Detection of fuel adulteration through Multivariate Analysis using Python Programming

Vimal Babu U[a], Nagamani M[b], Rama Krishna M[c], Anjaneya Prasad P[d], Tagore Finny P[d], Kumar Goud E[d], Vijay Cyril Y[d], Saidi Reddy M[d]

a,c(VFSTR University, Vadlamudi, Guntur, AP, INDIA)
b(UOH, Gachibowli, Hyderabad, Telangana. INDIA)
d(KG Reddy College of Engineering and Techonolgy, Hyderabad, Telangana, INDIA)
Corresponding Author; Vimal Babu

Abstract: The hydrocarbon fractions of the gasoline or Diesel can be detected at the in-situ with the application of the Internet of Things and can be accessed through the remote and data can be perceived through the smart phone. This will help in the finding the tailpipe exhaust, pollutants released into the air. The use of cheaper fuels in the transport field is causing more concern with respect to the health hazard. If the exhaust gas elements/composition is detected in advance, the environmental pollution can be contained to the limiting factors. This will help is limit the global warming so also the carcinogenic diseases, the health hazard in a long way. Multivariate Analysis (MVA) is one of the statistical principles of multivariate statistics, it refers to any statistical technique used to analyze data that arises from more than one variable, which involves observation and analysis of data, The main aim of MVA is to extract the data that is used to build a Model using classifiers with this exploring the data analysis by Automate Machine learning Algorithms, to explore the Machine learning Algorithms Python is used . Python language is one of the most flexible languages and can be used for various purposes. Python contains special libraries for machine learning namely scipy and numpy. In python, the libraries named pandas and matplotlib packages (for Data Analysis) are the libraries written for the Python programming language for data analysis and data manipulation. By taking the experimental dataset we are implementing the PCA (Principal component analysis) using Python code.

Keywords; Transport Fuels, Adulteration, Computational Techniques, Internet of Things, Multivariate Analysis, Principal component analysis, pandas, Machine learning.

Date of Submission: 05-09-2018   Date of acceptance: 21-09-2018

I.   Introduction:
The original fuel samples are randomly drawn from the petroleum retail outlets situated at Hyderabad city. The Hyderabad city is a concern for environmental pollution due to the increased population so also the use of automobile fuels. The original Gasoline sample is added with the definite cheaper hydrocarbon fuels (like Kerosene, Naphthalene, Mineral Turpentine Oil etc.) at periodical intervals and analyzed with the FTIR machine. The data collected is computed tithe the computational methods for the accuracy.

Fiber-optic sensors coupled with laser-induced fluorescence (LIF) spectroscopy and it is an expectant move towards the use of remote sensing of exhaust pollutants. The dissolved elements of the ocean water can be detected with the use of the smartphone on Beer-Lambert principle and other by based on evanescent field absorption from an unclouded U-bent sensing region of an optical fiber. The computational methods like PLS, partial least-squares regression method (PLS) is proved to be one of the best methods to forecast the ethanol content as well as the low cost, time reducing in evaluating the fuel composition. The neural networks agree to entirely classify the gasoline and light gas condensate fractions. The tailpipe emissions from the transport vehicles and the determination of a wide range of pattern recognition in petroleum-based fuels can be found from the vibrational spectrometry and chemometrics. Artificial neural network (ANN-MLP) approach based on the principal component analysis (PCA), which was believed to be efficient, has shown much inferior quality results. The electronic nose can be well developed to detect the transport fuel like gasoline and Diesel based on semiconductor sensors and the need to devise the portable kit with the e-Nose to test the fuel quality. Optical fiber sensing technology for Volatile Organic Compounds (VOC's) is more suitable than the mass spectrometry, gas chromatography.

The Datasets that contain more than one variable will be analyzed by some set of techniques in MVA and these techniques are especially valuable when compare with correlated variables and to understand the importance of experiments values and relationship among multivariate measurement and their structure we use

---

the conceptualized statistical study on each experiment. In data analysis, the computational multivariate methodology plays an important role because of the size and complexity of the underlying data sets. Many problems in the analysis of life science are multivariate in nature. The analysis of large multivariate data sets is a major challenge for life science research. The main important area of MVA is the statistical analysis of chemical data that is chemometrics and MVA has made easier to development of inexpensive, powerful analysis software and fast computers.

Materials and Experiments: The transport fuel of Gasoline is taken here in the experiment and doped with the adulterant fuel and the data is collected. This data is analyzed with the computational techniques.

Python libraries to implement Multivariate analysis

Numpy: It is one of the libraries for the python programming, adding plunk for a large collection of high-level mathematical functions to operate Multi-dimensional arrays and matrices.

Pandas: Pandas is a python Data Analysis Library for storing the data in the data frames and manipulate the data, and it also offers the data structures and operations for manipulating time series and numerical tables.

Matplotlib: This library is designed for the 2D plotting figures in python programming and its numerical extension numpy. And this library supports for the different 2 Dimensional plotting.

Sklearn (sci-kit-learn): It is one of the machine learning libraries in python and it features various clustering algorithms including support vector machines, classification, and regression, it is also exploited in k-maps, random forests and is designed to interoperate with python scientific and numerical.

Scipy: This library plunk for technical computing and scientific computing and it contains modules for integration, optimization, interpolation, linear algebra, image processing and special functions.

Seaborn: It is one of the data visualization libraries for python programming based on matplotlib. It mainly supports and provides informative statistical graphs and high-level interface attractive graphs.

Multivariate Data (MVD) analysis in python :

To analyze the multivariate data and plot the data into python For data analysis we will be using the Python Data Analysis Library which provides a number of useful functions for reading and analyzing the data, as well as a Data Frame storage structure, similar to that found in other popular data analytics languages, such as R. The data can be read in a pandas data frame using the read_csv()[5] function. The argument header=None tells the function that there is no header at the beginning of the file.

Before reading the dataset we have to import some libraries:

```
>>import numpy as np
>>import pandas as pd
>>import matplotlib as plt
>>>import sklearn
>>import seaborn as sns
```

Then we are Reading the dataset by the following code

```
>>df= pd.read_csv('data.csv')
```

Multivariate Data Plotting:

After reading a multivariate data set into python and next pace is to make a plotting the data, there are many ways of plotting a dataset we use some of the plotting techniques.

2D Data Analysis using Matrix Scatterplot:

Matrix scatterplot is one of the common ways of plotting multivariate data this plot shows each pair of variables plotted against each other, and to plot this we use the scatter_matrix()[5] function from the pandas library. We have to give some variables as input which we want to include in the plotting in the scatter_matrix() function. For example, that we just want to include the variables corresponding to the concentrations of the first six chemicals. These are stored in columns v1-v6 of the variable data. The parameter diagonal allows us to specify whether to plot a histogram ("hist") or a kernel Density Estimation("kide") for the variable. Given the dataset has to be converted into matrix form and to get the graphical view from the matrix we need to use python libraries i.e, matplotlib, seaborn as shown in the Fig(1.1) – Matrix Scatterplot for analysis of Multi-Variate Data in fuel adulteration detection
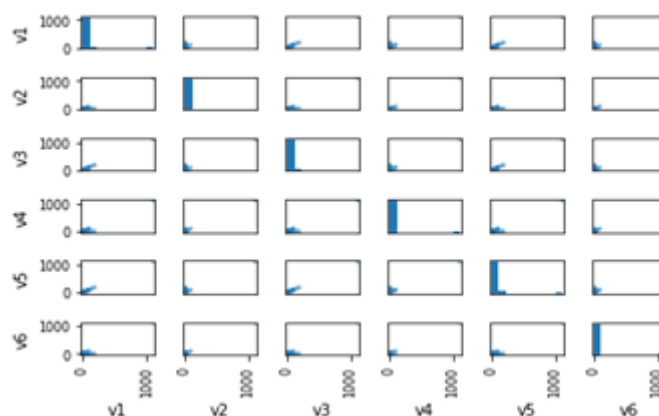
Fig(1.1) – Matrix Scatterplot for analysis of Multi-Variate Data in fuel adulteration detection

profile plot for multivariate Data:
Profile plot is the one of the plotting technique is used to shows the variations in each of the variables, by using this profile plot we can analyze the variations in the dataset of the each particularly selected variables as shown in the Fig(1.2) – Profile Plot for analysis of Multi-Variate Data in fuel adulteration detection
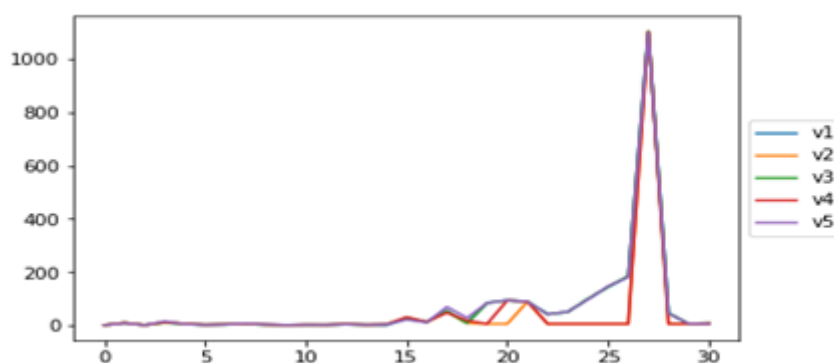

Fig (1.2) – Profile Plot for analysis of Multi Variate Data in fuel adulteration detection

From the above Fig (1.2) – Profile Plot[5] It is showing the mean and standard deviation for the selected variables from the dataset and the variables v5 has a lot higher of mean and standard deviation when compared to the other variables. To calculate the statistical summary for Multivariate Data that is mean and standard deviation for each of the variable in the multivariate dataset, this will be done using the mean() and std() functions in numpy and apply them to the data frame by using the member function apply. Pandas allow to do simple operations directly calling them as methods, for example, we could compute the means of a data frame 'df' by calling 'df.mean()'. An alternative option is to use the apply method of the pandas. DataFrame class, which applies the passed argument function along the input axis of the DataFrame. This method is powerful as it allows passing any function we want to be applied in our data.

Principal Component Analysis:
The main purpose of the Principal Component Analysis is to representation low-dimensional variable in the multivariate dataset. To perform the Principal Component Analysis on the experiment data (multivariate dataset) first we need to standardize the variables by using the function called scale(). After standardizing the variables do the Principal Component Analysis by using the PCA Class from the Sklearn.decompostion[5] package and fit method, which fits the model with the data X. The default solver is the Singular Value Decomposition ('SVD').
PCA is the one of the mathematics procedure that fits and transforms a number of correlated variables into a number of uncorrelated variables are called principal components. In the first principal component calculates for the variability in the data as possible, and next to each all succeeding component calculates for as much of the remaining variability as possible. This PCA is also similar to another multivariate procedure called Factor Analysis. They are often confused and many scientists do not understand the difference between the two

methods or what types of analysis they are each best suited as shown in Fig (1.3) – Principal Component Analysis
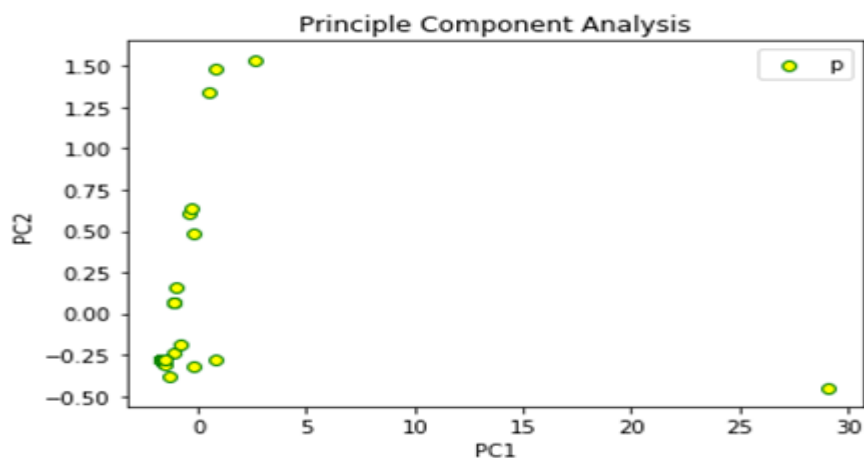

Fig (1.3) – Principal Component Analysis for analysis of Multi-Variate Data in fuel adulteration detection

Findings/Results and Discussion: The observed values from the FTIR machine, of different adulterer fuels samples, are analyzed with the computational techniques and yield good results in finding the composition of different parameters of the similar composition hydrocarbons doped samples.

Acknowledgment:

II. Conclusions

The exhaust coming from the transport automobile vehicles can be analyzed with the android application and can be monitored from the remote place. This will help the statutory authorities detect the vehicle causing the pollution. This will help in controlling the environmental pollution as well as the human health at large. In this, we implemented the principal component analysis using the python programming libraries and some statistical methods on the Empirical dataset (Experimental data) and we apply the all Data mining techniques To filter the Dataset. This Multivariate Data is used for the Big Data, IoT etc. To explore the Data Analysis.

References:

[1]. Iftak Hussain, Munmun Das, Kamal Uddin Ahmad, Pabitra Nath, Water salinity detection using a smartphone (2017), Sensors and Actuators B 239 (2017) 1042-1050, Sensors and Actuators B: Chemical
[2]. Helga G. Aleme, Let¡cia M. Costa, Paulo J.S. Barbeira, (2009), Determination of ethanol and specific gravity in gasoline by distillation curves and multivariate analysis, Talanta 78 (2009) 1422-1428,
[3]. Mohammadreza Khanmohammadi, Amir Bagheri Garmarudi, Miguel de la Guardia, (2012) Characterization of petroleum-based products by infrared spectroscopy and chemometrics, Trends in Analytical Chemistry, Vol. 35, 2012.
[4]. Helga G. Aleme, Let¡cia M. Costa, Paulo J.S. Barbeira, (2008) Fuel Determination of gasoline origin by distillation curves and multivariate analysis, Fuel 87 (2008) 3664-3668.
[5]. A Little Book of Python for Multivariate Analysis by Yiannis Gatsoulis is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. little-book-of-r-for-multivariate-analysis.readthedocs.org Based on a work at A Little Book of R for Multivariate Analysis by Avril Coghlan licensed under CC-BY-3.0.
[6]. N.K.L. Wiziack, A. Catini, M. Santonico, A. D'Amico, R. Paolesse, L.G. Paterno, F.J. Fonseca, C. Di Natale, (2009) A sensor array based on mass and capacitance transducers for the detection of adulterated gasoline, Sensors and Actuators B 140 (2009) 508-513
[7]. T. Babeva, A. Andreev, J. Grand, M. Vasileva, E. Karakoleva, B.S. Zafirova, B. Georgieva, J. Koprinarova, S. Mintova, (2017) Optical fiber-Ta2O5waveguide coupler covered with hydrophobic zeolite film for vapor sensing Sensors and Actuators B 248 (2017) 359-366.
[8]. S. Van der Walt, S.C Colbert, and G. Varoquaux. The NumPy array: A structure for efficient numerical computation. Computing in Science and Engineering, 11, 2011.
[9]. Vimal Babu. U, Ramakrishna. M, Nagamani. M., "Review on the detection of Adulteration in Fuel through Computational Technique"., Materials Today: Proceedings 4 (2017) 1723-1729.
[10]. Vimal Babu, Rama Krishna, Naga Mani(2016) " Air born pollutants toxins released from the tailpipe of the automobile in automobile fuel and its human hazards through computational Techniques"
[11]. Vimal Babu.Ua, Naga Mani.Mc, Rama Krishna.Mb, Tejaswini.Md(2017) "Data Preprocessing for Modelling the adulteration detection in Gasoline with BIS"., Materials Today: Proceedings 5 (2018) 4637-4645