

## **Classification of happiness in North Kalimantan using decision tree-based CHAID**

**Andika Veriyanto\***, Junezarra Thie Dea Giselle

BPS-Statistics of Kalimantan Utara Province, Tanjung Selor

\*Corresponding: andika.veriyanto@bps.go.id

Received: 19-01-2023, accepted: 24-03-2023

---

### **Abstract**

The concept of welfare is closely related to economic development. The 2021 happiness index in Indonesia shows 5 provinces with the highest index have a GRDP that tends to be low. This study aims to examine the factors that influence happiness in North Kalimantan as the youngest province, which, according to the SPTK 2021, has the second-highest happiness index. The study employs a quantitative method as well as inferential analysis with a decision tree based CHAID to classify happiness. This method is advantageous since it can handle both categorical and continuous variables, as well as data that is not regularly distributed. The happiness level is the dependent variable in this study, with 19 independent variables indicating various factors such as life satisfaction, affect, and eudaimonia. According to the report, 70.2 percent of households in North Kalimantan are in the happy category. According to the findings of this study, the majority of happy households are experiencing pleasure and satisfied with their income. In households where they lack of pleasure, there are other factors that explain happiness such as job satisfaction and not feeling stressed. The accuracy from the CHAID model is 78.85 percent, indicating the model is fit to classify happiness.

**Keywords:** Happiness index, CHAID, North Kalimantan, SPTK

**MSC2020:** 62C86

---

## **1. Introduction**

Well-being is measured in traditional economic development theory by: 1) Per capita income or equal to Gross National Income divided by the whole population [1] and 2) Economic growth which measured by Gross Domestic Product (GDP) or the total value of final goods and services produced by all economic units of a country [2] and it aims to enhance people's lives, yet some claim welfare has not increased [3]. Welfare indicators not only used to describe material prosperity, but rather lead to subjective well-being or happiness [4]. The Happiness Index becomes one way to quantify well-being by measuring non-material demands that cannot be measured by per capita income [5].

In the 1960s, happiness was merely described as life satisfaction [6] and later was defined as subjective well-being. Subjective well-being measures about how people feel and think

about life satisfaction, happiness, and psychological health [7]. Happiness is a real thing, while well-being is a construct. The authentic happiness theory sees happiness as an attempt to explain real things that measurable directly and well-being theory sees welfare as a construct of several measurable elements, but no single factor can define well-being [8]. The Organization for Economic Cooperation and Development (OECD) defines subjective well-being as a positive mental state including both positive and negative evaluations of life with the affective reactions to everything they have experienced [9].

The Gross National Happiness (GNH) is a multidimensional indicator first introduced in Bhutan consists of four pillars and nine domains to guide various policies, also considered more essential than GDP [10]. In 2011, OECD created Better Life Index to assess happiness by distinguished between current and future well-being [11] where the combination will increase happiness through time. The Easterlin Paradox shows increasing income does not improve well-being or happiness [12], [13] and has been tested in America for 30 years, where no improvement in life satisfaction despite significant money increases [14]. Economic expansion will not increase happiness in the long run, but will improve satisfaction at low GDP levels [15].

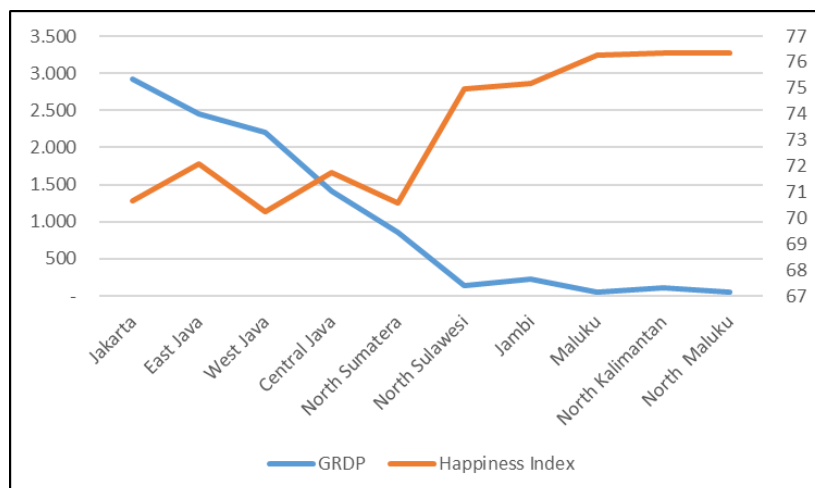


Figure 1. GRDP and happiness index 2021 (BPS, 2021)

In Indonesia, the happiness level measurement survey or survei pengukuran tingkat kebahagiaan (SPTK) carried out in 2014, 2017, and 2021 with adapted framework from OECD. Figure 1 shows five Indonesian provinces with the highest gross regional domestic product (GRDP) and happiness indices in 2021. Provinces with GRDP exceeding 1,500 trillion Rupiah score 70–72 points on the happiness index, while provinces with GRDP under 500 trillion Rupiah score 75–76 points. It shows the Easterlin Paradox, where greater GRDP does not lead to higher happiness index. Happiness is influenced by various factors, including demographic, socioeconomic, and geographic location while on the other side is a determinant of economic outcomes [16]. Considering the gap between GRDP and happiness level, the factors that influence happiness is critical

to investigate especially in provinces with high happiness index when GRDP is low. Given the diversity of culture, ethnicity, and other environmental factors in the five provinces with the highest happiness indices, the factors that contribute to happiness may vary among them.

North Kalimantan is the youngest province in Indonesia after separating from East Kalimantan in 2012. The territorial split strives to reduce disparities in development and other aspects of well-being. In the agenda of Nusantara Project or relocation the new Indonesian capital to East Kalimantan, North Kalimantan as a neighboring area and fraction from East Kalimantan becomes interesting for this study with its condition as the youngest province and low GRDP while its happiness index is always in the high category. North Kalimantan in 2017 had the 5th highest happiness index after East Kalimantan. While in 2021, North Kalimantan had the 2nd highest happiness index ahead of East Kalimantan at the 11th place. In regard to GRDP, in 2021 North Kalimantan had the 8th lowest GRDP while East Kalimantan had the 7th highest GRDP. This study examines the elements that influence happiness in North Kalimantan using a quantitative approach and decision tree-based CHAID (Chi-square Automatic Interaction Detection) as inferential analysis.

## **2. Methods**

This study utilizes raw data from SPTK 2021 using the household as a unit analysis. The dependent variable in this study is the evaluation of household life that describes the condition of the level of happiness. The happiness level scale (0-10) was divided into happy and unhappy categories. This happiness scale is obtained from "how happy are you with your life in general?". For the susceptibility modelling, this research used nineteen evaluative happiness conditioning variables. Each variable was also divided into two categories, one indicating households who do not feel the evaluated characteristic and the other indicating households who feel the evaluated characteristic.

The hypotheses used in this study are: 1) Indicators of life satisfaction (job satisfaction, income satisfaction, health satisfaction, family satisfaction, leisure satisfaction, social relationship satisfaction, environmental satisfaction, safety satisfaction, housing satisfaction, education satisfaction) have a positive effect on happiness; 2) Indicator of feeling pleasure have a positive effect on happiness, indicator of feeling anxious have a negative effect on happiness, and indicator of feeling stressed have a negative effect on happiness; 3) Indicators of eudaimonia (autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, self acceptance) have a positive impact on happiness. The proposed research framework can be seen in Figure 2.

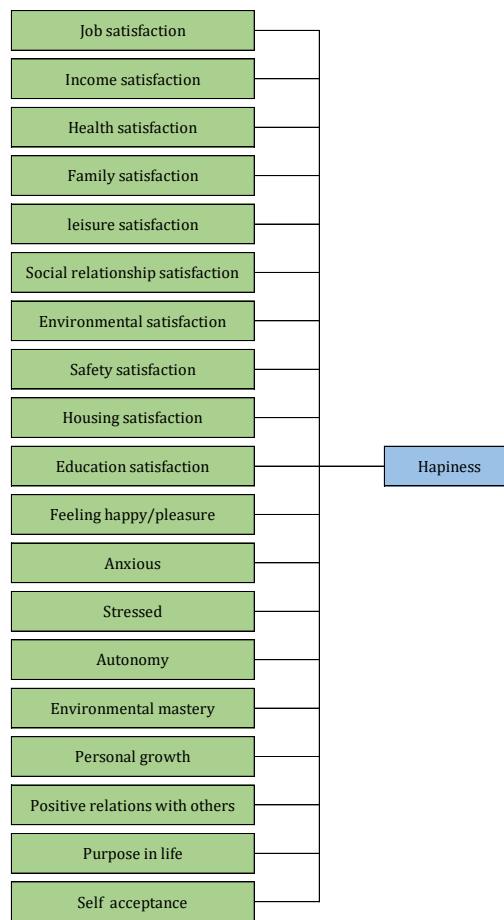


Figure 2. Research framework (BPS, 2021)

For prediction and classification, the decision tree method is commonly employed. Many regression and classification studies have used CHAID (Chi-square Automatic Interaction Detection) as one of the vital decision tree approaches to shape the results as a tree structure [17]. The decision tree is a hierarchical model that divides the heterogeneous input data set into groups that are homogenous in terms of the categories of the dependent variables [18]. CHAID approach is a multivariate analysis technique represented by a tree with branches representing predictors and discriminating sample groups. This approach is helpful since it can handle both categorical and continuous variables. In addition, CHAID also performs better when dealing with non-normally distributed data [19].

The CHAID algorithm stops when the classification iteration finds no significant chi-square value between the dependent and conditioning variables [19]. As a result, the nodes with the highest Chi-square values appear first in the tree, followed by the terminal nodes with lower Chi-square values. For the data of categorical nature, Pearson Chi-square will be considered as follow:

$$X^2 = \sum_{j=1}^J \sum_{i=1}^I \frac{(n_{ij} - m_{ij})^2}{m_{ij}} \quad (1)$$

where:

$$n_{ij} = \sum_{n \in D} fnI(x_n = i \cap y_n = j), \quad (2)$$

$n_{ij}$  is the observed cell frequency and  $m_{ij}$  is the estimated expected cell frequency for  $(x_n = i, y_n = j)$  following the independent model. The corresponding  $p$  value given by  $p = \Pr(x_d^e > x^2)$ .

Table 1. Confusion matrix for two-class classification

	Predicted Positive	Predicted Negative
Actual Positive	TP	FN
Actual Negative	FP	TN

A confusion matrix is one of the tools used to evaluate how effectively a classifier works [20]. The confusion matrix of size  $n \times n$  associated with a classifier contains predicted and actual classification information, where  $n$  is the number of possible classes [21]. As can be seen in Table 1, TP shows the number of rows labelled positive properly identified as positive by the classifier, FN shows the number of rows labelled positive mistakenly identified as negative by the classifier, FP shows the number of rows labelled negative mistakenly identified as positive by the classifier, TN shows the number of rows labelled negative properly identified as negative by the classifier. Furthermore, the confusion matrix can be used to derive the following calculations.

The classifier made the accuracy or the number of correct predictions. Accuracy is an excellent baseline criterion for evaluating the classifier's performance,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3)$$

Sensitivity or the number of actual positive cases predicted correctly as positive by the classifier. Sensitivity is also known as hit ratio, recall, or true positive recognition rate,

$$Sensitivity = \frac{TP}{TP + FN}. \quad (4)$$

Specificity or the number of actual negative cases predicted correctly as negative by the classifier. Specificity is also known as the True negative recognition rate,

$$Specificity = \frac{TN}{TN + FP} \quad (5)$$

### 3. Results

According to SPTK 2021, North Kalimantan holds the second-highest level of happiness in Indonesia after North Maluku [4]. North Kalimantan reached happiness index value of 76.33 or slightly different from North Maluku on the first rank which index value 76.34. Furthermore, according to the classification used in this study, 29.8 percent of households in North Kalimantan are in the unhappy category, whereas 70.2 percent of households are in the happy category (see Figure 3). The population of North Kalimantan in 2021 based on the results of the 2020 Population Census is 713,600 people. It means, approximately 500,000 people in North Kalimantan are in the happy category.

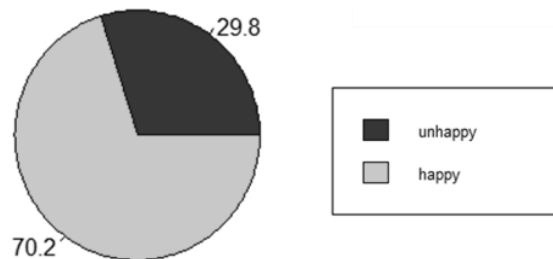


Figure 3. Pie chart of happiness category in North Kalimantan (Rstudio results, 2022)

In CHAID analysis, the first node (root node) represents the highest Chi-square value, and terminal nodes represent the end of the nodes or the most significant predictors. The root node for the happiness model in North Kalimantan is feeling pleasure, followed by nine nodes and 10 terminal nodes, as shown in Figure 4. The pleasure feeling is the most prominent characteristic that identifies happiness in North Kalimantan. In addition, financial satisfaction, job satisfaction, optimism, feeling stressed, health satisfaction, family satisfaction, and anxious feeling also have a substantial impact on happiness. At many nodes, the characteristics of income satisfaction emerge multiple times. This characteristic impacts a wide range of household types.

Based on the results of the CHAID analysis, there are 10 groups to describe happiness classification in North Kalimantan. In the first place, households with the characteristics of pleasure feeling and satisfied with their income tend to be classified as happy. This group contains 136 households with a 3.7 percent inaccuracy rate. The first group with the characteristics of pleasure feeling and satisfaction with their income has the lowest inaccuracy rate but the most observations. This suggests that most happy households in North Kalimantan experience feelings of pleasure and satisfaction with their income.

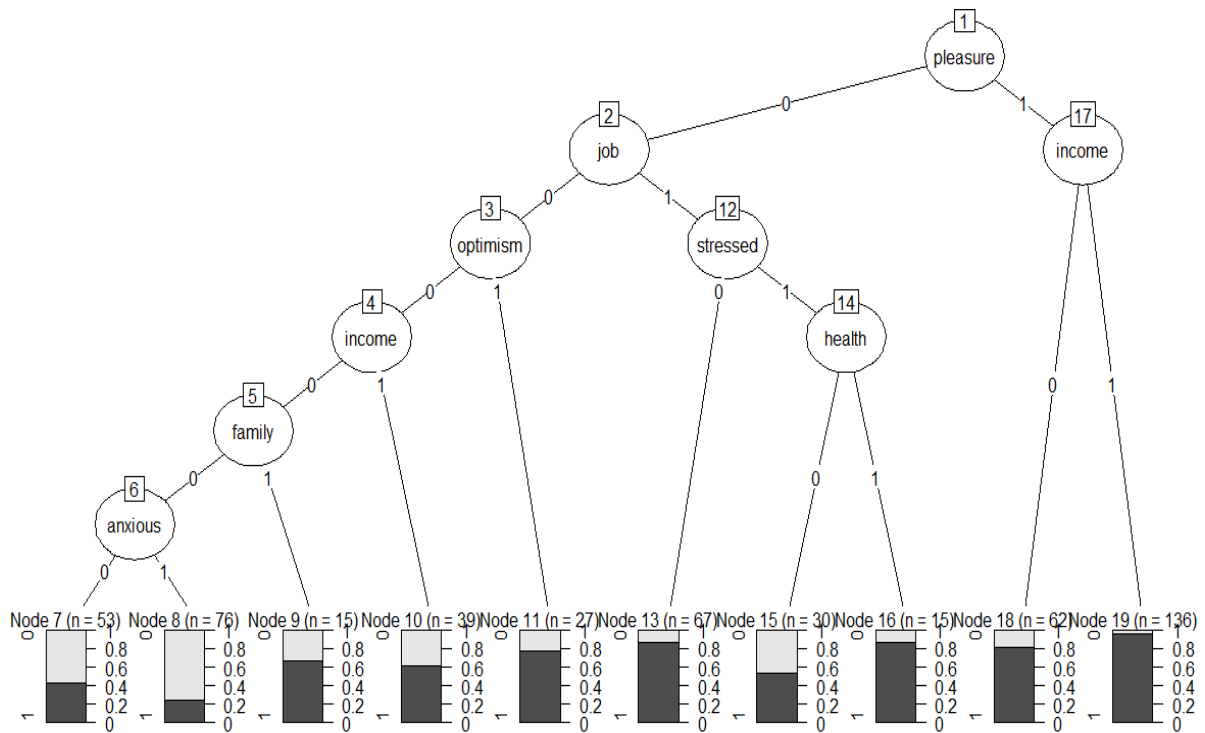


Figure 4. CHAID diagram (Rstudio results, 2022)

Other group shows households with the characteristics of pleasure feeling but not satisfied with their income tend to be classified as happy. This group contains 62 households with a 19.4 percent inaccuracy rate. However, this indicates households that are not satisfied with their income are still in the happy category as long as they are satisfied with their pleasure.

There is also a group of households with the characteristics of not feeling pleasure, satisfied with their job, and not feeling stressed tend to be classified as happy. This group contains 67 households with a 13.4 percent inaccuracy rate. This is the largest group with lowest inaccuracy rate among households who do not feel pleasure but classified as happy. The job satisfaction finding is in line with previous study where unemployment and job prospects affect well-being [22] also other study that find job satisfaction influence happiness [23], but the study in North Kalimantan did not find housing satisfaction to portray happiness. Apart from lack of pleasure, these households find satisfaction in their job and not feeling stressed. This indicates there are factors other than pleasure that can also explain happiness. The relationship between pleasure and happiness is still difficult to explain briefly because of its subjective experience and its relation with other hedonic and eudaimonic components [24], although other study finds it is hardly possible to see people claiming happiness while being devoid of pleasure [25]. Happiness is usually measured over a lengthy period. Pleasure, on the other hand, is more likely momentary. It is described as a type of sensation, feeling, or quality of experience that it shares with other such experiences [24]. While happiness is obtained by questioning the

quality of someone's life in general, the feeling of pleasure is obtained by asking, "how pleased are you in living your daily life?" through SPTK 2021.

Among the 10 groups, there are two groups reflecting unhappy households. Households with the characteristics of not feeling pleasure, not satisfied with their job, not optimistic about their life, not satisfied with their income, not satisfied with their family, and anxious tend to be classified as unhappy. This group contains 76 households with a 25 percent inaccuracy rate. This is the largest group with lowest inaccuracy rate among households who classified as unhappy. This indicates that most unhappy households in North Kalimantan experience feelings of not feeling pleasure, not satisfied with their job, not optimistic about their life, not satisfied with their income, not satisfied with their family, and anxious.

Job satisfaction appears in many groups in the results of this study. According to previous research, happy employees are described as more productive, satisfied with their jobs, have positive relationships with others, healthier, and more creative [16], while several factors that affect happiness among workers are include competence, business dedication, fairness promotion, and work-life balance [26]. Some of these factors can be the main key for stakeholders to increase happiness through job satisfaction.

The income findings in this study are in line with previous studies where people with higher income levels and a higher likelihood of being employed are happier [27]. Although the previous study found that education level has no impact on happiness, it generally has an essential role in case of job opportunities.

Table 2. Confusion matrix for the model

	Predicted Unhappy	Predicted Happy
Actual Unhappy	87	68
Actual Happy	42	323

Several output model evaluations can be derived from the confusion matrix in Table 2 to measure the quality of the model. The accuracy is 78.85 percent, indicating that 78.85 percent of observations were predicted correctly by the model. The sensitivity is 56.13 percent, which means 56.13 percent of the actual unhappy household was predicted correctly as an unhappy household by the model. Due to the small number of unhappy household samples, sensitivity has the potential to be of low value. While the specificity is 88.49 percent, which means 88.49 percent of the actual happy household was predicted correctly as a happy household by the model.



## 4. Conclusion

This study takes place in North Kalimantan as the youngest province with the second-highest happiness index after North Maluku. The results showed that 70.2 percent of households in North Kalimantan are in the happy category, while 29.8 percent of households are in the unhappy category. This corresponds to the fact that, as previously stated, North Kalimantan has the second-highest happiness index. It becomes clear that the majority of households are in the happy category. The findings suggest that feelings of pleasure and income satisfaction are the most important characteristics representing the happiness condition in North Kalimantan. In addition, other characteristics such as job satisfaction, optimism, feeling stressed, health satisfaction, family satisfaction, and anxious feeling have a substantial impact on happiness.

The key recommendations given to increase the happiness index are include looking at wider scope by reducing the number of unemployed and income inequality while, on the other hand, improving the quality of education. Happiness in North Kalimantan is more likely to be represented by income satisfaction. On the other hand, it is widely accepted that education has an important role in job opportunities. The higher the quality of education will more likely go along with the opportunity to get a better job. Thus, it creates better income and leads to better income satisfaction. The government, agencies, or policymakers also need to regularly expand job opportunities to absorb the growing workforce every year. It will reduce income disparity and promote happiness by reducing the number of unemployed and raising people's income. Furthermore, maintaining a fair workplace environment is also important to make employee more content with their job and satisfied with other aspects as well as their lives as a whole.

## References

- [1] M.P. Todaro and S.C. Smith, *Economic Development* (12<sup>th</sup>). 2015. [[GreenVersion](#)]
- [2] BPS, *Produk Domestik Regional Bruto Provinsi-Provinsi di Indonesia menurut Lapangan Usaha 2017-2021*. 2022. [[GreenVersion](#)]
- [3] B. Stevenson and J. Wolfers, "Economic growth and subjective well-being: Reassessing the Easterlin paradox," *Working Paper* **14282**, 2008. [[GreenVersion](#)]
- [4] BPS, *Indeks Kebahagiaan 2021*. 2021. [[GreenVersion](#)]
- [5] S.Y. Dewi, "Determinan indeks kebahagiaan di Indonesia," *Jurnal Ilmiah Mahasiswa FEB, Universitas Brawijaya*, vol. 8, no. 2, 2020. [[CrossRef](#)]
- [6] R. Veenhoven, "Happiness, Also Known as 'Life Satisfaction' and 'Subjective Well-Being,'" *Handb. Soc. Indic. Qual. Life Res.*, 2012. [[GreenVersion](#)]

- [7] R. Layard, “Measuring Subjective Well-Being,” 2011. [[GreenVersion](#)]
- [8] M.E.P. Seligman, *Flourish: A Visionary new understanding of happiness and well-being*. United States of America: Simon & Schuster, Inc., 2012. [[GreenVersion](#)]
- [9] OECD, *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing, 2013. [[GreenVersion](#)]
- [10] K. Ura, S. Alkire, and T. Zangmo, “GNH and GNH Index,” *Cent. Bhutan Stud.*, no. May, pp. 1–60, 2012. [[GreenVersion](#)]
- [11] M. Durand, “The OECD better life initiative: How’s life? and the measurement of well-being,” *Rev. Income Wealth*, vol. 61, no. 1, pp. 4–17, 2015. [[GreenVersion](#)]
- [12] L. Angeles, “A closer look at the Easterlin Paradox,” *J. Socio. Econ.*, vol. 40, no. 1, pp. 67–73, 2011. [[GreenVersion](#)]
- [13] V. Mentus and M. Vladisavljević, “Easterlin paradox revisited: Do increases in income bring higher levels of income satisfaction?,” *Sociologija*, vol. 63, no. 2, pp. 220–235, 2021. [[GreenVersion](#)]
- [14] A.E. Clark, P. Frijters, and M.A. Shields, “Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles,” *J. Econ. Lit.*, vol. 46, no. 1, pp. 95–144, 2008. [[GreenVersion](#)]
- [15] R.A. Easterlin and K.J. O’Connor, “The Easterlin paradox,” *Inst. Labor Econ.*, 2020. [[GreenVersion](#)]
- [16] M. Piekalkiewicz, “Why do economists study happiness?,” *Econ. Labour Relations Rev.*, vol. 28, no. 3, pp. 361–377, 2017. [[GreenVersion](#)]
- [17] G.V. Kass, “An exploratory technique for investigating large quantities of categorical data,” *Appl. Stat.*, vol. 29, no. 2, p. 119, 1980. [[GreenVersion](#)]
- [18] M. Milanović and M. Stamenković, “CHAID decision tree: methodological frame and application,” *Econ. Themes*, vol. 54, no. 4, pp. 563–586, 2016. [[GreenVersion](#)]
- [19] O.F. Althuwaynee, B. Pradhan, H.J. Park, and J.H. Lee, “A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping,” *Landslides*, vol. 11, no. 6, pp. 1063–1078, 2014. [[CrossRef](#)]
- [20] S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso, and R. Nooraeni, *Data Mining dengan R (Konsep Serta Implementasi)*. Bogor: In Media, 2018. [[GreenVersion](#)]
- [21] S. Visa, B. Ramsay, A. Ralescu, and E. van der Knaap, “Confusion matrix-based feature selection,” *Proc. 22nd Midwest Artif. Intell. Cogn. Sci. Conf. 2011*, vol. 710, pp. 120–127, 2011. [[GreenVersion](#)]

- [22] Z.B. Susanlı, “Life satisfaction and unemployment in Turkey: evidence from Life Satisfaction Surveys 2004–2013,” *Qual. Quant.*, vol. 52, no. 1, pp. 479–499, 2017. [[CrossRef](#)]
- [23] B. Fallahi, A. Mehrad, “A review of housing and job satisfaction and its link to life happiness,” *Online J. Couns. Educ.*, vol. 4, no. 2, pp. 40–50, 2015. [[GreenVersion](#)]
- [24] D.M. Haybron, “Happiness and Pleasure,” *Philos. Phenomenol. Res.*, vol. 62, no. 3, pp. 501–528, 2001. [[GreenVersion](#)]
- [25] M.L. Kringelbach, K.C. Berridge, “The neuroscience of happiness and pleasure,” *Soc Res*, vol. 77, no. 2, pp. 659–678, 2010. [[GreenVersion](#)]
- [26] S. Januwarsono, “Analytical of factors determinants of happiness at work case study on PT. PLN (Persero) Region Suluttenggo, Sulawesi, Indonesia,” *Eur. J. Bus. Manag.*, vol. 7, no. 8, pp. 9–17, 2015. [[GreenVersion](#)]
- [27] J. Cuñado and F. P. de Gracia, “Does education affect happiness? Evidence for Spain,” *Soc. Indic. Res.*, vol. 108, no. 1, pp. 185–196, 2012. [[CrossRef](#)]