TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 4, August 2020, pp. 1866~1873
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i4.14182

SWANSTAT: A user-friendly web application for data analysis using shinydashboard package in R

Rizal Bakri[1], Bagus Sartono[2], Hazan Azhari Zainuddin[3], Laode Ahmad Sabil[4]
[1]Department of Accounting, STIEM Bongaya, Indonesia
[2]Department of Statistics, IPB University, Indonesia
[1,3,4]Swanstatistics, Indonesia

Article Info

Article history:

Received Sep 23, 2019
Revised Mar 16, 2020
Accepted Apr 13, 2020

Keywords:

SWANSTAT
R programming language
Shiny server
Data science tools
Shinydashboard package

ABSTRACT

SWANSTAT is a user-friendly web application and free license that developed from the R programming language using shinydashboard Package. This research aims to create SWANSTAT was to streamline the routine workflow of data analysis so that users who are unfamiliar with R can perform the analysis interactively in a web browser with a cloud computing using a shiny server. The software development method used in this study is the SDLC with the waterfall model. The result of this research is the SWANSTAT software was developed using R by combining various packages and can be accessed online through various types of browsers on http://apps.swanstatistics.com. Besides, SWANSTAT consist of various features including the best visualization, the basis of statistical methods, help documents and tutorials. This research will continuously develop this application by enriching the latest statistical method, as well as improving the quality of features for data science needs.

Corresponding Author:

Rizal Bakri,
Department of Accounting,
STIEM Bangaya,
Makassar, Indonesia
Email: rizal.bakri@stiem-bongaya.ac.id

1. INTRODUCTION

Technological developments are more and more growing which is marked by the birth of variously sophisticated software as aids to facilitate people in doing their works. Nowadays, both commercial and open-source statistical software has been developed to analyze data that is too complex even though SPSS [1], JASP [2], SAS [3], MINITAB [4], STATA [5], and other software. However, some software is not yet free licenses and still requires user software installation on PC. The software which has been becoming popular and used by more statisticians and data scientist across nations is R that was first discovered by Ross Ihaka and Robert Gentleman in 1992 from the Statistics Department, University of Auckland, New Zealand [6]. The R software is open-source software that collects thousands of packages made by many contributors from over the world and has a high quality in processing large data. The users of R may utilize R studio and shinydashboard package to develop an application easily and interactively which may assist data analysts to handle their analytical problems without a need to build program script by themselves [7]. Based on the authors experience in conducting analytics training, many people face a big challenge in writing R programming due to their lack of knowledge in the language and programming structure. It is easier for them to use a click-and-drag application and not worrying about how the application handles the computation. However, it is not easy for

them to find a user-friendly software that is free (open source) and consists of various kinds of statistical analyses.

This research aims to design a more user-friendly application, free license and open source by utilizing the sophistication of R software, namely creating statistical tool interfaces from various existing packages that compiled in a dashboard using shinydashboard package. Then the developed application can be accessed online using a highly capable server with the help of a shiny server so users no longer need to install on PC. The software developed is named SWANSTAT. Besides, the SWANSTAT application is also equipped with a help document e.g. white papers and video tutorials to facilitate users' understanding of the utilization.

2. RESEARCH METHOD

2.1. Statistics tools

Statistics are now a popular topic in data science. Technological advances create large amounts of data and require complex analysis. At present, there are many analytical tools available that can be used to process data such as SPSS, SAS, JASP, STATA, MINITAB, R Studio, and others. Each of these applications has advantages and disadvantages. But the purpose of creating statistics tools is simply to make it easier for users to analyze complex data quickly.

2.2. Cloud computing

Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models [8].

2.3. Software development method

SWANSTAT software is developed by using the system development life cycle (SDLC) waterfall model with the following steps [9]:

- Requirement analysis

The first step is the requirements analysis stage which includes defining the problem, determining the purpose of software development, preparing all the materials needed for software development, and identifying the constraints that occur after the software is developed.

- Design

The second step is designing the SWANSTAT software design which includes dashboard design, data processing design, data analysis design, and output design analysis.

- Implementation

The third step is to compile the programming code to create a software dashboard, data analysis, and the results of the analysis. After that SWANSTAT software is installed on the cloud server so that it can be accessed online.

- Testing

The next step is to test the application. This stage is carried out to see the results of the analysis issued by the SWANSTAT software accessed online and evaluated by comparing the results with the results of the analysis in the R software (non-interface).

- Deployment/verification

The final stage is the application and verification of software. This stage is the application and verification by users or students related to the suitability of the needs with the development of application features that have been made.

3. RESULTS AND ANALYSIS

3.1. How to acces this software?

SWANSTAT has been developed to facilitate users who are not familiar to build program script and install the application on the desktop started in 2016. For a simplification reason, this application does not need to be installed. This application has been stored in a highly capable computing service that can be accessed online using a shiny server [10]. SWANSTAT can be accessed using many types of browsers on page http://apps.swanstatistics.com/swanstat. Figure 1 depicts how the SWANSTAT user interfaces look like after its URL successfully accessed.

3.2. Preparation and import data

Data preparation is an important stage before processing data using this application. The types of files supported by this application are that the ones with CSV extension (comma delimited) from Microsoft Excel.

For the next development, this application will be possible to be imported into various types of files e.g. SAS, MINITAB, SPSS, XLSX, and SQL. The initial step in using this application is to the data import by selecting the import data menu then searching for data using the browse button. After that, parameters are determined according to the data structures such as headers and separators. Then, upload data is selected on the select data active drop-down menu. Besides, SWANSTAT has also prepared data taken from the various package which are displayed on the menu as an exercise. Figure 2 shows the page for data import stages.

3.3. Example and execute

After the data is imported, the next step is to choose the right method based on the purpose of the analysis. To illustrate, in using iris to analyze the correlation between two quantitative variables, the first step is to choose the correlation menu then select the bivariate submenu. After that, all related quantitative variables are entered into the Select Multiple Variable boxes. In this research, we used the person correlation method. Visualization of the correlation analysis is displayed in the plot options tab menu. For the correlation analysis, the main package of SWANSTAT is corrplot [11]. Figure 3 shows the stages of correlation analysis using SWANSTAT. To run the application, the RUN button is selected then the analysis output is presented in the Summary and Plot tab menu as shown in Figure 4.

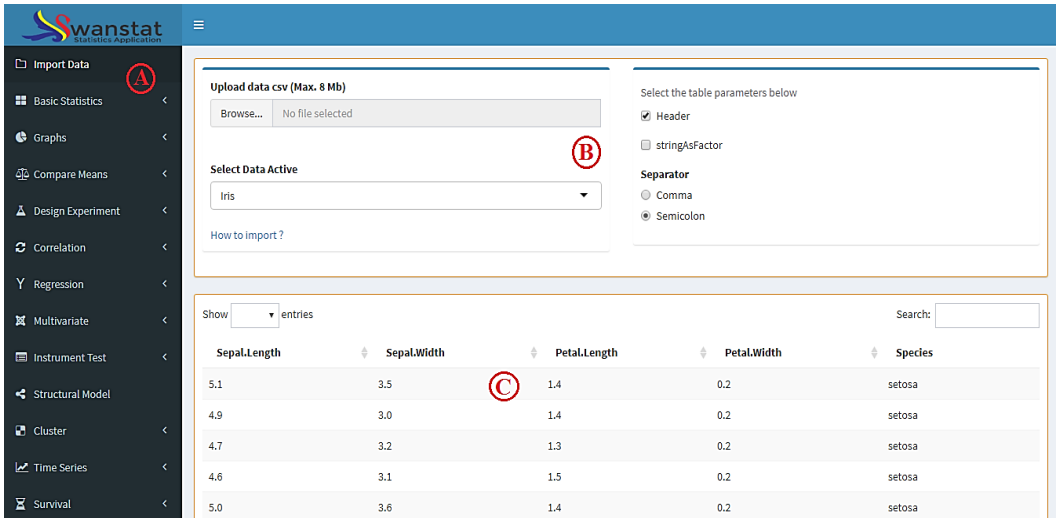

Figure 1. A screenshot of the SWANSTAT user interface;
(A) menus fir analyses, (B) upload or select data-driven, (C) data display

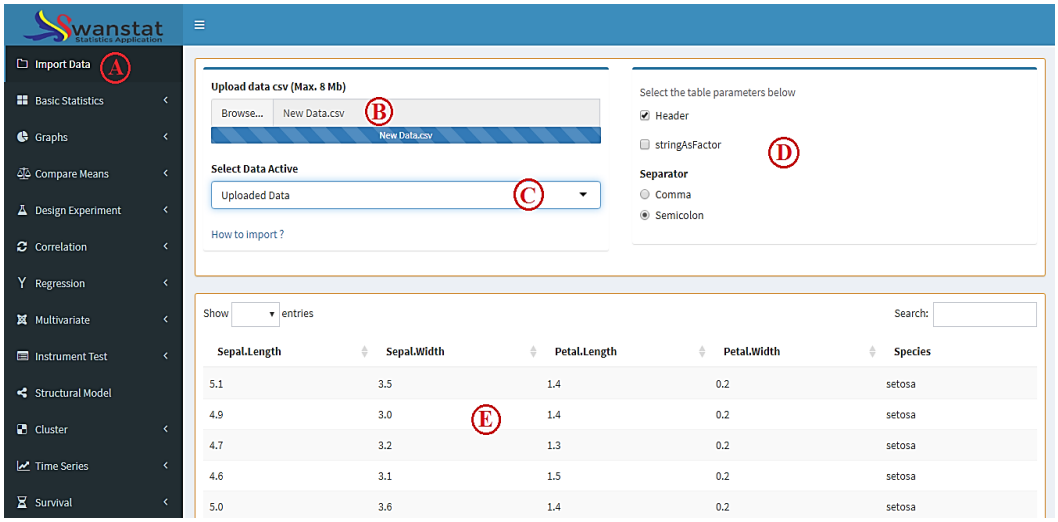

Figure 2. Upload data; (A) import data menu, (B) browse button,
(C) select data active menu, (D) options of data table, (E) data display

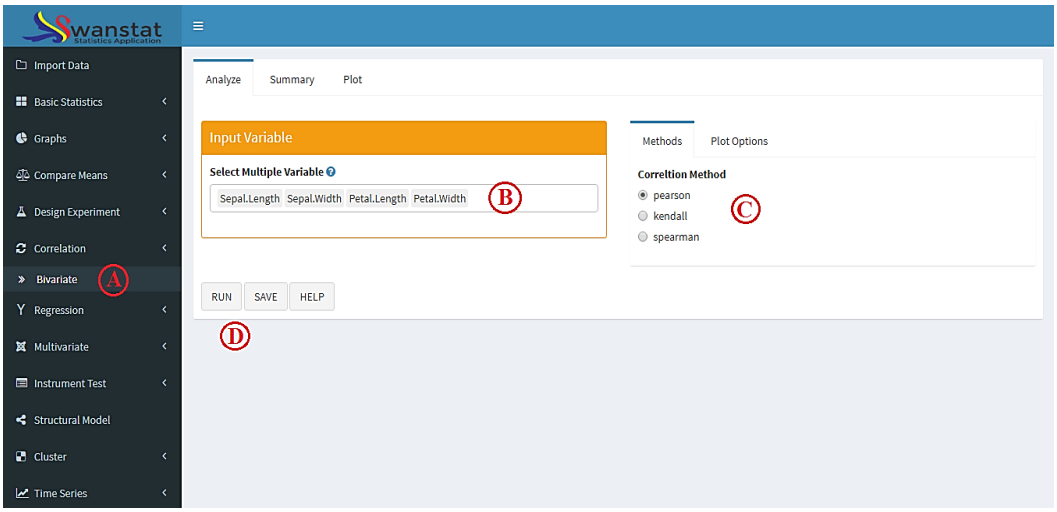

Figure 3. A classical correlation analysis in SWANSTAT; (A) menu for bivariate correlation analyses, (B) list of available variables that can be assigned to the analysis, (C) Correlation methods and additional plot options, (D) execute button for analysis, save output, and document help

Figure 4. The output of correlation analysis

3.4. Features of SWANSTAT

Each application certainly has its own advantages. By combining various packages inside the R, SWANSTAT is equipped with various features provided to facilitate users in operating and analyzing data. The features provided by this application include visualization, more statistical methods, help document, and save output report.

3.4.1. Visualization

Interesting visualization can help users to easily provide an interpretation of the analysis. This application is equipped with a visualization that combines various packages of Plots in R including graphics [6], ggplot2 [12], and plotly [13] package. Figures 5 and 6 shows some graphic forms available in different statistical tools in SWANSTAT.

3.4.2. Document help

Document help is designed to make users easily using this application. This document is provided in the form of articles and video tutorials. To access the document, SWANSTAT has prepared the HELP button at each analysis menu. When pressing the button, this application will be directed to the URL http://swanstatistics.com/category/swanstat/. Meanwhile, the tutorial video about the step by step of the use of this application can be accessed through the SWANSTAT channel on Youtube.

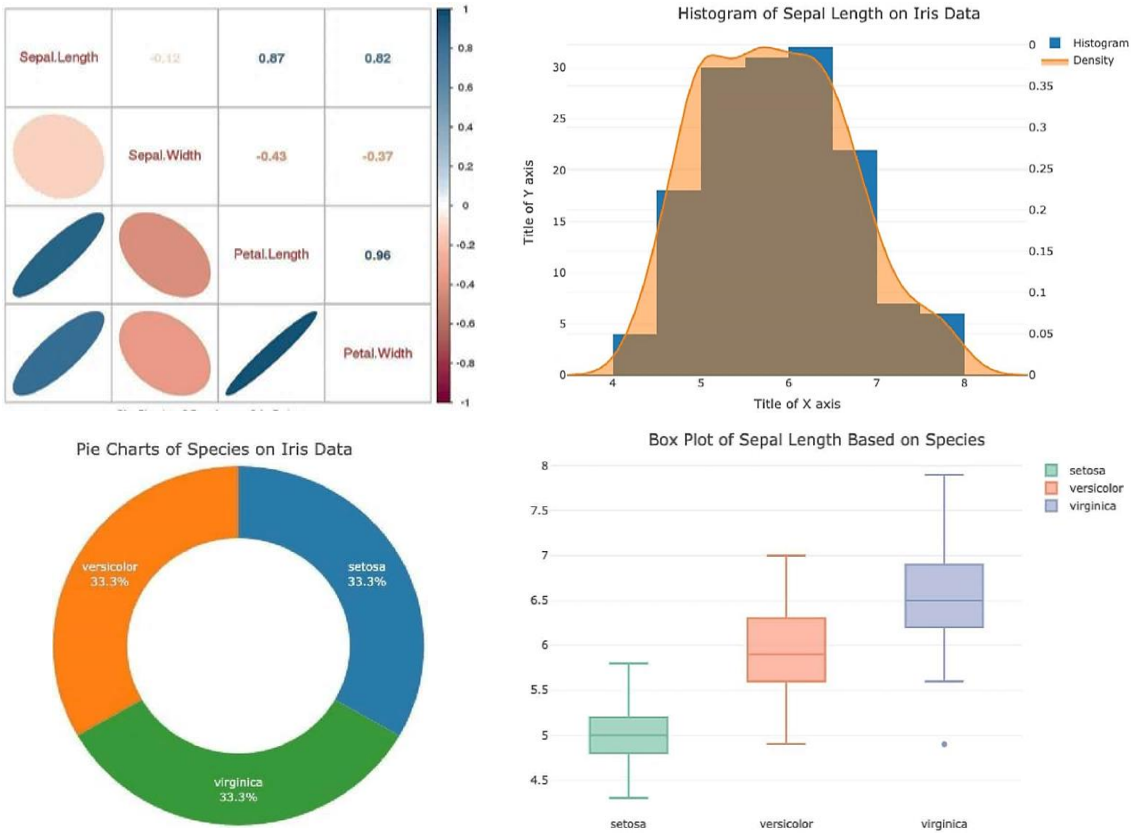

Figure 5. Correlation plot, Histogram, pie charts, and Box plot of SWANSTAT

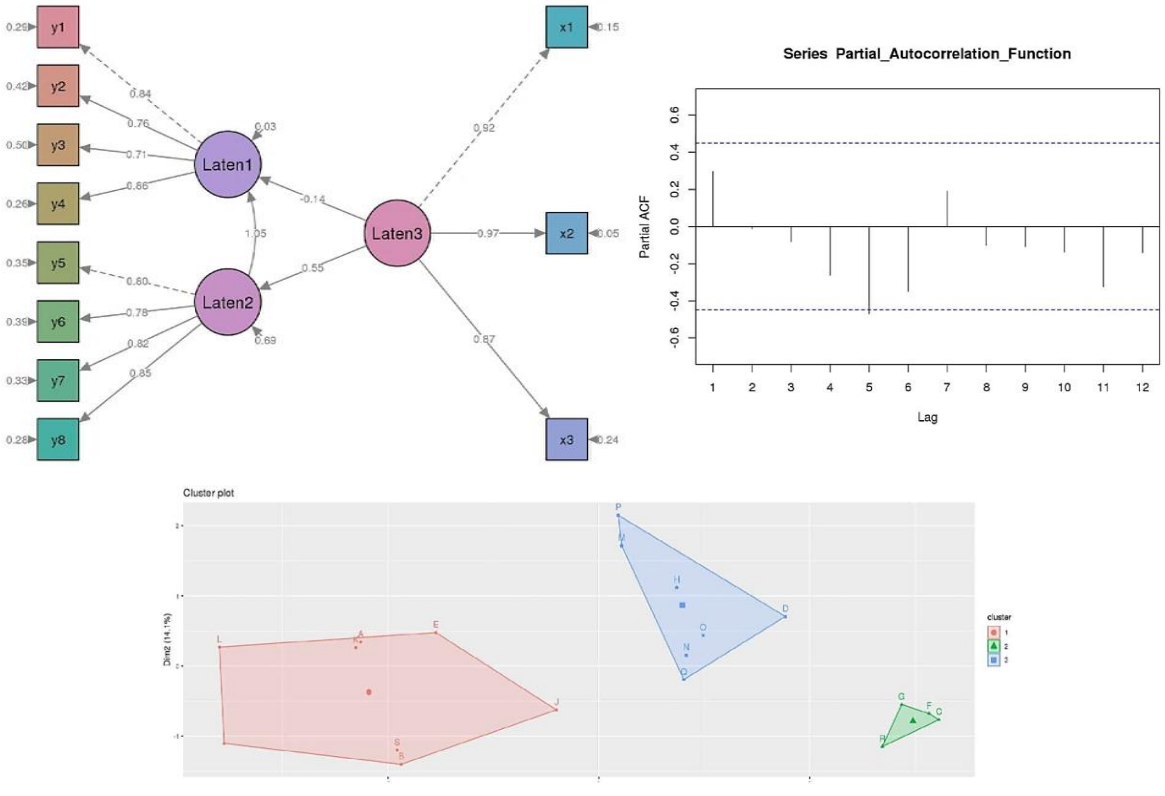

Figure 6. SEM plot, Cluster plot, and Time series plot of SWANSTAT

3.4.3. Save output report

Another important and helpful feature of this application is that the output from the analysis can be saved. The formats of document output are HTML and word with rmarkdown package [14]. The output of analysis using the SWANSTAT application is the same as the one produced by R Studio which is designed in Table 1. The application can be simply done by clicking the SAVE button after analyzing the data. Figure 7 shows the display box to save the output of the analysis. Select the document format then download it.

Figure 7. Save output box dialog

3.4.4. Available Analyses and R Packages used

The advancement of statistical methods leads to the development of this application to continue. The number of statistical tools in this application is continuously added based on the needs and scope of the analytical methods in R. SWANSTAT has been equipped with various statistical methods from some basic packages commonly used. An up-to-date list of the R packages used and their version is maintained on the project website http://apps.swanstatistics.com/r-package-list.html.

Table 1. Available analyses in SWANSTAT and R packages used

Function in SWANSTAT	R Packages	Function in SWANSTAT	R Packages
Development Components	shiny [15]	Multivariate	MVN [25]
	shinydashboard [7]		factoextra [26]
	rmarkdown [14]		FactoMineR [27]
	pander [16]	Instrument Test	psy [28]
	MASS [17]	Structural Model	lavaan [29]
Basics Statistics	stats [6]		semPLot [30]
	fBasics [18]		semTools [31]
	nortest [19]		semMediation [32]
Graphs	graphics [6]	Cluster Analysis	cluster [33]
	ggplot2 [12]		e1071 [34], gclus [35]
	plotly [13]		dendextend [36]
Compare Means	webr [20]	Time Series	forecast [37]
Design of Experiment	stats [6]		tseries [38]
Correlation	corrplot [11]	Survival Analysis	survival [39], survminer [40]
Regression	lmtest [21]		
	car [22], caret [23]		
	nnet [24]		

4. DISCUSSION AND FEATURE WORK

This research is the beginning of the application development plan. We have tried to develop statistical applications that can be accessed online. The features we have created include import data with a CSV file extension, visualization of various package graphics, some basic statistical methods, document help in the form of articles and video tutorials, etc. future work will develop more statistical methods with more interesting features. We, the development team, are opening opportunities for programmers and statisticians to contribute to the development of this software.

ACKNOWLEDGMENTS

The idea to develop this application is initiated from the public‘s anxiety over the use of coding-based software. The team expressed the gratitude thanks to the Lecturers of Department of Statistics of IPB University for supporting the development of this application, especially to Mr. Prof. Dr. Asep Saefuddin, Mr. Prof. Dr. Khairil Anwar Notodiputro, Prof. Dr. Hari Wijayanto, M.Si., Mr. Dr. Anang Kurnia, and Mr. Dr. Farit Mochamad Afendi who have acted as the advisors. We also say thanks to all the contributors of packages used in developing this application.

REFERENCES

[1] Field A., "Discovering Statistics using IBM SPSS statistics," London: SAGE, 2013.
[2] Love J., Selker R., Marsman M., Jamil T., Dropmann D., Verhagen J., Ly A., Gronau Q. F., Smira M., Epskamp S., Matzke D., Wild A., Knigth P., Rouder J. N., Morey R. D., Wagenmakes E. J., "JASP: Graphical Statistical Software for Common Statistical Design," *Journal of Statistical Software*, vol. 88, no. 2, pp. 1-17, 2019.
[3] SAS Institut, "SAS/STAT user's guide version 6," *Institute Incorporated*, 1990.
[4] Meyer, Ruth, Krueger D., "Minitab guide to statistics," *Prentice Hall PTR*, 2001
[5] Hamilton, Lawrence, C., "Statistics with Stata v 12," *Cengage Learning*, 2012.
[6] R. Core Team, "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing*, 2013.
[7] Chang, Winston, Barbara B. R., "Shinydashboard: create dashboards with'Shiny," *R package version 0.7. 1*, 2018. [Online]. Available: https://cran.r-project.org/web/packages/shinydashboard/shinydashboard.pdf. Accessed: March, 2019.
[8] Mell, Peter, Grance T. "The NIST definition of cloud computing," *NIST*, 2011.
[9] Bakri R., Data U., Astuti N. P., "Auto Sales Forecasting Software Based on Computational Intelligence Website to Optimize Product Marketing Strategy Management (in Bahasa: Aplikasi Auto Sales Forecasting Berbasis Computational Intelligence Website untuk Mengoptimalisasi Manajemen Strategi Pemasaran Produk)," *Jurnal Sistem Informasi Bisnis*, vol. 9 no. 2, pp. 244-251, 2019.
[10] R. Studio Team, "Shiny Server Open Source v1.4.0 Administrator's Guide," *R Studio, Northern Ave, Boston*, 2016.
[11] Wei T., Simko V., "An Introduction to corrplot Package," *R package version 0.84*, 2017, Available: https://cran.r-project.org/web/packages/corrplot/corrplot.pdf. Accessed: March, 2019.
[12] Wickham, Hadley, Winston C., "ggplot2: Create elegant data visualisations using the grammar of graphics," *R package version 2.1, 2016*, Available: https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf. Accessed: March, 2019.
[13] Sievert, Carson, et al., "plotly: Create Interactive Web Graphics via 'plotly. js'," *R package version 4.1*, 2017, Available: https://cran.r-project.org/web/packages/plotly/plotly.pdf. Accessed: April, 2019.
[14] Xie, Yihui, Joseph J., Allaire, Garrett G., "R markdown: The definitive guide," *Chapman and Hall/CRC*, 2018.
[15] Chang, Winston, et al., "Shiny: web application framework for R," *R package version 1.5*, 2017. [Online]. Available: https://cran.r-project.org/web/packages/shiny/shiny.pdf. Accessed: April, 2019.
[16] Daróczi, Gergely, Roman T., "Pander: An R Pandoc Writer," *R package version 0.3 7*, 2013. [Online]. Available: https://cran.r-project.org/web/packages/pander/pander.pdf. Accessed: April, 2019.
[17] Ripley, Brian, "MASS: support functions and Datasets for Venables and Ripley's MASS," *R package version 7.3-23*, 2011. [Online]. Available: https://cran.r-project.org/web/packages/MASS/MASS.pdf. Accessed: April, 2019.
[18] Wuertz, Diethelm, T. Setz, Y. Chalabi, "Fbasics: Rmetrics-markets and basic statistics," *R package version 3042.89*, 2017. [Online]. Available: https://cran.r-project.org/web/packages/fBasics/fBasics.pdf. Accessed: April, 2019.
[19] Gross, Juergen, Uwe L., "Nortest: Five omnibus tests for testing the composite hypothesis of normality*," R package version 1.0-4*, 2015. [Online]. Available: https://cran.r-project.org/web/packages/nortest/nortest.pdf. Accessed: April, 2019.
[20] Moon K. W., "webr: Data and Functions for Web-Based Analysis," *R package version 0.1.0*, 2018. [Online]. Available: https://cran.r-project.org/web/packages/webr/webr.pdf. Accessed: May, 2019.
[21] Hothorn T., Zeileis A., Farebrother R. W., Cummins C., "lmtest: Testing Linear Regression Models," *R package version 0.9-36*, 2018. [Online]. Available: https://cran.r-project.org/web/packages/lmtest/lmtest.pdf. Accessed: May, 2019.
[22] Fox, John, Sanford W., "An R companion to applied regression," London: SAGE, 2018.
[23] Kuhn M., "Caret: Classification and Regression Training," *R package version 6.0-80*, 2017. [Online]. Available: https://cran.r-project.org/web/packages/caret/caret.pdf. Accessed: May, 2019.
[24] Ripley B., "nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models," *R package version 7.3-12*, 2016. [Online]. Available: https://cran.r-project.org/web/packages/nnet/nnet.pdf. Accessed: May, 2019.
[25] Korkmaz S., Goksuluk D., Zararsiz G., "MVN: An R Package for Assessing Multivariate Normality," *The R Journal*, vol. 6, no 12, pp 1510–162, 2014.
[26] Kassambara A., Mundt F., "Factoextra: Extract and Visualize the Results of Multivariate Data Analyses," *R package version 1.0.5*, 2017. [Online]. Available: https://cran.r-project.org/web/packages/factoextra/factoextra.pdf. Accessed: May, 2019.
[27] Lê S., Josse J., Husson F., "FactoMineR: A Package for Multivariate Analysis," *Journal of Statistical Software*, vol. 25, no. 1, pp. 1–18, 2008.
[28] Falissard B., "psy: Various procedures used in psychometry," *R package version 1.1*, 2012. [Online]. Available: https://cran.r-project.org/web/packages/psy/psy.pdf. Accessed: May, 2019.
[29] Rosseel, Yves. "Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA)," *Journal of statistical software*, vol. 48, no. 2, pp 1-36, 2012.
[30] Epskamp, Sacha, Simon S., "semPlot: Path diagrams and visual analysis of various SEM packages' output," *R package version 1.1.1*, 2019. [Online]. Available: https://cran.r-project.org/web/packages/semPlot/semPlot.pdf. Accessed: May, 2019.
[31] Jorgensen D. T., Pornprasertmanit S., Schoemann M. A., Rosseel Y., "semTools: Useful tools for structural equation modeling," *R package version 0.5-1*, 2018. [Online]. Available: https://cran.r project.org/web/packages/semTools/semTools.pdf. Accessed: May, 2019.
[32] Moon K. W., "semMediation: Visualize the Mediation Effect for Structural Equation Modelling," *R package version 0.1.1*, 2017.

[33] Maechler M., Rousseeuw P., Struyf A., Hubert M., "cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.," *R package version 2.0.7-1*, 2018. [Online]. Available: https://cran.r-project.org/web/packages/cluster/cluster.pdf. Accessed: May, 2019.
[34] Meyer D., Dimitriadou E., Hornik K., Weingessel A., Leisch F., "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien," *R package version 1.7-1*, 2019. [Online]. Available: https://cran.r-project.org/web/packages/e1071/e1071.pdf. Accessed: May, 2019.
[35] Hurley C., "gclus: Clustering Graphics," *R package version 1.3.2*, 2019. [Online]. Available: https://cran.r-project.org/web/packages/gclus/gclus.pdf. Accessed: May, 2019.
[36] Galili T., "Dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering," *Bioinformatics*, vol. 31, no. 22, pp. 3718-3720, 2015.
[37] Hyndman, Rob J., Yeasmin K., "Automatic time series for forecasting: the forecast package for R," *No. 6/07 Clayton VIC, Australia: Monash University, Department of Econometrics and Business Statistics*, 2007.
[38] Trapletti, Adrian, Kurt H., B. LeBaron, "Tseries: time series analysis and computational finance," *R package version 0.10-11,* 2007. [Online]. Available: https://cran.r-project.org/web/packages/tseries/tseries.pdf. Accessed: May, 2019.
[39] Terry M. T., Patricia M. G., "Modeling Survival Data: Extending the Cox Model," New York: *Springer*, ISBN 0-387-98784-3, 2000.
[40] Kassambara, Alboukadel, Marcin K., P. Biecek, "Survminer: Drawing Survival Curves using'ggplot2'," *R package version 0.3 1*, 2017. [Online]. Available: https://cran.r-project.org/web/packages/survminer/survminer.pdf. Accessed: May, 2019.

BIOGRAPHIES OF AUTHORS

Rizal Bakri. Lecturer in the Department of Accounting at STIEM Bongaya Makassar. The mastery of statistical and computational theory is obtained through master education in the Department of Statistics of IPB University and a bachelor in the Departement of Mathematics of UNM Makassar. He has good experience in the field of information technology and computing by designing various statistical methods in the form of website based interface software with R. He is also active as a statistical and computing consultant in several industries and a trainer in various data science training using R, Python, and SAS software.

Bagus Sartono. Lecturer in the Department of Statistics at IPB University who has teaching experience in statistics and data science courses. Strengthening the theory obtained through education until Ph.D. in the Universitet Antwerp, Belgium. While his ability in the applied field is mostly gained from experience in various companies and organizations. He is also active as a trainer in various quantitative and analytical methods training, including the use of SAS, R, and Python software.

Hazan Azhari Zainuddin. Consultant and programmer at Swanstatistics corporation, Indonesia. The mastery of statistical and computational theory is obtained through master education in the Department of Statistics of IPB University and a bachelor in the Departement of Mathematics Education at UNM Makassar. He has good experience in the field of computing by designing various statistical methods in the form of website based interface software with R programming language. He is also active as a statistical consultant in several industries and a trainer in various data science training.

Laode Ahmad Sabil. Consultant and programmer at Swanstatistics corporation, Indonesia. The mastery of statistical and computational theory is obtained through master education in the Department of Statistics of IPB University and a bachelor in the Departement of Mathematics Education at UNM Makassar. He has good experience in the field of computing by designing various statistical methods in the form of website based interface software with R programming language. He is also active as a statistical consultant in several industries and a trainer in various data science training.