



UK Car Accidents

CS 210 Group Project

19 MAY 2019 / GIRGINDILARA / LEAVE A COMMENT

Project Evaluation

What were the difficulties you encountered during the project?

The first thing that was encountered during the project was to arrange the space of `param_grid` in `GridSearchCV`. When the space was large, it took time to get the output and probably we should wait a lot. First, the following code was written:

```
param_grid = {  
  
    "max_depth": np.arange(1, 10),  
  
    "min_samples_split": np.arange(2, 20),  
  
    "min_samples_leaf": np.arange(2, 20)  
  
}
```

Then, because of the reason mentioned above, it was attempted to get the output by lessening the space. After trying various combinations, the following code was used to get the output:

```
param_grid = {  
  
    "max_depth": np.arange(1, 4),  
  
    "min_samples_split": np.arange(2, 4),  
  
    "min_samples_leaf": np.arange(2, 4)  
  
}
```

The second problem that was encountered during the project was to find the proper hypothesis because features of your hypothesis is significant in terms of the implementation of your code. You should select the suitable one for the flow of the subsequent steps.

If you were given sufficient amount of resources, what additional datasets would you utilize?

The relationship between the accidents and causes of accidents were examined and were tried to find were there any correlation between any of these conditions. Accident Severity, road conditions etc. are the attributes we have in our dataset. In addition to these attributes, we could have been examined the contribution of accidents to the supplier economy. Moreover, speeding, drunk driving and reckless driving are also shown as the main reasons for the car accidents. If we have additional dataset or resources which contains these data, we can demonstrate the relationship between the accidents and these attributes in more accurate way.

Compare the machine learning algorithms you used, in terms of performance and applicability to your dataset.

Decision Tree and Random Forest were used as two machine learning algorithms. Because the Decision Tree is a subset of Random Forest, doing one of these machine learning algorithms eased the other's implementation.

In terms of performance, Decision Tree is better. This arises from the implementation of Random Forest since Random Forest is tested more than one Decision Tree. That is, by using the same feature, it can be create more than one decision tree. It can be obtained more than one better solution from this more than one decision tree. It cannot be figured out at first glance which one will give the better solution. By stating the better solution, it meant to increase the prediction accuracy of the selected attribute. Briefly, all of these implementations are done for this purpose. For this reason, we came up a reduced run time in Decision Tree compared to Random Forest.

What improvements could have been done in your project?

In the stage of hypothesis testing, we could not find any correlation. We could not try enough to find any correlation. It can be good to find any correlation or use a dataset which we can find a correlation.

Bunu paylaş:



Be the first to like this.

Published by girgindilara

[View all posts by girgindilara](#)

 Genel

 Edit

Leave a Reply

Enter your comment here...

PREVIOUS

Linear Regression

BLOG AT WORDPRESS.COM.