

Analysis of CSR Program Against Regional Inequality in Bogor Regency Using K-Means and Random Forest Algorithms

Muhammad Ardiansyah Rizkiyanto^{*1}, Sahril Sabirin², Arief Wibowo³

^{1,2} Magister of Computer Science, Faculty of Information Technology, Budi Luhur University, Indonesia

³ Master of Computer Science, Faculty of Information Technology, Budi Luhur University, Indonesia

E-mail: ¹2311600379@student.budiluhur.ac.id, ²2311600585@student.budiluhur.ac.id, ³arief.wibowo@budiluhur.ac.id

Abstract

Bogor Regency is vast and has significant economic and social potential. Collaboration between businesses and local governments is essential to achieve regional development goals. Corporate Social Responsibility (CSR) plays a role in sustainable economic, enhancing the quality of life for the community. CSR can be implemented independently by companies or supported by the CSR Support Group (TF-TJSL). The Bogor Regency is divided into three development areas, the Western, Central and Eastern, with regional inequality reflected in a Williamson index of 0.731. CSR has the potential to reduce these inequalities through positive contributions. This study analyzes CSR programs on regional inequality in Bogor Regency using data mining technology with K-Means and Random Forest algorithms. The K-Means algorithm shows the optimal result with the best silhouette score at K=2 with a score of 0.76268, reflecting a clear separation between clusters representing regional inequality. The Random Forest algorithm shows excellent classification ability with an accuracy of 0.985 and other evaluations of precision, recall, and f1-score are almost perfect, which indicates its effectiveness in classifying data into three clusters according to development areas. The regression model evaluation results are also good, with a very low MSE (0.003961), indicating minimal prediction error.

Keywords — CSR, K-Means, Random Forest, Regional Inequality

1. INTRODUCTION

Bogor Regency consists of 40 sub-districts and 435 villages/urbans with an area of 299,178 Ha ^[1]. With an area 4 times larger than the Special Region of Jakarta, Bogor Regency has great economic and social potential. This requires effective collaboration between companies and local government to achieve regional development goals. Corporate Social Responsibility (CSR) is a company's commitment to participate in sustainable economic development in order to improve the quality of life and the environment that benefits both the company itself, the local community and society in general ^[2].

Regional development is an effort to improve the welfare of the community and the progress of a region ^{[3], [4]}. These efforts are outlined in regional planning documents, such as the Work Plan (Renja), the Regional Medium-Term Development Plan (RPJMD), and the

Regional Long-Term Development Plan (RPJPD). The three planning documents must be prepared synergistically and complement each other to achieve regional development goals ^[5]. The annual Renja must refer to the five-year RPJMD, while the RPJMD must be prepared based on the twenty-year RPJPD. Thus, regional development can run systematically, purposefully, and continuously.

Bogor Regency is led by the Regent as the Regional Head who has a vision and mission to be achieved. Based on the vision, mission, and program of the Regional Head, the RPJMD document will be prepared. To realize the vision and mission outlined in the achievement of target indicators of goals and objectives, it needs to be done with the right strategy and based on clear policy direction ^[6]. CSR program is expected to make a positive contribution to the regional development that has been prepared. CSR programs can run effectively and efficiently if there is good coordination and communication between the government and the private sector (companies) ^[7], ^[8]. The implementation of CSR can be carried out with an independent pattern and a pattern facilitated by the CSR Facilitation Team (TF-TJSL) ^[9]. An independent pattern means that the company can carry out CSR programs independently according to a predetermined agenda, both in the immediate environment of the operational area and in other areas outside the operational area ^[9]. The programs that have been implemented are then reported to the Bogor Regency Government. The facilitated pattern means that the CSR actor submits a request for TF-TJSL facilitation to synergize programs that support the achievement of Bogor District Government performance, especially outside the company's operational area and which are not budgeted in the APBD/APBN ^[9].

Understanding the effectiveness of CSR programs can help identify potential improvements, design more effective strategies and strengthen collaboration between the business sector, government and communities to achieve sustainable regional economic development goals ^[10]. The beneficiaries of the CSR programs expect the programs to continue and be sustainable as they are considered to help the community in improving the quality and welfare of life ^[11].

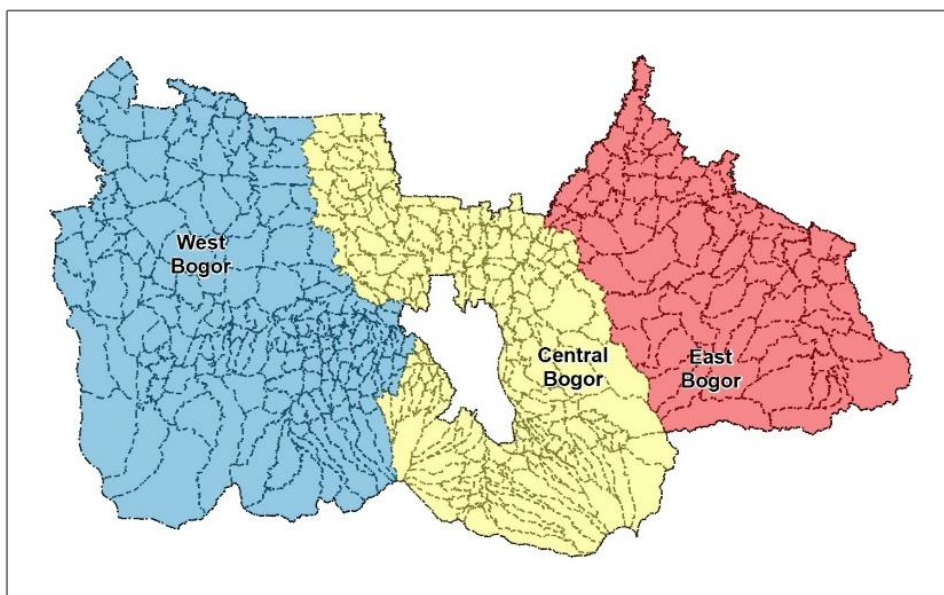


Figure 1. Map of Development Areas (WP) of Bogor Regency

Bogor Regency as one of the buffer areas of the Special Region of Jakarta, has enormous regional potential such as tourism, agriculture, plantations, mining, and others. The development of each potential region is centered on development areas in Bogor Regency. Bogor Regency is mapped into 3 development areas (WP), namely Western WP, Central WP, and Eastern WP ^[1]. The Western WP consists of 14 sub-districts, namely Cigudeg, Jasinga, Sukajaya, Nanggung, Parungpanjang, Tenjo, Rumpin, Leuwiliang, Cibungbulang, Leuwisadeng, Pamijahan, Dramaga, Ciampea, and Tenjolaya. The Central WP consists of 19 sub-districts, namely Cibinong, Bojonggede, Sukaraja, Babakan Madang, Citeureup, Tajurhalang, Parung, Kemang, Ciseeng, Rancabungur, Gunungsindur, Cigombong, Caringin, Cijeruk, Tamansari, Ciomas, Ciawi, Cisarua, and Megamendung. The Eastern WP consists of 7 sub-districts, namely Cileungsi, Gunung Putri, Klapanunggal, Jonggol, Cariu, Sukamakmur, and Tanjungsari. Industrial designation areas are spread across the three WPs with 2 sub-districts in the Western WP, 10 sub-districts in the Central WP, and 4 sub-districts in the Eastern WP ^[1]. From this distribution, it is known that only 14.29% of sub-districts in the Western WP are industrial designation areas, 52.63% of sub-districts in the Central WP, and 57.14% of sub-districts in the Eastern WP. This shows that the number of companies in the Western WP has the smallest number compared to the Central and Eastern WPs.

Tabel 1. Williamson Index by Development Area in Bogor Regency

WP	PDRB (Rp million)	Population (people)	PDRB per Capita (Rp thousand)	(Yi-Y)^2	fi/n	(Yi-Y)^2*(fi/n)	
Eastern	75,054,498.71	1,066,202.00	70.39	1,709.62	0.19	332.05	
Central	68,533,596.81	2,816,705.00	24.33	22.24	0.51	11.41	
Western	15,864,775.90	1,606,629.00	9.87	367.57	0.29	107.58	
Total	159,452,871.41	5,489,536.00	29.05			451.04	
						[(Yi-Y)^2 * (fi/n)]^0.5	21.238
						Williamson Index = [(Yi-Y)^2 * (fi/n)]^0.5 / (PDRB Per Capita)	0.731

Source: Results of Economic Analysis in the Bogor District RTRW Revision Study, 2022

Bogor Regency has a very large area with diverse topography, from lowlands to mountains, making regional development interventions in all regions cannot be fulfilled optimally, which can result in inequality between regions. Regional inequality is a common problem found in many regions ^[12]. Regional inequality can be measured using the Williamson Index ^[13]. Based on Table 1, it is known that Bogor Regency has a Williamson Index of 0.731 points, which indicates moderate regional inequality. This condition can be seen based on the Gross Regional Domestic Product (PDRB) per capita of each WP, where the Western WP has a gap with the Eastern WP and the Central WP. A large gap between one region and another can lead to many complex social and economic problems, such as creating a welfare gap in the various regions ^[14].

Given these regional inequalities, the CSR program has great potential to help reduce regional inequalities in Bogor Regency by making positive contributions to various aspects of community life in disadvantaged areas. Through well-targeted initiatives, companies can help improve the quality of education, health, infrastructure, economy and environment in areas that require special attention. Therefore, it is necessary to conduct an analysis of CSR program that have been carried out by companies, so that an evaluation of the implementation of CSR programs can be conducted. The results of the analysis and evaluation can assist the Bogor

Regency Government in formulating policies or strategies in cooperating with companies to succeed in regional development.

Based on these problems, this research will analyze the CSR program with a systematic and data-based approach using data mining technology. The technology used is using the K-Means and Random Forest algorithms. The K-Means algorithm can group data into clusters based on similarities between data ^[15]. Each cluster can be analyzed to identify the main characteristics that distinguish one cluster from another, making it easier to understand the different CSR strategies among companies. The Random Forest algorithm is able to deal with unbalanced data ^[16], where some categories of CSR may be more numerous than others.

2. RESEARCH METHOD

The research method used in this research is a quantitative method with data analysis techniques using the K-Means and Random Forest Algorithms. The data used is secondary data in the form of data on the implementation of CSR programs in Bogor Regency from January 2019 - July 2024 with a total of 1,991 rows. The data consists of 16 columns, namely nama_perusahaan, id_perusahaan, alamat_perusahaan, tahun, program_tjsl, nilai_anggaran, volume, satuan, kategori, jenis_rencana_kerja, id_kecamatan, nama_kecamatan, id_desa, nama_desa, alamat, and keterangan.

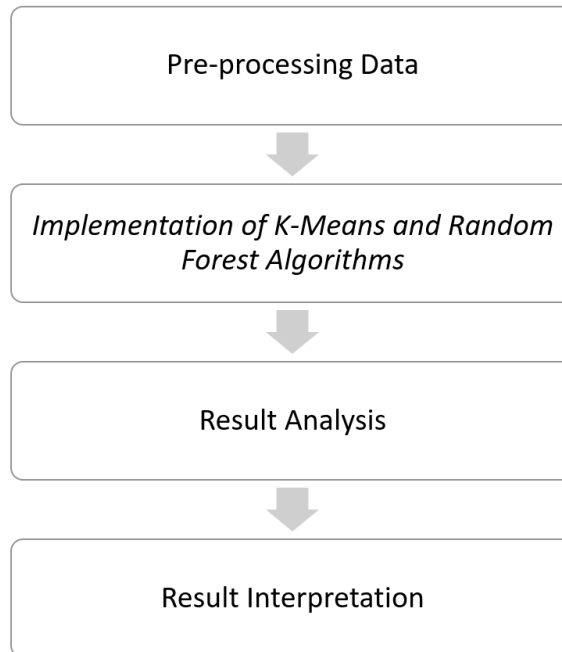


Figure 2. Research Method

2.1. *Pre-processing Data*

This stage ensures that the data used in the analysis is clean, relevant, and ready to be processed by the algorithm. In this stage, data cleansing, data transformation, and data reduction will be performed.

2.1.1. Data Cleansing

Data cleansing will handle missing values by removing data that has many missing values and handle data duplication by removing duplicate data that can cause bias in the analysis. This process will remove data that has null values in the features to be used and will also remove duplicate data rows.

2.1.2. Data Transformation

Data transformation will convert categorical data into numerical format (label encoding). Label encoding works by identifying all category values in a feature, then assigning each category a unique numerical value. After that, the category values are converted into numerical values according to the mapping that has been created. This technique allows category data that was originally text to be converted into numbers so that it can be processed by machine learning models.

2.1.3. Data Reduction

Data Reduction is a process in data processing that aims to simplify data by reducing the number of features, samples, or dimensions without losing significant information. One technique that is often used in data reduction is the removal of irrelevant features.

2.2. Implementation of K-Means and Random Forest Algorithms

This stage processes data that has gone through the pre-processing stage using the K-Means and Random Forest algorithms. K-Means algorithm will perform classification. Random Forest algorithm will perform classification and regression. This implementation uses the python programming language by using the scikit-learn library.

2.3. Result Analysis

This stage will analyze the results of each algorithm used, K-Means and Random Forest algorithms.

2.3.1. K-Means Algorithm

The K-Means algorithm will evaluate the classification model using the Silhouette Score which can be calculated using formula (1) ^[17].

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

2.3.2. Random Forest Algorithm

The Random Forest algorithm will use a confusion matrix consisting of 4 components, namely True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), for classification model evaluation. TP is correct positive predictions (the model predicts positive, and it is actually positive). FP is incorrect positive predictions (the model predicts positive, but it is actually negative). TN is correct negative predictions (the model predicts negative, and it is actually negative). FN is incorrect negative predictions (the model predicts negative, but it is actually positive) ^[18].

		Predicted class	
		Positive	Negative
True class	Positive	TP	FN
	Negative	FP	TN

Figure 3. Confusion Matrix

These components will be used to calculate accuracy which can be calculated using formula (2), precision using formula (3), recall using formula (4), and f1-score using formula (5) ^[18]. For regression model evaluation using mean squared error (MSE) which can be calculated using formula (6) ^[18].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

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

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \quad (6)$$

2.4. Result Interpretation

The result interpretation stage is an important step to understand and convey the meaning of the analysis that has been done. In the K-Means algorithm, it can be assessed how well the objects are clustered by checking the Silhouette Score value. A value close to 1 indicates good clustering. In the Random Forest algorithm, the accuracy and balance of the model in predicting the class can be assessed. So that from the interpretation of the results of these two algorithms, recommendations can be made based on the characteristics of the cluster found.

3. RESEARCH RESULTS AND DISCUSSION

In this research, the data has been successfully pre-processed. The first process is data cleansing by deleting data containing null values in the year, category, id_kecamatan, and id_desa columns. Then the removal of duplicate data rows. Next, the data transformation process is carried out by converting categorical data into numeric (label encoding) on columns of kategori, jenis_rencana_kerja, and kecamatan (converted into WP based on the division of development areas). Finally, data reduction is performed by selecting or removing irrelevant data, namely the nilai_anggaran, nama_perusahaan, alamat_perusahaan, program_tjsl, volume, satuan, nama_kecamatan, nama_desa, alamat, and keterangan. This data pre-processing makes the data clean, the data has reduced the number of rows from 1,991 to 1,000 rows of data.

```
data = data.dropna(subset=['tahun'])
data = data.dropna(subset=['kategori'])
data = data.dropna(subset=['id_kecamatan'])
data = data.dropna(subset=['id_desa'])
data.drop_duplicates(inplace=True)
```

Figure 4. Data Cleansing Remove Null and Duplicate Data

```
kategori_mapping = {
    'Lingkungan Hidup': 1,
    'Sosial': 2,
    'Pendidikan': 3,
    'Kesehatan': 4,
    'Lainnya': 5
}
data['kategori'] = data['kategori'].map(kategori_mapping)
```

Figure 5. Data Transformation Convert “Kategori” Data from Categorical to Numeric Format

```

wp_mapping = {
  'Nanggung':1,
  'Leuwiliang':1,
  'Leuwisadeng':1,
  'Pamijahan':1,
  'Cibungbulang':1,
  'Ciampea':1,
  'Tenjolaya':1,
  'Dramaga':1,
  'Rumpin':1,
  'Cigudeg':1,
  'Sukajaya':1,
  'Jasinga':1,
  'Tenjo':1,
  'Parung Panjang':1,
  'Ciomas':2,
  'Tamansari':2,
  'Cijeruk':2,
  'Cigombong':2,
  'Caringin':2,
  'Ciawi':2,
  'Cisarua':2,
  'Megamendung':2,
  'Sukaraja':2,
  'Babakan Madang':2,
  'Citeureup':2,
  'Cibinong':2,
  'Bojong Gede':2,
  'Tajurhalang':2,
  'Kemang':2,
  'Ranca Bungur':2,
  'Parung':2,
  'Ciseeng':2,
  'Gunung Sindur':2,
  'Sukamakmur':3,
  'Cariu':3,
  'Tanjungsari':3,
  'Jonggol':3,
  'Cileungsi':3,
  'Klapanunggal':3,
  'Gunung Putri':3
}
data['wp'] = data['nama_kecamatan'].map(wp_mapping)

```

Figure 6. Data Transformation Convert “Kecamatan” Data from Categorical to Numeric Format

```

jenis_rencana_kerja_mapping = {
  '-':0,
  'Rencana Kerja TJSJ': 1,
  'Rencana Kerja Perusahaan': 2
}
data['jenis_rencana_kerja'] = data['jenis_rencana_kerja'].map(jenis_rencana_kerja_mapping)

```

Figure 7. Data Transformation Convert “Jenis Rencana Kerja” Data from Categorical to Numeric Format

```

data = data.drop('nilai_anggaran', axis=1)
data = data.drop('nama_perusahaan', axis=1)
data = data.drop('alamat_perusahaan', axis=1)
data = data.drop('program_tjstl', axis=1)
data = data.drop('volume', axis=1)
data = data.drop('satuan', axis=1)
data = data.drop('nama_kecamatan', axis=1)
data = data.drop('nama_desa', axis=1)
data = data.drop('alamat', axis=1)
data = data.drop('keterangan', axis=1)
    
```

Figure 8. Data Reduction Remove Irrelevant Features

The K-Means algorithm produced the best value with 2 clusters with cluster 1 consisting of 83% data and cluster 2 consisting of 17% data. The CSR program in Western WP has 12.90% of the data in cluster 1 and 87.10% in cluster 2, Central WP has 84.55% of the data in cluster 1 and 15.45% in cluster 2, and Eastern WP has 85.92% of the data in cluster 1 and 14.08% in cluster 2. These results show that Eastern WP and Central WP have data distribution patterns that are more similar to each other than Western WP.

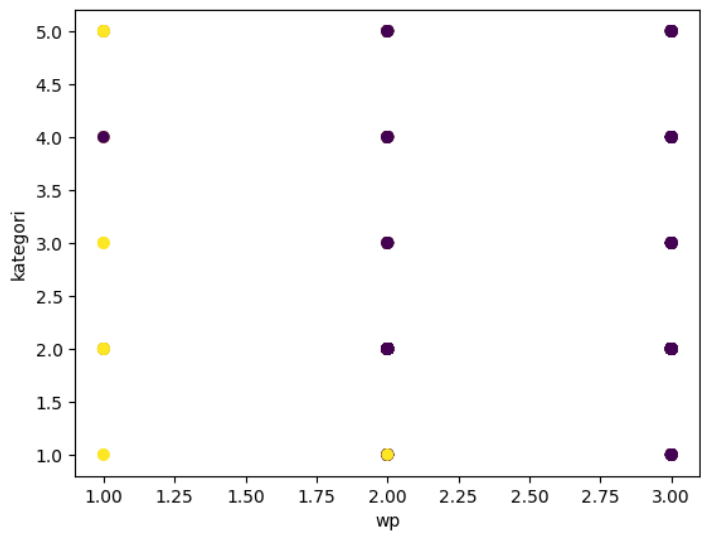


Figure 9. K-Means Clustering With K = 2

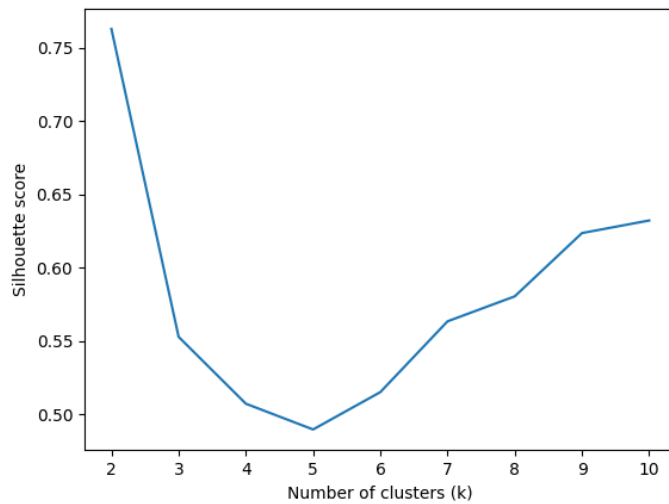


Figure 10. Silhouette Score Analysis

To measure how well the objects in a cluster are separated from objects in other clusters, the silhouette score is calculated. From the calculation of each number of clusters, the best silhouette score is obtained at $K = 2$ with a score of 0.76268 and the worst silhouette score at $K = 5$ with a score of 0.48971.

The implementation of Random Forest algorithm uses 80% data as training data and 20% data as testing data. Based on the data, this algorithm classifies the data into 3 clusters according to WP, where cluster 1 (Western WP) consists of 4 data (1.68%), cluster 2 (Central WP) consists of 122 data (51.26%), and cluster 3 (Eastern WP) consists of 112 data (47.06%). The Random Forest algorithm is also capable of performing regression to model and analyze the relationship between variables. Figure 11 shows that there are some points that are far from the reference line, indicating that the Random Forest model does not always provide accurate predictions. However, there are also some points that are very close to the line, indicating that the model has some correct predictions.

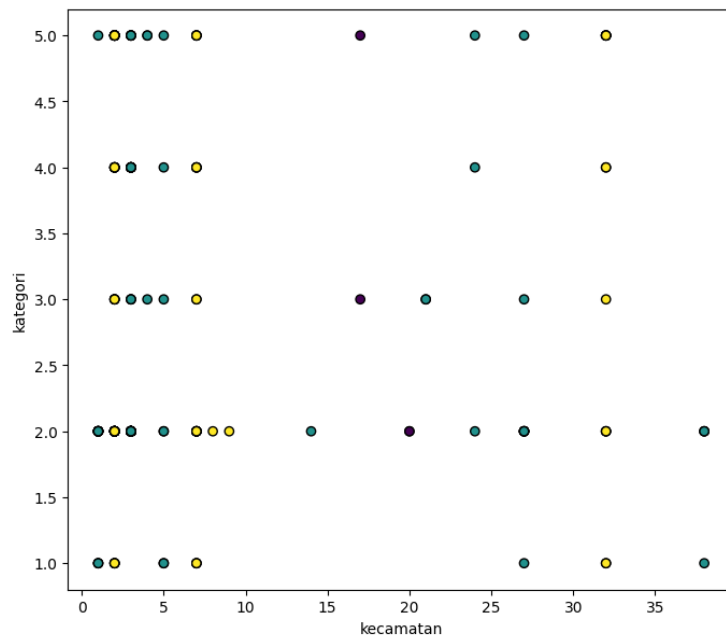


Figure 11. Random Forest Classification

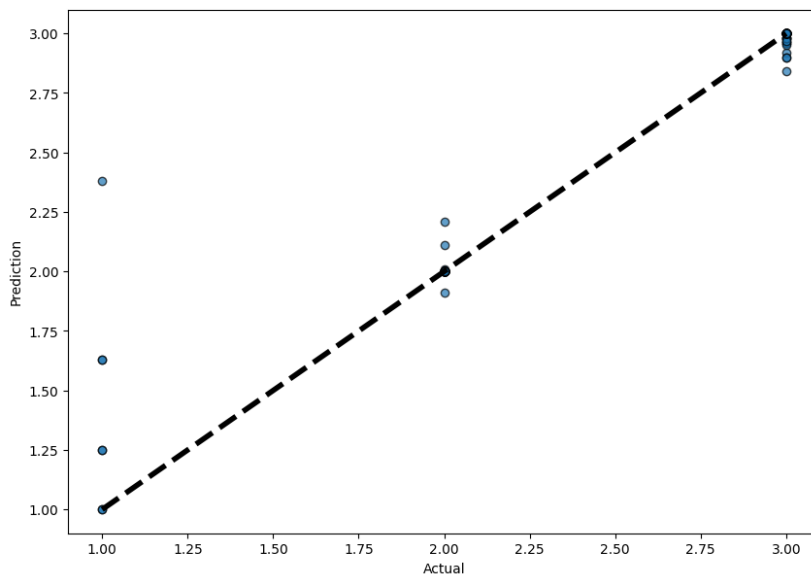


Figure 12. Random Forest Regression

To evaluate the performance of the Random Forest algorithm classification model, a confusion matrix is used. The resulting confusion matrix in Figure 12 illustrates class 0 (Western WP) with TP as many as 4, FP as many as 0, FN as many as 3, and TN as many as 193. Class 1 (Central WP) with 103 TP, 3 FP, 0 FN, and 94 TN. Class 2 (Eastern WP) with 90 TP, 0 FP, 0 FN, and 107 TN. From the confusion matrix, the accuracy, precision, recall, and f1-score values are calculated. The accuracy value obtained is 0.985, precision is 0.985459, recall is 0.985 and f1-score is 0.985001. Then, to evaluate the performance of the Random Forest algorithm regression model, the mean squared error (MSE) is used. The MSE value obtained from this model is 0.003961.

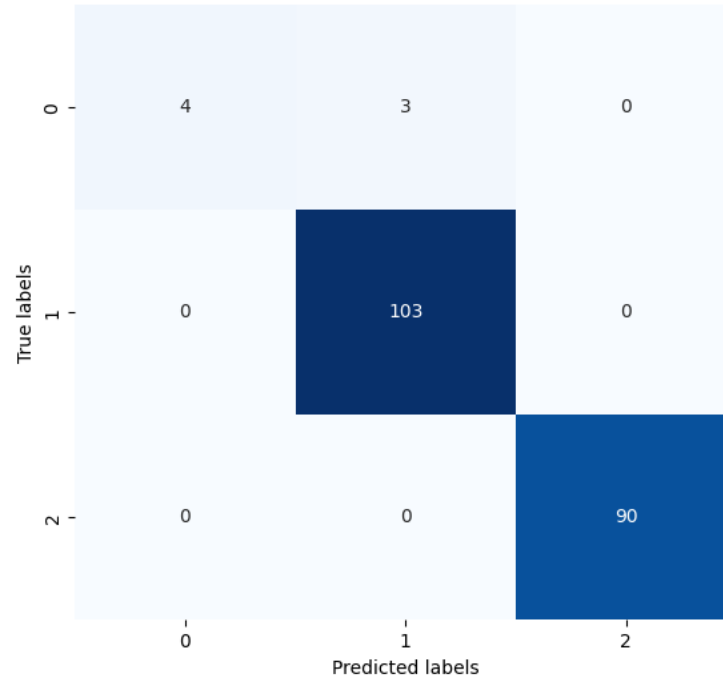


Figure 13. Confusion Matrix of Random Forest Result

The K-Means algorithm that has the best silhouette score is with the number of clusters of 2 ($K=2$). The resulting 2 clusters illustrate that the objects in the cluster are quite clearly separated between cluster 1 and cluster 2. This is in line with the condition of regional inequality in Bogor Regency, where there is a gap between the Western WP with the Eastern WP and the Central WP. The implementation of the Random Forest algorithm on this data uses 80% of the data as training data and 20% as testing data. The classification results show that this algorithm is able to classify the data well into 3 clusters based on WP. Performance evaluation of the classification model using confusion matrix showed excellent values with accuracy of 0.985, precision of 0.985459, recall of 0.985, and f1-score of 0.985001. These values indicate that the Random Forest model is very effective in classifying the data into the correct clusters. In addition, evaluation of the regression model using mean squared error (MSE) resulted in a value of 0.003961, which indicates that the model has a very small prediction error.

The results of the analysis using the K-Means and Random Forest algorithms provide a clear picture of regional inequality in Bogor Regency, as well as practical guidance for stakeholders to take appropriate action. By knowing that there are two main clusters describing the inequality between the Western WP, Eastern WP and Central WP, stakeholders can plan more targeted development, focus on the development of disadvantaged regions and direct CSR programs more precisely. The Random Forest classification model shows a remarkable ability to classify data into appropriate groups, which can also be used to accurately monitor

and evaluate program effectiveness. With this data-driven approach, decision-making becomes smarter and more scalable, allowing for more effective policy adjustments to address inequality. Collaboration between local governments and the private sector is also essential to reduce inequality between regions through CSR programs that are based on in-depth data analysis. The implementation of these measures is expected to bring real and equitable benefits to all communities in Bogor Regency.

4. CONCLUSION

This research has successfully conducted data pre-processing which includes data cleansing, data transformation, and data reduction. These processes are important to ensure the quality of data used in analysis and modeling. The K-Means algorithm showed optimal results with the best silhouette score at a cluster count of 2, reflecting a clear separation between clusters representing regional inequality in Bogor Regency. In addition, the Random Forest algorithm showed excellent classification capability with an accuracy of 0.985 and other evaluation values (precision, recall and f1-score) close to 1, indicating its effectiveness in classifying the data into three clusters by region. The regression model evaluation results also performed well, with a very low MSE (0.003961), indicating minimal prediction error. Thus, this research provides insight into the relationship between the implementation of corporate social responsibility and the condition of regional inequality in Bogor Regency through the application of the K-Means and Random Forest algorithms. These results are expected to be taken into consideration in decision-making or policy of the Bogor Regency Government in cooperating with companies to succeed in regional development by maximizing the role of TF-TJSL so that every region in Bogor Regency benefits from the implementation of CSR. Stakeholders can use this information to plan more targeted development, focusing on less developed regions. The information from the data can help identify priority regions that need immediate intervention and design more effective and efficient policies.

5. SUGGESTED

In future research, it is recommended to include more variables related to social and economic conditions, such as education level, health access, or unemployment rate, which may provide a more holistic picture of regional inequality. Further research could also focus more on evaluating the impact of specific CSR programs in different regions of Bogor Regency, by analyzing whether these programs have significantly reduced inequality.

6. REFERENCES

- [1] Kabupaten Bogor, “Peraturan Daerah Kabupaten Bogor Nomor 1 Tahun 2024 Tentang Rencana Tata Ruang Wilayah Kabupaten Bogor Tahun 2024-2044,” *Peraturan Daerah Kabupaten Bogor*, 2024.
- [2] Kabupaten Bogor, “Peraturan Daerah Kabupaten Bogor Nomor 1 Tahun 2023 Tentang Tanggung Jawab Sosial dan Lingkungan Perusahaan,” *Peraturan Daerah Kabupaten Bogor*.
- [3] M. Teja, “Pembangunan Untuk Kesejahteraan Masyarakat di Kawasan Pesisir,” *Aspirasi*, vol. 6, no. 1, pp. 63–76, 2015.

- [4] A. Rahman, "Identifikasi Strategi Peningkatan Kesejahteraan Masyarakat di Kecamatan Sungaiambawang Kabupaten Kuburaya Provinsi Kalimantan Barat," *Jurnal Manajemen Pembangunan*, vol. 5, no. 1, pp. 17–36, 2018.
- [5] A. Agustar, I. W. Syarfi, and E. Wati, "Sinkronisasi Antara Perencanaan Nagari dengan Perencanaan Pembangunan Daerah di Kabupaten Solok," *Jurnal Ilmiah Wahana Bhakti Praja*, vol. 11, no. 1, pp. 18–46, May 2021, doi: 10.33701/jiwbp.v11i1.1573.
- [6] Kabupaten Bogor, "Peraturan Daerah Kabupaten Bogor Nomor 3 Tahun 2021 Tentang Perubahan Atas Peraturan Daerah Nomor 4 Tahun 2019 tentang Rencana Pembangunan Jangka Menengah Daerah Tahun 2018-2023," *Peraturan Daerah Kabupaten Bogor*.
- [7] L. Trisnawati, A. Syamsuadi, S. Hartati, and I. Reskiyanti, "Koordinasi Pemerintah dan Swasta dalam Program Corporate Social Responsibility (CSR) School Improvement di Kabupaten Pelalawan," *Journal of Governance and Local Politics*, vol. 03, no. 02, pp. 115–123, 2021.
- [8] B. Suryadi, P. Adaro, F. Ilmu Sosial dan Ilmu Politik, and U. Lambung Mangkurat, "Sinergitas Pemerintah Daerah Kabupaten Tabalong dan PT Adaro Indonesia dalam pelaksanaan Program CSR," *JPP: Jurnal Administrasi Publik dan Pembangunan*, vol. 4, no. 1, pp. 10–18, 2022.
- [9] Bupati Bogor, "Surat Edaran Bupati Bogor tentang Pengoordinasian Pelaksanaan TJSL Perusahaan di Kabupaten Bogor Tahun 2023 - 2028," 2023.
- [10] Nasfi, Sattar, N. Y. Simbolon, P. H. Susanti, A. Rosid, and Handayani, "Analisis Efektivitas Program Corporate Social Responsibility (Csr) Dalam Mendorong Mendorong Pembangunan Ekonomi Wilayah Kota Padang Panjang Provinsi Sumatera Barat," *Journal of Human And Education*, vol. 4, no. 1, pp. 10–18, 2024.
- [11] Hildawati and D. Afrizal, "Kepuasan Masyarakat Terhadap Pelaksanaan Program Corporate Social Responsibility (CSR) PT. Pertamina RU II Dumai," *JIANA: Jurnal Ilmu Administrasi Negara*, vol. 19, no. 2, p. 153167, 2021.
- [12] M. Nasution, "Ketimpangan Antar Wilayah & Hubungannya dengan Belanja Pemerintah: Studi di Indonesia," *Jurnal Budget*, vol. 5, no. 2, p. 84102, 2020.
- [13] I. Al Majiid and Woyanto Neni, "Analisis Ketimpangan Wilayah dan Potensi Ekonomi di Kawasan Kedungsepur Tahun 2017-2021," *BISECER (Business Economic Entrepreneurship)*, vol. VI, no. 1, pp. 25–37, 2023.
- [14] Wahidin, T. Yuniarti, and E. Astuti, "Kajian Tingkat Kemajuan Daerah Kabupaten/Kota dan Ketimpangan Wilayah Antar Kabupaten/Kota di Pulau Lombok Propinsi Nusa Tenggara Barat," *ELASTISITAS: Jurnal Ekonomi Pembangunan*, vol. 4, no. 1, pp. 14–24, 2022.
- [15] S. N. Mayasari and J. Nugraha, "Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022," 2023.
- [16] R. O'Brien and H. Ishwaran, "A random forests quantile classifier for class imbalanced data," *Pattern Recognit*, vol. 90, pp. 232–249, Jun. 2019, doi: 10.1016/j.patcog.2019.01.036.
- [17] S. P. Lloyd, "Least Squares Quantization in PCM," 1982.
- [18] L. Breiman, "Random Forests," *Kluwer Academic Publishers*, vol. 45, pp. 5–32, 2001.