v. 4, n. 4, 2024

International Journal of Human Sciences Research

USE OF THE KNN ALGORITHM TO PREDICT FUEL CONSUMPTION IN CARGO TRANSPORT VEHICLES USED IN MINING

João Jacob de Ávila Neto

Mario Sergio da Luz
UFTM

Abstract: The paper describes the use of the K-*Nearest Neighbor* (KNN) algorithm, applied in programming language, with the objective of predicting fuel consumption in cargo transport vehicles used in mining. For this development, the data available in a specific dispatch system for this type of application was used. Mining is an important sector of the Brazilian economy, as it operates by supplying raw materials to various areas of the industry. Among the operating costs of companies in the sector, an important portion goes to fuel consumption, this happens due to the constant need to move material, often over long distances. Tools that allow a prediction of fuel consumption value can contribute both to better planning and to the optimization of the vehicle allocation process. This fact becomes even more significant when most companies already have systems that provide the necessary information for this implementation, but they do not use them.
Keywords: KNN; Dispatch; Mining; Regression.

INTRODUCTION

The pursuit of a more detailed understanding of raw data comprises data analysis. It contemplates the use of several techniques, ranging from the already known statistical methods to modern solutions based on Artificial Intelligence (AI) and Machine Learning (ML) (DE AMO, 2004).

The adoption of modern tools to perform data analysis in order to assist management and decision-making in activities within industrial environments is gaining more and more strength. Specifically within the mining sector, the great complexity of its projects ends up making its process expensive, so that small improvements in the process end up bringing great benefits to companies in the sector.

In mining, an important portion of the expenses for maintaining the operation goes towards fueling equipment for transporting materials, this occurs since, in most cases, the area where the raw material is extracted and the processing area, where the ore begins to be transformed into a final product, they are separated by considerable distances, they have different paths that can be taken, in addition to having great dynamics on the ground, requiring a great deal of work in the logistics of this equipment. To help the allocation process, then, electronic dispatch systems are used, consisting of several solutions, between software and hardware, to help this process and, mainly, the production control. These systems make it possible to obtain a large amount of information, in addition to production variables, which are often not used by the company (ALMEIDA; DE CASTRO NEVES; DE FIGUEIREDO, [n.d.]).

Despite the characteristics of mining, the sector is strategic, having an important role for the country, as it is responsible for the production of raw materials for several other types of industries. In the Brazilian economy, according to the Brazilian Institute of Mining (IBRAM, 2019), in the Activity Report from July 2018 to June 2019, Brazilian mineral production represents 1.4% of all GDP Brazil, and according to IBGE employs around 195,000 workers directly. The importance of the sector is even more evident when one observes its relevance in Brazilian exports. In 2018, the country exported more than 409 million tons of mineral goods, and generated foreign exchange of US$ 29.9 billion, which represented 12.5% of Brazil's total exports, and 36.6% of the trade balance.

Based on the previous information, the objective of this work is to use the KNN (K-Nearest Neighbor) algorithm for a software application for predicting the fuel consumption of freight vehicles used in mining, using the input the data provided by the Electronic fleet management systems used in this type of application. Bearing in mind the importance of controlling fuel consumption in the mining sector, and the possibilities of using this type of application to make inferences, it becomes interesting that it can be used in various stages of vehicle allocation.

MINING AND DISPATCH SYSTEMS

The entire cycle of loading, transport and unloading of ore comprises the dispatch, and its management can be done manually, based only on the decision of a human being without the aid of any tool, but it can also be carried out with the help of programs and equipment specific to help the allocation of trucks, this type of solution is called electronic dispatch systems. The electronic solution is often used due to the complexity of decision-making regarding the allocation of vehicles, since in most cases the area where extraction takes place and the processing area are separated by considerable distances, there are different routes that can be used by vehicles, these being highly dynamic routes, undergoing daily changes, and finally, the regime used by many companies for uninterrupted mine exploration. (ALVARENGA, 1997)

C.A.N.

In order for the dispatch systems to work, several technologies are used to acquire information. One of them is to obtain data through the CAN network (Controller Area Network - CAN) of the vehicles, allowing the obtaining of data from various sensor values present in the vehicles, such as speed, acceleration and engine temperature.

CAN is a network specified by the *International Organization for Standardization* (ISO), originally developed by BOSCH and made available in the mid-1980s.

It is a multi-master synchronous serial message transmission system with a maximum transmission rate of 1 M bits per second (Mbps). Unlike other transmission protocols, the CAN Bus does not send large blocks of data, and it does not work with a master device under the supervision of the central bus. In a CAN network, messages are short and transmitted to the entire network (broadcast), which allows data consistency in each node of the system. (Instruments, 2016)

GPS

Another source of information for dispatch systems comes from the use of the global positioning system. (*global positioning system* - GPS) which, in addition to providing location (latitude and longitude), also provides data such as altitude and vehicle course. In electronic dispatch systems, the positioning of cargo vehicles within a mine is of great importance, as this type of information enables production to be monitored, in addition to adding more data to those already provided by CAN.

The GPS system provides data for free and continuously, in any position on earth, this happens passively, that is, the data is only received. The information is transmitted via radio signal, preferably a minimum of three different satellites to obtain a correct position. As soon as the data is received from the satellite, the receiving system makes the data generally available in the form of RS-232 serial communication (EL-RABBANY, 2002).

The most used standard in GPS communication is defined in the specification of the *National Marine Electronics Association* (NMEA). The idea of NMEA is the formation of blocks consisting of several characters, starting with $GP followed by the specific name of the frame, commas delimiting the different variables available. Below is an example of an RMC block (ASSOCIATION, 2002).

GPS provides a wide variety of blocks, specifically, for dispatch systems the main ones are GGA and RMC, some of the main data provided by them can be seen in Table 1. Despite the popularization of GPS, there are other positioning systems by satellites, also called constellations, are the GLONASS, BEIDU and Galileo constellations.

Block	Parameter	Position in block
GGA	Latitude	3
	Longitude	4
	Altitude	9
	Horizontal dilution (HDOP)	8
	Fixed date	6
	Number of satellites used	7
RMC	Speed	7
	North South	4
	East West	6
	Course	8
	Date	9

Table 1 – Parameters most used in dispatch systems.

Source: From the author (2019).

KNN

One of the algorithms, the KNN (K-*Nearest Neighbor*), emerged as a classification technique, basically working by classifying objects by the distance between the class and the unknown value, for this it uses as a basis examples passed to the program in the training phase. In addition to classification, the KNN algorithm can also be used for regression, that is, to identify a pattern within a database, and from this pattern a prediction can be obtained. In this algorithm, when used for prediction, this is done based on the distance in the closest values, the so-called nearest neighbors. ("Nearest Neighbors", [n.d.]).]

There are several different ways to calculate this distance. The simplest is the Euclidean distance:

$$d(x,y) = \sqrt{(x-y)^2} \quad (1)$$

Another measure also used is the Minkowsky distance:

$$dw(p,q) = (\sum_{n=1}^{n} |x-y|^{q})^{\frac{1}{q}} \quad (2)$$

By the library used in this work, scikit-learn, the regression can be implemented in two ways:

a) KNeighborsRegressor: An integer value, usually represented by k, which represents the number of nearest neighbors of each point used to obtain the point of interest in the regression ("Nearest Neighbors", [n.d.]);

b) RadiusNeighborsRegressor: The basis on neighbors is obtained from a fixed radius from the point of interest. ("Nearest Neighbors", [n.d.])

EXPERIMENTAL METHODOLOGY

Data were collected from a Mercedes Actros 4844 model vehicle. The sample was taken from 03/01/2018 08:42:10 to 03/06/2018 23:07:50. The data file was taken in CSV format in a total of 25805 lines, containing information from several sources, mainly CAN and GPS. The columns referring to date, time, RPM, speed, altitude and fuel information were then selected. The RPM variable represents the number of rotations per minute of the vehicle's engine. The accelerator corresponds to the percentage, from zero to one hundred, of the activation of this pedal by the driver. The fuel variable represents the value consumed by the vehicle, and is provided in the form of a floating point value, a number with one decimal place. This decimal place varies in values between 0 and 5, this change happens at each consumption of 0.5 liters of fuel by the vehicle.

Variable	Source
Date	GPS
Hour	GPS
Altitude	GPS
kmhour	GPS
RPM	CAN
Fuel	CAN
Speed	CAN

Table 2 – Data set used.

Source: From the author.

The implementation of the solution was developed in python language, along with the libraries and versions described below:

a) scipy: 1.2.1;
b) numpy: 1.15.2;
c) matplotlib: 3.0.0;
d) pandas: 0.23.4;
e) sklearn: 0.21.3;

RESULTS

The fuel variable, as it is obtained directly from the CAN protocol, does not allow its use for forecasting measures, since its value only increases during vehicle use. So, to forecast fuel consumption, a variable was created, here called rate. This new variable was created by grouping the different fuel consumption values, which counted the number of occurrences within each of these values, according to the example in table 3. As each variation in the fuel value occurs after the consumption of 0.5 liters of fuel, this value was divided by the total number in each grouping, thus making it possible to visualize where fuel consumption occurred more quickly. The value obtained was associated with data obtained directly from the vehicle, as can be seen in Figure 1.

Fuel	Number of clustered points
106312.5	7
106313.0	29
106313.5	9
106314.0	8

Table 3 - Number of occurrences at different fuel values.

Source: From the author (2020).

The variable rate represents the variations in fuel consumption, with this it is possible to verify under which conditions, values of other variables, can influence the increase in fuel consumption. Also in figure 1, it is possible to see that the ascent of the truck in altitude is related to increases in the value of the rate variable.

Using the variables altitude, RPM and speed as input, and the rate variable as output, an algorithm for prediction was implemented using the KNN technique. It was tested in two implementations, the first using KNeighborsRegressor, the parameters used are shown in figure 20 and the prediction result in figure 21. The determination of the cluster number was done via distortion graph, figure 22. In both implementations, for training, 25804 data were used, and in the forecast, the last 445. In addition, the Minkowsky distance was used in both implementations.

In the second implementation, using RadiusNeighborsRegressor, the parameters used are in figure 23 and the result in sequence, in figure 24. The determination of the radius (variable *radios*) was obtained experimentally, testing from number 7 to number 1, the value that presented the the best result was with the value 2. The second implementation presented the best results, visually, when compared to the first.

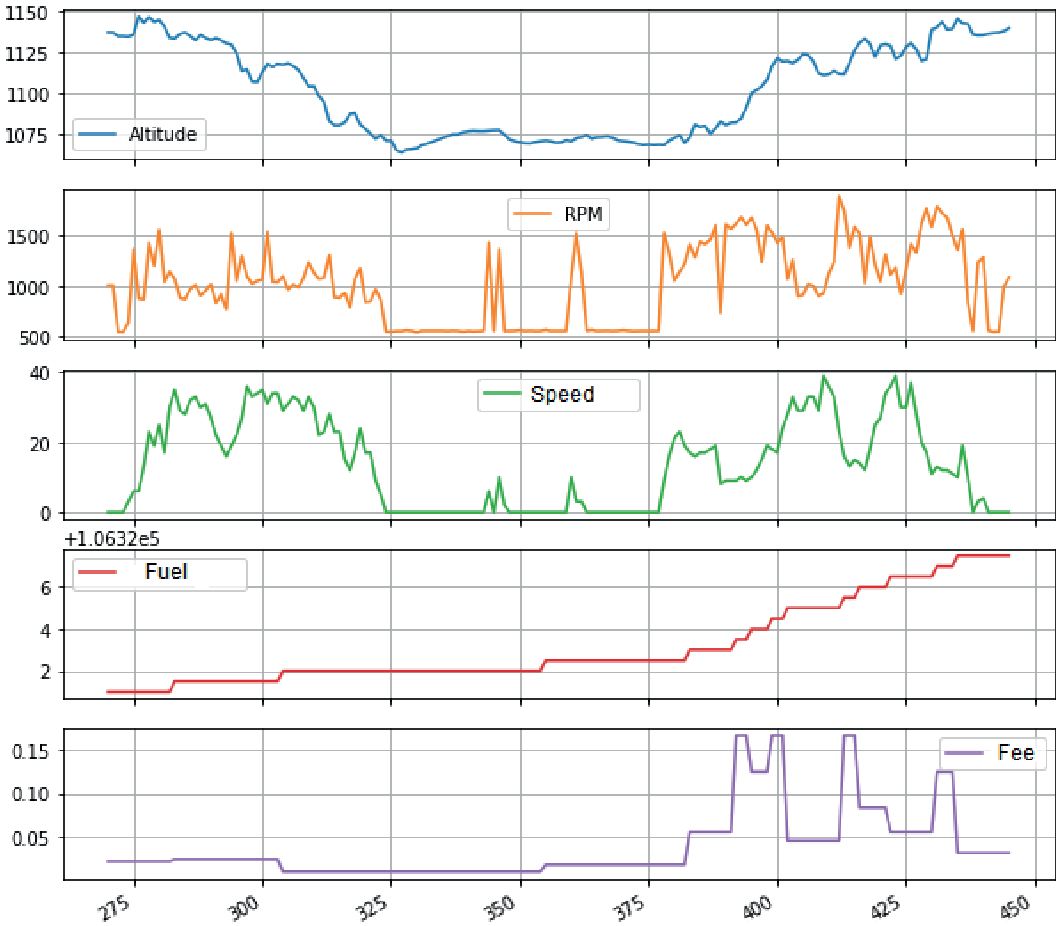

Figure 1 - Behavior of variables within a cargo vehicle trip. Source: From the author (2020).

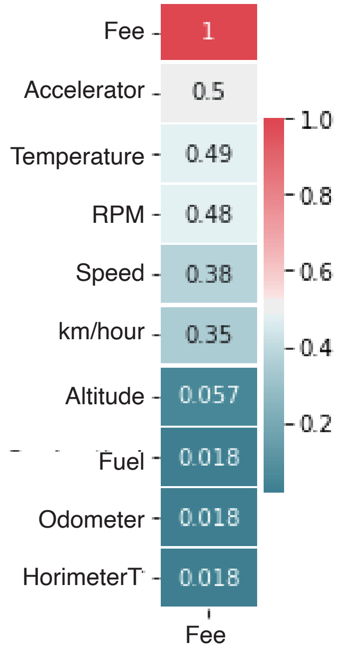

Figure 2 – Correlation (Spearman) of the rate variable.

Source: From the author (2020).

```
KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
                    metric_params=None, n_jobs=None, n_neighbors=Ellipsis, p=2,
                    weights='uniform')
```

Figure 3 - Paraments used in KNN prediction. Source: From the author (2020)

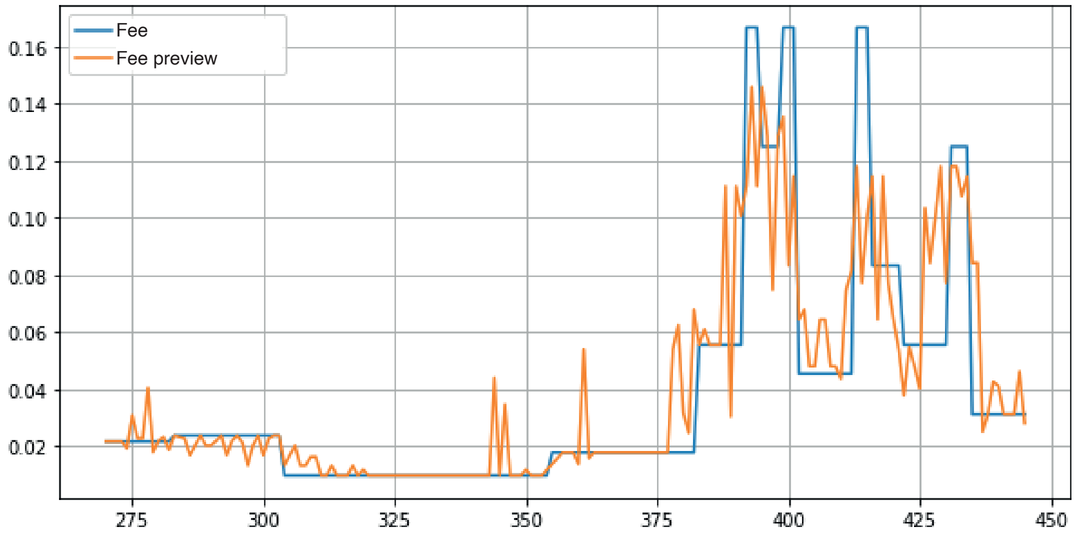

Figure 4 - Prediction of the knn rate variable (k=5). Source: From the author (2020)

```
RadiusNeighborsRegressor(algorithm='auto', leaf_size=30, metric='minkowski',
                         metric_params=None, n_jobs=None, p=2, radius=Ellipsis,
                         weights='uniform')
```

Figure 5 - Parameters used in KNN prediction (Radios=1.0). Source: From the author (2020).

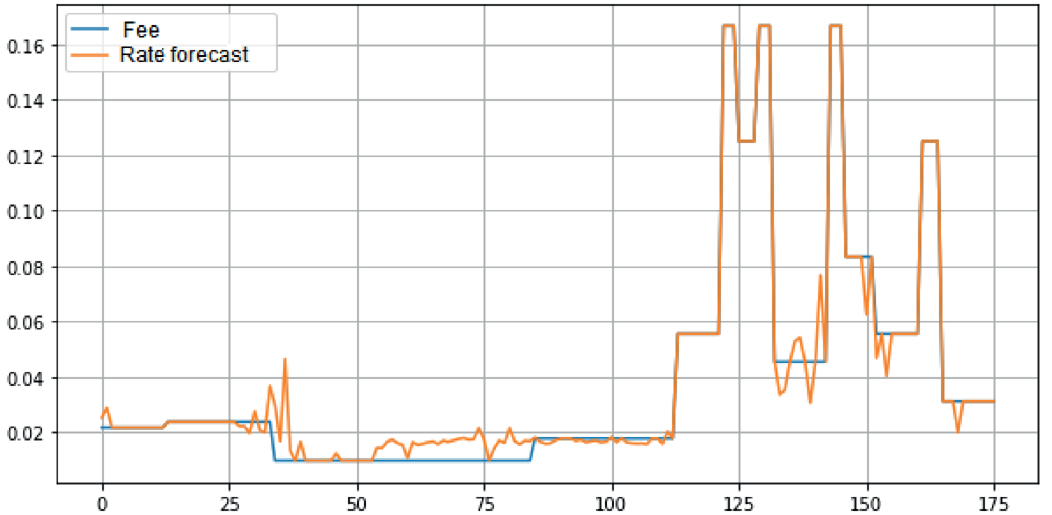

Figure 6 - Prediction of the knn rate variable (radios=2). Source: From the author (2020).

CONCLUSION

This article presented a study on the KNN technique for prediction, the objective was to analyze data from dispatch systems used in mining in order to analyze the possibility of this technique in predicting the fuel consumption of freight vehicles. An important part of the work was to create a way of analyzing the fuel, since the original format provided by the CAN network does not allow for the analysis of its variation over time. Once the way in which the fuel is analyzed was established, the algorithm was then developed to predict consumption using the variables available through the system as input. behavior of the relationships of different variables within this process. It is believed that the results obtained in this work open up new possibilities for further studies related to the area.

REFERENCES

ALMEIDA, C. M.; DE CASTRO NEVES, T.; DE FIGUEIREDO, A. C. L. Estudo de viabilidade econômica do uso de correias transportadoras nas operações de lavra de uma mina de minério de ferro. [s.d.].

ALVARENGA, G. B. Despacho ótimo de caminhões numa mineração de ferro utilizando algoritmo genético com processamento paralelo. Belo Horizonte, MG, Brasil: Master‘s thesis, 1997.

ASSOCIATION, N. M. E. NMEA 0183--Standard for interfacing marine electronic devices. [s.l.] NMEA, 2002.

DE ALMEIDA GUIMARÃES, A. O Protocolo CAN Bus nas Aplicações Off-Road: Uma Análise Comparativa entre os Padrões Existentes. 2001.

DE AMO, S. Técnicas de mineração de dados. Jornada de Atualizaçao em Informatica, 2004.

EL-RABBANY, A. Introduction to GPS: the global positioning system. [s.l.] Artech house, 2002.

INSTRUMENTS, T. Introduction to the controller area network (CAN). SLOA101A. Dallas, TX, 2016.

Nearest Neighbors. Disponível em: . Acesso em: 3 jul. 2020.