

# Pythagoras Tree Applied For Determined Instagram Usage Habit Decision

Erlin Windia Ambarsari  
Universitas Indraprasta PGRI  
Jakarta, Indonesia  
[erlinunindra@gmail.com](mailto:erlinunindra@gmail.com)

Herlinda  
Universitas Indraprasta PGRI  
Jakarta, Indonesia  
[linda\\_said72@ymail.com](mailto:linda_said72@ymail.com)

Submitted: Mar 1, 2020  
Accepted: Mar 15, 2020  
Published: Apr 1, 2020

**Abstract**— In previous studies, Pythagoras Tree constructed using the Regression Method, namely ID3 of Standard Deviation Reduction (SDR). The study using SDR for Classification, which uses the Coefficient of Variant (CV) rather than Information Gain. Data obtained from previous research about Instagram Usage Habit. The result of the study that SDR is useful for the classification method for constructing Pythagoras Tree. However, the target attribute is must use a numerical variable to gain the Standard Deviation and the Mean. Empty data does not affect calculations. Although instances must be discarded, thereby reducing the amount of data. For the case of the Instagram Usage Habit itself, not getting the right pattern to deciding due to the data obtained is less. The reason is Instagram Usage Habit's attributes has ambiguous. Therefore, the rule becomes unclear. However, Construct Pythagoras Tree successfully done because the decision can track based on the trunk of the tree.

**Keywords**—Pythagoras Tree; ID3; Standard Deviation Reduction; Classification

## I. INTRODUCTION

Pythagoras, as we know that it has a right triangle, and the algebraic equation is known as  $c^2=a^2+b^2$ . Therefore, Pythagoras has several theorems to prove the equation. However, according to (Teia, 2018) that Pythagoras has a connection to a geometrical context. Pythagoras can assume as a term of the area: The hypotenuse square's area, which opposite the right angle, is equal to the sum of the square's area on two legs.

Pythagoras' geometric when the areas have looping until it can establish the fractal. A fractal tree which is known as Pythagoras Tree. In Data Mining, Pythagoras Tree able to used as a Decision Tree (Ambarsari, Ar Rakhman Awaludin, Suryana, Hartuti, & Rahim, 2019), which alternative visualization for hierarchy (Beck, Burch, Munz, Di Silvestro, & Weiskopf, 2015).

The concept of Pythagoras Tree is the split of data from hypotenuse's area of a square into the legs of the square area. The purpose of building the Pythagoras Tree is to simplify data association explores.

However, construct data in Pythagoras Tree is need the Decision Tree method, which divides two kinds: Classification and Regression. The Classification method focused on the variables which have category or characters, and Regression for the variables was numeric or continuous. Therefore, in the study about Decision Tree have several methods able to use, such as C4.5 (Handrianto & Farhan, 2019) as a preferred algorithm to Classification, which developed from ID3. Besides, CART, which algorithm uses Classification and Regression at the same time. Also, C5.0 barely used because of its still new algorithm, which expanded of C4.5 (Thariqa, Sitanggang, & Syaufina, 2016).

In the study of (Windia Ambarsari, Avrizar, Doni Sirait, Dwiasnati, & Rahim, 2019) used ID3, which usually used as Classification of Information Gain. The study utilizes Standard Deviation Reduction, which used Coefficient of Variant (CV) to construct Regression for Decision Tree. However, especially for the case, we experiment with used CV for Classification, which the data processing obtained from (Herlinda, 2019).



(Herlinda, 2019) to clarify the relation of Instagram to self-esteem towards the user. Result of the study that the influence of Instagram usage on self-esteem is low because it was not able to predict self-esteem level. Processing data on the study used a question and Likert scale to discover relationships. However, the research for Pythagoras Tree applied utilizes a data raw by removing the item obtained by the Likert level. Another element used in the study of (Herlinda, 2019) becomes attributes for Classification Instagram Usage Habit.

## II. LITERATURE REVIEW

### A. ID3 of Standard Deviation Reduction

As mention before, Standard Deviation Reduction (SDR) used for construct Regression on Decision Tree. Also, SDR based on standard deviation ( $\sigma$ ) has beneficial for a numerical dataset on conditional attributes nor target attributes. When  $\sigma$  has zero, instances in a dataset becomes homogeneous. The equation of  $\sigma$  is (Windia Ambarsari et al., 2019):

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (u_i - \mu)^2} \quad (1)$$

where

$$\mu = \frac{1}{n} \sum_{i=1}^n u_i \quad (2)$$

Meanwhile, the construction of Decision Tree by SDR is by splitting datasets, which purpose to verify the connection between two attributes, namely target and predictor (conditional). Predictor as a determination of Decision Tree establishing.

The standard deviation is calculated recursively for developing each branch, which its subtracted with the previous standard deviation for obtaining SDR. The attribute selected with the most substantial SDR is the best choice in determining the branch. The equation of SDR as follows (Windia Ambarsari et al., 2019):

$$SDR = \sigma - \sigma(T, P) \quad (3)$$

where

$$\sigma(T, P) = \left(\frac{n_1}{n} \times \sigma(P)_1\right) + \left(\frac{n_2}{n} \times \sigma(P)_2\right) \quad (4)$$

Decision Tree stopped construction when the Coefficient of Variant (CV) is less than the threshold ( $<0.1$ ).

$$CV = \frac{\sigma(P)}{\mu} \times 100\% \quad (5)$$

### B. Pythagoras Tree

The following theory presented in the article of (Ambarsari et al., 2019), fundamentals of Pythagoras Tree as follow:

- 1) Pythagoras based on the terms of area, which is the square area of the hypotenuse is equal to the sum of the square area of opposite and adjacent.
- 2) As mention before, the equation of Pythagoras is  $c^2 = a^2 + b^2$ . It is mean that to construct a decision tree based on Pythagoras, which dataset  $c^2$  and subset of  $a^2$  and  $b^2$ . Subsets split become legs, based on SDR.

$$Area1 = (a + b)^2 \quad (6)$$

$$Area2 = c^2 + 4 \left(\frac{1}{2} ab\right) \quad (7)$$

$$Area1 = Area2$$

$$(a + b)^2 = c^2 + 4 \left(\frac{1}{2} ab\right)$$

$$a^2 + 2ab + b^2 = c^2 + 2ab$$

$$a^2 + b^2 = c^2 \quad (8)$$

- 3) Construct Pythagoras Tree also determined by angle to distinguish opposite and adjacent, which the equation is:

$$\sin \theta = \frac{\text{opposite}}{\text{hypotenuse}} \quad (9)$$

The Illustration of Pythagoras Tree as follows below:

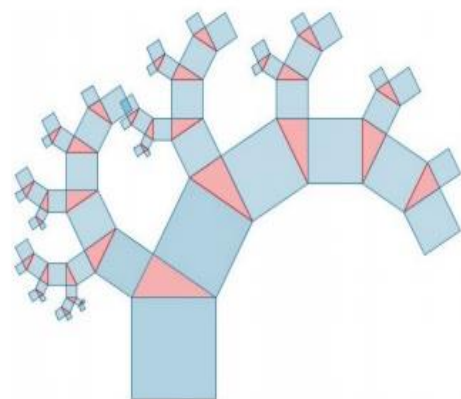


Fig. 1 Illustration of Pythagoras Tree (da Costa Reis, 2015)

## III. PROPOSED METHOD

Classification with regression method for constructing Pythagoras Tree to Instagram Habit Usage datasets have several steps as follows:

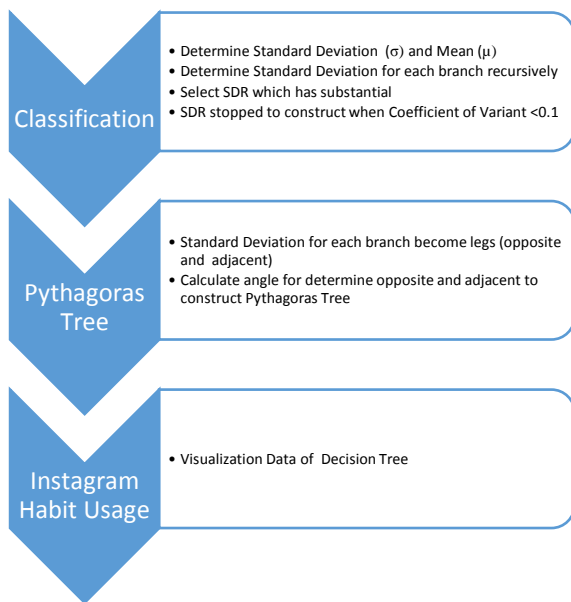


Fig. 2 Several Steps to Construct Instagram Habit Usage's Pythagoras Tree

Instagram Habit Usage's datasets as attributes that have Characters or Categories calculated the Standard Deviation ( $\sigma$ ) and the Mean ( $\mu$ ) for the beginning. Calculation Standard Deviation ( $\sigma$ ) itself by subtracting each attribute target with a mean ( $\mu$ ), each subtraction of the values given by the power of two.

Split Instagram Habit Usage's datasets to verify the relation to each attribute of target and predictor. The construct decision tree is doing some training by comparing SDR, whichever is more significant. When the  $CV > 0.1$ , the dataset needs subdividing. Splitting datasets based on SDR become leg determination for construct Pythagoras Tree, especially the size of the square area. Also, build branches for construct Pythagoras Tree depends on the angle  $\theta$  because it determines which one of two legs is opposite and adjacent.

**IV. RESULT AND DISCUSSION**

The study attaches 33 datasets, and 7 attributes (Initials, Age, Propose to use Instagram, Download Content, Frequency of using Instagram, The duration of using Instagram, Cumulative of Using Instagram) taken from (Herlinda, 2019), which the result of constructing Pythagoras Tree as in Fig. 3:

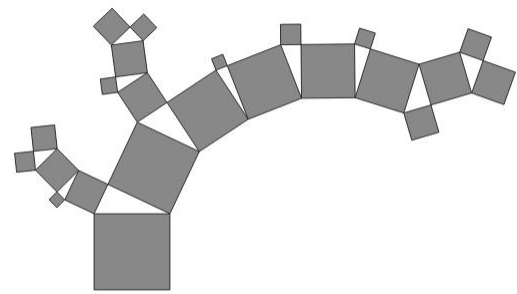


Fig. 3 Instagram Habit Usage's Pythagoras Tree

The first is to determine attribute as the target; the study obtains Age attribute. Observe in Fig. 4, in Instagram Habit Usage's datasets, there is empty data. The results in empty data not counted. Therefore, the total number of n that should be 33 datasets becomes 32.

Initials	Age	Propose to use Inst.	Download Content	Frequency of using Inst.	Duration of using Inst.	Cumulative of Using Inst.
A1	20	Communicate L...	Another person...	1 to 3 times	< 5 minutes	< 0.5 hour
A2	21	Uploading edit...	Video cinematic	1 to 3 times	< 5 minutes	< 0.5 hour
A3	19	Communicate L...	Group Photo	1 to 3 times	5 to 15 minutes	0.5 to 1 hour
A4	19	?	?	1 to 3 times	30 to 60 minutes	1 to 2 hours
A5	19	Uploading Con...	Selfie	1 to 3 times	30 to 60 minutes	1 to 2 hours
A6	18	Communicate L...	Selfie	> 9 times	30 to 60 minutes	1 to 2 hours
A7	19	Uploading Con...	Objects Photo	4 to 6 times	5 to 15 minutes	0.5 to 1 hour
A8	19	Communicate L...	Selfie	> 9 times	15 to 30 minutes	1 to 2 hours
A9	18	Posting and Inf...	Objects Photo	4 to 6 times	5 to 15 minutes	2 to 4 hours
A10	19	Surfing	Objects Photo	1 to 3 times	15 to 30 minutes	2 to 4 hours
A11	21	Surfing	Selfie	1 to 3 times	< 5 minutes	< 0.5 hour
A12	19	Communicate L...	Selfie	7 to 9 times	30 to 60 minutes	1 to 2 hours
A13	19	Surfing	Selfie	1 to 3 times	15 to 30 minutes	0.5 to 1 hour
A14	18	Uploading Con...	Selfie	?	> 60 minutes	4 to 6 hours
A15	19	Surfing	Selfie	> 9 times	5 to 15 minutes	4 to 6 hours
A16	18	Posting and Inf...	Objects Photo	4 to 6 times	15 to 30 minutes	1 to 2 hours
A17	18	Communicate L...	Selfie	1 to 3 times	15 to 30 minutes	< 0.5 hour
A18	22	Uploading Con...	Selfie	> 9 times	30 to 60 minutes	0.5 to 1 hour
A19	18	Uploading Con...	Selfie	7 to 9 times	5 to 15 minutes	0.5 to 1 hour
A20	19	Communicate L...	Video Content	4 to 6 times	15 to 30 minutes	1 to 2 hours
A21	20	Uploading Con...	Objects Photo	1 to 3 times	< 5 minutes	< 0.5 hour
A22	19	Communicate L...	Another person...	4 to 6 times	5 to 15 minutes	0.5 to 1 hour
A23	18	Posting and Inf...	Objects Photo	1 to 3 times	15 to 30 minutes	0.5 to 1 hour
A24	20	Surfing	Objects Photo	1 to 3 times	15 to 30 minutes	< 0.5 hour
A25	20	Communicate L...	Objects Photo	4 to 6 times	> 60 minutes	2 to 4 hours
A26	19	Uploading Con...	Selfie	7 to 9 times	< 5 minutes	0.5 to 1 hour
A27	21	Uploading Con...	Objects Photo	> 9 times	15 to 30 minutes	0.5 to 1 hour
A28	19	Surfing	Group Photo	1 to 3 times	5 to 15 minutes	< 0.5 hour
A29	20	Uploading Con...	Design Picture	4 to 6 times	> 60 minutes	2 to 4 hours
A30	21	Surfing	Never Download	1 to 3 times	5 to 15 minutes	< 0.5 hour
A31	20	Communicate L...	Objects Photo	1 to 3 times	< 5 minutes	< 0.5 hour
A32	19	Communicate L...	Selfie	1 to 3 times	< 5 minutes	0.5 to 1 hour
A33	18	Communicate L...	Selfie	4 to 6 times	15 to 30 minutes	0.5 to 1 hour

Fig. 4 Empty Data In Instagram Habit Usage's datasets

Calculation of the mean ( $\mu$ );  $(20+21+19+19+...+19+18)/32=19.28125$ , and the Standard Deviation ( $\sigma$ ) =  $\sqrt{(0.516601563^2 + 2.954101563^2 + 0.079101563^2 + ... + 1.641601563^2)} = 1.067543178$ .

Split the datasets to verify the relation of target and predictor attributes. Conduct training to select attributes that have the highest SDR values. For example, we separate the attribute of Cumulative of Using Instagram, which has  $SDR = 0.099426768$ .

TABLE 1. CUMULATIVE OF USING INSTAGRAM ATTRIBUTE

Cumulative of Using Instagram				
Datasets	Standard Deviation	Average	Total	CV
<0.5 h	0.942809	20	9	4.714045
0.5 to 1 h	0.978019	19	23	5.14747
1 to 2 hrs				
2 to 4 hrs				
4 to 6 hrs				

In Table 1. we split data of <0.5 hours, in which the total sample is 9 instances, and the remaining data is 23—this separate data used as the legs of Pythagoras, namely 9 and 23. As mention before, a total of 32 instances (9+23) means hypotenuse. 23 is adjacent (the value of adjacent is usually higher than the opposite), and 9 is the opposite. Therefore, the calculation angle  $\theta$  found by  $\sin \theta = \sqrt{9/32}=32.02776011$ . Afterward, do the training again as before, for example, we split data from <0.5 hours.

Based on Tabel 2, Instances that Cumulative of Using Instagram is <0.5 hour, which has divided is:

- 1) Initials: A1, A2, A11, A17, A21, A24, A28, A30, A31
- 2) Age: 18 to 21
- 3) Propose to use Instagram: Communicate Interactively (CI), Uploading Edited Video (UV), Uploading Content (UC), Surfing (Sf)
- 4) Download Content: Another person's photo, Video cinematic, Selfie, Objects Photo, Group Photo, Never Download
- 5) Frequency of using Instagram: 1 to 3 times
- 6) The duration of using Instagram: <5 minutes, 5 to 15 minutes, 15 to 30 minutes

TABLE 2. SPLIT DATA BASED ON DOWNLOAD CONTENT

Initials	Age	Propose to use Instagram	Download Content	Frequency of using Instagram	The duration of using Instagram	Cumulative of Using Instagram
A1	20	CI	Another person's photo	1 to 3 times	< 5 mins	<0.5 h

A2	21	UV	Video cinematic	1 to 3 times	< 5 mins	<0.5 h
A11	21	Sf	Selfie	1 to 3 times	< 5 mins	<0.5 h
A17	18	CI	Selfie	1 to 3 times	15 to 30 mins	<0.5 h
A21	20	UC	Objects Photo	1 to 3 times	< 5 mins	<0.5 h
A24	20	Sf	Objects Photo	1 to 3 times	15 to 30 mins	<0.5 h
A28	19	Sf	Group Photo	1 to 3 times	5 to 15 mins	<0.5 h
A30	21	Sf	Never Download	1 to 3 times	5 to 15 mins	<0.5 h
A31	20	CI	Objects Photo	1 to 3 times	5 to 15 mins	<0.5 h

The split data of Table 2 calculated as in Table 1, which SDR= 0.283181521. The result is the split of Download Content, which the classification as one is 1 Another person's photo, 2 Selfie, 3 Objects Photo, and 1 Group Photo; the other is 1 Video Cinematic and 1 Never Download. The total is 9 instances, which separated as in Table 3.

TABLE 3. DOWNLOAD CONTENT ATTRIBUTE

Download Content				
Datasets	Standard Deviation	Average	Total	CV
Another person's photo	0.880631	19.71429	7	4.466967
Selfie				
Selfie				
Objects Photo				
Objects Photo				
Group Photo	0	21	2	0
Objects Photo				
Video cinematic				
Never Download				

Based on Table 3, that CV in Video Cinematic and Never Download has zero. However, it is not <

0.1, and the Standard Deviation is 0. It means instances in a dataset becomes homogeneous. The construct Pythagoras become stopped. The illustration of constructing Pythagoras Tree based on data split as follows:

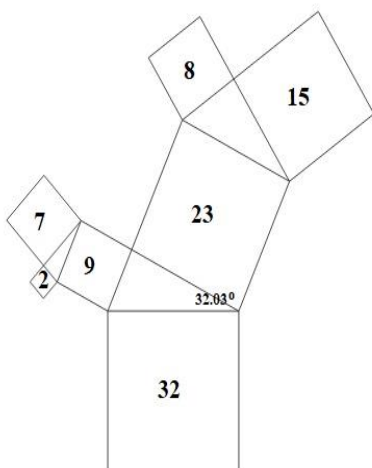


Fig. 5 Data Have Separated in Pythagoras Tree

The result of the construct Pythagoras Tree in Fig. 5 till gains the decision for Instagram Usage Habit as view in Fig. 6:

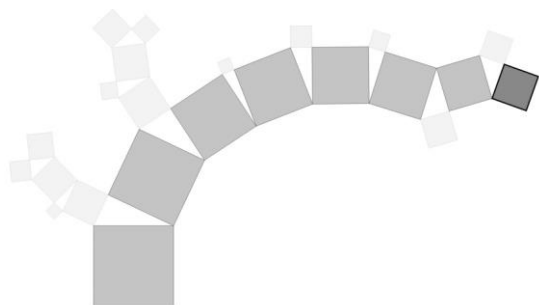


Fig. 6 Instagram Usage Habit Decision Track

The decision of Pythagoras Tree obtained from the trunk of the tree. The end of the tree's trunk itself (said to be a leaf, but has a more substantial area) is a rule used as a decision tree, in which the mean value is equal to the target value. Therefore, the result for Instagram Usage Habit Decision, which can track based on Pythagoras Tree, is Age 19 years. Tracing is convenient to do when the Pythagorean Tree construction is complete, therefore its capable of forms a rule. Below is the Decision Tree rule for Age 19 years in the Instagram Usage Habit:

Mean: 19.000
Standard deviation: 0.000
3 samples
Frequency of using Instagram: 1 to 3 times, 4 to 6 times or 7 to 9 times
Cumulative of Using Instagram: 1 to 2 hours, 2 to 4 hours or 4 to 6 hours
The duration of using Instagram: 15 to 30 minutes or 30 to 60 minutes
Download Content: Objects Photo or Selfie
The Propose to use Instagram: Communicate Interactively, Posting and Information or Surfing

Fig. 7 Instagram Usage Habit Rule

Base on Fig. 7 that the Instagram Usage Habit, not getting the right pattern to deciding due to the data obtained, is less. The reason is for ages 19 years that frequency of using Instagram between 1 to 3 times, 4 to 6 times, or 7 to 9 times, for example. Therefore, the rule of the Decision Tree becomes ambiguous. However, in Construction of Pythagoras Tree, successfully done.

## V. CONCLUSION AND SUGGESTION

Based on observations from the result and discussion section, SDR is useful for the classification method to constructing Pythagoras Tree. However, the target attribute is must use a numerical variable to gain Standard Deviation ( $\sigma$ ) and Meant ( $\mu$ ). Empty data does not affect calculations. Although instances must be discarded, thereby reducing the amount of data.

Although the target for Age 19 years able classified by the Pythagoras Tree. In the study, the Instagram Usage Habit Decision has ambiguous attributes. Therefore, the rule becomes unclear, although attributes are ambiguous, in the development of Pythagoras Tree successfully done because the decision can track based on the trunk of the tree.

Beside discuss of Pythagoras Tree construction, it needs a discussion about Pruning in Pythagoras Tree and how to develop Random Forest Pythagoras Tree for further research.

## VI. REFERENCES

- Ambarsari, E. W., Ar Rakhman Awaludin, A., Suryana, A., Hartuti, P. M., & Rahim, R. (2019). BASIC CONCEPT PYTHAGORAS TREE FOR CONSTRUCT DATA VISUALIZATION ON DECISION TREE LEARNING. *Journal of Applied Engineering Science*, 17(4), 468–472. <https://doi.org/10.5937/jaes17-21960>

- Beck, F., Burch, M., Munz, T., Di Silvestro, L., & Weiskopf, D. (2015). Generalized pythagoras trees: A fractal approach to hierarchy visualization. *Communications in Computer and Information Science*. [https://doi.org/10.1007/978-3-319-25117-2\\_8](https://doi.org/10.1007/978-3-319-25117-2_8)
- da Costa Reis, J. N. (2015). Uma Árvore de Pitágoras Explorando os Fractais no Ensino Médio. *Ciência e Natura*, 37(3), 411–418. <https://doi.org/10.5902/2179460X14636>
- Handrianto, Y., & Farhan, M. (2019). C.45 Algorithm for Classification of Causes of Landslides. *Sinkron*, 4(1), 120–127. <https://doi.org/10.33395/sinkron.v4i1.10154>
- Herlinda, H. (2019). Hubungan Antara Durasi Penggunaan Instagram Per Hari Dan SELF ESTEEM Pada Remaja Akhir. *Seminar Nasional Teknologi*, 559–564.
- Teia, L. (2018). The Pythagorean geometric gear. *Australian Senior Mathematics Journal*, 32(1), 54–64.
- Thariqa, P., Sitanggang, I. S., & Syaufina, L. (2016). Comparative analysis of spatial decision tree algorithms for burned area of peatland in Rokan Hilir Riau. *Telkomnika (Telecommunication Computing Electronics and Control)*, 14(2), 684–691. <https://doi.org/10.12928/telkomnika.v14i2.3540>
- Windia Ambarsari, E., Avrizar, R., Doni Sirait, E., Dwiasnati, S., & Rahim, R. (2019). Regression Tree Role for Interpret Monetizing of Game Live Streaming. *Journal of Physics: Conference Series*, 1424, 012014. <https://doi.org/10.1088/1742-6596/1424/1/012014>