Maintaining Religious Harmony through Predicting the Level of Lawlessness using Linear Regression

Adam Wardana
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
adamwardana25@gmail.com

Dekis Aldamawan
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
dekisaldamawan12@gmail.com

M. Rival Ridautal Lillah
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
rivallmuhammad2@gmail.com

Mochamad Anwar Wirahadi
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
mochamadanwar403@gmail.com

Mohamad Rasyid Ridho
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
mohamadrasyidridho11@gmail.com

Naufal Rizqullah
Program and Partnership
Google Developer Student Clubs
Bandung, Indonesia
naufalrizqullah23@gmail.com

Abstract— Crimes are basically committed due to low human faith and ultimately disrupt harmony, including inter-religious harmony. Therefore, the purpose of this study is to predict and determine the number of law violations that occurred in Yogyakarta Province in 2022 as a case study using a linear regression algorithm based on data ranging from 2012 to some of the data on law violations in 2022. The method used is the forecasting method in which the stages of data collection, data preprocessing, data modeling, and prediction processes are carried out. The accuracy of the linear regression model is 90%. However, there is a difference in accuracy between the training data and the test data. The training data gets an accuracy of 99% while the accuracy achieved by the test data is at 94% which is 5% lower than the training data. The prediction results using the linear regression model shows the number of the data on law violation that will occur in 2022 with correlation between crimes.

Keywords-forcasting; lawlessness prediction; linear regression; religious harmony

I. INTRODUCTION

Indonesia is a country whose people adhere to different religions, including Islam, Catholicism, Protestantism, Hinduism, Buddhism and Confucianism. The diversity of religions adhered to is tied to the values of Pancasila as the basis of the state to create harmony between religious communities [1]. The low faith of a person so that committing a crime can damage inter-religious harmony [2].

Crime is often interpreted as behavior that violates the rule of law because of which a person can be punished. Crime occurs when someone violates the law either directly or indirectly or is a form of negligence that can result in punishment. In the province of Yogyakarta, the annual violations of law vary. The number of such violations can be calculated year-to-year and can be predicted using many methods. Therefore, on this occasion we will use a forecasting method to predict the number of law violations that will occur in 2022.

There are various forecasting methods that have been used to predict the number of law violations, including Double

Maintaining Religious Harmony through Predicting the Level of Lawlessness using Linear Regression
Adam Wardana, M. Rival Ridautal Lillah, Mohamad Rasyid Ridho, Dekis Aldamawan, Mochamad Anwar Wirahadi, and Naufal Rizqullah
Khazanah Journal of Religion and Technology
Online ISSN: 2987-6060

Exponential Smoothing (ES) [3], Triple Exponential Smoothing [4], Fuzzy Logic [5], Exponential Smoothing [6]. In addition, the use of forecasting methods to predict the number of law violations has been practiced in various places such as the Indonesian seas [3], Probolinggo Regency, and Batam City.

Yogyakarta is a province that is frequently visited by the wider community for various activities such as study tours, holidays, or just for recreation because of its uniqueness. The more people there are, the greater the possibility of law violations that will occur. A system for predicting the number of law violations is needed for the development of public order and security so that Yogyakarta Province is always safe and peaceful. In this research a prediction system has been built for the number of law violations in Yogyakarta Province using a linear regression algorithm. Based on this explanation, this study aims to predict the level of crime to maintain religious harmony with a case study in Yogyakarta as a machine learning model that was built.

II. RELATED WORKS

The forecasting method using a linear regression algorithm has been applied in various studies including predicting the supply of tablet-type drugs [7], predicting long-term electricity usage [8], and predicting the number of students enrolled in Maranatha Christian University [9]. In addition, there are several research that using regressions to predict crime, such as: (1) using a machine learning model, predict the trends and patterns of crime in Bangladesh [10]; (2) an analysis and prediction model for crimes using many variables [11]; (3) organizational crime and formal employment case in Colombian regression discontinuity evidence [12]; (4) examining crime and the shared housing industry using a spatially weighted regression method is part of the sharing economy [13]; (5) detailed neighborhood-level trends in crime during the Chicago COVID-19 epidemic [14]; and (6) fuzzy rough nearest neighbor and sequential minimal optimization with logistic regression for detecting credit card fraud [15].

III. RESEARCH METHODS

A. Research activities

The research method is the rule used in collecting data from various information, by carrying out several steps, namely by conducting a literature study and studying several references related to this research, as well as collecting data that is useful for obtaining solutions to the problem formulation. The stages of research are a series of processes of how research is carried out. These stages are provided in the Figure 1 that begin with identifying the problem until conclude the prediction.
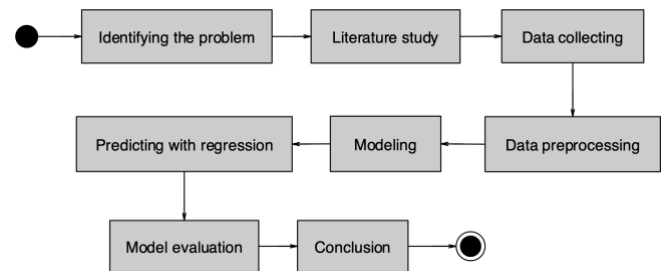

Fig 1. Research activities

- Identification of problems

Formulation or identification of the problem is the basic (initial) thing that needs to be done in research. With a problem, we can conduct research to find a solution to the problem we identified earlier. In this study, the problem to be solved is how to predict the number of law violations that will occur in 2022 in Yogyakarta Province using a linear regression model.

- Study of literature

At this stage, various research is carried out in related journals or journals that have the same discussion related to the research we are conducting.

- Data collection

The data used in this study is official data sourced from http://bappeda.jogjaprov.go.id/. The data was selected over the past 11 years (2012 – 2022). The data contains the types of crimes with the number of each of these crimes. Then, to carry out and produce a predictive output, the data will be divided into input and output categories.

- Data Preprocessing

Data preprocessing aims to transform data into a format that is simpler and more effective in obtaining more accurate values and reducing computation time for large-scale problems, thereby making data values smaller without changing the information they contain.

- Modeling

At this stage, a new model of linear regression is created for further predictions to be made on the data.

- Prediction with Regression

In the prediction stage, the model is used to train training data and validate data testing.

- Model Evaluation

Accuracy calculations are carried out in training and testing, so that you can see the differences in the accuracy of the 2 stages (training and testing).

- Conclusion Drawing

Conclusions are drawn based on the predicted data which is then compared with the original data.

B. Linear regression

In general, there are two kinds of relationships between two or more variables, which are usually called the form of the relationship and the closeness of the relationship. To determine the form of the relationship then used regression analysis. As for the closeness of the relationship can be known by correlation analysis. There are also those who say that linear regression is a statistical method used to form a

Maintaining Religious Harmony through Predicting the Level of Lawlessness using Linear Regression
Adam Wardana, M. Rival Ridautal Lillah, Mohamad Rasyid Ridho, Dekis Aldamawan, Mochamad Anwar Wirahadi, and Naufal Rizqullah
Khazanah Journal of Religion and Technology
Online ISSN: 2987-6060

model of the relationship between the dependent (Y) variable and one or more independent (X) variables [16]–[19], with the aim of estimate and predict the population mean or the mean value of the dependent variable based on the known values of the independent variables. The results of the regression analysis are in the form of coefficients on each variable X (independent). The coefficient is obtained by predicting the value of the Y (dependent) variable with an equation. The regression coefficient is calculated with two objectives at once, to minimize the deviation between the actual value and the estimated value of the Y (dependent) variable based on existing data. This regression analysis consists of two equation models: simple linear regression analysis and multiple linear regression analysis. Formula (1) and (2) are a formula for simple and multiple linear regression.

$$Y = a + bX \qquad (1)$$
$$Y = a + b1X1 + b2X2 + \cdots + bnXn \qquad (2)$$
$$a = (\Sigma y - b\Sigma x)/n \qquad (3)$$
$$b = (n\Sigma xy - (\Sigma x)(\Sigma y))/n\Sigma x^2 - (\Sigma x)^2 \qquad (4)$$

With: Y is response variable or dependent variable; X is variable predictor or causal factor variable (independent); a is constant (formula 3); b is magnitude of the response caused by the predictor (formula 4); and n is amount of data.

IV. RESULT AND DISCUSSION

A. Data Pre-processing

We do not process the data that we obtain from the official website of the Yogyakarta provincial development planning agency, instead we will carry out the data pre-processing stage first. Data pre-processing is a process that converts raw data into a form that is easier to understand [20], [21]. This process is important because raw data often does not have a regular format. In addition, so that we can also see the data from a statistical point of view so that the data can show the correlation between one variable and another.

- Data Cleaning

At this stage the raw data that we have obtained is re-selected and then incomplete, irrelevant, and inaccurate data is deleted or eliminated. By doing this stage, we can avoid misunderstandings when analyzing the data. From this stage we obtain data on the types of crimes per year that we will use. The following is a table image of the results of data cleaning or data cleaning. Table I presents the total data of crime from 2012 until 2021 that used to predict crime in 2022.

TABLE I. TYPE OF CRIME FROM 2012-2022

Type of Crime	Total Crime by Year										
	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Theft by Weighting	962	872	741	710	356	872	606	479	335	370	122
Violent Theft	222	174	212	221	105	174	153	77	63	48	8
Motor Vehicle Theft	543	743	608	512	246	743	451	389	148	106	49
Ordinary Theft	719	671	622	702	302	671	751	553	581	695	232
Mild Persecution	423	398	451	434	209	398	336	278	44	286	115
Severe Persecution	38	44	48	55	42	44	66	86	30	36	6
Murder	9	12	10	7	2	12	2	4	13	26	5
Rape	20	21	21	7	7	21	11	8	22	16	5
Adultery	36	34	24	27	14	34	39	37	28	37	12
Decency / Obscene	27	39	50	40	30	39	55	60	59	56	19
Embezzlement / Fidusa	524	539	565	599	289	539	500	343	397	421	149
Fraud / Underhanded Acts	1364	1403	1296	1548	672	1403	950	839	561	617	256
Domestic violence	92	122	165	123	86	122	133	128	81	80	29
Terrorist	0	0	1	0	0	0	0	0	0	0	0
Drugs	257	336	389	406	277	385	508	511	270	357	70
Human trafficking	0	2	2	1	1	1	0	0	0	0	0
Cyber Crime	45	25	9	23	3	6	8	11	6	5	2
Illegal Logging	9	3	7	2	3	3	0	1	0	0	0
Illegal Fishing	0	0	0	0	0	0	0	0	0	0	0
Illegal Mining	0	0	0	0	5	0	6	5	1	3	0
Corruption	3	7	3	3	4	6	7	5	1	1	0
Oil and Gas / BBM	3	0	0	0	1	0	0	8	1	0	0
Counterfeit money	4	0	2	1	1	4	1	1	4	2	0
Anarchist Demonstration	0	0	5	0	0	0	0	0	2	0	0
Mass Riot	0	0	0	0	0	0	0	0	0	0	0
Sara	0	0	0	0	0	3	2	0	0	0	0
Serious Human Rights Violations	0	0	0	0	0	0	0	0	0	0	0
Human Rights Violations Against Children	0	0	0	0	0	0	0	0	0	0	0

- Data Transformation

At this stage, all data will be uniform in data type. The dataset that we used initially contained several variables of the type of object or character variables. However, after this process was carried out, some of the data types of crimes were uniformly typed into integers or numbers. The following is a table of variable data types that we use in this research process.

- Correlation Between Variables

The last stage in the data pre-processing process that we did was to see the correlation between variables. In this research, we are looking for a correlation between each case

Maintaining Religious Harmony through Predicting the Level of Lawlessness using Linear Regression
Adam Wardana, M. Rival Ridautal Lillah, Mohamad Rasyid Ridho, Dekis Aldamawan, Mochamad Anwar Wirahadi, and Naufal Rizqullah
Khazanah Journal of Religion and Technology
Online ISSN: 2987-6060

of law violation that occurred in the province of Yogyakarta. The result shows that most law violations will increase if other law violations occur. For minority cases, there are three types of law violations which will decrease if other law violations occur. The three types of law violations are cases of law violation in the form of rape, illegal mining, and illegal logging. It was found that illegal logging cases will decrease if illegal mining cases increase and vice versa. In addition, it was also found that there was a correlation between the two cases and rape cases. We found that cases of illegal mining and illegal logging increased, so rape cases decreased.

B. Data Indexing

After carrying out the data pre-processing process, indexing will then be carried out where the process will initiate the variables that will be used in the prediction process using a linear regression algorithm. This indexing is needed to determine position data in Cartesian coordinates so that the result will be obtained in the form of predictive data on cases of law violations that will occur in 2022. In this indexing process, we determine that the first variable is data on legal violations that we have while the second variable is data on legal violations that will be predicted. After this indexing process, the process of predicting law violations in 2022 is carried out using a linear regression algorithm. After modeling, prediction results were obtained for law violations in Yogyakarta Province in 2022 in the form of 20 cases of motor vehicle theft, 1 case of circulation of counterfeit money, 111 cases of drugs, 10 cases of serious maltreatment, 235 fraud/fraudulent acts.

C. Prediction Results

This section will discuss how to predict the level of lawlessness in the province of Yogyakarta in 2022 using a linear regression model. To predict the level of law violations in Yogyakarta Province with a linear regression algorithm, training data and data tests are carried out, where the size of the test data is 40% of the total data in the dataset so that the accuracy obtained for both is 99% and 94%. However, the difference in the accuracy of these two data, especially the test data which is lower than the training data, indicates that there is an important change or difference in the understood data (Pyle et al., 1999). In the object of this study, data on criminal acts was taken 9 years back, from 2012 to 2021.

Linear regression is a machine learning approach that is used to test the extent of a causal relationship between variables (Kamal A, 2012). The causal variable is usually written with X and the effect or response variable will be written with Y. After pre-processing the data and splitting the data, predictions are made on the X test data using X and Y train data. so that predictions are obtained in the form of arrays with predicted values: [25, 1, 111, 1, 11, 243, 1, 1, 0, 0, 3, 17]. After the predictions are entered into the table, the predicted level of law violations in the future is obtained 543 743 608 512 246 743 451 389 148 106 49 617 256 1364 1403 1296 1548 672 1403 950 839 561 in 2022, namely in the form of theft of motor vehicles as many as 25 cases 1 case, 111

cases of drugs, 1 case of illegal logging, 11 cases of serious maltreatment, 243 cases of fraud/fraudulent acts, 1 case of mass riots, 1 case of terrorists, 3 cases of cybercrime, and violent theft as many as 17. Table II shows the comparation result between the real data and prediction result of criminal case in 2022.

TABLE II. PREDICTION RESULT

Type of Crime	Total Crime in 2022	
	Real	Prediction
Violent Theft	8	243
Motor Vehicle Theft	49	1
Severe Persecution	6	25
Fraud / Underhanded Acts	256	1
Terrorist	0	3
Drugs	70	11
Illegal Logging	0	111
CyberCrime	2	1
Corruption	0	1
Counterfeit money	0	17
Mass Riot	0	0
Sara	0	0

V. CONCLUSION

Based on the results of the analysis that has been done, it can be concluded that in general the forecasting method can be done using a linear regression algorithm, using clear stages. Starting from data collection where in the object of this research, crime data were taken from the previous 9 years, from 2012 to 2021. Then proceed with data pre-processing, data modeling, and finally the prediction process with a linear regression algorithm. By using a linear regression algorithm, the predicted results of law violations in 2022 are in the form of 25 cases of motor vehicle theft, 1 case of counterfeit money, 111 cases of drugs, 1 case of illegal logging, 11 cases of serious abuse, fraud/fraudulent acts 243 cases, 1 case of mass riots, 1 case of terrorists, 3 cases of cybercrime, and 17 cases of violent theft. and increase religious harmony. Future research may use algorithms or other machine learning methods.

REFERENCES

[1] D. Septian, "Pemahaman Nilai-Nilai Pancasila Dalam Memperkuat Kerukunan Umat," *TANJAK: Journal of Education and Teaching*, vol. 1, no. 2, pp. 155–168, 2020.

[2] K. Mayasaroh, "Toleransi Strategi Dalam Membangun Kerukunan Antarumat Beragama Di Indonesia," *Al-Afkar, Journal For Islamic Studies*, pp. 77–88, 2020.

[3] H. Lumaksono, H. Hozairi, B. Buhari, and M. Tukan, "Prediksi Jumlah Pelanggaran Hukum di Laut Indonesia Menggunakan Metode Double Exponential Smoothing," *Mnemonic: Jurnal Teknik Informatika*, vol. 3, no. 1, pp. 17–23, 2020.

[4] A. D. W. Sumari, R. Y. A. Pratama, and O. D. Triswidrananta, "Sistem Prediksi Tingkat Kriminalitas Menggunakan Metode Triple Exponential Smoothing: Studi Kasus Pada Polres Kabupaten Probolinggo," *Jurnal Teknik Informatika Vol*, vol. 13, no. 2, p. 171, 2020.

[5] Y. Yenni and I. Utnasari, "Prediksi tingkat kriminalitas menggunakan fuzzy logic," *Indonesian Journal of Computer Science*, vol. 8, no. 2, pp. 164–175, 2019.

[6] E. Chua and G. Tumibay, "Crime data forecasting using exponential smoothing," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1.1, pp. 69–75, 2020.

Maintaining Religious Harmony through Predicting the Level of Lawlessness using Linear Regression
Adam Wardana, M. Rival Ridautal Lillah, Mohamad Rasyid Ridho, Dekis Aldamawan, Mochamad Anwar Wirahadi, and Naufal Rizqullah
Khazanah Journal of Religion and Technology
Online ISSN: 2987-6060

[7] Z. Muttaqin and E. Srihartini, "Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet," *JSiI (Jurnal Sistem Informasi)*, vol. 9, no. 1, pp. 12–16, 2022.

[8] A. Hartawan, "Rancang Bangun Sistem Prediksi Penggunaan Listrik Jangka Panjang Menggunakan Metode Regresi Linier Berbasis WEB (Studi Kasus PT. PLN (Persero) Unit Induk, Distribusi Bali)," *PETIK: Jurnal Pendidikan Teknologi Informasi Dan Komunikasi*, vol. 8, no. 1, pp. 35–41, 2022.

[9] S. E. C. Wijaya and A. Widjaja, "Prediksi Jumlah Mahasiswa yang Masuk ke Universitas Kristen Maranatha Menggunakan Regresi Linier," *Jurnal STRATEGI-Jurnal Maranatha*, vol. 4, no. 1, pp. 175–184, 2022.

[10] A. A. Biswas and S. Basak, "Forecasting the trends and patterns of crime in Bangladesh using machine learning model," in *2019 2nd international conference on intelligent communication and computational techniques (ICCT)*, 2019, pp. 114–118.

[11] S. Shukla, P. K. Jain, C. R. Babu, and R. Pamula, "A multivariate regression model for identifying, analyzing and predicting crimes," *Wirel Pers Commun*, vol. 113, pp. 2447–2461, 2020.

[12] G. Khanna, C. Medina, A. Nyshadham, and J. A. Tamayo, "Formal employment and organized crime: Regression discontinuity evidence from Colombia," National Bureau of Economic Research, 2019.

[13] Y.-H. Xu, L. Pennington-Gray, and J. Kim, "The sharing economy: A geographically weighted regression approach to examine crime and the shared lodging sector," *J Travel Res*, vol. 58, no. 7, pp. 1193–1208, 2019.

[14] G. M. Campedelli, S. Favarin, A. Aziani, and A. R. Piquero, "Disentangling community-level changes in crime trends during the COVID-19 pandemic in Chicago," *Crime Sci*, vol. 9, pp. 1–18, 2020.

[15] A. S. Hussein, R. S. Khairy, S. M. M. Najeeb, and H. T. ALRikabi, "Credit Card Fraud Detection Using Fuzzy Rough Nearest Neighbor and Sequential Minimal Optimization with Logistic Regression.," *International Journal of Interactive Mobile Technologies*, vol. 15, no. 5, 2021.

[16] A. I. Permatasari and W. F. Mahmudy, "Pemodelan regresi linear dalam konsumsi Kwh listrik di Kota Batu menggunakan algoritma genetika," *DORO Repos. J. Mhs. PTIIK Univ. Brawijaya*, vol. 5, no. 14, pp. 1–9, 2015.

[17] S. I. Bangdiwala, "Regression: simple linear," *Int J Inj Contr Saf Promot*, vol. 25, no. 1, pp. 113–115, 2018.

[18] K. Park, R. Rothfeder, S. Petheram, F. Buaku, R. Ewing, and W. H. Greene, "Linear regression," in *Basic Quantitative Research Methods for Urban Planners*, 2020, pp. 220–269. doi: 10.4324/9780429325021-12.

[19] T. M. H. Hope, "Linear regression," in *Machine Learning: Methods and Applications to Brain Disorders*, 2019, pp. 67–81. doi: 10.1016/B978-0-12-815739-8.00004-3.

[20] Anonymous, "Analisis Data Jadi Lebih Mudah dengan Data Preprocessing," *algorit.ma*, 2022. https://algorit.ma/blog/data-preprocessing-adalah/

[21] A. Oliver, "Bikin Data Lebih Mudah Dibaca, yuk, Kenalan dengan Data Preprocessing," *glints.com*, 2021. https://glints.com/id/lowongan/data-preprocessing-adalah/#.YwMrfXZBxD8 (accessed Aug. 22, 2022).