Addressing Poverty and Citizens Welfare with Analysis of Economic Progress within Indonesia and its Provinces

Vian Sebastian Bromokusumo
Department of Computer Science and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia
viansebastianbromokusumo@mail.ugm.ac.id

Sasha Annabel
Department of Computer Science and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia
sashaannabel@mail.ugm.ac.id

Nabilla Nafisatuz Zahrah
Department of Computer Science and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia
nabillanafisatuzzahrah@mail.ugm.ac.id

***Abstract*—This research paper is intended to retrieve analytical information on Indonesian workers' economy and welfare for the goal of addressing the national state of poverty.**

Keywords—clustering, regression, poverty line, trend, wage, minimum wage

I. Introduction

Economic disparity among different regions within a country poses significant challenges to achieving sustainable development. Indonesia, in particular, is a vast archipelago with diverse socioeconomic landscapes where regional disparities in worker welfare are pronounced [1]. Understanding these disparities is crucial for the formulation of effective policies aimed at fostering inclusive economic growth. This project leverages machine learning techniques such as clustering and regression to analyze the welfare of workers across Indonesia's 34 provinces, utilizing data on wages, minimum wages, poverty lines, and expenses. By examining these variables, the project aims to provide insights that can inform government policies and interventions, in alignment with the United Nations Sustainable Development Goal (SDG) 1: No Poverty. Ensuring no poverty is foundational for improving worker welfare, as it directly influences employment opportunities, productivity, and overall economic well-being.

II. Problem

Though Indonesia experiences general economic growth as a developing country, there are known disparities in worker welfare across its provinces. However, it is unclear which provinces perform better economically, especially with variations in wages, minimum wages, poverty lines, and living expenses [2]. Understanding the four main factors mentioned above will lead to a better view on the standard of living and level of poverty of workers in each Indonesian province. Here, the methods for analysis in this project are curated specifically to identify patterns and trends in the worker welfare data, thereby highlighting provinces that are lagging behind for the government's attention.

As poverty is the main topic of this project, the authors have created 2 main hypotheses that are considered vital to be addressed :

1. As a developing country, the ratio of average monthly wage of citizens and the poverty line are predicted to increase or stay constant steadily in the predicted future.
2. Provinces with a high poverty line are due to the high average expenses yet low average wages of their citizens. Essentially, lower income than outcome.

III. Proposed Research Methods

Considering the problems mentioned above, the team attempts to create a program with two main sections, as is observable in the flowchart below.

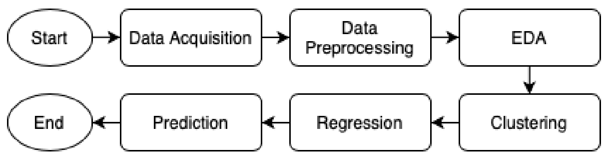

fig 1 : project flowchart

The focus is to perform both descriptive and prescriptive analytics:

1. Unsupervised clustering process to segment provinces and recognize how three economic features (average wage, average expenses, and poverty line of the citizens in the particular province) are related to each other.
2. Regression to acquire predictions on the poverty line, expenses, minimum wage, and wage projections of the next five years (until year 2028).

A. Data Acquisition

The entire dataset was retrieved from Kaggle [3], and the information within it was taken from the official site of Badan Pusat Statistik Indonesia (Central Bureau of Statistics Indonesia). It is in the form of four different .csv files each containing specific data on provinces, year, period of survey, region type (suburban, urban, etc.), and other features. Each csv files contain a table, and the tables contained in the dataset consist of the Garis Kemiskinan (Poverty Line)

Table, the Pengeluaran (Expenses) Table, the UMP (Minimum Wage) Table, and the Upah (Wage) Table.

1. The Poverty Line table is defined by the formula shown in fig 2 [11] below:

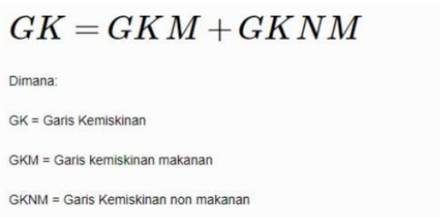

fig 2 : poverty line formula

Here, GK = Poverty Line, GKM = Food Poverty Line, and GKNM = Non-Food Poverty Line.

In essence, the poverty line is the minimum amount of money needed to fulfill both food needs (PLF) of 2100 kcals and non-food base needs such as housing and others. In all tables, the region is divided into village, urban, and suburban areas, and the period of survey is divided into March and September. The type consists of total (PL), food poverty line (PLF), and non-food poverty line (PLNF).

2. The expenses table consists of the expenses per capita data based on provincial, region, type and year. Both the region and type is the same as the poverty line table.

3. The minimum wage and wage table consists of corresponding data based on provincial and year data. The minimum wage is the minimum wage per month, and the wage is in rupiah per hour.

B. Data Preprocessing

After checking the tables briefly, some problems arise in data formats, especially on the Poverty Line and Expenses data. These two tables contain information such as type, region, and periodes. On the other hand, the table format the team would like to use would only need to include columns Province, Year, and the particular value (Wage, Expenses, etc.).

Therefore, for the preprocessing methods, the data is aggregated to achieve the final format. For both expenses and poverty line tables, only the "TOTAL" type is considered as it has correctly summed all other types in preceding rows. As for the regions, they are aggregated to the sum, to get the value of provincial output. Then the data is filtered to start from 2015 only, as some tables contain data as far as 2004 while others do not. The filtering decision is to uniform the analysis. Finally, the columns are renamed the columns for easier use.

province	year	pov_line_avg	pov_line_max
ACEH	2015	799732.0	830738.0
ACEH	2016	842996.7	873458.0

table 1.0: final format of Poverty Line Data

province	year	expenses
ACEH	2015	2403737.0
ACEH	2016	2601520.0

table 1.1: final format of Expenses Data

province	year	min_wage
ACEH	2002	330000.0
ACEH	2003	425000.0

table 1.2: final format of Minimum Wage Table

province	year	wage
ACEH	2015	11226
ACEH	2016	13627

table 1.3: final format of Wage Data

C. Exploratory Data Analysis

Once the data condition is guaranteed, a simple Exploratory Data Analysis (EDA) is performed. The output visualizations can be observed below.

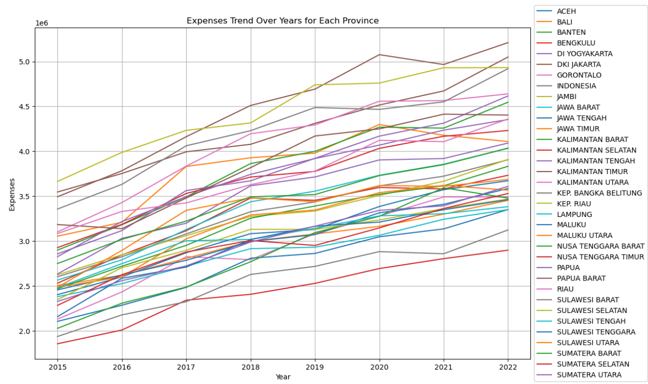

fig 3.0: Trend of Average Monthly Expenses for Every Province

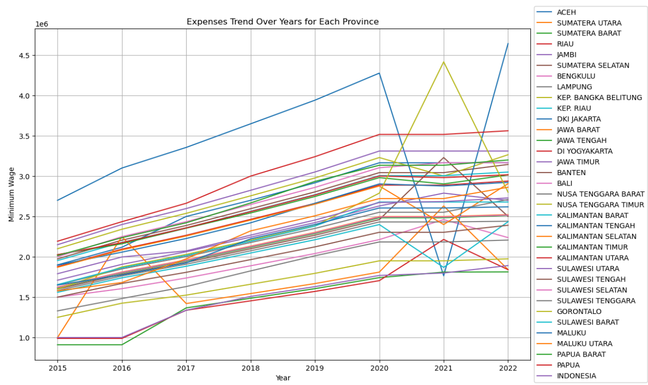

fig 3.1: Trend of Average Monthly Minimum Wage for Every Province

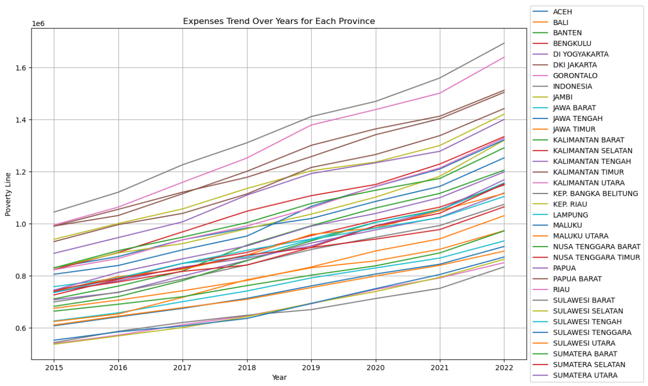

fig 3.2: Trend of Poverty Line for Every Province

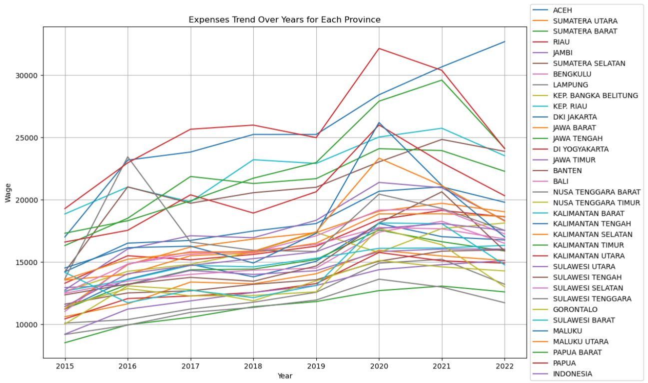

fig 3.3: Trend of Average Monthly Wage for Every Province

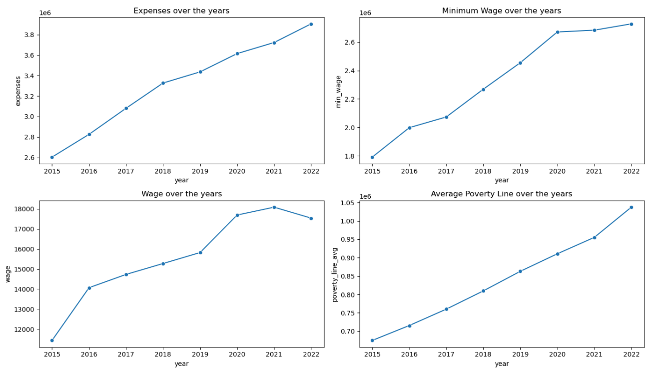

fig 3.4: overall trend Indonesia

It can be observed from the figures 3.0 - 3.4 that all expenses, poverty line, minimum wage, and wage have an upward trend, with an especially smooth trend on expenses and poverty line. This shows that in general, the future living costs have continued to increase throughout the years. On the other hand, the minimum wage and especially wage have more fluctuations and are more volatile in general.

In the pairplot visualization on fig. 3.5, all features have a rather positive correlation trend with each other, as they are slightly clustered and skewed positively. This shows that they have a relatively linear relationship.

Further proving the observation from before, in fig. 3.6, it is shown that each feature has a relatively high correlation with each other, on the emphasis on expenses with poverty line, wage, and minimum wage.

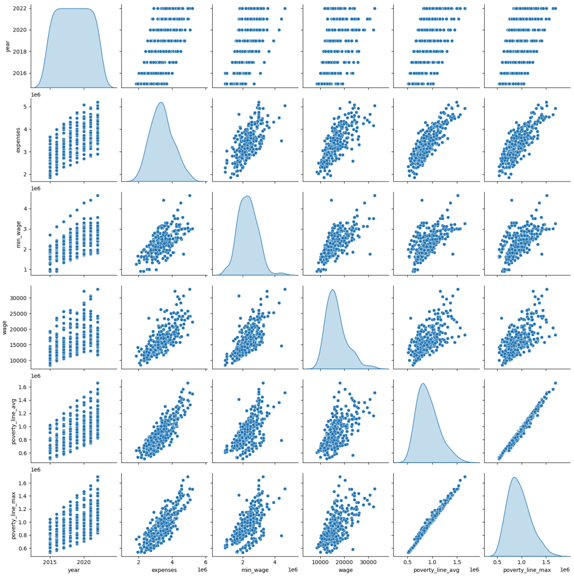

fig. 3.5: pairplot of features

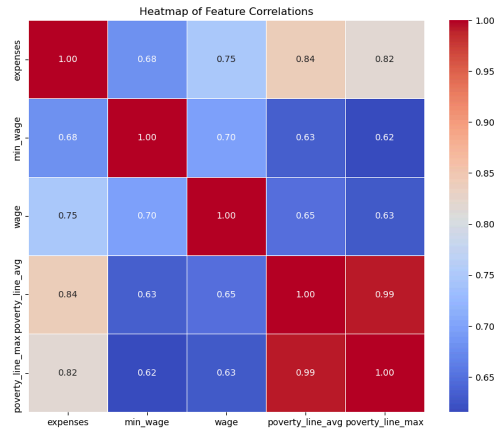

fig 3.6: correlation matrix

D. Clustering

The purpose of clustering is to see how a program automatically groups all 34 provinces and how each group/cluster's three economic features (average wage, average expenses, and poverty line) are distinct to other clusters'. To determine which clustering methods to be used, however, it is important to first observe the data point plots of all provinces. Since there are 3 features to consider, the plot will be in a 3D space, as seen in fig.4.1.

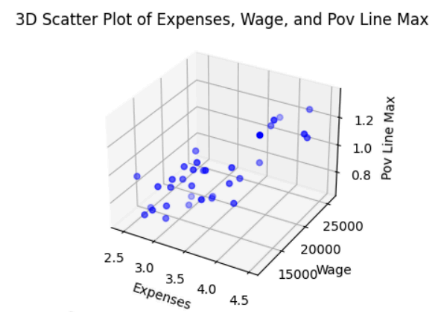

fig 3.7: correlation matrix

As seen from the scatter plot, the distribution of the 34 data points (each data point representing a province) would be most suitable with K-Means clustering. The density is too similar, thus the DBScan clustering algorithm will not be ideal. Next, the number of clusters (*k*) is found by applying the Elbow Method, before then following with the K-Means clustering itself.

E. Regression

To achieve the prediction results, regression is done on the dataset. The choice of applying polynomial regression instead of linear regression is because of the ability of the polynomial regression to capture intricate and nonlinear relationships between features. This is possible because of the incorporation of polynomial terms in the regression equation [12].

IV. Final Results and Discussion

A. Clustering Results

The visualization of data points after clustering can be seen in fig. 4.0. Although the cluster that each data point belongs to has been obtained, the "meaning" behind each cluster is still unclear. Therefore, a boxplot of the clustering results are retrieved, to see the relationships between the three features (average wages, average expenses, and poverty line) for each cluster.

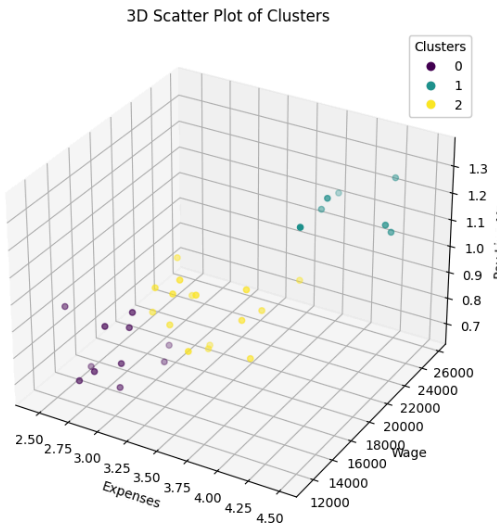

fig 4.0: 3D Cluster plot after K-Means Clustering

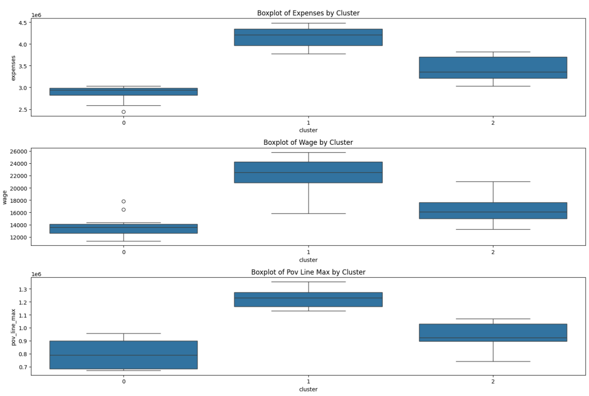

fig 4.1: Cluster interpretation plot

Fig. 4.1 is the boxplot that is used for interpretation of each cluster. According to the plot, cluster 0 consistently has the lowest average expenses, lowest average wages, and lowest poverty line out of all. Meanwhile, cluster 1 has the highest range in the three features, while cluster 2 is in the medium range for all. Derived from these findings, it can be said that the features are all directly related. Now, to summarize the results of how provinces are grouped, the team renames the clusters to reflect the actual meanings :

- Tier 1 Provinces → Highest in Average Expenses, Highest in Average Wages, and Highest in Maximum Poverty Line
- Tier 2 Provinces → Medium in Average Expenses, Medium in Average Wages, and Medium in Maximum Poverty Line
- Tier 3 Provinces → Lowest in Average Expenses, Lowest in Average Wages, and Lowest in Maximum Poverty Line

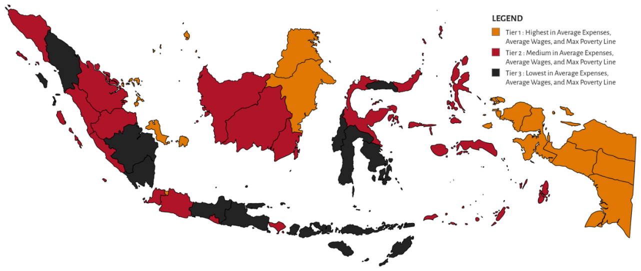

fig 4.2: Cluster Mapping of Indonesian Provinces, labeled by color.

It is notable that the most ideal (hypothetical) economical situation is to have high wages, low expenses, and low poverty line. Tier 1 provinces, which are highest in all three economic features, would mean that they are perhaps regions that give promising work to the population, yet require a high spending of living. In these Tier 1 provinces, it is likely that there is a huge gap between the families of lower and upper socioeconomic classes. Examples of provinces in Tier 1, taken from the map, are Kalimantan Timur (East Kalimantan), DKI Jakarta, Papua, and some few others who are widely known as either major urban cities or famous development sites. Meanwhile, Tier 3 provinces, which are lowest in all features, may be regions where the disparity in economy (income and expenditures) between families of different social classes are less obvious and more intertwined.

B. Regression Results

The following trends are results of the polynomial regression.

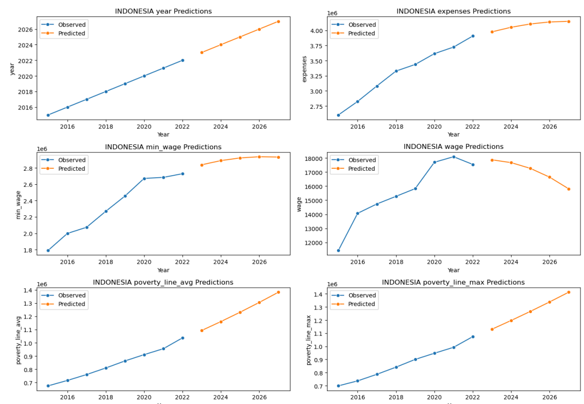

fig 4.3: Prediction on Indonesia

Unlike clustering, which focuses on how performance of each province and the groups they are categorized to, the regression is intended to focus on the predictions of average wages, average expenses, and poverty line in Indonesia as a whole.

From fig. 4.2, it is clear that Indonesian citizens' expenses and poverty line continues to increase, however the average wage is observed to decrease. This is an incredibly alarming prediction— a rise in the poverty line would mean a higher minimum spending for a good standard of living, but a decreasing wage means lower income for said spending. As such, the main point is the model expects that there will be a situation where the citizens' actual income may not be able to accommodate the rise in minimum spending, unless the standard of living decreases (life in poverty is more widespread).

As mentioned, however, such conclusion is for Indonesia as a nation. Now, regressions are also performed on the provinces of each cluster (Tier 1, Tier 2, and Tier 3) to see which Tier will perform better in terms of average wages, which is seen to decrease over time when referred to from the national trend.

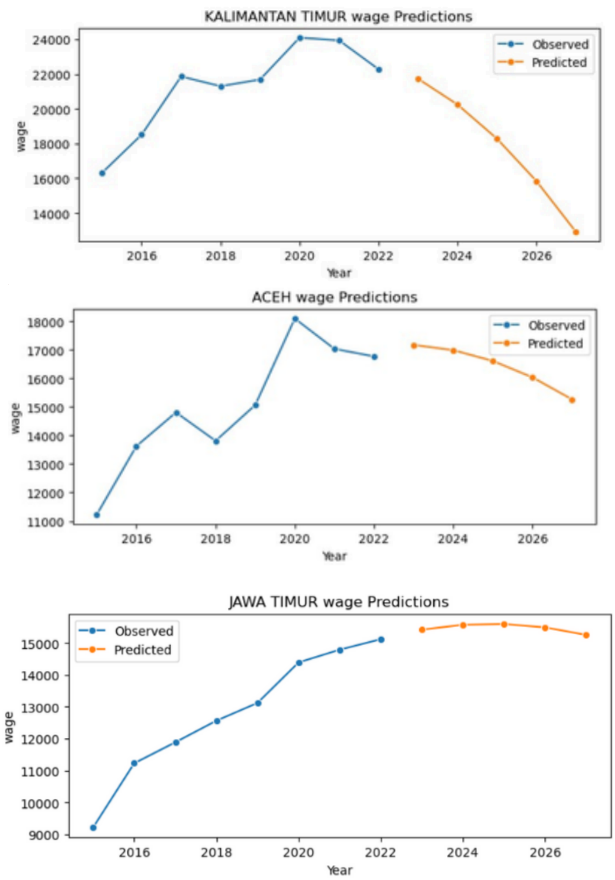

fig 4.3: Performance of Provinces in Tier 1, Tier 2, and Tier 3.

Kalimantan Timur is an example province which is part of Tier 1, while Aceh is part of Tier 2 and Jawa Timur part of Tier 3. Comparing the three provinces with one another, Kalimantan Timur has the highest rate of wage decrease, followed by Aceh and Jawa Timur, the latter of which has a nearly stagnant rate of wages (most ideal).

It is quite a surprising finding since Tier 3 provinces are known to have the lowest wages nationally, yet would be able to cope the best with increase in average expenses and rising poverty line by having a steady average wage for the next few years. On the other hand, Tier 1 and Tier 2 provinces, who are arguably in a better economic state as of the current moment, would experience greater struggle due to a predicted jump to lower wages.

This reaffirms the findings from the clustering process, where being in Tier 3 as a province does not mean having the worst economical state. Having a low average wage that can handle the low poverty line and low average expenses can be better in the long term, compared to having extreme high wages against a high poverty line and average expenses.

V. Conclusion

First, we predict from the hypothesis that the ratio of average monthly wages to the poverty line would remain constant or increase as inflation happens in the world. However, the danger is when the ratio between wages and poverty line worsens. It is predicted that many Indonesian provinces will experience a decrease in average wages for their populations, while the poverty line increases. This means that by 2028, the required money to live well is higher, but the money available to do so is less. While average wages are predicted to decline in Tier 1 provinces, which have the highest wages as of now, will experience the biggest rate in wage decrease.

Secondly, from the clustering result (map), it seems that provinces with highest wages are also highest in poverty lines. The advanced regions (mostly Tier 1) may have harder economic challenges where the economic disparity is high and the gap between bottom and top of the economic class is bigger than regions with low wages and low poverty lines.

In conclusion of this project, there is a great inequality in worker welfare between provinces in Indonesia, emphasizing the critical necessity for government support in areas facing the most severe economic challenges. The relationships between wages, poverty lines and living expenses are proving crucial in informing any strategy that seeks to raise living standards or reduce poverty in Indonesia.

Acknowledgment

The authors are grateful to Mrs. Yunita Sari, S.Kom., M.Sc., Ph.D., the university lecturer on Machine Learning, for her insights and teachings that made this project possible.

References

[1] European Commission "Reducing inequalities is essential to ensure sustainable development benefits all, especially those furthest behind" [Online] Available at: "https://international-partnerships.ec.europa.eu/news-and-events/stories/reducing-inequalities-essential-ensure-sustainable-development-benefits-all-especially-those_en"

[2] Article "Kemiskinan di Indonesia Maret 2023" [Online] Available: "https://www.bps.go.id/id/pressrelease/2023/07/17/2016/profil-kemiskinan-di-indonesia-maret-2023.html"

[3] Dataset "Kesejahteraan Pekerja Indonesia" [Online] Available: "https://www.kaggle.com/datasets/rezkyyayang/pekerja-sejahtera

[4] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[5] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[6] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[7] K. Elissa, “Title of paper if known,” unpublished.

[8] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[9] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[10] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[11] “Garis Kemiskinan Formula” [Online] Available: https://pusaka.magelangkab.go.id/metadata/indikator/detailIndikator/17

[12] Mo, H. (2023). Comparative analysis of linear regression, polynomial regression, and ARIMA model for short-term stock price forecasting. In Proceedings of the 2nd International Conference on Financial Technology and Business Analysis (DOI: 10.54254/2754-1169/49/20230509). Syracuse University, U.S.

Roles

Vian Sebastian Bromokusumo

1. Literature review
2. Data Acquisition, Preparation, and Cleaning
3. EDA
4. Regression

Sasha Annabel

1. Clustering analysis
2. Poster Design
3. Report

Nabilla Nafisatuz Zahrah

1. Report
2. Poster Design
3. Powerpoint

Google Colab Link

[1] Data Preprocessing, EDA : https://drive.google.com/file/d/19-L1FPyWWkx-SrnaZkRzFbBn6I4-N7o_/view?usp=sharing

[2] Clustering : https://colab.research.google.com/drive/1hxfmDPQRM8bHLMHlk5gyReUeztG6mq17?usp=sharing

[3] Regression : https://drive.google.com/file/d/1c9f5Awtrm_k241CXT8u3dP5AI5-wcAuu/view?usp=sharing

Poster & PPT Link

https://drive.google.com/file/d/1mxyuVF0yO-hktJ_3KUU299tHvHwflQFU/view?usp=sharing

All-In Drive Folder Link

https://drive.google.com/drive/folders/1jBStCyWQ9-GhWC69nSxJS03nwxEMiksa?usp=sharing