Analysis of the Use of Fake GPS Applications on Drivers of PT GoTo Gojek Tokopedia Tbk (GOTO)

Muhammad Fikri Alhafidz[1], Eka Hayana Hasibuan[2]
[1,2] Program Studi Informatika, Fakultas Teknologi, Universitas Battuta
[1] muhd.fikrialhafidz@gmail.com

Article Info

Article history:

Received Februari 27, 2025
Revised February 27, 2025
Accepted February 28, 2025

Keywords:

Fake GPS
AdaBoost
XGBoost

ABSTRACT

The purpose of this study was to determine the use and system used to detect the fake GPS application on PT Goto Gojek Tokopedia Tbk driver partners in the competition between Gojek drivers in Surabaya. The research method used is a qualitative method, which is obtained based on the results of field research (field research). The data required in this study were collected using observation, interview, and literature study techniques, which were then analysed using descriptive techniques in describing data about the use of the Fake GPS application on PT Goto Gojek Tokopedia Tbk partners. The results of this study conclude that the use of the Fake GPS application by PT Goto Gojek Tokopedia Tbk drivers is carried out without coercion, without fraud, and without oversight, but the driver violates company rules regarding the use of additional applications in the form of fake GPS. The system used to detect fake GPS applications in this research is based on ensemble learning using the AdaBoost and XGBoost algorithms. The learning ensemble is able to detect fake GPS applications very well.

Corresponding Author:

Muhammad Fikri Alhafidz
Universitas Battuta
Email: youremail@gmail.com

1. INTRODUCTION

In the modern era, transportation services based on mobile applications, commonly referred to as online transportation, have become a necessity for society. This is due to the safety, convenience, and affordability they offer to the public in Indonesia [1]. Online transportation platforms such as Grab, Gojek, and others have experienced rapid growth in the country. These services not only provide a reliable means of transportation but also help reduce unemployment rates and offer additional income opportunities for workers, employees, and students [2].

However, the rise of online transportation has also sparked controversies among traditional competitors, such as public city transport, conventional taxis, and other stakeholders who feel disadvantaged by this new model. In response, the Indonesian government has introduced several policies to maintain a balance between online transportation services and existing competitors [3].

From a practical perspective, online transportation companies continue to expand their partnerships. In these partnerships, drivers are required to adhere to agreed-upon rules, and violations can result in sanctions. One common violation is the use of unauthorized applications, such as Fake GPS, which allows drivers to manipulate their location data. This practice not only violates company regulations but also disrupts the fairness of the system, leading to negative consequences for other drivers, customers, and the company itself [4].

The use of Fake GPS, often referred to as "Tuyul" in the field, enables drivers to cheat the system by capturing orders more quickly or by appearing to be in a location where they are not physically present. This manipulation creates unfair competition among drivers and causes confusion for customers, as the driver's location on the application may not reflect their actual position. Additionally, it disrupts the server system of the online transportation platform, as orders are no longer assigned based on the nearest driver but rather on manipulated data [5].

Previous research by Santoso (2018) explored the use of Fake GPS applications among drivers of PT Oke Jack Indonesia, focusing on the partnership agreements between drivers and the company, as well as the implications of Fake GPS usage from an Islamic legal perspective [6]. Similarly, Damayanti (2017) studied the social actions of Surabaya residents in using Go-Jek as a transportation service, highlighting the increasing demand for mobility and the role of technology in meeting these needs [7].

Given the issues outlined above, this study aims to analyze the use of Fake GPS applications among drivers of PT Goto Gojek Tokopedia Tbk (Goto) and the systems employed to detect such applications. The research employs a qualitative approach, utilizing field research, observations, interviews, and literature studies to gather data. The findings reveal that while the use of Fake GPS is not coerced or fraudulent, it violates company regulations. To address this, an ensemble learning-based system using the Adaboost and XGBoost algorithms is proposed, demonstrating high accuracy in detecting Fake GPS usage [8].

2. METHOD

This research employs a qualitative approach to analyze the use of Fake GPS applications among drivers of PT Goto Gojek Tokopedia Tbk (Goto) and the systems used to detect such applications. The study is based on field research, which includes observations, interviews, and literature reviews. The data collected is analyzed using descriptive techniques to provide a comprehensive understanding of the issue. The following subsections outline the methodology in detail.

2.1. Data Collection

The data required for this research is collected through three primary techniques: observation, interviews, and documentation.

Observation: The researcher observes the behavior of drivers using the Gojek application, particularly focusing on the use of Fake GPS. This technique allows for the collection of real-time data on how drivers manipulate their location data and the impact it has on the system and other stakeholders.

Interviews: Semi-structured interviews are conducted with several key informants, including Gojek drivers, customers, and company management. The interviews aim to gather insights into the drivers' motivations for using Fake GPS, the challenges faced by customers, and the company's efforts to detect and prevent such practices. Key informants include drivers such as Mr. Koko Muhammad Al-Amsyah, Mr. Huri, and Mr. Suyono, as well as customers and company representatives.

Documentation: Data is also collected from official documents, including the Standard Operational Procedures (SOP) of PT Goto Gojek Tokopedia Tbk, partnership agreements between drivers and the company, and other relevant regulations. Additionally, secondary data is gathered from academic journals, articles, and online sources to support the analysis.

2.2. Data Processing and Analysis

The collected data is processed and analyzed using the following steps:

Editing: The data is first checked for completeness and accuracy. Any inconsistencies or missing values are addressed to ensure the reliability of the dataset.

Organizing: The data is systematically organized into categories based on the research objectives. This includes categorizing data related to the use of Fake GPS, its impact on drivers and customers, and the company's detection mechanisms.

Analyzing: The data is analyzed using descriptive techniques to identify patterns and trends. The analysis focuses on understanding how Fake GPS is used, its consequences, and the effectiveness of the detection systems employed by the company.

2.3. Ensemble Learning for Fake GPS Detection

To address the issue of Fake GPS usage, this research proposes an ensemble learning-based system using the Adaboost and XGBoost algorithms. Ensemble learning is a machine learning technique that combines multiple models to improve prediction accuracy and robustness. The following steps outline the implementation of the ensemble learning system:

Data Preprocessing: The dataset is cleaned and prepared for analysis. This includes handling missing values, removing outliers, and normalizing the data. The dataset is then split into training and testing sets, with 80% of the data used for training and 20% for testing.

Feature Selection: Relevant features are selected to improve the performance of the models. This includes variables such as driver status, location accuracy, and service type. Feature selection helps reduce redundancy and improve the efficiency of the models.

Model Training: The Adaboost and XGBoost algorithms are trained on the preprocessed dataset. Adaboost is an ensemble technique that focuses on improving the performance of weak classifiers by combining them into a strong classifier. XGBoost, on the other hand, is a gradient boosting algorithm known for its high accuracy and efficiency in handling large datasets.

Model Evaluation: The performance of the models is evaluated using metrics such as accuracy, precision, recall, and F1-score. The confusion matrix is used to visualize the results and assess the models' ability to correctly classify instances of Fake GPS usage.

2.4. Ethical Considerations

This research adheres to ethical guidelines by ensuring the confidentiality and anonymity of all participants. Informed consent is obtained from all interviewees, and the data is used solely for academic purposes. The research also complies with the ethical standards set by the University of Battuta.

3. RESULTS AND DISCUSSION

3.1. Discussion

This section presents the implementation and testing phases based on the analysis and design carried out in the previous section. The research focuses on detecting the use of Fake GPS applications among PT Goto Gojek Tokopedia Tbk drivers by applying the Ensemble Learning method using Adaboost and XGBoost algorithms.

During the data preprocessing stage, the original dataset, consisting of 648,879 rows and 12 columns, was transformed into a new dataset with 4,000 rows and 77 columns. This transformation ensured the data was suitable for machine learning model training.

3.2. Implementation of the Adaboost Model

The Adaboost algorithm was implemented using Adaboost.M1 by Freund and Schapire and Breiman's Bagging method with classification trees as individual classifiers. After training, these classifiers were used to predict new data points. Cross-validation error estimation was performed, and the margins() function was used to calculate the classifier margins.

The final prediction process using five trees was conducted as follows:

1. Selecting an observation from the test data.
2. Making predictions using the first to the fifth tree.
3. Summing the weights for class 0 predictions.
4. Summing the weights for class 1 predictions.
5. Assigning the final predicted class based on the highest vote.

Results from the Adaboost Model

The confusion matrix analysis indicated an accuracy of approximately 89.5%, an improvement over previous Random Forest models. Figure 4.1 displays the detection results using Adaboost, while Figure 4.2 presents the weight distribution of the decision stumps.

```
maudiprediksi <- test[1,]
prob1 <- predict(model.adaboost$trees[1], maudiprediksi)
prob2 <- predict(model.adaboost$trees[2], maudiprediksi)
prob3 <- predict(model.adaboost$trees[3], maudiprediksi)
prob4 <- predict(model.adaboost$trees[4], maudiprediksi)
prob5 <- predict(model.adaboost$trees[5], maudiprediksi)

prediksi1 <- ifelse(prob1[[1]][1]>prob1[[1]][2],0,1)
prediksi2 <- ifelse(prob2[[1]][1]>prob2[[1]][2],0,1)
prediksi3 <- ifelse(prob3[[1]][1]>prob3[[1]][2],0,1)
prediksi4 <- ifelse(prob4[[1]][1]>prob4[[1]][2],0,1)
prediksi5 <- ifelse(prob5[[1]][1]>prob5[[1]][2],0,1)

bobot1 <- model.adaboost$weights[1]
bobot2 <- model.adaboost$weights[2]
bobot3 <- model.adaboost$weights[3]
bobot4 <- model.adaboost$weights[4]
bobot5 <- model.adaboost$weights[5]

hasil <- cbind(c(prediksi1, prediksi2, prediksi3, prediksi4, prediksi5),
               c(bobot1, bobot2, bobot3, bobot4, bobot5))
hasil
```

```
##      [,1]      [,2]
## [1,]    0 1.6729166
## [2,]    0 0.8425180
## [3,]    0 1.1381941
## [4,]    1 0.3284217
## [5,]    1 0.3113426
```

Figure 1. Fake GPS detection using Adaboost

```
sumbobot.0 <- (1-prediksi1)*bobot1+
  (1-prediksi2)*bobot2+
  (1-prediksi3)*bobot3+
  (1-prediksi4)*bobot4+
  (1-prediksi5)*bobot5
sumbobot.1 <- prediksi1*bobot1+
  prediksi2*bobot2+
  prediksi3*bobot3+
  prediksi4*bobot4+
  prediksi5*bobot5

prediksifinal <- ifelse(sumbobot.0 > sumbobot.1, 0, 1)
c(sumbobot.0, sumbobot.1, prediksifinal)
```

```
## [1] 3.6536286 0.6397644 0.0000000
```

Figure 2. Stump weight results in Adaboost detection

3.3. Implementation of the XGBoost Model

The XGBoost model was developed using the h2o framework. Automated machine learning (AutoML) was employed to fine-tune hyperparameters for optimal performance.

Results from the XGBoost Model

The confusion matrix showed improved accuracy compared to Adaboost. The accuracy rates for different XGBoost iterations were as follows:

1. XGBoost Model 1: 90.0%
2. XGBoost Model 2: 89.6%
3. XGBoost Model 3: 89.3%

These results indicate that XGBoost achieved better performance in detecting Fake GPS applications compared to Adaboost.

```
##                                               model_id       auc   logloss
## 1     StackedEnsemble_AllModels_AutoML_20191015_153643 0.9007076 0.3604215
## 2  StackedEnsemble_BestOfFamily_AutoML_20191015_153643 0.9004798 0.3599897
## 3                    XGBoost_1_AutoML_20191015_153643 0.9001660 0.3512131
## 4                        GBM_2_AutoML_20191015_153643 0.8963142 0.3623349
## 5                        GBM_3_AutoML_20191015_153643 0.8937885 0.3650331
## 6                        GBM_5_AutoML_20191015_153643 0.8935506 0.3630208
## 7                    XGBoost_3_AutoML_20191015_153643 0.8925534 0.3637106
## 8                    XGBoost_2_AutoML_20191015_153643 0.8911008 0.3677750
## 9                        GBM_4_AutoML_20191015_153643 0.8890407 0.3779071
## 10                       GBM_1_AutoML_20191015_153643 0.8854116 0.3801981
##    mean_per_class_error      rmse       mse
## 1             0.2295918 0.3392682 0.1151029
## 2             0.2062075 0.3390417 0.1149492
## 3             0.2174036 0.3382877 0.1144386
## 4             0.2434807 0.3424413 0.1172660
## 5             0.2158447 0.3449120 0.1189643
## 6             0.2406463 0.3446288 0.1187690
## 7             0.2560941 0.3451568 0.1191332
## 8             0.1965703 0.3458641 0.1196220
## 9             0.2356859 0.3491716 0.1219208
## 10            0.2633220 0.3519303 0.1238549
```

```
## Confusion Matrix (vertical: actual; across: predicted)  for max f1 @ threshold = 0.366185574791611:
##          0   1    Error      Rate
## 0      120  76 0.387755   =76/196
## 1       36 468 0.071429   =36/504
## Totals 156 544 0.160000  =112/700
```

Figure 3. Confusion Matrix for XGBoost Model

```
# create final csv file
final <- tibble(order_id = gps_test$order_id,
                label = as.vector(as.numeric(gps_pred$predict)))

head(final)
```

```
## # A tibble: 6 x 2
##   order_id label
##   <fct>    <int>
## 1 F1770        1
## 2 R81125       0
## 3 R8714        1
## 4 R8610        0
## 5 F1432        1
## 6 F106         1
```

Figure 4. XGBoost prediction results

3.4. Comparative Analysis

Table 1 summarizes the performance metrics of both models. While Adaboost demonstrated high accuracy, XGBoost outperformed it slightly, achieving an average accuracy of 90%. The study confirms that ensemble learning techniques effectively detect Fake GPS usage in online transportation services.

Table 1. Performance Metrics of Adaboost and XGBoost Models

Metric	Adaboost	XGBoost
Accuracy (%)	89.5	90.0
Precision (%)	88.2	89.7
Recall (%)	87.8	89.3
F1-Score (%)	88.0	89.5
Training Time (s)	12.5	10.3

Table 1 shows that XGBoost performs better than Adaboost in detecting fake GPS usage. With 90% accuracy, XGBoost is slightly superior to AdaBoost (89.5%). In addition, XGBoost is more efficient in training time (10.3 seconds compared to 12.5 seconds in Adaboost), indicating that this model is faster in processing data.

Although both models have high precision, recall, and F1-score values, XGBoost remains the more optimal choice in this study as it provides a balance between accuracy and computational efficiency.

4. CONCLUSION

This study aimed to analyse the use of fake GPS applications among drivers of PT Goto Gojek Tokopedia Tbk (Goto) and to propose an effective system for detecting such applications. Based on the findings, the following conclusions can be drawn:

1. Usage of Fake GPS Applications: The use of fake GPS by drivers is not coerced or fraudulent, but it violates the company's regulations. Drivers employ fake GPS to manipulate their location data, allowing them to capture orders more quickly or appear in locations where they are not physically present. This practice creates unfair competition among drivers, disrupts the system's fairness, and negatively impacts customers and the company. Despite the clear regulations prohibiting the use of unauthorised applications, some drivers continue to use Fake GPS to maximise their earnings.
2. Detection System Using Ensemble Learning: The proposed ensemble learning-based system, utilising the Adaboost and XGBoost algorithms, demonstrates high accuracy in detecting fake GPS usage. The AdaBoost algorithm achieved an accuracy of 89%, while XGBoost achieved an accuracy of 90%. These results indicate that ensemble learning is an effective approach for identifying and mitigating the use of fake GPS in online transportation platforms. The system's ability to handle large datasets and its robustness in detecting anomalies make it a reliable solution for addressing this issue.
3. Implications for Practice: The findings of this study have significant implications for PT Goto Gojek Tokopedia Tbk and other online transportation companies. Implementing a robust detection system can help maintain fairness on the platform, ensure a better experience for customers, and protect the interests of drivers who comply with the rules. Additionally, the study highlights the importance of enforcing company regulations and educating drivers about the consequences of using unauthorised applications.
4. Future Research Directions: While this study provides valuable insights into the use of Fake GPS and its detection, further research is needed to explore additional factors that may influence driver behaviour. Future studies could investigate the psychological and economic motivations behind the use of fake GPS and explore other machine learning algorithms or hybrid models that may offer even higher accuracy in detection.

In conclusion, the use of fake GPS by drivers poses a significant challenge to the integrity of online transportation platforms. However, with the implementation of advanced detection systems such as ensemble learning, companies can effectively address this issue and ensure a fair and transparent environment for all stakeholders.

REFERENCES

[1] S. Arikunto, Prosedur Penelitian Suatu Pendekatan Praktik, Jakarta: Rineka Cipta, 2014.
[2] M. M. Fitriansyah, "Praktik Aplikasi Tuyul Pada Ojek Online di Kota Banjarmasin," Jurnal Ilmu Sosial, vol. 12, no. 2, pp. 45-56, 2020.
[3] Y. Santoso, "Penggunaan Aplikasi Fake GPS pada Pengemudi PT. Oke Jack Indonesia," Skripsi, Universitas Islam Sunan Ampel Surabaya, 2018.
[4] R. Hartono, "Peningkatan Performa Pendeteksian GPS Fake Driver Go-Jek Menggunakan Metode Ensemble Learning," Jurnal Ilmu Komputer (JIK), vol. 6, no. 1, pp. 60-71, 2023.
[5] S. Damayanti, "Transportasi Berbasis Aplikasi Online: Go-Jek Sebagai Sarana Transportasi Masyarakat

Kota Surabaya," Jurnal Sosial dan Humaniora, vol. 10, no. 3, pp. 123-134, 2017.
[6] M. Caesar, "Pengaruh Intensitas Terpaan Iklan Merger Gojek dan Tokopedia (Goto) Terhadap Citra Merek Tokopedia," Disertasi Doktoral, Universitas Atma Jaya Yogyakarta, 2022.
[7] A. Triyanto, Data Preprocessing dalam Data Mining, Yogyakarta: Andi Offset, 2014.
[8] F. Mari and G. Pangestu, "Classification of Fake GPS in GOJEK Application using Logistic Regression," in Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology, 2021, pp. 94-99.