Jurnal Sistem Informasi dan Informatika (Simika) *P-ISSN: 2622-6901*
Vol. 8 No. 2 Tahun 2025 *E-ISSN: 2622-6375*

IMPLEMENTATION OF KERNEL COMBINATION GAUSSIAN PROCESS REGRESSOR IN LOYALTY PREDICTION (CASE STUDY: ONLINE MOTORCYCLE TAXI)

Luqna Aziziyah[1], Dwi Arman Prasetya[2], Trimono Trimono[3]
Program Studi Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur
Jl. Rungkut Madya, Gn. Anyar, Kec. Gn. Anyar, Kota Surabaya
e-mail: *[1]**luqna25aziziyah@gmail.com**, [2]arman.prasetya.sada@upnjatim.ac.id,
[3]trimono.stat@upnjatim.ac.id.

Abstract
In the application-based transportation industry, customer loyalty is a crucial factor affecting service sustainability. This study aims to analyze and predict customer loyalty in online motorcycle taxi services in Surabaya using the Gaussian Process Regressor (GPR) with a kernel combination approach. Data were collected through a survey of 467 students from public universities in Surabaya, considering service quality, price, and innovation factors. The analysis process includes data processing, validation, cleaning, and modeling using Gaussian Process Regression techniques. The results indicate that the kernel combination in GPR effectively captures complex non-linear patterns in survey data, with low Root Mean Squared Error (RMSE) and R^2 values close to 1. These findings suggest that the proposed approach can provide accurate customer loyalty predictions. This study contributes to developing strategies for online motorcycle taxi service providers to enhance user experience and maintain market share. The findings highlight the importance of applying machine learning models to understand customer behavior and support data-driven business decision-making.

Keyword: *Customer Loyalty, Data Analysis, Gaussian Process Regressor, Online Motorcycle Taxi, Service Quality*

INTRODUCTION

The development of information technology has brought about major changes in various aspects of life, including the transportation sector. One of the prominent innovations is the application-based online motorcycle taxi service which is now the main choice for urban communities because it offers convenience, speed, and efficiency, especially for individuals with high mobility (Chasanah et al., 2024). In Indonesia, especially in Surabaya City which is the center of education with the largest number of state universities in East Java, the demand for this service is very high among students. This condition creates a potential and competitive market for online motorcycle taxi service providers (Kuswanto et al., 2020).

In the increasingly competitive online transportation industry, customer loyalty is a crucial factor in maintaining business sustainability. Loyal customers tend to use services more often, recommend them to others, and provide useful feedback to the company. Therefore, understanding the factors that influence loyalty such as service quality, price, and innovation is important to maintain and increase market share (Iqbal et al., n.d.; Fadhila, 2024; H. M. Nguyen et al., 2024).

Conventional approaches such as linear regression are often used to analyze customer loyalty. However, this approach has limitations in capturing complex and non-linear data patterns. As analytical technology advances, machine learning-based approaches are increasingly being applied to understand customer behavior in more depth and make predictions with a higher level of accuracy. One of the prominent methods in non-linear regression is the Gaussian Process Regressor (GPR), which is not only able to handle complex data but also provides an estimate of the uncertainty of the prediction results (Shi et al., 2022). Several studies have shown that the combination of several basic functions (kernels) in GPR, such as

RBF, Rational Quadratic, Constant, and White, can improve model performance. Each kernel has a role such as RBF captures general patterns, Rational Quadratic reflects intermediate variations, Constant maintains model stability, and White helps handle high noise in data (Pan et al., 2021;Nguyen et al., 2021).

However, the application of this method is still mostly limited to engineering fields such as machine performance prediction or groundwater modeling, with structured technical measurement data. Research using GPR on interval-scale survey data (0–10), such as in customer loyalty studies, is still very limited. In fact, this type of data is commonly used in customer behavior research and often contains noise in the data (Deniz & Bülbül, 2024).

In response to this gap, this study proposes the application of Kernel Combination Gaussian Process Regressor (KCGPR) by combining four types of kernels to model and predict online motorcycle taxi customer loyalty. The data used comes from an interval-scale survey (0–10) of students from five state universities in Surabaya.

This study provides a new contribution in customer loyalty modeling based on survey data, by utilizing a probabilistic non-linear regression approach that is able to accommodate data uncertainty and complexity. The results obtained are expected to be the basis for companies to design more targeted customer retention strategies, increase satisfaction, and strengthen competitiveness in the online transportation industry.

RESEARCH METHOD

Generating accurate predictions requires a systematic and organized approach at every stage of the analysis. Each step in the process plays a crucial role in ensuring that the developed model can effectively capture patterns in the data and produce relevant predictions. Therefore, clear stages ranging from data collection, data processing, to modeling and evaluation are essential in building a reliable prediction system (Dudek & Baranowski, 2022; Deniz & Bülbül, 2024).

This study aims to predict online motorcycle taxi customer loyalty using the Gaussian Process Regressor (GPR) approach with a kernel combination. The research methodology consists of several key stages: data collection, validation and reliability testing, data preprocessing, modeling, and model evaluation.

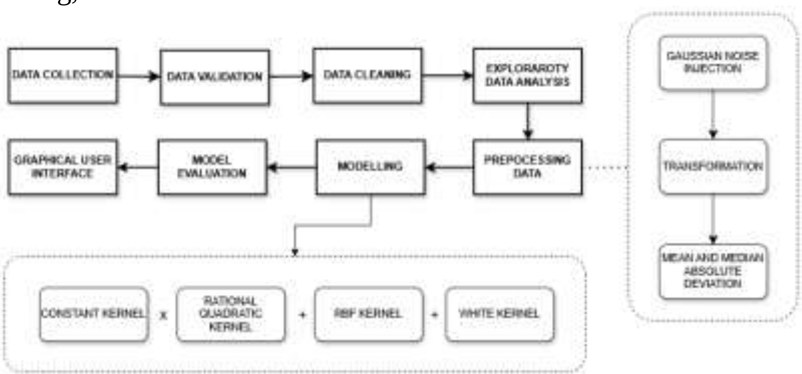

Figure 1. Research FlowChart

1. Data and Respondents

   Data collection in this study was conducted through a survey using an interval scale-based questionnaire. The survey was designed to measure factors that affect customer loyalty to online ojek services in Surabaya city. The selection of the survey method is based on its ability to collect quantitative data efficiently and allow measurement of respondents' perceptions of various aspects of the service. The minimum number of respondents required is determined using the Slovin formula with a margin of error (e) of 5%.

$$n = \frac{N}{1 + N(e)^2} \tag{1}$$

The questionnaire was designed using a 0-10 interval scale, which provides a wider variety of answers and allows for a more detailed assessment of the respondents' level of satisfaction or agreement with each indicator. The questionnaire consisted of a total of 20 questions divided into four main sections, with an average completion time of approximately 10–15 minutes. The indicators in the questionnaire reflect the following constructs:

a. Service Quality: ServiceQuality_Driver, ServiceQuality_UI, ServiceQuality_Vehicle (Keni, 2020).
b. Price: Price_Affordable, Price_Transparency, Price_Satisfaction (Widnyani et al., 2020).
c. Inovation: Innovation_Service Development, Innovation_Ease, Innovation_Response (Ni Made Widnyani et al., 2020).
d. Loyality: Loyalty_Satisfaction (Rinjani, 2024).

2. Data Validation

Data validity was tested through two main approaches, namely Kaiser-Meyer-Olkin (KMO) for data suitability for factor analysis, and Cronbach's Alpha to measure the internal consistency of the indicators used in each variable. In addition, Exploratory Factor Analysis (EFA) was also applied to verify the factor loading of the indicators in the questionnaire used (Stephanie Glen, 2009; Jim Frost, 2024).

***Table 1**. KMO Value Aceptance Rate*

KMO Value	Level of Acceptance
KMO ≥ 90	Marvelous
0.80 ≤ KMO < 0.90	Meritorius
0.70 ≤ KMO < 0.80	Average
0.60 ≤ KMO < 0.70	Mediocre
0.50 ≤ KMO < 0.60	Terrible
KMO < 0.50	Unacceptable

A good KMO is generally above 0.6, and the higher the KMO value, the stronger the suitability of the data to be analyzed using factor analysis techniques. The formula of KMO.

$$\mathrm{MO_j} = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i\neq j} r_{ij}^2 + \sum_{i\neq j} u} \tag{2}$$

Reliability testing by calculating the Cronbach's Alpha value for each variable tested with a formula as follows

$$\alpha = \frac{N * \bar{c}}{\bar{v} + (N-1) * \bar{c}} \tag{3}$$

3. Data Cleaning

This step in the cleaning process, in addition to the process of dealing with missing data and duplication, identifying outliers that may exist in the dataset is quite important. Outliers can be defined as values that are significantly different from other values in the dataset and can significantly affect the analysis results. Outlier detection itself uses the Interquartile Range (IQR) method which has the following formula (Pritha Bhandari, 2023)

$$IQR = Q3 - Q1 \tag{4}$$
$$Lower\ Bound = Q1 - (1.5\ \times IQR) \tag{5}$$
$$Upper\ Bound = Q3 + (1.5\ \times IQR) \tag{6}$$
$$n < Lower\ Bound\ \ \vee\ n > Upper\ Bound \tag{7}$$

Handling is done using the Winsorizing technique because it can replace outlier values with values that are closer to the upper limit or lower limit of the distribution by writing the Winsorized Mean Formula expressed as follows (Marshall Hargrave, 2023).

$$Winsorized\ Mean = \frac{x_n \dots x_{n+1} + x_{n+2} \dots x_n}{N} \tag{8}$$

4. Exploratory Data Analyst

The Exploratory Data Analysis (EDA) stage is carried out to understand the basic patterns, data distribution, and relationships between variables in the dataset. The results of this stage become the basis for decision-making at the Data Preprocessing stage, where appropriate data processing techniques can be applied to prepare the data before further analysis. Variance and standard deviation are calculated to measure data variation.

$$Varians\ (\sigma^2) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2 \tag{9}$$

$$Standard\ Deviation\ (\sigma) = \sqrt{\sigma^2} \tag{10}$$

$$CV = \frac{\sigma}{\mu} \tag{11}$$

Low values in these metrics indicate data homogeneity, which can hinder the model's ability to capture complex patterns. For such cases, Noise Injection is recommended to increase data variation. Kernel Density Estimation (KDE) is used to analyze the distribution of data across variables, providing smooth visualizations of probability distributions. For non-normal or skewed distributions, a Gaussian kernel function is applied (Setiawan, 2022).

$$K\ (u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \tag{12}$$

Skewness and Kurtosis calculations help identify asymmetry or peaked distributions, signaling the need for data transformation to enhance the distribution for better modeling

5. Data Prepocessing

The results of the Exploratory Data Analysis (EDA) stage provide important information that is used in the Data Preprocessing stage to improve data quality before entering the modeling stage (Andraz Krzisnik, 2021). To address data homogeneity and enhance model accuracy, the Gaussian Noise Injection technique is applied by adding random noise with a normal distribution (mean = 0, standard deviation = 0.3) to columns with low variation. This process enriches the dataset, increasing its variability and improving the model's ability to recognize complex patterns, particularly in predicting customer loyalty.

In addition, data transformation is performed using the RobustScaler technique, which scales the features based on the median and interquartile range (IQR), ensuring uniformity across the dataset for more efficient processing. The transformation is calculated as follows.

$$X_{new} = \frac{X - X_{median}}{IQR} \tag{13}$$

Furthermore, the Mean and Median Absolute Deviation (MAD) are calculated for each factor (Quality of Service, Price, Innovation, and Loyalty) to assess the spread of data. The MAD provides a more robust measure of variability than the standard deviation, making it less sensitive to outliers

$$MAD\ (QL) = median\ (|QS_i - median(KL)|) \tag{14}$$

$$MAD\ (P) = median\ (|P_i - median(H)|) \tag{15}$$

$$MAD\ (I) = median\ (|I_i - median(I)|) \tag{16}$$

This comprehensive preprocessing ensures the dataset is optimized for subsequent modeling, improving predictive performance

6. Modeling
   Gaussian Process Regression (GPR) is a method that not only provides predictions in the form of average values but also estimates the uncertainty in each prediction. This ability makes GPR very suitable for data analysis that has a high level of noise or uncertainty. In this study, GPR is used by combining four main kernels which can be explained in the following equation
$$k(x, x') = C \cdot K_{RationalKuadratic}(x, x') + K_{RBF}(x, x') + K_{WhiteNoise}(x, x') \tag{17}$$
$$Y(x) \sim GP(\mu(x), k(x, x')) \tag{18}$$
$$Y = f(QS, P, I) \tag{19}$$
   In the context of customer loyalty prediction, GPR models a probabilistic relationship between training data (X,y) and new data (x). This model is designed to predict the latent loyalty value (Y) based on a combination of input features. This approach provides advantages in data-driven decision making, especially when the data has a high degree of variability

7. Model Evaluation
   Several evaluation metrics are commonly used for regression models, including Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2). In this study, three main metrics are used to evaluate the Gaussian Process Regressor (GPR) model.
$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_1 - \hat{y})^2} \tag{20}$$
$$R^2 = 1 - \left[\frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}\right] \tag{21}$$
$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \hat{Y}_i}{Y_i}\right| \tag{22}$$
   Using 3 types of evaluation is because they have their own advantages. For instance, MAPE helps understand the error in terms of percentage (Chicco et al., 2021). R^2 provides insight into how well the model explains the variability of the data. RMSE emphasizes the importance of large errors and provides an absolute measure of the prediction error.

8. Graphical User Interface
   Streamlit was chosen to implement loyalty prediction for several strong reasons that support its effectiveness and ease of application development. Streamlit can be easily integrated with many Python libraries such as Pandas, Matplotlib, and Scikit-learn. This allows developers to utilize existing data analysis tools and combine them with loyalty prediction models. With the ability to run machine learning models directly within the interface, Streamlit allows users to get loyalty prediction results based on simple and intuitive data input. Thus, Streamlit is an ideal choice for building user interfaces in loyalty prediction applications, providing a good user experience as well as efficiency in development.

RESULT AND DISCUSS

1. Data Collection and Validation
   With an estimated population of around 130,000 public students who are active users of online motorcycle taxi services, the calculation of the number of samples uses the Slovin formula as follows

$$n = \frac{130.000}{1 + 130.000(0.05)^2} = 399 \tag{23}$$

With rounding, the minimum number of respondents needed is 400 students. To anticipate the possibility of invalid or incomplete data, the total respondents collected were 467 students. The random sampling technique was used to randomly select samples from the population of active users over the past 6-12 months with the criteria of using the service more than once a month. This approach ensures that each user has an equal chance of being selected, so that population representation can be guaranteed.

The KMO (Kaiser-Meyer-Olkin) test in this study resulted in a value of 0.899, which indicates that the data is very suitable for factor analysis. Furthermore, the Cronbach's Alpha value for each variable shows the following results.

Table 2. Cronbach's Alpha Value for Each Variable

Variable	**Service Quality**	**Price**	**Innovation**	**Loyalty**
Value	0.79	0.82	0.84	0.78

Based on the general criteria for Cronbach's Alpha values, all variables in this study showed an adequate level of internal consistency, with values of more than 0.7, so they can be considered reliable for the measurement of these variables

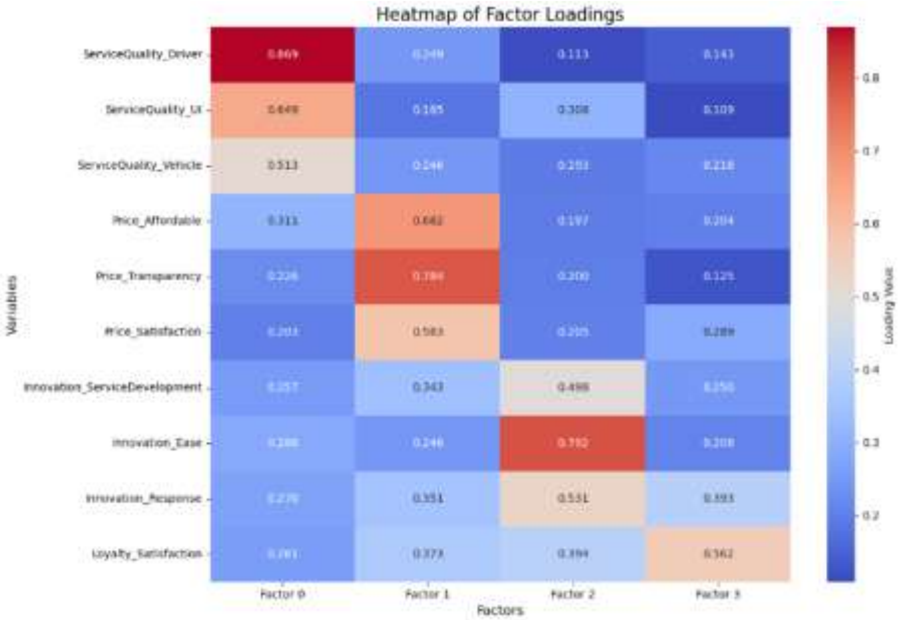

Figure 2. Heatmap Factor Loadings

In Figure 2, all factor loadings for the indicators of service quality, innovation, price, and loyalty are above the threshold value of 0.5, which indicates that the indicators are valid for measuring the factors they represent. Based on the results of the validation carried out, the data used in this study can be considered valid and reliable for further analysis.

2. Data Cleaning

After removing unnecessary columns such as student identities and blank grades, regarding overcoming outliers contained in the data, the Winsorizing method was chosen to be applied to all columns that were detected to have outliers. The results are visualized in Figure 3 so that a before and after comparison can be seen.

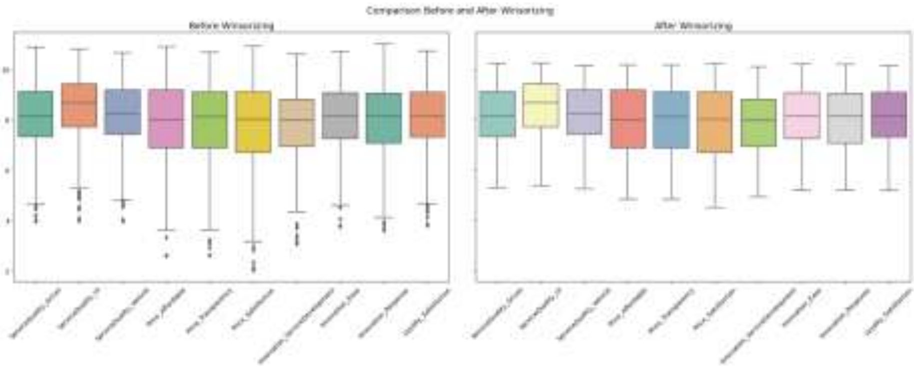

Figure 3. Comparison Before and After Winsorizing

3. Exploratory Data Analyst

As a result of the descriptive analysis of the data, several important characteristics were found to be the basis for decision making in the preprocessing process. The variance and standard deviation show that some variables have a fairly low spread of data, so the contribution to the analysis may be limited.

Table 3. Statistical Summary Table

Variable	Variance	Std. Deviation	Coefficient of Variation
ServiceQuality_Driver	1.889837	1.374713	0.169794
ServiceQuality_UI	2.113720	1.453864	0.174002
ServiceQuality_Vehicle	1.943149	1.393969	0.171041
Price_Affordable	3.024740	1.739178	0.222459
Price_Transparency	3.040557	1.743719	0.222248
Price_Satisfaction	3.545441	1.882934	0.245280
Innovation_Service Development	2.438136	1.561453	0.201325
Innovation_Ease	2.038342	1.427705	0.177371
Innovation_Response	2.087271	1.444739	0.182202
Loyalty_Satisfaction	2.052927	1.432804	0.176969

Looking at the coefficient of variation (CV), most of the variables have low CV values, which are below 0.25. This value indicates a fairly high degree of homogeneity in the data, meaning that the relative variability between data values within the variable is low compared to the mean. This can reduce the ability of the model to significantly differentiate the influence between variables.

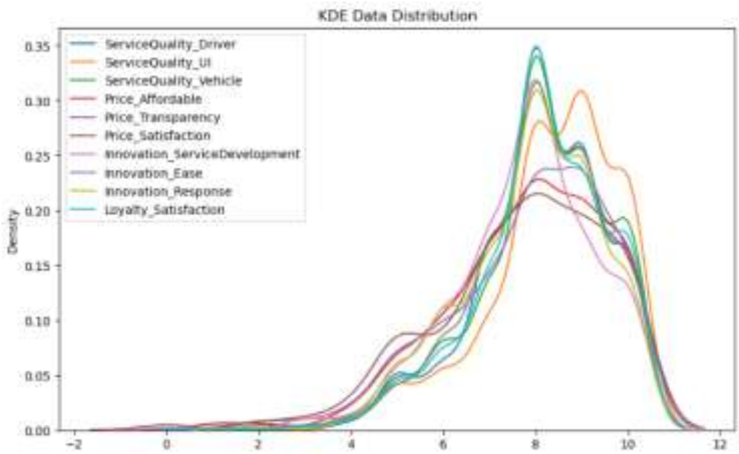

Figure 4. KDE Data Distribution

P-ISSN: 2622-6901
E-ISSN: 2622-6375

Table 4. Statistical Distribution Table

Variable	Skewness	Kurtosis
ServiceQuality_Driver	-0.826230	1.323688
ServiceQuality_UI	-1.181946	1.848878
ServiceQuality_Vehicle	-0.670537	0.118706
Price_Affordable	-0.890863	1.047467
Price_Transparency	-0.910914	0.699256
Price_Satisfaction	-0.953820	1.192107
Innovation_Service Development	-0.883763	1.750300
Innovation_Ease	-0.824544	1.088876
Innovation_Response	-0.596497	0.088952
Loyalty_Satisfaction	-0.904112	1.463323

Graphically, all variables have negative skewness, indicating that the data distribution is left-skewed. Most variables have a kurtosis close to 1, indicating that the data distribution tends to be close to a normal distribution but with a slightly lower skewness. Due to the high degree of homogeneity and lack of variability in the data distribution, a noise injection step was performed at the preprocessing stage. The purpose of noise injection is to increase data variability so that the model can be more sensitive to differences between variables. This injection is done by adding small controlled noise to avoid significant changes in the original data pattern

4. Data Prepocessing

The results of the Exploratory Data Analysis (EDA) stage indicate a problem with too low data variation (high homogeneity), which may limit the model's ability to recognize more complex patterns.

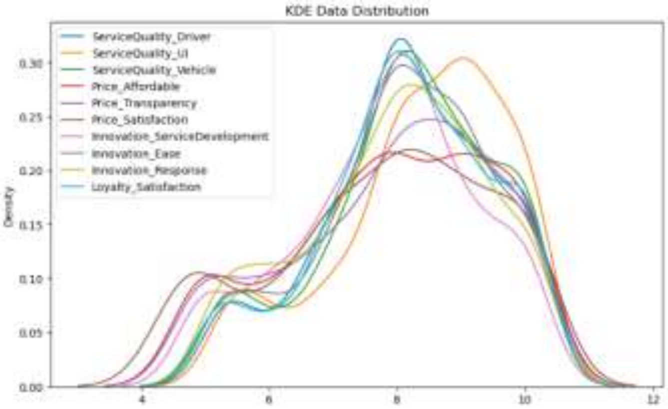

Figure 5. KDE Data Distribution After Injection

After noise injection, the data distribution becomes more symmetrical and closer to a normal distribution, which can help improve the model's performance in analyzing the relationship between variables. The addition of noise also causes the data to become more dispersed, reducing problems that may arise from distributions that are too narrow or too sharp. This process is expected to improve the model's ability to capture hidden patterns and reduce the possibility of overfitting caused by data that is too homogeneous. Ratings were calculated based on the mean and median absolute deviation (MAD) of each indicator. The analysis results were visualized using boxplots to show the distribution of each indicator's assessment on each factor. The box represents the interquartile range (IQR), the median line shows the middle value

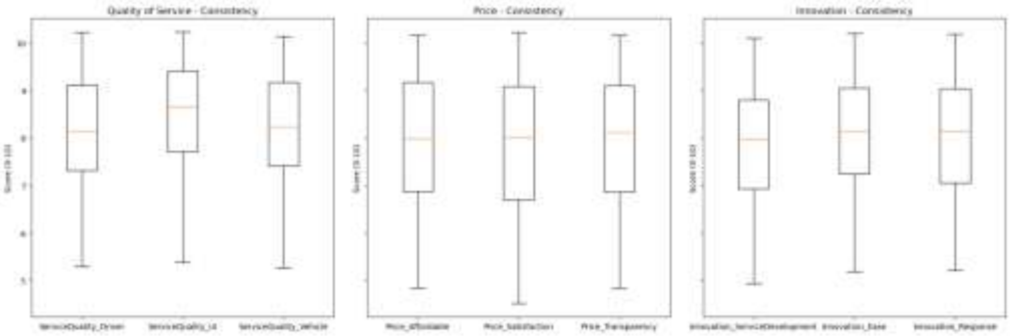

Figure 6. Boxplot of Consistency Distribution

The ServiceQuality_ui indicator shows the highest level of assessment consistency, with a narrower distribution range than Service Quality_driver and Service Quality_responsiveness. This reflects that customers have a more uniform perception of the user interface. In general, MAD provides information that most indicators have good consistency of assessment. This means that the data for these indicators is quite stable and can be directly used for the next steps, such as normalization or aggregation. In performing optimization, transformations can be performed.

5. Model and Evaluate

The combined kernel used in this model is the sum of several basic kernels, which are used to model complex relationships in the data and handle noise. Not only with kernel optimization, but also model hyperparameters using alpha = 0.2 to control the noise level or model confidence in the training data. So as to obtain the optimal value of various kernels in performing calculations.

$$3.6 ** 2 * RationalQuadratic(alpha = 1e + 05, lenght\_scale = 11.9) + RBF(length\_scale = 8.85e + 03) + WhiteKernel(noise\_level = 0.324). \tag{24}$$

These results show that the GPR model that has been optimized with the combined kernel provides excellent predictions, both on training and testing data. The model shows a good ability to capture data patterns, with high R^2 and relatively low prediction error. Here are examples of predictions on some data.

Table 5. Prediction On Data

Description	R2- Score	RMSE	MAPE
Training	0.8817	0.9835	9.0217
Testing	0.8221	0.9225	9.3346

These results show that the optimized GPR model with the joint kernel provides excellent predictions on both training and testing data. The model showed a good ability to capture the pattern of the data, with a high R^2 and relatively low prediction error. The optimized kernel parameters show that the model has adapted well to the scale and structure of the data, enabling accurate and stable predictions.

Table 6. Comparison of Actual, Predicted, Error, and Squared Error Values of the GPR Model

Actual	Predicted	Error	Squared_Error
7.022709	7.862341	-0.839632	0.704982
7.886981	8.128299	-0.241318	0.058235
9.015297	8.063623	0.951674	0.905683
9.066922	8.908658	0.158263	0.025047
9.195046	8.191987	1.003058	1.006126

Overall, the model seems to be able to provide fairly accurate predictions, with some larger errors at some points. Larger squared errors indicate more significant errors, as seen in rows 1, 3, and 5. Despite some large errors, the model still shows fairly good predictions at most points (such as rows 2 and 4), indicating that the overall performance of the model is still quite good. Coupled with the use of streamlit as a user interface for applying the GPR algorithm with a combination of kernels in making predictions. As a result, it looks like this

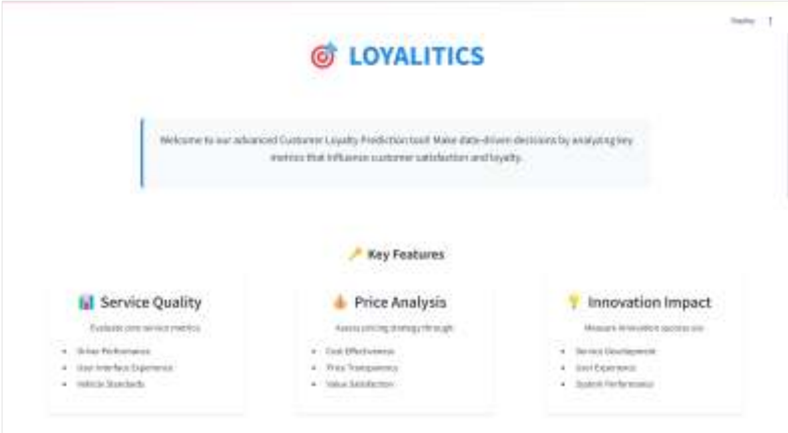

Figure 7. HomePage Streamlit Loyality Prediction

Where from filling in the input from the user, it will produce predictions like the following

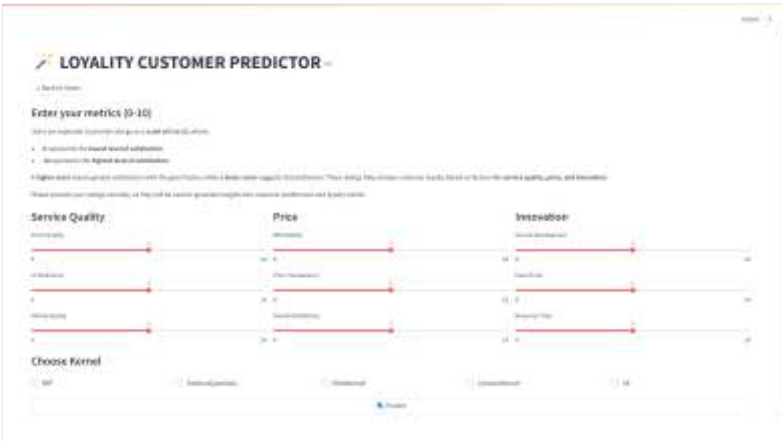

Figure 8. Customer Input for Loyality Prediction

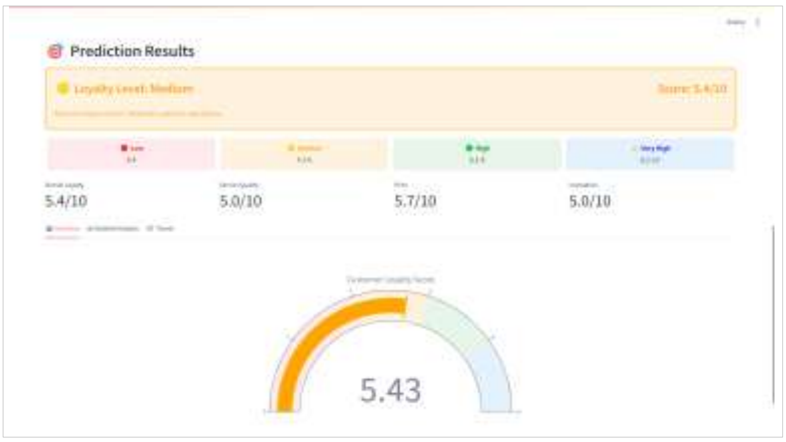

Figure 9. Customer Loyality Level

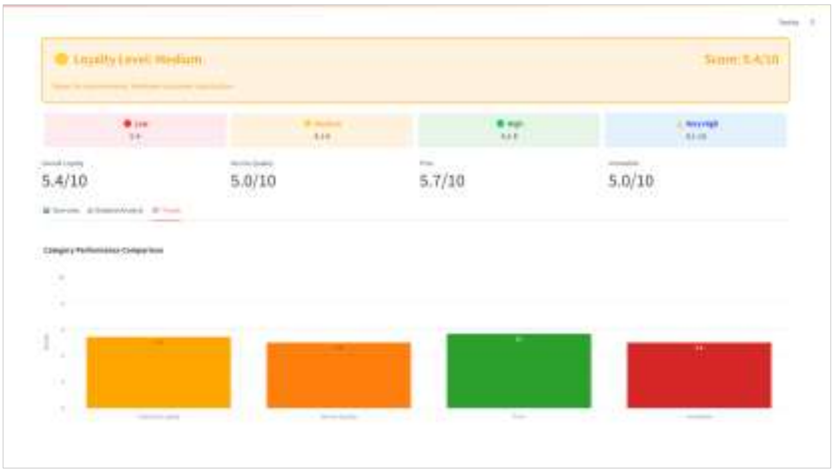

Figure 10. Customer Average Loyality by Category

CONCLUTION

This research successfully applies the Kernel Combination Gaussian Process Regressor (KCGPR) to predict customer loyalty in online motorcycle taxi services, with a focus on university students in Surabaya. Through a systematic approach, this research identifies and analyzes factors that affect customer loyalty, including service quality, price, and innovation. Using survey data processed through various stages, from data collection and validation to modeling and evaluation, we can produce an accurate and relevant model. The analysis results show that the optimized GPR model with a combination of kernels is able to capture complex patterns in the data, resulting in high R^2 values and relatively low prediction errors. The use of noise injection techniques also proved effective in increasing data variation, so that the model can be more sensitive to differences between variables. In addition, model evaluation using metrics such as RMSE and MAPE showed good performance on both training and testing data. The implementation of Streamlit as a user interface allows users to easily input data and get real-time customer loyalty predictions. Thus, this research not only contributes academically in the development of loyalty prediction methods, but also offers practical solutions for online motorcycle taxi service providers in improving customer experience and maintaining market share in the midst of intense competition. Overall, this research confirms the importance of utilizing machine learning techniques, especially Gaussian Process Regressor, in understanding and predicting customer behavior. The findings are expected to be the basis for further research and the development of more effective marketing strategies in the application-based transportation industry.

REFERENCE

Andraz Krzisnik. (2021, January 9). *How To Make Gaussian Noise On Image*. Epoch House.

Chasanah, N., Ferriswara, D., & Listyawati, L. (2024). Analysis of the Quality of Maxim and Indrive Online Motorcycle Taxi Services for Students of Dr. Soetomo University Surabaya. *International Journal of Multicultural and Multireligious Understanding (IJMMU), 11*(1). https://doi.org/10.18415/ijmmu.v11i1.5285

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science, 7*, 1–24. https://doi.org/10.7717/PEERJ-CS.623

Deniz, E., & Bülbül, S. Ç. (2024). Predicting Customer Purchase Behavior Using Machine Learning Models. *Information Technology in Economics and Business*. https://doi.org/10.69882/adba.iteb.2024071

Dudek, A., & Baranowski, J. (2022). Gaussian Processes for Signal Processing and Representation in Control Engineering. *Applied Sciences (Switzerland)*, *12*(10). https://doi.org/10.3390/app12104946

Iqbal, M., Setiawan, H., Akhsani, R., Prayoga, S., & Kusumawati, A. (n.d.). Analisis Kualitas Layanan E-Health Surabaya: Pendekatan Integratif E-Servqual Dan Importance Performance Analysis (Ipa). *Jurnal Sistem Informasi Dan Informatika (Simika) P-ISSN*, *7*(2), 2024.

Jim Frost. (2024). *Cronbach's Alpha: Definition, Calculations & Example*. Statistics By Jim.

Keni, J. D. (2020). Prediksi Kualitas Pelayanan Dan Kepercayaan Terhadap Loyalitas Pelanggan: Kepuasan Pelanggan Sebagai Variabel Mediasi. In *Jurnal Manajerial dan Kewirausahaan: Vol. II* (Issue 1).

Kuswanto, A., Sundari, S., Harmadi, A., & Hariyanti, D. A. (2020). The determinants of customer loyalty in the Indonesian ride-sharing services: offline vs online. *Innovation and Management Review*, *17*(1), 75–85. https://doi.org/10.1108/INMR-05-2019-0063

Ni Made Widnyani, Vitalia Carla Rettobjaan, & A.A. Ngurah Bagus Aristayudha. (2020). Pengaruh Harga, Promosi Dan Inovasi Terhadap Loyalitas Pelanggan Gojek (Studi Kasus Pada Universitas Bali Internasional). *Jurnal Ilmiah Manajemen Dan Bisnis*, *5*(2), 75–92. http://journal.undiknas.ac.id/index.php/manajemen

Marshall Hargrave. (2023, July 19). *Winsorized Mean: Formula, Examples and Meaning*. Investopedia.

Nguyen, K., Krumm, J., & Shahabi, C. (2021). *Gaussian Process for Trajectories*. http://arxiv.org/abs/2110.03712

Pan, Y., Zeng, X., Xu, H., Sun, Y., Wang, D., & Wu, J. (2021). Evaluation of Gaussian process regression kernel functions for improving groundwater prediction. *Journal of Hydrology*, *603*. https://doi.org/10.1016/j.jhydrol.2021.126960

Pritha Bhandari. (2023, June 21). *How to Find Interquartile Range (IQR) | Calculator & Examples*. Scribbr.

Rinjani, S. F. (2024). Analisis Pengaruh Kepuasan Pelanggan dan Kualitas Layanan terhadap Loyalitas Pelanggan Dalam Industri Jasa Gojek. In *Jurnal Multidisiplin West Science* (Vol. 03, Issue 06).

Setiawan, A. (2022). *Penentuan Distribusi Skewness Dan Kurtosis Dengan Metode Resampling Berdasar Densitas Kernel (Studi Kasus Pada Analisis Inflasi Bulanan Komoditas Bawang Merah, Daging Ayam Ras Dan Minyak Goreng di Kota Semarang). Prosiding Seminar Nasional Sains Dan Pendidikan Sains, 240-246*

Shi, X., Jiang, D., Qian, W., & Liang, Y. (2022). Application of the Gaussian Process Regression Method Based on a Combined Kernel Function in Engine Performance Prediction. *ACS Omega*, *7*(45), 41732–41743. https://doi.org/10.1021/acsomega.2c05952

Stephanie Glen. (2009). *Kaiser-Meyer-Olkin (KMO) Test for Sampling Adequacy*. Statistics How To. https://www.statisticshowto.com/kaiser-meyer-olkin/