**An Interactive Web Tool for Understanding the Vehicle Behaviour**

**Details**

First name, Surname       Hongyang Liu

University                       Halmstad University

Course Name                  Project in Computer Science and Engineering

Supervisors                     Yuantao Fan, Reza Khoshkangini

**Introduction**

Traveller sometimes face the mechanical problems when buses systems overwork or administrators of bus companies could not find abnormal operation immediately.  These phenomena would reduce the efficiency of buses system.

With the increasing number of signal data for public buses, many abnormal operations of the transportation could be found via the data analysis. Through the anomaly detection techniques,  administrators of bus companies could even predict the mechanical problems.

The general goal of this project was therefore to help administrators to detect unusual bus operations via an interactive web interface.

* Related Work

Stacey Ronaghn(2018) wrote an article regarding a way to detect outliers or unusual behaviour via machine learning algorithm. Her objective was to introduce the anomaly detection techniques which  includes: The use-cases are detecting abnormal behaviour of equipment in a manufacturing plant using data such as temperature, pressure and humidity; The most common anomaly detection algorithms was for anomaly detection: K-Means, One-class Support Vector Machines, and Autoencoders[1].

* Contribution

There are different types of bus operations like GPS, Selected gear, Vehicle speed, etc. These data could provide support for data scientists and fleet operators. However, it is hard for an administrator who knows little about bus troubleshooting skills to understand the behaviour of city bus just through pure data. To solve the problem, the project generally could be split into two parts including making a visualization tool and data analysis.

The objective of this project is firstly to make an interactive website for administrators of bus company. This would contribute to them understanding the bus data.

After that, administrators could select any specific data from the website and label them for further analysis.

* + Visualization tool
    - Interactive website should have functions of comparing different kinds of bus signal features and locate the position of the vehicle.
    - Interactive website should provide users the filter functions
    - Interactive website’s tools are easy and simple for users
  + Data analysis
    - Data analysis part has the function that labels the data.
    - Supervised learning algorithms should be combined with the website interface.
    - The training data patterns should be applied for prediction.
* Limitations
  + In the back-end web interface tool, we have much improvements that could be applied. However, because of the lack of time,  some machine learning algorithms have not been used in the project. For example, there are only two features including timestamp and the feature chose by users form dropdown list. The two-class algorithms cannot solve multi-features problems.
  + Because the data is massive, the computer cannot handle the data immediately. The system still need some time to solve its efficiency.

**Methods**

* Tools

To accomplish the web based application, the objective is that users could see the results of the vehicle operation including the map and some other operation like engine speed.

And then users could choose what they would like to analyse. They could put these data into labelling model part.

Users next could label the data through the buses’ behaviour. After labelling, the algorithms would calculate the pattern between input data and output data. After that, at the application part, users could choose best one to use other data and the system would give a results. The results may be helpful for users to analyse the operation of the vehicle like whether the abnormal ratio in the whole routine. There are four sections/parts which should be achieved.

* + data visualization

Users could choose calendar, dropdown and slider bar to filter the number of  data. The results would be plotted immediately.

* + labelling

Users could select data and label it. They can also do it on other data as well.

* + modelling the evaluation

Users could classify the data to find normal and abnormal or usual or unusual rows.

* + labelling automatically

The system would plot the evaluation of data analysis. Accuracy is a metric to do evaluation.

* Procedures
  + Filter data

In this task, we should filter the data to train the algorithm, the data comes  from the original signals which detected by automatic censors. The signal change with time series. Because the number of data is massive, users could acquire the data via choosing the components like calendar, dropdown and slider.

* + Plot data

The data chose by last procedure should be plotted instantly with scatter diagram

* + Select data

Users select the data which used for training.

* + Label data

Users could classify the data to meet their requirements like classification.

* + Algorithm

Some supervised learning algorithms should be used for data classification like KNN, Neural Network, SVM and so on.

* + Evaluation

The system plot the results of the data analysis. Accuracy is a metric for performance measure.

* + Application

Users could choose the censor data and then use the best algorithm which trained in last procedure to predict the classification of the censor data.

* Materials

The proposed web functions were developed using Dash and Scikit-learn. To build an interactive website, we compared several program languages. Python is good for fast-prototyping, this could reduce the complexity of the program and save developer amount of time. Python language has some other advantages, the compatibility of some machine learning libraries like Scikit-learn, improving the efficiency of the program

* + Data processing

To improve the speed of websites,  data can be transferred from database to csv file.  To select the data what user need, the web system use the select box and  stored techniques to accomplish the function.

* + l

Plotly is a modern analytic apps for web interface which combined python and some front-end language.

* + Dash

Dash is a productive Python framework for building web applications. Through a couple of simple patterns, Dash abstracts away all of the technologies and protocols that are required to build an interactive web-based application which could be shared through URL.

* + Algorithm

Supervised learning algorithm like KNN, SVM, Binary Tree and Neural Network .

* + - KNN
    - SVM
    - Binary Tsaree
    - Neural Network
  + Scikit-learn

Scikit-learn is a free software machine library for the python programming language. It features various classification, regression and clustering algorithms[4].

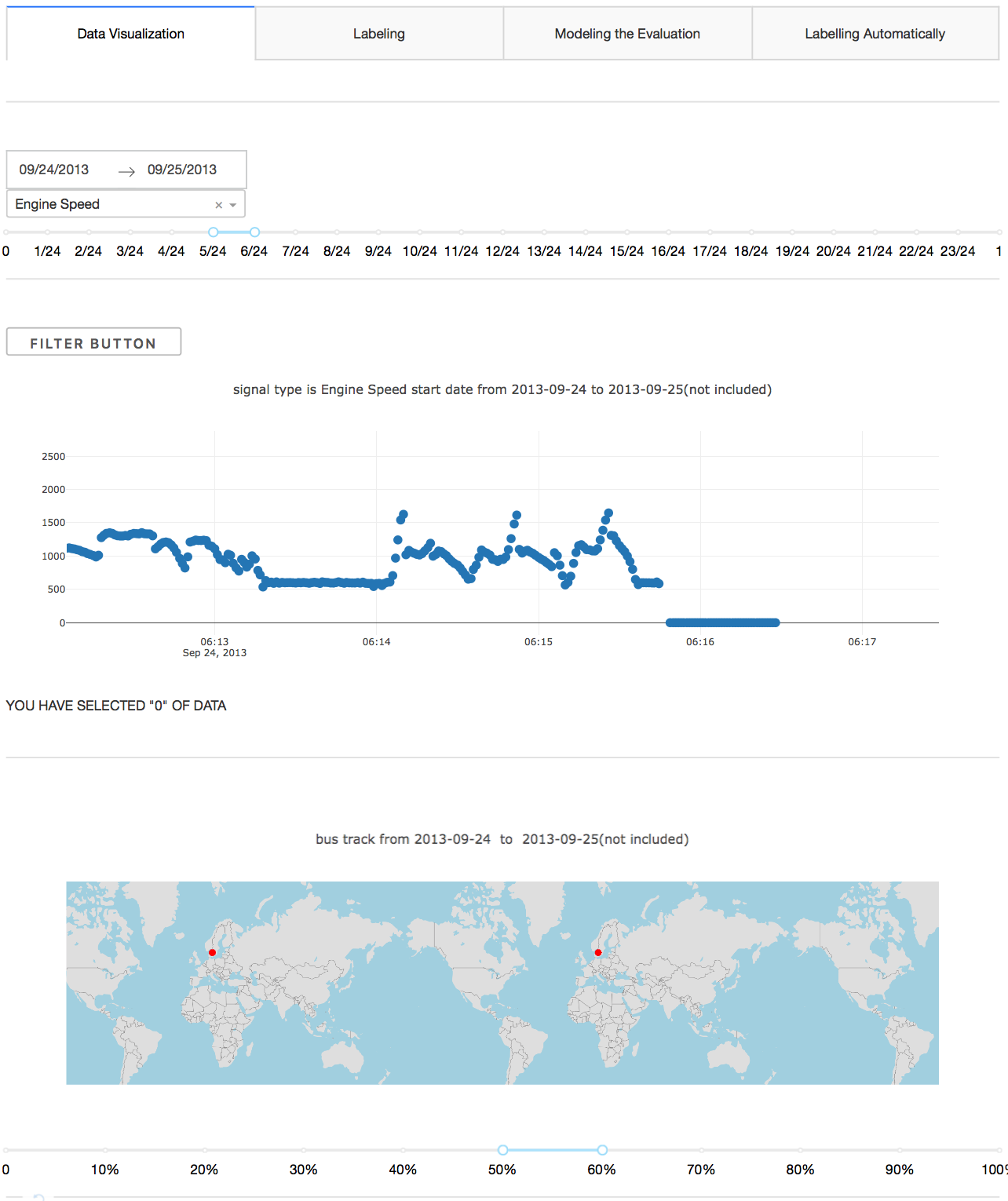
Scikit-learn also has some metrics methods to evaluate the performance of the system, like ROC curve, accuracy, FN rate, etc. These measurement would help the users to choose the best method to apply in the application[6]

* + PyCharm

PyCharm is an intergraded development environment. It is helpful for computer programming, especially for the Python language[7].

**Results**

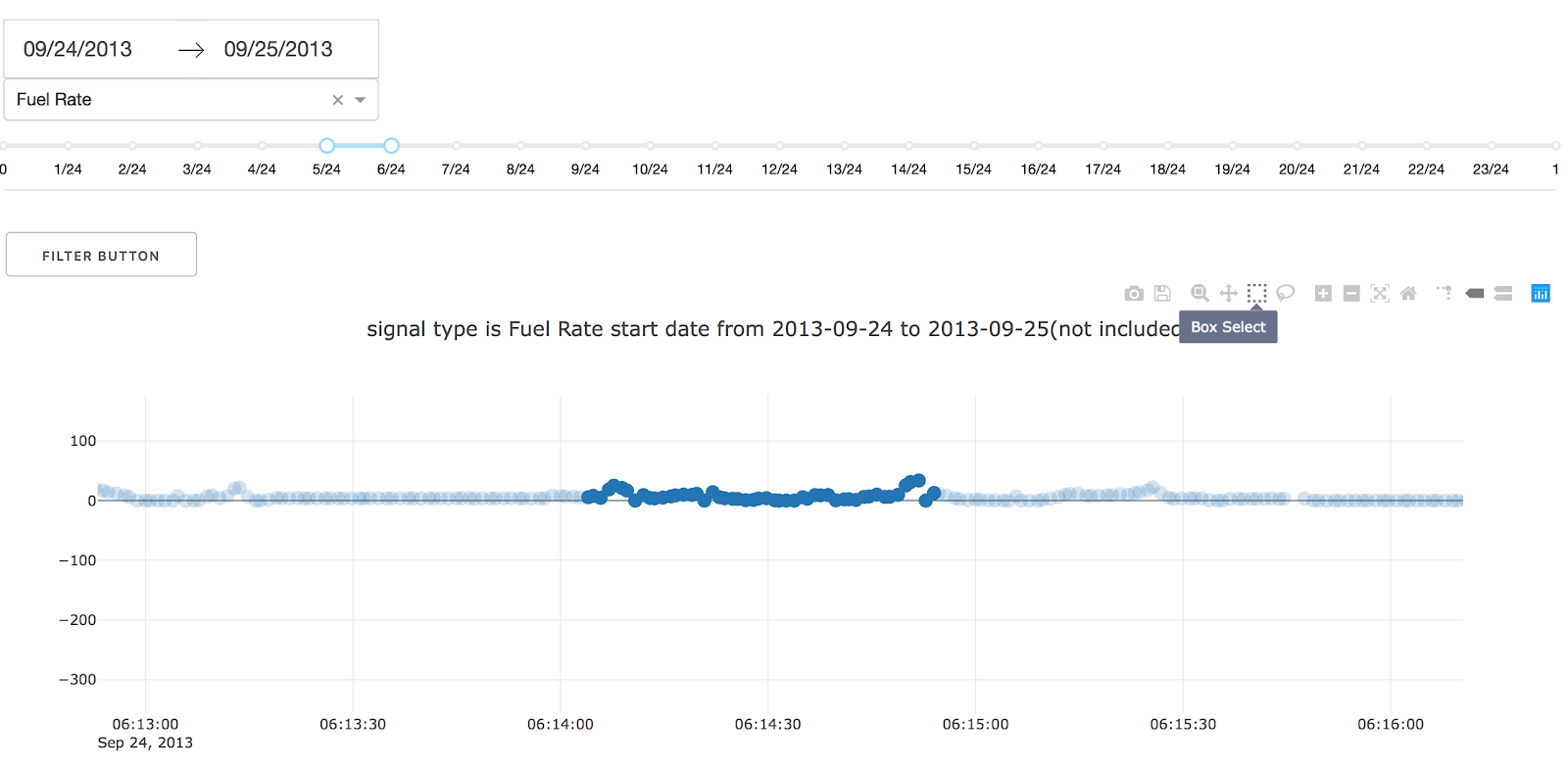
In this part, there are several features to describe four tabs of functions including data visualization, labelling, modelling the evaluation and labelling automatically.



**Figure 1 data visualization**

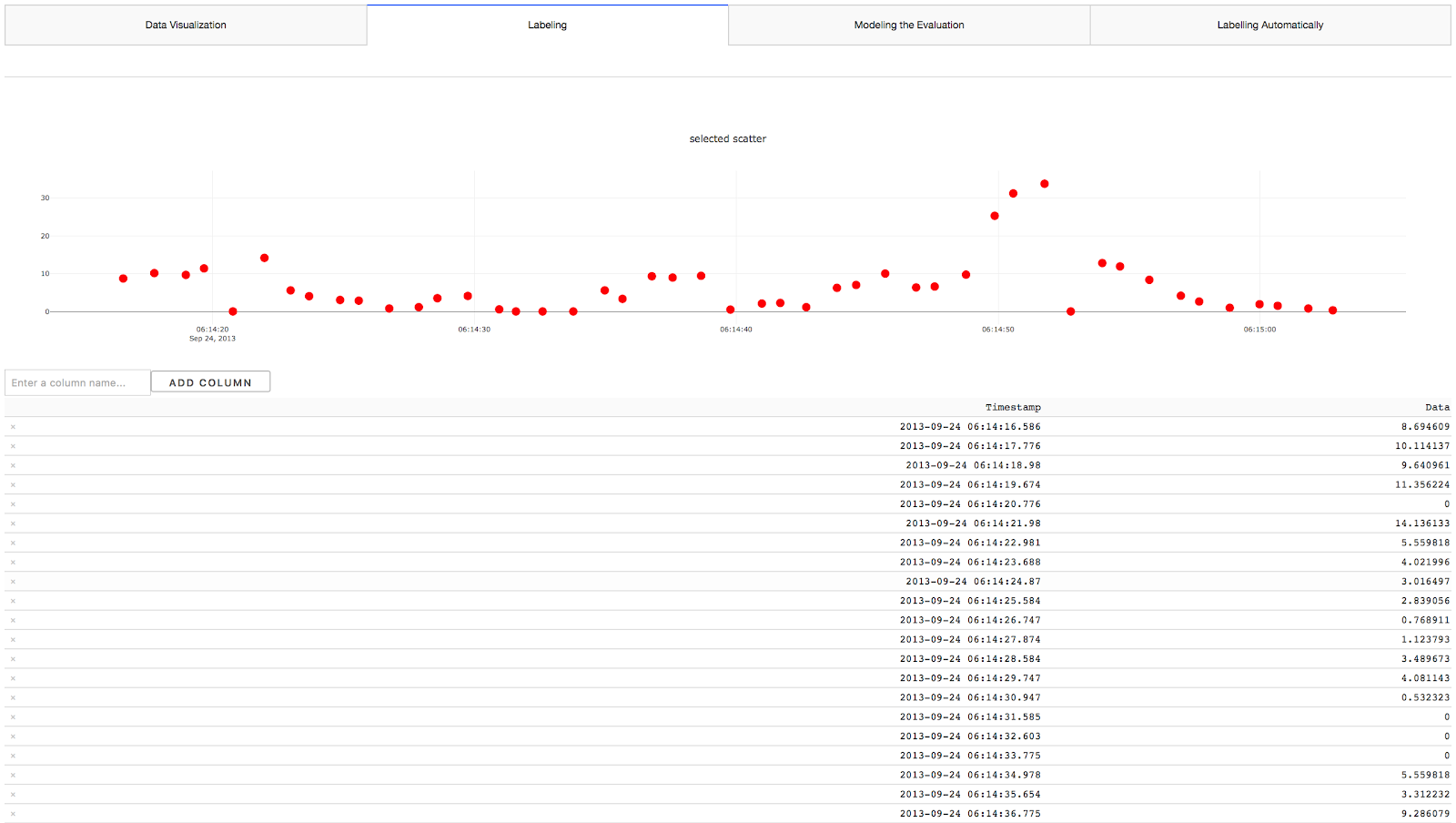
**Figure 1**: Users could filter the data via some components. In figure one, users choose different types of data and plot them. They can also see the track of the bus in the map. Users could choose calendar, dropdown and slider bar to filter the number of  data. The results would be plotted immediately.

Interactive website was built to illustrate the signal scatter data which could be limited through time duration and signal type. Users could also click map tab, the bus track would be showed with timestamp, longitude and latitude.



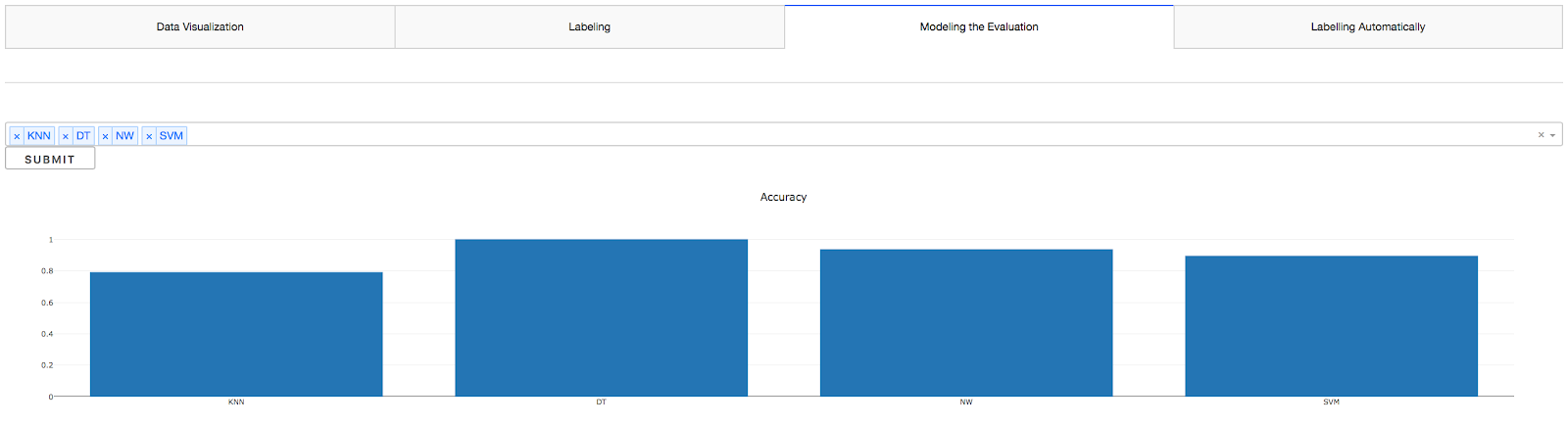
**Figure 2 Selected data to label**

**Figure 2**: Figure 2 also describes the function of data visualization part. Users can select the data according to their demand. It is fuel rate start date from 2014-09-24 to 2013-09-25. It is one day’s data.  The slider split the data into 24 segments, the filter just plots them from 5/24 to 6/24 and also the data would be stored in the table in next tag. Users could also apply the pattern to other data type.



**Figure 3 Labelling**

**Figure 3:**  It is labelling data part. Users could classify the data to find normal and abnormal or usual or unusual rows according to the selected scatter which chose by users in data visualization part. This is the procedure of labelling data on supervised learning.



**Figure 4 Modelling Evaluation**

**Figure 4**: Users could choose the algorithms what they decide to use. And then, the system algorithms would handle the data and calculate the accuracy using the supervised learning algorithms. This means the ratio which the right data divides the whole data. The system would plot the evaluation of data analysis. Accuracy is a metric to do evaluation.



**Figure 5 Modelling Evaluation**

**Figure 5**: In this part, after users has trained the algorithm to get the best algorithm. The system would apply the same pattern and algorithm to predict some data selected by users as input. It would give us an results which showed on the table. This is helpful for users to do data analytics.

**Discussion and Conclusion**

* Problem and solutions

During the process of making website, I meet the problem how to improve the speed of website delay, as the number of data is massive.

To handle this problem, database data was converted to csv format. This method improves the efficiency of website[9].

* Conclusion

In the data visualization part, The website was made through Dash which was a productive framework for building web application based on python language. As a basic interactive website, the interface should contain input and output widgets so that users could easily use the interface to input information and get the results from the web applications. The interface should contribute users to understand what is needed and match content with some interactive tools like calendar, sliders maps and buttons etc.

Users could manually choose any part of time series using website tools and choose a specific way like labelling, storing or comparing to handle data. And then, the website could automatically store the work for further study.

In the machine learning part, I combined machine learning knowledge with interactive website to analyse the data and understand the behaviour of the buses. We could explore the behaviour of the buses from the data of buses with the usage of supervised learning. For example, whether the buses run in the city or country could be considered as labels for the machine learning above. If the bus is running average 80% time spent on the countries and another 20% time spent on the cities. While the administrator find a bus only run 60% time spent time on the countries and 40% time on the cities.

To sum up, if the data is abnormal, the professional would be reminded. As a wrap up, we need evaluate the web interface and compare different algorithms to address the problem.

Consequently, the project is helpful for users to understand the behaviour of the vehicle and  detect abnormal behaviour of equipment in engine speed, engine temperature, etc.

**References:**

[1]Title: Machine Learning Trying to detect outliers or unusual behavior;  Author: Stacey, Ronaghan <https://medium.com/@srnghn/machine-learning-trying-to-detect-outliers-or-unusual-behavior-2d9f364334f9>

[2]PyCharm:

<https://en.wikipedia.org/wiki/PyCharm>

[3]plotly

[https://plot.ly](https://plot.ly/)

[4] scikit-learn

<https://en.wikipedia.org/wiki/Scikit-learn>

[5] Dash Introduction, Dash User Guide, What’s Dash?

[https://dash.plot.ly](https://dash.plot.ly/)

[6] scikit-learn library

<https://scikit-learn.org/stable/index.html>

[7]Machine Learning

[https://www.coursera.org](https://www.coursera.org/)

[8]O Svensson1, S Thelin2 , S Byttner and Y Fan Indirect Tire Monitoring System - Machine Learning Approach

[9]Sqlite 3 converted to csv file

<https://blog.csdn.net/pylduck/article/details/78250732>