futuhul Hadi and Evi Ramadhani

AIP Conference Proceedings **2668**, 070003 (2022); https://doi.org/10.1063/5.0111784

SHOW ABSTRACT

No Access . October 2022

Clustering on 'Z' e-learning application's reviews using self organizing maps and K-means methods

Nur Anisa Damayanti and Irhamah Irhamah

AIP Conference Proceedings **2668**, 070004 (2022); https://doi.org/10.1063/5.0113103

SHOW ABSTRACT

No Access . October 2022

Assessing the optimum condition of multivariate second order response surface model through the asymptotic inference of the eigenvalues

Wayan Somayasa, Ruslan Ruslan and Desak Ketut Sutiari

AIP Conference Proceedings **2668**, 070005 (2022); https://doi.org/10.1063/5.0112594

SHOW ABSTRACT

No Access . October 2022

Spatial panel data model of poverty in Bengkulu province with fixed effect and spatial error autocorrelation

Dyah Setyo Rini, Sigit Nugroho, Dian Agustina, Ratu Eva Febriani, Tommy Jomecho, Diah Permata Syaiful and Sanita Grasia Rumahorbo

AIP Conference Proceedings **2668**, 070006 (2022); https://doi.org/10.1063/5.0112039

SHOW ABSTRACT

No Access . October 2022

Spatial autoregressive quantile regression as a tool for modelling human development index factors in 2020 East Java

Annisa B. Tribhuwaneswari, Alfisyahrina Hapsery and Widya Kharisma Rahayu

AIP Conference Proceedings **2668**, 070007 (2022); https://doi.org/10.1063/5.0112828

SHOW ABSTRACT

No Access . October 2022

Does a nonlinear nexus between crude oil prices, exchange rates, inflation, and Bangladesh's performance of remittances?

Rehana Parvin

AIP Conference Proceedings **2668**, 070008 (2022); https://doi.org/10.1063/5.0111819

SHOW ABSTRACT

No Access . October 2022

Time series analysis on amount palm oil production data in East Kalimantan

Meiliyani Siringoringo, Ika Purnamasari, Sri Wahyuningsih, Rito Goejantoro and Surya Prangga

AIP Conference Proceedings **2668**, 070009 (2022); https://doi.org/10.1063/5.0111753

SHOW ABSTRACT

No Access . October 2022

Mapping the health quality in Sumenep using K-Medoids Algorithm

Iis Dewi Ratih, Mike Prastuti, Zakiatul Wildani, Sri Pingit Wulandari, Wahyu Wibowo, Sri Mumpuni Retnaningsih, Amara Deviana Chaniago, Mumtazah Nurul 'Aini and Farich

AIP Conference Proceedings **2668**, 070010 (2022); https://doi.org/10.1063/5.0111821

SHOW ABSTRACT

No Access . October 2022

Attitudes towards COVID-19 vaccines to support the achievement of government targets: A case study of Bontang city

Darnah Andi Nohe, M. Fathurahman Suyitno, Rito Goejantoro and Memi Nor Hayati

AIP Conference Proceedings **2668**, 070011 (2022); https://doi.org/10.1063/5.0113139

SHOW ABSTRACT

No Access . October 2022

Geographically temporally weighted regression model for GIS mapping of influence COVID-19 in East Kalimantan

Sifriyani Sifriyani, Idris Mandang and Fidia Deny Tisna Amijaya

AIP Conference Proceedings **2668**, 070012 (2022); https://doi.org/10.1063/5.0111808

SHOW ABSTRACT

No Access . October 2022

Estimation of nonparametric ordinal logistic regression

based on local scoring algorithm

Marisa Rifada, Nur Chamidah and Ratih Ardiati Ningrum

AIP Conference Proceedings **2668**, 070013 (2022); https://doi.org/10.1063/5.0111771

SHOW ABSTRACT

No Access . October 2022

Forecasting premium adequacy to claim paid ratio in life insurance industry with COVID-19 effect using multilayer perceptron neural network

Siti Maghfirotul Ulyah, Marisa Rifada, Elly Ana, Christopher Andreas, Ilma Amira Rahmayanti and Salsabylla Nada Apsariny

AIP Conference Proceedings **2668**, 070014 (2022); https://doi.org/10.1063/5.0111946

SHOW ABSTRACT

No Access . October 2022

Measuring the efficiency of economic growth in East Java using data envelopment analysis

Zakiatul Wildani, Wahyu Wibowo and Sri Pingit Wulandari

AIP Conference Proceedings **2668**, 070015 (2022); https://doi.org/10.1063/5.0111755

SHOW ABSTRACT

No Access . October 2022

Sentiment analysis of IPOT application reviews using naïve Bayes method

Nur Rochmah Kurniasari and Wahyu Wibowo

AIP Conference Proceedings **2668**, 070016 (2022); https://doi.org/10.1063/5.0116704

SHOW ABSTRACT

No Access . October 2022

Sentiment analysis of *SOCO by* Sociolla e-commerce application review on Google Play Store site using naïve Bayes classifier method

Rega Andika Rahmawati and Wahyu Wibowo

AIP Conference Proceedings **2668**, 070017 (2022); https://doi.org/10.1063/5.0115598

SHOW ABSTRACT

No Access . October 2022

Sentiment analysis data by ShopeePayLater Twitter's opinion using naïve Bayes classifier

Fauzillah Indriani and Wahyu Wibowo

AIP Conference Proceedings **2668**, 070018 (2022); https://doi.org/10.1063/5.0111732

SHOW ABSTRACT

BROWSE VOLUMES

No Access . October 2022

Forecasting of analytic residential price index using ARIMA Box-Jenkins

Adi Prayogo Muhammad, Wahyu Wibowo and Mike Prastuti

AIP Conference Proceedings **2668**, 070019 (2022); https://doi.org/10.1063/5.0116915

SHOW ABSTRACT

No Access . October 2022

The impact of the Covid-19 pandemic and Joe Biden's winning speech to IDX composite using multi input intervention analysis

Maria Novita and Irhamah Irhamah

AIP Conference Proceedings **2668**, 070020 (2022); https://doi.org/10.1063/5.0124110

SHOW ABSTRACT

No Access . October 2022

Synthetic minority over-sampling technique nominal continous logistic regression for imbalanced data

Iis Dewi Ratih, Sri Mumpuni Retnaningsih, Islahulhaq Islahulhaq and Vivi Mentari Dewi

AIP Conference Proceedings **2668**, 070021 (2022); https://doi.org/10.1063/5.0111804

SHOW ABSTRACT

No Access . October 2022

Forecasting gold based on ensemble empirical mode decomposition and Elman Recurrent Neural Network

Ardityan Purbo Adhitama, Heri Kuswanto and Irhamah Irhamah

AIP Conference Proceedings **2668**, 070022 (2022); https://doi.org/10.1063/5.0111698

SHOW ABSTRACT

Study of features importance level identification of machine learning classification model in sub-populations for food insecurity

Cite as: AIP Conference Proceedings **2668**, 070003 (2022); https://doi.org/10.1063/5.0111784
Published Online: 11 October 2022

Endang Yuliani, Bagus Sartono, Hari Wijayanto, et al.

Study of Features Importance Level Identification of Machine Learning Classification Model in Sub-Populations for Food Insecurity

Endang Yuliani[1,a)], Bagus Sartono[1], Hari Wijayanto[1], Alfian Futuhul Hadi[2], and Evi Ramadhani[3]

[1] *Department of Statistics, IPB University, Bogor, West Java, Indonesia*
[2] *Department of Mathematics, University of Jember, Jember, East Java, Indonesia*
[3] *Department of Statistics, Syah Kuala University, Banda Aceh, Aceh, Indonesia*

[a)] Corresponding author: passsipbendang@apps.ipb.ac.id

Abstract. To explain a complicated machine-learning model, data scientists work a lot with identifying the importance of predictor features of the model. Shapley Additive Explanation (SHAP) and Permutation Feature Importance (PFI) are popular methods useful to measure the feature-importance levels. This research examines the utilization of both techniques to reveal the contribution of predictors in a model to classify the food-insecurity status of Indonesian households. Food insecurity is a condition in which a person does not have protected access to safe and nutritious food in sufficient quantities for normal growth and development and active and healthy life. Instead of identifying significant predictors for a population in general, the study is interested in identifying each sub-population: urban and rural areas in West Java Province. A random forest algorithm was implemented to generate a model using both complete data and separate data. A follow-up analysis was then conducted by applying SHAP for both types of data and PFI for separate data. In general, the two approaches using SHAP resulted in quite similar feature importance levels. Meanwhile, the results of SHAP and PFI are relatively different.

INTRODUCTION

Big data is a term for data that is large, has a fast growth rate, and usually consists mainly of unstructured data. Therefore, machine learning techniques are needed in the big data analysis process. This technique produces predictive models with excellent accuracy [1]. In addition, this technique can capture linear patterns to provide additional information that generally fails to be captured by the classical linear model approach [2] and produce more satisfactory predictions. One of the most widely used supervised machine learning prediction models is a classification tree-based model. One of these models is a random forest which can produce better accuracy performance than some other algorithms because it is a development of the ensemble tree method whose formation is a combination of several classification trees [3].

Popular interpretation methods used are PFI based on classification tree [3] and SHAP [4]. These two interpreters have different algorithms for calculating the feature importance. SHAP places more emphasis on the attribution of features and provides local information or information from an individual. SHAP can describe the effect of each predictor to obtain an individual's guess. Meanwhile, PFI is based on the importance score derived from the error rate of the original model. A feature that does not contribute to the performance of a model, then changing the data structure will have no significant effect. However, a significant feature will significantly affect the model's performance if the data structure is changed. These two methods produce the level of feature importance, which can be a solution in interpreting the black-box model that is useful for the government in making policies and determining the direction of development. One of the focuses of world development, including Indonesia, is food insecurity. Food insecurity is when a person does not have protected access to safe and nutritious food in sufficient quantities for average growth

The 3rd International Conference on Mathematics and Sciences (The 3rd ICMSc)
AIP Conf. Proc. 2668, 070003-1–070003-10; https://doi.org/10.1063/5.0111784
Published by AIP Publishing. 978-0-7354-4214-6/$30.00

and development and active and healthy life [5]. Food insecurity is the second purposes of the Sustainable Development Goals (SDGs) of 17 other purposes that the government has agreed.

According to the March 2020 Susenas data, as many as 21.6% of households are food insecure, of which 26.48% are in rural areas and 18.91% in urban areas. Rural and urban area are sub-populations that have different characteristics. So it is possible to have different levels of feature importance. Complete data and separate data approaches are used to identify the importance of features in the sub-populations. [6] found that the variables that affect food insecurity are having low levels of education, less social capital, weak social networks, low household income, and being unemployed. [7] found that households with younger, less-educated household heads were more likely to suffer food insecurity.

Based on the problems above, we need a method to interpret the black box model (random forest), namely SHAP and Permutation Feature Importance. Furthermore, the complete data and separate data approaches are expected to provide additional insight into the interpretation of sub-populations.

LITERATURE REVIEW

Food Insecurity Experience Scale

FAO launched the Voices of Hungry project in 2013 to develop a methodology for measuring the severity of food insecurity, namely Food Insecurity Experience Scale (FIES). FIES measures the severity of food insecurity at the household or individual level, whose value depends on yes/no answers to 8 questions regarding respondents' access to adequate food [8]. In 2017 Indonesia used FIES for the first time in the National Socio-Economic Survey (Susenas) questionnaire for the household level. The experience of global food insecurity begins with a feeling of worry about not having enough food. In conditions that do not improve, adjustments will be made to the quality of nutrition and types of food. This condition indicates a state of mild food insecurity. However, the ongoing food insecurity situation causes respondents to skip meals and have food portions that are not following what is needed. This condition indicates a moderate food insecurity situation. The worst condition is when hungry but have no food and even experience not eating all day. This condition suggests a state of severe food insecurity.

Random Forest

The Random Forest (RF) algorithm is the development of the Classification and Regression Tree (CART) method by applying the bootstrap aggregating (bagging) and random feature selection [3]. The RF algorithm is an algorithm that is suitable for classifying big data, and in the RF algorithm, there is no pruning of features as in the decision tree algorithm. The RF method combines many trees, unlike a single tree which only consists of one tree to make classification and class predictions. In RF, tree formation is done by training sample data. The way to take data samples is using Sampling with replacement. The selection of features used for split is taken randomly. Classification is run after all trees are formed. The determination of the classification in this RF is taken based on the votes from each tree, and the majority votes are the winners.

The following is a algorithm for constructing a Random Forest on a data set consisting of n observations and p predictor features [3][9]:

1. Perform random sampling of size n with recovery on the data cluster. This stage is the bootstrap stage.
2. Using the bootstrap example, the tree is constructed until it reaches its maximum size (without pruning). Then, arrange a tree based on these data, but in each separation process, randomly select $m \ll p$ predictor features and do the best separation. This stage is the stage of random feature selection.
3. Repeat steps 1 and 2 k times to form a forest consisting of k trees.
4. Perform a composite estimate based on k trees. For the classification case, a majority vote is used, which is the category or class that appears most often due to predictions from k classification trees.

Classification Model Evaluation

The classification resulting from a method is expected to classify all data correctly, but it is undeniable that the performance of a system cannot be 100% accurate and correct. Generally, the measurement of the performance of the classification model performs using a confusion matrix, namely cross-tabulation between the response feature data included in the prediction and observation classes [10]. For example, for cases with two classes, the cross-tabulation formed is as follows:

TABLE 1. Confusion matrix

Actual	**Predicted**	
	Positive	**Negative**
Positive	True Postive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

The performance of a classification model can be measured through three values, namely accuracy, F1-Score, and AUC. Accuracy is the percentage of the model that is correct in making predictions. The F1-score is a comparison of the average precision with weighted recall. The AUC value measures discriminatory performance by estimating the output probability from a randomly selected sample. The AUC value is 0-1, the greater the AUC value, the stronger the classification used.

Permutation Feature Importance (PFI)

Permutation Importance is an algorithm to obtain information on the feature importance by permutating (rearranging data sets) the features used in training prediction models. The permutation feature importance measurement was introduced by [3] for random forest algorithm. Then, [11] proposed an agnostic model version of feature importance called the reliance model. The inputs determining feature importance are trained to model f, matrix variable X, response vector y, error size $L(y, f)$. Thus, the Permutation Feature Importance Algorithm is :

1. Estimate the original model error $e^{orig} = L\big(y, f(X)\big)$
2. For each feature $j = 1, \dots, p$ do:
   - Generate feature matrix X^{perm} by permuting feature j in the data X. his breaks the association between feature j and true outcome y.
   - Estimate error $e^{perm} = L\big(Y, f(X^{perm})\big)$ based on the predictions of the permuted data.
   - Calculate permutation feature importance $FI^j = \frac{e^{perm}}{e^{orig}}$ Alternatively, the difference can be used: $FI^j = e^{perm} - e^{orig}$
   - Sort features by descending FI

If a feature does not contribute much to the performance of a model, then changing the arrangement of the data will have no significant effect. On the other hand, a feature that has a significant contribution will significantly affect the model's performance if the data structure is changed.

Shapley Value

The Shapley value created by Lloyd Shapley (1953) is an in-game method of determining the rewards to a player "fairly" by considering the player's contribution to the accumulated value. The player is referred to as the predictor, while the rewards is the estimated value. Thus, all possible combinations of players or predictor must be evaluated with and without the jth predictor to calculate Shapley's value. If M is the number of players, $S \subseteq \mathcal{M} = \{1, \dots, M\}$ is a subset consisting of S players and assumed a characteristic function or a contribution function $v(S)$ which can map a subset of players to a number, then the Shapley Value for the jth player is:

$$\varphi_j(v) = \sum_{S \subseteq \mathcal{M} \setminus \{j\}} \frac{|S|!\,(M - |S| - 1)}{M!} \big(v(S \cup \{j\}) - v(S)\big) \tag{1}$$

Shapley Additive Explanation

SHAP is an interpretation method for individual predictions based on the optimal game theory of Shapley Values [12]. SHAP aims to explain the prediction of an individual x by calculating the contribution of each feature. The SHAP interpretation formula is:

$$g(z') = \emptyset_0 + \sum_{j=1}^{M} \emptyset_j z'_j \quad (2)$$

Where g is the interpretation model, $z' \in \{0,1\}^M$ is the (simplified) coalition vector, M is the maximum coalition size, and $\emptyset_j \in \mathbb{R}$ is the Shapley value. The value of $g(z')$ is calculated for all observations, so the size of the Feature Importance is the sum of all observations:

$$I_J = \sum_{i=1}^{n} \left|\emptyset_j^{(i)}\right| \quad (3)$$

One type of SHAP is treeSHAP which is used for machine learning based on classification trees, such as random forest.

MATERIALS AND METHODS

Source of Data

The data from the National Socio-Economic Survey (Susenas) Kor 2020 (March) of West Java Province, including 24,769 sample households. The level of food insecurity is a response variable in this study. These predictors are presented in Table 2 below:

TABLE 2. Research predictors

Features Name	Measurement Scale	Features Name	Measurement Scale
Education of Household Head	Ordinal	Grantee of Health Insurance National Program	Nominal
Vulnerable Household Head	Nominal	Grantee of Health Insurance Local Program	Nominal
Number of Family Members Having Saving Account	Ratio	Grantee of Scholarship Social Program	Nominal
Number of Family Members Illiterate	Ratio	Roof Types	Nominal
Main Income From the Transferee	Nominal	Floor Types	Nominal
Ownership of Land	Nominal	Wall Types	Nominal
Internet Access	Nominal	House Size	Ratio
Access to Outpatient Treatment	Nominal	Electricity	Nominal
Grantee of Non Cash Social Assistance	Nominal	Types of Cooking Fuel	Nominal
Grantee of Hopeful Family Program	Nominal	Drinking Water Source	Nominal
Grantee of Prosperous Family Program	Nominal	Decent Drinking Water	Nominal
Grantee of Social Assistance From Local Goverment	Nominal	Decent Sanitation	Nominal

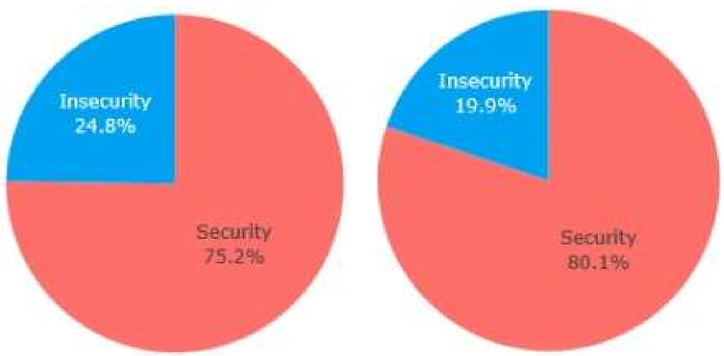
Procedure of Analysis

There are three step of analysis carried out in this study. First step is the data pre-processing, then the model formation, and the model interpretation. The analysis and processing data is using Python software.

1. Data Pre-processing
   a. Data preparation
      At this step, several things are carried out, namely: aggregation of the individual to household-level data, discarding observations containing missing values, "Not Answering" codes and "Don't Know"codes for the eight FIES questions, and categorization of food insecurity consisting of "Vulnerable" and "not".
   b. Data exploration
2. Model formation
   a. Complete data
      First, dividing training data (70%) and testing data (30%), then balancing training data with SMOTE technique, and doing modeling using random forest algorithm.
   b. Separate data
      For each area, divide the training data (70%) and test data (30%), then balance the training data using the SMOTE technique and modeling using the random forest algorithm.
3. Model Interpretation
   Interpretation for the separated data using Permutation Feature Importance (PFI), while SHAP is applied to complete and separate data.

RESULT AND DISCUSSION

Of the 25091 sample households in the March 2020 SUSENAS of West Java province, as many as 322 households stated they did not know (code 8) or refused to answer (code 9) at least one of the 8 FIES questions so that only 24769 households will be analyzed in this study.



(a) (b)

FIGURE 1. (a) Comparison of response variable classes in rural areas. (b) Comparison of response variable classes in urban areas

The comparison of food security and food insecurity households based on area status is shown in Fig. 1(a) and 1(b). In rural areas, the number of food security is 6631 households (75.2%), while food insecurity is 2182 households (24.8%). In urban areas, number of food security is 12787 households (80.1%), while food insecurity is 3169 households (19.9%). The random forest algorithm is used in classification modeling by carrying out a 10-fold cross-validation process to obtain stable model performance by looking at the optimal hyperparameter values. The hyperparameters used are ntree (the number of trees formed), mtry (the number of predictor), and the criterion (gini or entropy). The optimal hyperparameters are shown in Table 3.

TABLE 3. Optimal hyperparameters

Hyperparameter	Models		
	Rural	Urban	Complete
ntree	152	181	265
mtry	5	5	5
criterion	gini	gini	entropy

The comparison of the goodness of the random forest algorithm for the three models is shown in Table 4. The metrics we use to evaluate the model are AUC, accuracy, and F1-Score. Accuracy is used to determine the percentage of households correctly predicted to be food insecure and not food insecure from all households. In addition, accuracy is used because the dataset has a very close (symmetric) number of false negative and false positive data. The F1-score is used to find the optimal combination of precision and recall. The data in this study were unbalanced and focused on sensitivity and specificity, so the AUC was appropriate to evaluate the model.

TABLE 4. Model goodness measure

Models	AUC	Accuracy	F1-Score
Rural	83.3%	74.87%	75.5%
Urban	87.68%	79.40%	80.1%
Complete	85.71%	77.69%	78.0%

Feature Importance for Rural Area

Figure 2 represent feature importance for rural areas using PFI. Based on the PFI method, the predictors that characterize food insecurity in rural areas are house size, education of household head, number of family members having a saving account, decent sanitation, and floor types. The feature above is the most influential, and the lower the influence decreases.

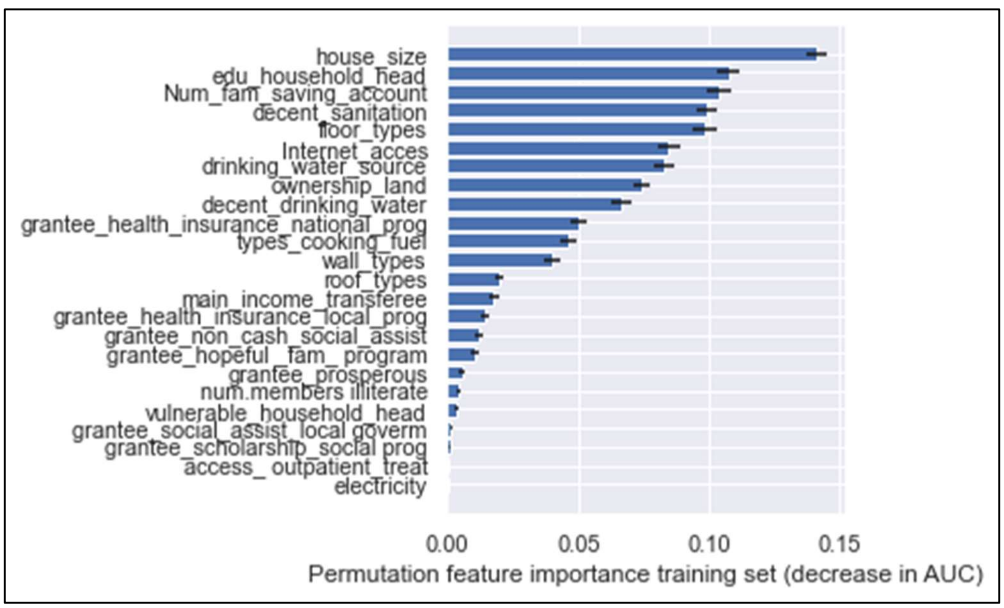
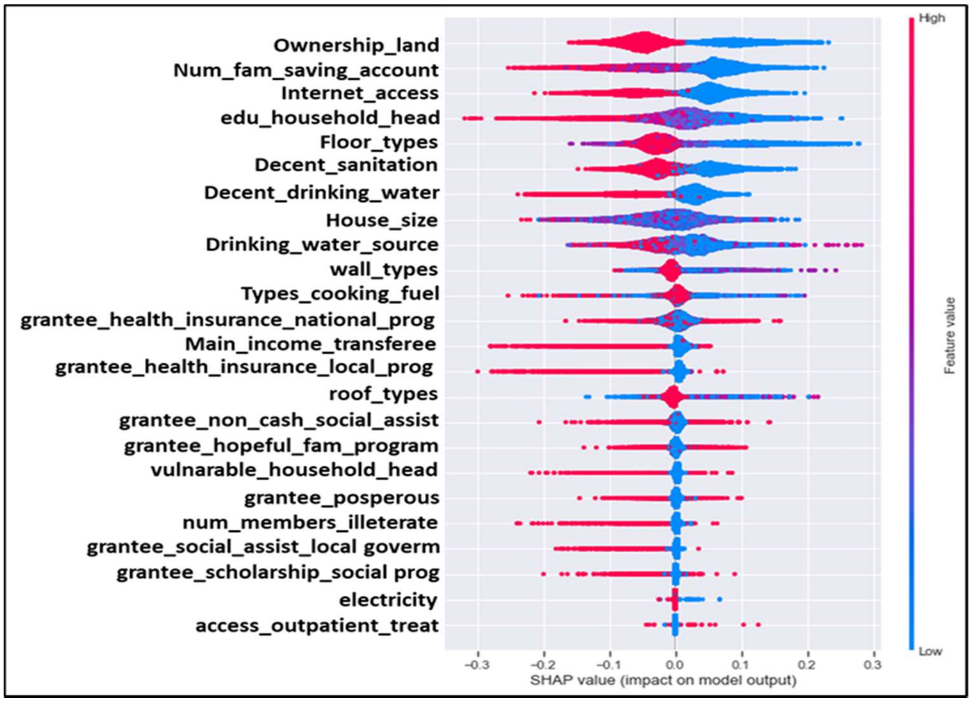
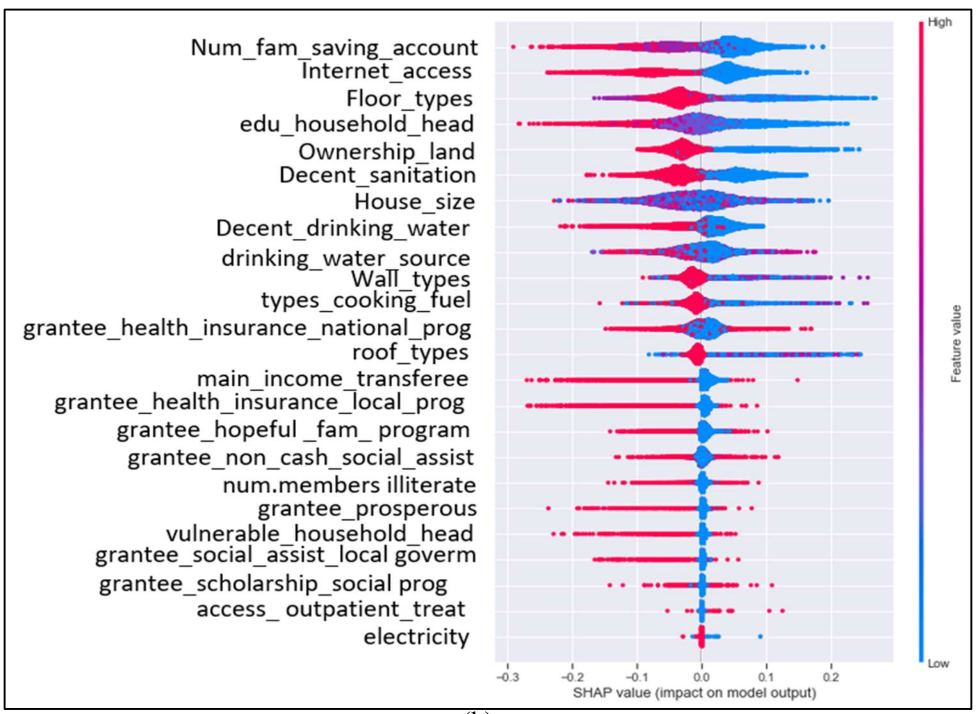


FIGURE 2. Feature importance using PFI in rural areas



(a)



(b)

FIGURE 3. (a) SHAP summary plot for complete data in rural areas, (b) SHAP summary plot for separate data in rural areas

Figure 3 represents the importance of using SHAP for complete and separate data. It can be seen that for these two approaches, the top five features are the same, namely ownership of land, the number of family members having a saving account, internet access, education of household head, and floor types. The x-axis is the SHAP value, and the y-axis is the variables listed in rank order. The predictor with the top order means that it has the most significant contribution to the prediction. The blue color indicates a low variable value, the closer to red the higher the variable value. The dots on the SHAP summary plot represent households (individuals).

The first feature importance in complete data is ownership of land. If the household owns land assets, then the value of Shapley gets smaller. It means that the probability of households being categorized as food insecure is getting smaller. On the other hand, if the household does not own land assets, the value of Shapley will increase. It means that the possibility of households entering the food insecurity category is higher.

The first feature importance in separate data is the number of family members having a saving account. In this predictor, if more household members have savings accounts, the Shapley value will be lower. It means that the more household members who have savings accounts, the lower the probability of a household being categorized as food insecure. On the other hand, if fewer household members have savings accounts, then the value of Shapley will be higher. It means that the probability of a household being categorized as food insecure is higher. The next most feature important is internet access. If the value of the internet access for each household is higher, then the Shapley value will be smaller. It means that the probability of a household being categorized as food insecure is getting smaller. Conversely, if the value on the internet access for each household is lower, then the Shapley value will be greater. It means that the probability of a household being categorized as food insecurity is higher.

The interesting in here is the grantee of health insurance national program. If households receive assistance from this program, then Shapley's value will be even greater. It means that the opportunity for households to be categorized as food insecure is getting bigger. On the other hand, if households do not receive assistance from this program, then Shapley's value will decrease. This means that the opportunity for a household to enter the food insecurity category is getting smaller. It becomes two opposite sides.

Feature Importance for Urban Area

Figure 4 represent feature importance for urban areas using PFI. Based on the PFI method, the features that characterize food insecurity in urban areas are house size, education of household head, number of family members having a saving account, ownership of land, and drinking water source.

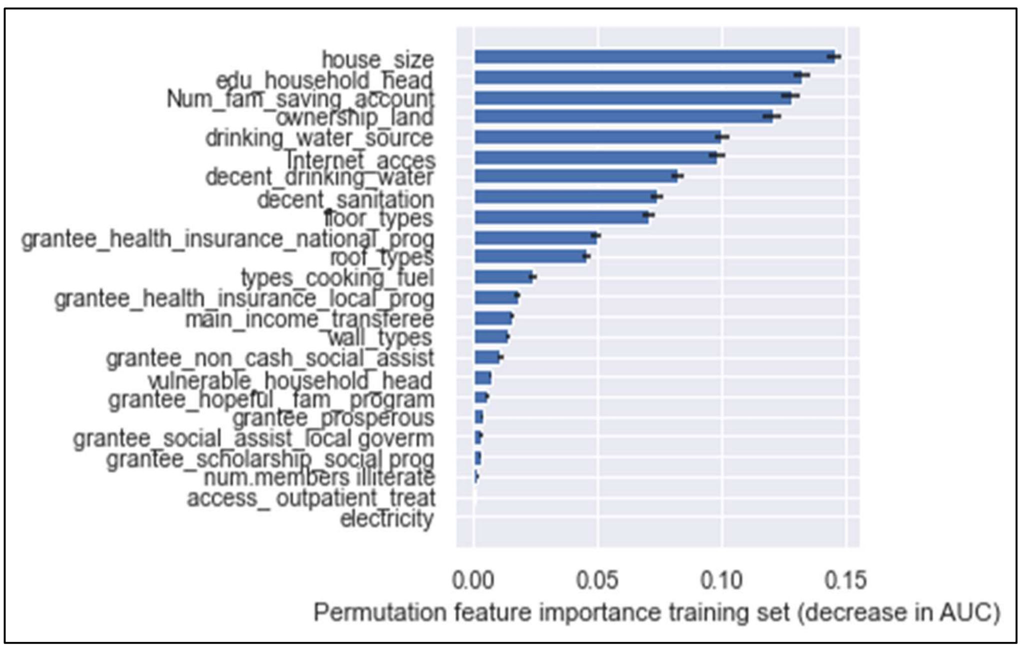
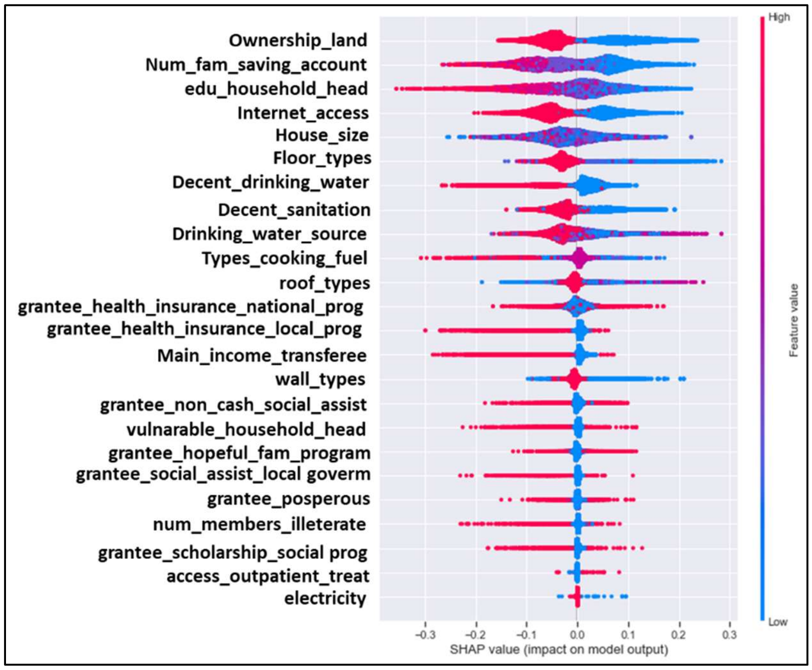
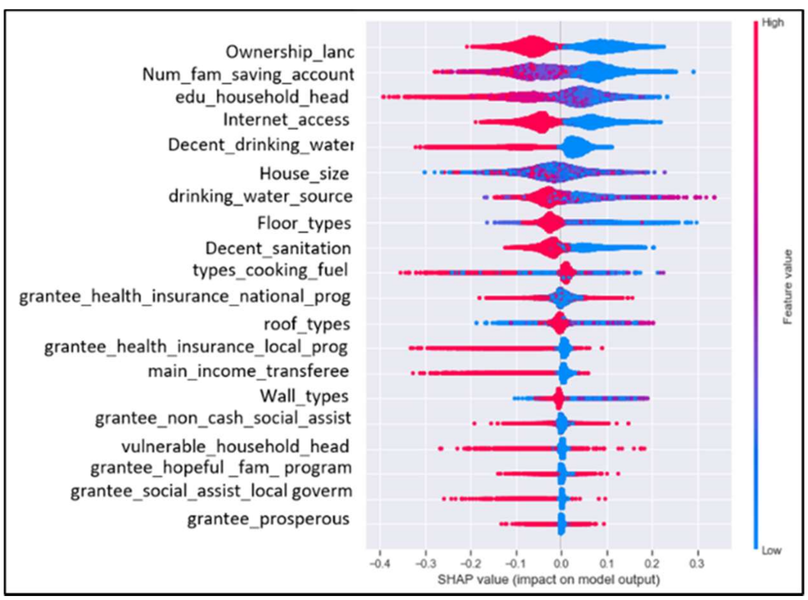


FIGURE 4. Feature importance using PFI urban areas



(a)



(b)

FIGURE 5. (a) SHAP summary plot for complete data in urban areas, (b) SHAP summary plot for separate data in urban areas

Figure 5 represents the importance of using SHAP for complete and separate data. It can be seen that for these two approaches, the top four features are the same, namely ownership of land, number family members having saving account, education of household head, and internet access. Based on the SHAP summary plot, the first feature importance is ownership of land. If the household owns land assets, then the value of Shapley gets smaller. It means that the probability of households being categorized as food insecure is getting smaller. On the other hand, if the household does not own land assets, the value of Shapley will increase. It means that the possibility of households entering the food insecurity category is higher. The next most feature importance for separate data in urban is the number of family members having a saving account. In this predictor, if more household members have savings accounts, the Shapley value will be lower. It means that the more household members who have savings accounts, the lower the probability of a household being categorized as food insecure. On the other hand, if fewer household members have savings accounts, then the value of Shapley will be higher. It means that the probability of a household being categorized as food insecure is higher.

The next most feature importance is education level of the household head. The higher the education of the head of the household, the lower Shapley's score will be. It means that the higher the education of the head of the household, the lower the probability of a household being categorized as food insecure. On the other hand, the lower the education level of the household head, the higher Shapley's score will be. This means that the possibility of households entering the food insecurity category is higher.

CONCLUSION

In general, the two approaches using SHAP resulted in a quite similar feature importance levels. Meanwhile, SHAP and PFI are relatively different. In rural areas, the features of the education of household head, number of family members having saving account, and floor types are features importance in the three models in the top five. In urban areas, the features of the education of household head, number of family members having a saving account, and ownership of land feature importance in the three models in the top five. The computational process takes a long time, especially for the SHAP method. We recommend using complete data in analyzing the level of features importance. This study only uses the random forest algorithm, which is an ensemble tree method. Therefore, it is necessary to compare with other methods that are not tree-based. The SHAP method can be applied to models not based on a classification tree using the KernelSHAP.

ACKNOWLEDGEMENT

Research reported in this publication was supported by the Ministry of Research and Techonogy/National Agency of Research and Innovation - Republic of Indonesia under Award Number 1/E1/KP.PTNBH/2021. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Agency of Research and Innovation.

REFERENCES

1. J. Boelaert and É. Ollion. *Revue Française de Sociologie.* 59(3), 475-506 (2018).
2. S. Mullainathan and J. Spiess. *Journal of Economic Perspectives.* 31(2), 87-106 (May 2017).
3. J. Breiman. *Random forests, Machine learning.* 45(1), 5-32 (October 2001).
4. E. Strumbelj and I. Kononenko, *The Journal of Machine Learning Research.* 11, 1-8 (March 2010).
5. Food and Agriculture Organization of The United Nation, [IFAD] International Fund for Agricultural Development, [WFP] World Food Programme. The state of food insecurity in the world 2013: The multiple dimensions of food security. (FAO, Roma, 2013).
6. M. D. Smith, M. P. Rabbitt, and A. Coleman-Jensen. *World Development.* 93, 402-12 (May 2017).
7. D. Magaña-Lemus, A. Ishdorj, C. P. Rosson, and J. Lara-Álvarez. *Agricultural and Food Economics.* 4(1), 1-20 (December 2016).
8. Food and Agriculture Organization The United Nation. The Food Insecurity Experience Scale: Measuring food insecurity through people’s experiences. (FAO, Roma, 2017).
9. L. Breiman and A. Cutler. Manual–Setting Up. Using. And Understanding Random Forests V4.0 [Internet], downloaded on 13 May 2021, available at https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf
10. M. Kuhn and K. Johnson. Applied Predictive Modeling. (Springer, New York, 2013).
11. A. Fisher, C. Rudin, and J. Dominici. *J. Mach. Learn. Res.* 20(177), 1-81(December 2019).
12. S. M. Lundberg, and S. I. Lee. “*A Unified Approach to Interpreting Model Predictions*”, in *Proceedings of the 31st International Conference on Neural Information Processing Systems.* (2017). 4768-4777.