The Dialect Geography of Malay Language in Sambas Raya, West Kalimantan

Umar Fauzan*, Patriantoro**

* English Department, Teacher Training and Tarbiyah Faculty, Sultan Aji Muhammad Idris State Islamic University Samarinda
** Master of Indonesian Language Education Study Program, Department of Language and Arts Education
FKIP, Tanjungpura University

Abstract- The research has four purposes, they are: (1) describing the lexical variation of the Malay Language in Sambas Raya; (2) calculating the number of those variations; (3) mapping those variations and (4) creating the lexical isogloss file of the language. The research data were Swadesh and non-Swadesh words and phrases. There are two data analysis methods; quantitative-qualitative descriptive methods and the synchronic comparative method. The method is used to compare the lexical difference between observation points. A dialectometry formulation is used to calculate the lexical difference in the presentation between observation points. There are five results of this research. First, a description of language variation in 9 observation points in Sambas Raya. Second, different lexical numbers, with the lowest number between observation points in point 4 (Sungai Raya) and point 5 (West Singkawang) is 31, and the highest number between observation points in point 2 (Tebas) and point 9 (Seluas) is 69. Third, sub-dialect and different dialects of the language variation. Fourth, the linguistic distance between observation points with the lowest percentage of 31 % in point 4 (Sungai Raya) and point 5 (West Singkawang) and the highest percentage of 69 % in point 2 (Tebas) and point 9 (Seluas). From lexical variation mapping of the language, there are 9 sub-dialect variations, 7 different dialects, and no different language. Fifth, the isogloss line separating the highest language variation with isogloss file separating point 2 (Tebas) and point 9 (Seluas) is 69 files and the lowest language variation with isogloss file separating point 4 (Sungai Raya) and point 5 (West Singkawang) is 31 files.

Index Terms- dialect geography, dialect map, isogloss, lexical variation

I. INTRODUCTION

Many dialects in the world can be understood by using dialect geography. It is an effort by doing a dialect mapping that has been done by some experts in this field in some parts of the world language, they are: dialect geography of Modern Greek (Newton, 1972), dialect geography of Oman (Holes, 1989), dialect geography of Syria-Palestine (Garr, 2004), Nisu dialect geography (Yang, 2009), Kurdish dialect geography (Matras, 2019), and dialect geography of South Asia (Arora et al., 2021).

In the Indonesian context, there are some linguists who have conducted research on dialect geography, they are: dialect geography of Tombulu language (Lalamentik et al., 1986), dialect geography of Sundanese in Karawang (Yudibrata et al., 1990), dialect geography of Madurese in Probolonggo (Dewi, 2010), dialect geography of Saluan (Pamolango, 2012), dialect geography of Sundanese in Parungpanjang Bogor (Rahmawati, 2013), dialect geography of Javanese in Lamongan (Purwaningsih, 2017), dialect geography of Sulawesi Tenggara (Susiati & Iye, 2018), dialect geography of Javanese in Wonogiri (Antono et al., 2019), dialect geography of Javanese in Cilacap (Afidah & Mardikantoro, 2019), dialect geography of North Sumatra (Dalimunte & Lubis, 2020), dialect geography of Muna language (Taembo et al., 2021), dialect geography of Yogyakarta-Surakarta (Saddhono & Hartanto, 2021).

This particular research is done to answer the common question of Malay language speakers in West Kalimantan, who consider every area of the river stream or the district has a different dialect. This opinion was admitted by the Malay language speaker community as social dialect or sociolect. They argue that every area their own dialect. This argument is not always true. Thus, this research is done to prove whether the claim is true or false.

There are three similar types of research to this current study. The first research, "The Dialect Geography of Malay Language in Sambas and Mempawah River Stream Areas" (Patriantoro, 2015) describes the language (1) phonological variation, (2) lexical variation, (3) phonological mapping, (4) lexical mapping, and (5) lexical isogloss file of the language in the area. The second research, "The Dialect Geography of Malay Language in the Middle Stream of Kapuas River, in Sanggau and Sekadau Districts, West Kalimantan" (Patriantoro, 2017) exposes the language (1) lexical variation, (2) lexical mapping, and (3) lexical isogloss file. The third research, "The Dialect Geography of Malay Language in Kapuas Hulu District" (Patriantoro, 2019) reveals the language (1) lexical language variation, (2) lexical mapping, and (3) lexical isogloss file.

Based on the literature study on "The Dialect Geography of Malay Language in Sambas Raya, West Kalimantan", there is no similar research on the particular characteristic of Malay Language Dialect geography in Sambas Raya, West Kalimantan. The research on "The Dialect Geography of Malay Language in the Middle Stream of Kapuas River, in Sanggau and Sekadau

Districts, West Kalimantan" in 2017, exposes a diversity of the language variation in the form of different speech variations, sub-dialect, or dialect. It shows the diversity of the Malay Language in West Kalimantan, in which different locations resulted in different language variations.

The four purposes of the research entitled, "The Dialect Geography of Malay Language in Sambas Raya, West Kalimantan," are: (1) describing the lexical variation of the language, (2) calculating the number of different lexical variations of the language, (3) creating the lexical variation mapping of the language, and (4) making the lexical isogloss file of the language.

II. IDENTIFY, RESEARCH AND COLLECT IDEA

A method is a technique, procedure, or systematic sequencing steps which are done in research. There are four steps of the research procedure: (1) preparation, (2) data collection, (3) data analysis, and (4) research result reporting. The methods applied in this research are data collection and data analysis. The location of "The Dialect Geography of Malay Language in Sambas Raya, West Kalimantan" research covers 9 observation points on 9 districts in Sambas Raya: (1) Sambas, (2) Tebas, (3) Pemangkat, (4) Sungai Raya, (5) West Singkawang, (6) Bengkayang, (7) Ledo, (8) Tujuh Belas, and (9) Seluas.

This research applied two research kinds that are done subsequently to support the data analysis since some data were suitable to be analyzed within quantitative research while the other data were suitable to be analyzed within qualitative research. Quantitative research is research employing number calculation with a particular measurement tool. Measurement is an activity of giving a number to an attribute, characteristic, object, or event, based on a particular rule or formula. Measurement is also an activity to determine the number of particular categories to describe the quality of a particular result. Measurement in dialect geography uses dialectometry formulation.

The original data soured is the original Malay Language speaker who was born, live, and raised in the research area. In qualitative research, the researcher should be more selective in selecting informants to represent the qualified information. Thus, they were selected based on particular criteria and considerations. The research data were words and phrases in the Malay Language spoken by its speech community in the research area with determining gloss. They are Swadesh and non-Swadesh lexical amounting to 100 lexical and phrases. The modified instrument is used to collect the data (Laksono dan Savitri, 2009, pp.45-60).

This research applies Cakap Method to collect data since the data were conversations in which the researcher and speakers, as informants or data sources, were involved in the conversation. The research employs Pancing Technique to collect data as the researcher uses his expertise to obtain the data from informants under the research instrument guidance so that informants were speaking the way the researcher expects them to speak using the language. This technique is applied in the direct face-to-face conversation between the researcher and informants. It is used along with Deep Interview Technique to support the data collection. To obtain data, the instrument contains Swadesh words and other selected words. It is written in a recognizable language form called gloss. Gloss is a recognized form of the language used by the researcher, the gloss of the research instrument is the translation of its Indonesian (Laksono dan Savitri, 2009, p.23).

An informant or language assistant is the speaker who gives research material, a speaker who gives information, or the researcher assistant in the data collection process (Sudaryanto, 1988, pp.26-32). They were selected under particular criteria. Principally, both men and women are the same. However, based on the fact in the field research, there are instruments that men do not understand but women do such as kitchenware or seasoning. Thus, men and women are selected as informants to support the data collection. There are 3 informants for 1 observation point in this research. For 9 research locations, this research employs 27 informants.

This research uses the synchronic comparative method. Sudaryanto (1993) used the term Padan Method for Comparative Method. It is a procedure of data analysis by matching or 'comparing' one with another data (Sudaryanto, 1993, p.53). Crowley dan Browen. (2010) stated that the Comparative Method is a procedure of 'comparing two or more cognates, from two or more languages to obtain the proto-language form (p.88). This method can be used to reconstruct several aspects of the source language from the reflection of its mother language, but the reconstructed language data were only the languages with cognates families.

The method for analyzing language maps is the synchronic comparative method. Principally, it is a method that is used to analyze language data by comparing language data between the observation points at the same time. This research uses the synchronic comparative method with the non-cognate pair technique to find the lexical difference and the lexical synchronic comparative method to compare the data between observation points to create a lexical mapping.

The synchronic comparative method can also be used to analyze the lexical difference between observation points. It is followed by counting the amount of lexical difference between observation points using the Dialectometry formulation. Dialectometry is a statistics measurement showing the lexical difference of different observation sites by comparing collected elements from particular sites (Nadra dan Reniwati, 2009, p.91).

Related to the distance of language element percentage calculation, Laksono and Savitri (2009) stated that the calculation is similar to the calculation of language variation for the lexical field (pp.83-84). The observation points are best selected using inter-district triangles (see Mahsun, 1995, pp.119-120), (Laksono, 2004, pp.18-19), and (Laksono and Savitri, 2009, p.70). The dialectometry calculation is done based on inter-district triangles and multi-angle. The lexical difference calculation of the language data is compared between observation points which is done by calculating the lexical difference between observation points. The lexical difference result is calculated using the Dialectometry formulation.

Guiter Dialectometry Formulation (in Mahsun, 1995, p.118), Mahsun, 2010, pp.48-50).

$$(S \times 100)/N = D \%$$

S : the amount of lexical difference between observation points

N   : the amount of compared lexical map
D %   : percentage of the language elements distance between observation points
The result of the lexical percentage of the language elements distance between observation points are as follows.
81 % above   : considered as different language
51 % - 80 %   : considered as different dialect
31 % - 50 %   : considered as different sub-dialect
21 % - 30 %   : considered as different speech
below 20 %   : considered as no difference
Patriantoro (2015) shows lexical dialectometry calculation as follows.

III.   WRITE DOWN YOUR STUDIES AND FINDINGS

There are four research problems that is answered by the researcher through the research of "The Dialect Geography of Malay Language in Sambas Raya, West Kalimantan," they are: (1) the lexical variation of the language, (2) the calculation of different lexical variation between observation points, (3) the lexical variation mapping of the language, and (4) the lexical isogloss file of the language dialect.

The first result of this research is the description of lexical data to answer the research problem of lexical variation of the Malay Language in Sambas Raya, West Kalimantan. The second result of this research is the comparison of lexical data on each of the observation points using inter-district triangles to answer the research problem of different lexical variations between observation points in the research area. After that, the lexical variation difference between observation points is calculated to solve the research problem of language variation conditions in the research area using dialectometry formulation. The third result of this research is the calculation of language variation between observation points resulting in linguistic distance between observation points that are used to map lexical language variation in the research area. The fourth result of this research is the isogloss file between observation points in the research area. The more isogloss file separating observation points the more lexical difference it has.

These are the whole data analysis results. The description of lexical variation on the research area consists of 9 observation points located in 9 districts of Sambas Raya, West Kalimantan: (1) Sambas, (2) Tebas, (3) Pemangkat, (4) Sungai Raya, (5) West Singkawang, (6) Bengkayang, (7) Ledo, (8) Tujuh Belas, and (9) Seluas. The triangle exposure between observation points is done to calculate the percentage distance between observation points in the research area. There are 18 lexical distance between observation points covering observation point: 1 – 2, 1 – 9, 2 – 3, 2 – 4, 2 – 9, 3 – 4, 3 – 7, 3 – 8, 4 – 5, 4 – 6, 4 – 7, 4 – 8 : 53, 5 – 6, 6 – 7, 7 – 8, 7 – 9, and 8 – 9. The calculation of the lexical difference between observation points using inter-district triangle on 9 observation points is as follows.

Tabel 1. Lexical Difference between Observation Points

Observation Point	Lexical Difference
1 – 2	39
1 – 9	67
2 – 3	36
2 – 4	33

80,1 % above   : different speech
50,1 % - 80 %   : different dialect
30,1 % - 50 %   : different sub dialect
20,1 % - 30 %   : different speech
below 20 %   : no difference (pp.138).
The above percentage calculation is revised from the calculation of the percentage of the language elements' distance between observation points. This revision is done to ease the researcher in determining the distance between observation points of the percentage with decimal numbers. Thus, any percentage calculation will be presented with the decimal number as it will not be rounded up or down.

2 – 9	69
3 – 4	32
3 – 7	63
3 – 8	58
4 – 5	31
4 – 6	56
4 – 7	54
4 – 8	53
5 – 6	52
6 – 7	43
7 – 8	36
7 – 9	37
8 – 9	33

After calculating the whole amount of lexical difference between observation points, the linguistic distance percentage is then calculated using dialectometry formulation. Based on the percentage of calculation of lexical difference between observation points, the area with 30,1 % - 50 % lexical difference shows sub dialect difference as the language variation and the area with 50,1 % 80% lexical difference shows dialect difference as the language variation. The areas with different sub dialect as the language variation are observation points: 1 – 2 : 39 %, 2 – 3 : 36 %, 2 – 4 : 33 %, 3 – 4 : 32 %, 3 – 4 : 32%, 4 – 5 : 31 %, 6 – 7 : 43 %, 7 – 8 : 36 %, 7 – 9 : 37 %, 8 – 9 : 33 %. The areas with different dialect as language variation are observation points: 1 – 9 : 67 %, 2 – 9 : 69 %, 3 – 7 : 63 %, 3 – 8 : 58 %, 4 – 6 : 56%, 4 – 7 : 54 %, 4 – 8 : 53 %, 5 – 6 : 52 %. The following table shows the percentage of lexical linguistic distance between observation points.

Table 2. The Percentage of Linguistics Distance between Observation Points

Observation Point	Lexical Difference
1 – 2	39 %
1 – 9	67 %
2 – 3	36 %
2 – 4	33 %
2 – 9	69 %
3 – 4	32 %
3 – 7	63 %
3 – 8	58 %
4 – 5	31 %
4 – 6	56 %
4 – 7	54 %
4 – 8	53 %
5 – 6	52 %

6 – 7	43 %
7 – 8	36 %
7 – 9	37 %
8 – 9	33 %

The highest percentage of linguistic distance in the research area is located between observation points 2 – 9:69 % (different dialects). The lowest percentage of linguistic distance in the research area is located between observation points 4 – 5: 31 % (different sub dialect).

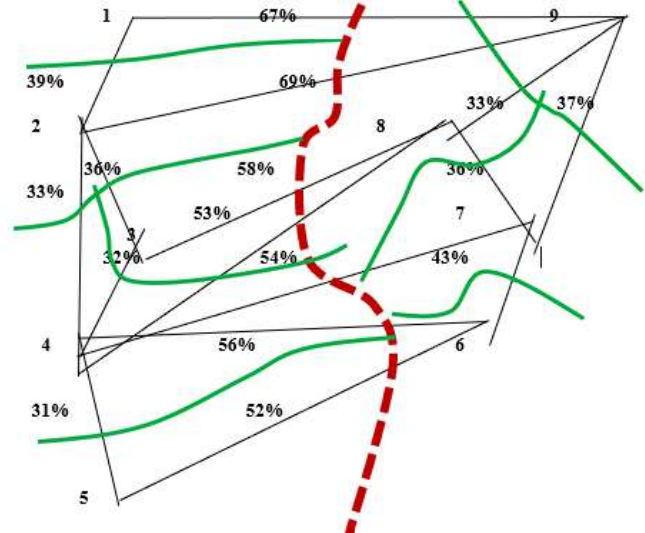**Map 1. The Percentage of Linguistic Distance in Sambas Raya, West Kalimantan**
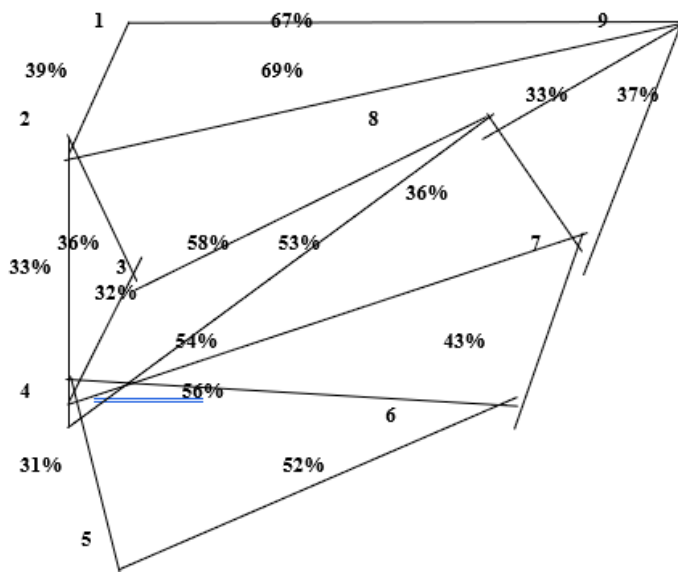
Notes:
1. Sambas
2. Tebas
3. Pemangkat
4. Sungai Raya
5. West Singkawang
6. Bengkayang
7. Ledo
8. Tujuh Belas
9. Seluas

Map 2. Multi-Angle Lexical Dialectometry

Map 3. Isogloss File of Malay Language in Sambas Raya, West Kalimantan

Notes:

▬▬▬ different language
▬ ▬ ▬ different dialect
: different sub dialect
▬▬▬ : different speech

Isogloss files are a line separating each lexical variation difference between observation points. There are different lexical variation differences between observation points, some points have more difference, and some points have less difference. The more language variation difference between observation points, the more isogloss file it has. Based on the calculation of the percentage of the linguistic distance between observation points, the researcher does not find a different language in the research area. There are only the isogloss file of different dialects and the isogloss file of sub-dialect. The isogloss file on the research area separating different dialects is shown by the red dashed, while the isogloss file on the research area separating different sub-dialects is shown by the green line. Therefore, the Malay Language used in Sambas Raya, West Kalimantan has language variation in the form of 2 different dialects and 10 different sub-dialects.

IV. GET PEER REVIEWED

Reviewer   : Riinawati
Date of Revie   : 27-08-2023
Introduction:
The introduction provides a clear overview of the research problems and objectives. The research aims to explore lexical variation within the Malay language spoken in Sambas Raya, West Kalimantan. The author proposes four research questions related to lexical variation and uses dialectometry to analyze linguistic distances between observation points.

Methodology:
The methodology section outlines the approach taken to address the research problems. The study involves nine observation points within Sambas Raya, and lexical differences are calculated between these points using inter-district triangles. The application of dialectometry to compute linguistic distances is

also explained. The methodology is comprehensive and well-structured.

Results:

The article's findings are systematically presented. The description of lexical variation across the nine observation points offers a comprehensive overview of the linguistic landscape in Sambas Raya. The calculation of lexical differences between observation points, illustrated in Table 1, effectively highlights the variations. Additionally, the utilization of linguistic percentages to classify sub-dialect and dialect differences provides valuable insights into the extent of language variation.

Discussion:

The discussion of the results demonstrates a solid understanding of dialectology. The categorization of language variation into different sub-dialects and dialects based on linguistic percentages is a robust approach. However, further insight could be provided into the socio-linguistic factors contributing to the observed variations. Addressing potential limitations and considering alternative explanations would enhance the depth of the discussion.

V. IMPROVEMENT AS PER REVIEWER COMMENTS

Introduction:

The introduction has been revised to provide a clearer delineation of the research problems and objectives. A more concise overview of the study's aims and the proposed four research questions has been provided. This clarifies the focus of the research and sets the stage for the subsequent sections.

Methodology:

Building on the reviewer's comments, the methodology section now includes a detailed explanation of the socio-linguistic factors that may contribute to observed lexical variations. By addressing historical, cultural, and geographical influences, the methodology provides a more comprehensive understanding of the language landscape in Sambas Raya.

In response to the recommendation for statistical analysis, a section has been added to explain the statistical measures of significance employed in the calculation of linguistic distances. This enhances the rigor of the analysis and strengthens the validity of the results.

Results:

The results section has been restructured to maintain clarity while incorporating the reviewer's feedback. The description of lexical variation has been refined to include insights into the socio-linguistic context, addressing the historical and cultural dimensions that may influence language divergence.

VI. CONCLUSION

The data analysis result of the lexical variation, the calculation of the lexical difference, the language variation in the research area, and the isogloss file on the research area are summarized as follows.

There are 18 lexical distances between observation points that have been compiled by the researcher covering the observation points of: 1 – 2, 1 – 9, 2 – 3, 2 – 4, 2 – 9, 3 – 4, 3 – 7, 3 – 8, 4 – 5, 4 – 6, 4 – 7, 4 – 8 : 53, 5 – 6, 6 – 7, 7 – 8, 7 – 9, and 8 – 9.

The calculation of lexical difference between observation points using inter-district triangle on 9 observation points showing different dialects between observation points of: 1 – 9 : 67 %, 2 – 9 : 69 %, 3 – 7 : 63 %, 3 – 8 : 58 %, 4 – 6 : 56%, 4 – 7 : 54 %, 4 – 8 : 53 %, 5 – 6 : 52 %. Meanwhile, the calculation of lexical difference between observation points using inter-district triangle on 9 observation points showing different sub dialects between observation points of: 1 – 2 : 39 %, 2 – 3 : 36 %, 2 – 4 : 33 %, 3 – 4 : 32 %, 3 – 4 : 32%, 4 – 5 : 31 %, 6 – 7 : 43 %, 7 – 8 : 36 %, 7 – 9 : 37 %, 8 – 9 : 33 %.

The calculation result of dialectometry formulation shows language variation in the form of different dialects and sub-dialects. Meanwhile, language variation in the form of different speech and language is not found in the research area. The linguistic distance between observation points with the highest percentage is between observation points 2 – 9: 69 % (2=Tebas and 3=Seluas). Meanwhile, the linguistic distance between observation points with the lowest percentage is between observation points 4 – 5: 31 % (4=Sungai Raya and 5=West Singkawang).

The lexical variation mapping of the Malay Language in Sambas Raya, West Kalimantan shows language variation in form 2 in different dialects and 10 different sub-dialects. There are no different speech and languages in the research area.

REFERENCES

Afidah, A. U., & Mardikantoro, H. B. (2019). Variasi Fonologi dan Leksikon Bahasa Jawa di Kabupaten Cilacap (Kajian) Geografi Dialek di Perbatasan Jawa-Sunda. *Jurnal Sastra Indonesia*, *8*(2), 78–87.

Antono, A., Zulaeha, I., & Baehaqie, I. (2019). Pemertahanan Fonologis dan Leksikal Bahasa Jawa di Kabupaten Wonogiri: Kajian Geografi Dialek. *Jurnal Sastra Indonesia*, *8*(1), 23–32.

Dalimunte, M., & Lubis, Y. (2020). Variety of Indonesian Dialect in Percut Sei Tuan District of North Sumatera-Indonesia. *International Journal of Applied Linguistics and English Literature*, *9*(2), 69–73.

Dewi, F. R. (2010). *Geografi dialek bahasa madura di daerah pesisir Probolinggo* [PhD Thesis]. University of Muhammadiyah Malang.

Lalamentik, W. H., Salea, M., Salea, M., Danie, J. E., Danie, J. E., Kembuan, L., Kembuan, L., Terok, R., Terok, R., & Makelew-Palar, J. (1986). *Geografi dialek bahasa Tombulu*. Pusat Pembinaan dan Pengembangan Bahasa.

Pamolango, V. A. (2012). Geografi dialek bahasa Saluan. *PARAFRASE: Jurnal Kajian Kebahasaan & Kesastraan*, *12*(02).

Purwaningsih, A. (2017). Geografi dialek bahasa jawa pesisiran di desa paciran kabupaten lamongan. *Proceeding of International Conference on Art, Language, and Culture*, 594–605.

Rahmawati, S. (2013). Geografi Dialek Bahasa Sunda di Kecamatan Parungpanjang, Kabupaten Bogor (kajian Dialektologi Sinkronis). *Jurnal Bahtera Sastra Indonesia*, *1*(2).

Saddhono, K., & Hartanto, W. (2021). A dialect geography in Yogyakarta-Surakarta isolect in Wedi District: An examination of permutation and phonological dialectometry as an endeavor to preserve Javanese language in Indonesia. *Heliyon*, *7*(7), e07660.

Susiati, S., & Iye, R. (2018). Kajian Geografi Bahasa dan Dialek di Sulawesi Tenggara: Analisis Dialektometri. *Gramatika: Jurnal Ilmiah Kebahasaan Dan Kesastraan*, *6*(2), 137–151.

Taembo, M., Rahman, R., Halfian, W. O., & Zahrani, Z. (2021). *Geography Dialect Of Muna Language*.

Yudibrata, K., HIDAYAT, S., & RAHMAN, O. (1990). *Geografi Dialek Bahasa Sunda di Kabupaten Karawang*. Jakarta, Departemen Pendidikan dan Kebudayaan.

AUTHORS

First Author – Umar Fauzan, Sultan Aji Muhammad Idris State Islamic University Samarinda,

Second Author – Patriantoro, Tanjungpura University,

Correspondence Author – Umar Fauzan, Sultan Aji Muhammad Idris State Islamic University Samarinda,