Driver Predicting Behavior Based on Accelerometer and Gyroscope Sensors with K-Means Algorithm Method

Michael Sitorus¹, Melisa², Desty Herninda³
¹,²,³Information System Department, BRI Institute
¹michael.sitorus@bri-institute.ac.id, ²melisa260202@gmail.com, ³destyaherninda36908@gmail.com

Abstract

The transportation with an increasing population level in Indonesia will also affect the smoothness of traffic. However, with the existence of transportation that is owned by someone, it is often misused in driving. Aggressive driving behavior is a major factor in traffic accidents. As reported by the AAA Foundation for Traffic Safety, 106,727 dangerous accidents – 55.7 percent of the over past four year period involved drivers engaging in one or more acts of aggressive driving. Therefore, to predict driver behavior, the K-Means algorithm used the Orange Application. Using the driver behavior dataset that has been recorded with Accelerometer and Gyroscope Sensor-based applications, the results obtained where the k value is the number of cluster 1 there are 814 items, cluster 2 has 997 items, cluster 3 has 1273 items. From this test, it was found that proportion of predicted driver behavior with Aggressive qualifications was 96.4%, Normal was 92.2% and from driver behavior with Low qualifications was 74.3%. The accuracy rate of research using the K-Means algorithm to determine predictions is 0.861 or 86.1%. The results of predictions can help prevent accidents or other risks that will occur in the future.

Keywords: Accelerometer, Gyroscope, Sensor, K-Means Algorithm, Predictions

1. Introduction

In the 4.0 era, there have been several modern transportations that have sprung up in Indonesia. The more transportation with an increasing population level in Indonesia will also affect the smoothness of traffic. However, with the existence of transportation that is owned by someone, it is often misused in driving. As we see and know, everyone has a different driving behavior.

Driver behavior can be categorized according to several types of speed, namely aggressive, normal, slow. Where all three have their respective advantages and disadvantages. Many people may think that people who like to drive at high speed are dangerous [1]. Meanwhile, driving at high speed also increases the risks that could happen to motorists on the road. Such as the risk of crime, accidents, robbery, to get a ticket. We can actually minimize this risk.

Aggressive driving behavior is a major factor in traffic accidents. As reported by the AAA Foundation for Traffic Safety, 106,727 fatal accidents – 55.7 percent of the total – over the past four year period involved drivers engaging in one or more acts of aggressive driving [2]. This study was conducted over four years, and the most common aggressive acts were:

1. Erratic or reckless driving
2. Not obeying traffic signs or signs
3. Illegal turns
4. Speed up
5. Failure to grant the right of way
6. Illegal passing
7. Tailgating

Aggressive driving also causes countless no dangerous accidents that often result in serious injury. It's best to stay calm behind the wheel.

The problems raised have been researched with the title “Driver Behavior Scoring Application Using Fusion Sensors” [3]. In this study. Researchers reiterate the problem of driver behavior that can increase mortality. By predicting the behavior of drivers using applications based on Accelerometer and Gyroscope sensors with the K-Means Algorithm Method.

Traffic issues is a vital thing that needs to be studied on it [4]. With the existence of sophisticated technology in the current era, it makes it easier for humans to be able to record driver behavior data while driving through an application based on Accelerometer and Gyroscope sensors. The data will later help in predicting driver behavior in driving. The results of these predictions can prevent accidents or other risks that will occur in the future.
2. Research Methods

The characteristics of drivers are important in understanding their behavior on the road. These characteristics encompass various factors, including the driver's natural abilities, learning capacity, and motives. By examining a driver's motives and attitudes, it becomes possible to gain insights into why they behave the way they do behind the wheel.

Accelerometer is a device used to measure acceleration and the effect of gravity on acceleration [5]. When the speed of an object changes, the object accelerates. Acceleration or what is often called acceleration is the rate of change of velocity with respect to time. Just like speed, acceleration is also divided into two, namely average acceleration and instantaneous. Acceleration is the result of the derivative of velocity with respect to time and velocity is obtained from the derivative of displacement with time.

Meanwhile, A gyroscope is a navigation tool that can used to measure the orientation of the device. Manner The work of the gyroscope is to measure angular velocity and how quickly it surrounds an axis [6]. Accelerometer and gyroscope sensors are sensors that are often used to determine the type of movement [7] and angular position [8]. The workings of this system can be seen in Figure 1.

3. Results and Discussions

Research methods are basically a scientific way to obtain data with a specific purpose and usefulness. The scientific way means research activities that is based on scientific features, namely rational, empirical, and systematic. Rational means research activities carried out in a reasonable way, so that it is affordable by human reasoning. Empirical means the ways that done it can be observed by the human senses, so that others can observe and know the means used. Systematic means a process that used in research that uses certain steps that is logical [9].

Methods are steps or ways to achieve a certain goal. The research method can be explained as a way to achieve research objectives, in this case the steps are carried out with a scientific approach (rational, empirical and systematic). In general, research methods are classified into several types, including qualitative, quantitative, and combination. So based on the description of the theory above, it can be concluded that the research method is sharing ways, steps, and models for conducting a study. The method used by the author as a means and guide in this study is to use the K-Means Algorithm.

K-Means is one of the attempted non-hierarchical data clustering methods partition the existing data into the form of one or more clusters or groups so that the data which have equally detrimental characteristics in the same cluster and data which have different characteristics from other groups [10].

A well-known algorithm in the clustering method is K-means because its fairly simple algorithm and efficient[11]. The K-means algorithm uses a 14 iterative process to get a cluster database. It takes the desired number of initial clusters as input and produces the final centroid point as output.

DOI: https://doi.org/10.31004/riiggs.v2i1.29
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)
The K-means method will choose pattern k as the starting point of the centroid randomly. The number of iterations to reach the cluster centroid will be influenced by the initial cluster centroid candidates randomly. So that we get a way of developing the algorithm by determining the centroid cluster which is seen from the high initial data density in order to get higher performance.

K-Means Clustering is a method in data mining, which is used to perform grouping large amounts of data. Where from each data is in one group, has characteristics that same or different characteristics with the data in other groups [12]. In its completion, the K-Means algorithm will produce a centroid point which is the goal of the K-Means algorithm. After the K-Means iteration stops, each object in the dataset becomes a member of a cluster. The cluster value is determined by searching all objects to find the cluster with the closest distance to the object. The K-means algorithm will group data items in a dataset into a cluster based on the shortest distance [13].

The initial centroid value chosen randomly which is the starting point, will be calculated the distance to all data using the Euclidean Distance formula. Data that has a short distance to the centroid will form a cluster. This process continues until there is no change in each of the 15 groups [14].

Steps in conducting clustering using the K means algorithm as follows [15]:

1. Specify the Number of clusters k.
2. Initialize the cluster center K. It can done in different ways. But the most numerous ways used is with determine randomly. Center of the cluster is assigned an initial value with random numbers.
3. Allocate all data/objects to in the nearest cluster. Closeness Two data/objects can be obtained based on the distance of the two objects aforementioned. To be able to calculate the distance of all data to the center point cluster then can use the Euclidean distance theory formulated below:

$$D(ij) = \sqrt{(X_{ki} - X_{ij})^2 + (X_{2i} - X_{2j})^2 + \cdots + (X_{ki} - X_{kj})^2}$$

where:

- $D(ij) = $ Distance of data to i to center Cluster J
- $X_{ki} = $ Data to i attribute data to k
- $X_{kj} = $ Center point to j on attribute to k

4. Recalculate cluster center with cluster membership that now. The cluster center is the average of all objects/data in specific clusters. If desired could also use the median of the cluster. So the average (mean) is not the only measure wearable.

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \cdots + X_{nk})$$

where:

- $R_k = $ New average.
- $N_k = $ Number of training patterns on the cluster (k).
- $X_{nk} = $ Pattern to (n) that becomes cluster section (k).

5. Reassign each object using the center of the new cluster. If the center cluster does not change anymore then The clustering process is complete. Or Go back to step number 3.

![Flowchart K-Means Algorithm](image-url)

The Research Process on Driver Behavior with the K-Means Algorithm is carried out using the Orange application. In Figure 3 is a Research Scheme using Orange.
4. Conclusion

4.1. Research Dataset

This study uses a driver behavior dataset by taking 2 data samples per second. This data includes Speed (x, y, z axis in meters per second squared (m/s²)), Rotation (X, Y, Z axis in degrees per second (°/s)), Classification label (Aggressive, Normal, Slow), Timestamp (time in seconds) All existing data is the result of an application recorder designed on the basis of accelerometer and gyroscope sensors.

4.2. Process

At this stage, the data that has been input is then clustered to determine the prediction results with the classification or category of Aggressive, Normal, Slow driver behavior. Before clustering the data will be entered into the preprocessing after getting the results of the data table that has been processed, then it can then be entered into a bar chart or distribution.

4.3. Results

The clustering results from the preprocessing and processing stages can then be processed with the K-Means Algorithm. At the processing stage of the K-Means Algorithm, 3 clusters are taken. To get the results of the Confusion Matrix, testing and scoring of the data that has been processed is first carried out. The test results and scoring use the results of the Logical Regression test which can be seen in Figure 7.
Figure 7. Test Results and Scoring
Where in Figure 8 and Figure 9, 3 clusters are taken according to the K-Means method where the k value is the number of cluster 1 there are 814 items, cluster 2 has 997 items, cluster 3 has 1273 items.

Figure 8. Confusion Matrix Results
From the figure below it can be seen that the total data is 3084 items. From this test, it was found that the proportion of predicted driver behavior with the Aggressive classification was 96.4%, Normal was 92.2% and from driver behavior with Low classification was 74.3%.

Figure 9. Percentage Rate Proportion of Predictions
The accuracy rate of research using the K-Means algorithm to determine predictions of driver behavior is 0.861 or 86.1%.

4. Conclusion
The conclusion, the research has produced several important findings. Firstly, the driver behavior data can be divided into three distinct clusters: Aggressive, Normal, and Low. Secondly, testing conducted using the Orange application resulted in an accuracy level of 86.1%, based on a recall value of 83.5% and a precision score of 86.1%. However, it is worth noting that while the system is capable of predicting driver behavior, it is not always precise. This lack of precision may be attributed to the values generated by the sensors in the data recording application. Overall, these findings suggest that technology can be used to gain insights into driver behavior, but further research is needed to improve the accuracy of these predictions

Reference


