# Prototype model for big data predictive analysis in logistics area with Apache Kudu

**Liliya MILEVA[1], Pavel PETROV[2], Plamen YANKOV[3], Julian VASILEV[4], Stefka PETROVA[5]**

[1] University of Economics, Varna, Bulgaria
l.mileva@ue-varna.bg

[2] University of Economics, Varna, Bulgaria
petrov@ue-varna.bg

[3] University of Economics, Varna, Bulgaria
yankov.plamen@ue-varna.bg

[4] University of Economics, Varna, Bulgaria
vasilev@ue-varna.bg

[5] University of Economics, Varna, Bulgaria
s.petrova@ue-varna.bg

**Abstract.** Logistics is area, which is evolving rapidly, generating a lot of data lately. There are several problems that, even when partially digitized, require additional work. At the same time there is a need of big data analyses. These analyses are represented by machine learning and statistical analyses. One important problem is these with missing data for delivery in the whole process of transportation, other is about empty freight transportation. The purpose of this paper is to present an ICT prototype model with analytical tool Apache Kudu in logistics area, which will contribute to resolve such problems in area. Content analysis and systematic approach are used. Statistical methods and statistical data are used. Growth rates of empty road transport are calculated. The presented data are for Bulgaria and its neighbour’s county, excluding Romania, because of the missing data. Croatia is also included in the analysis. Growth rate analysis indicates problems with empty fright transport in some countries. The tendency for Bulgaria is to reduce empty road freight transport, even though it is better for logistics organizations to work for optimizing the process of delivery.

**Key words:** logistics, machine learning, big data, big data analyses, predictive analysis, Apache Kudu, Apache Hadoop, logistics software, growth rates;

## 1. Introduction and literature review

**Acknowledgments**. This work is supported by the "Digitization of the Economy in an Environment of Big Data" Project - BG05M2OP001-1.002-0002.

Organizations worldwide, use a variety of methods, models and techniques for data analysis (Calzon, 2021). They use them to create and restructure their business models as well as to avoid certain problem areas in their storage, incompatibility due to diversity, visualization problems, lack of tracking and analysis in real time, which in turn lead to large financial losses and other issues.

At present, more and more big data and their problems are entering science (Kuyumdzhiev, 2020). Previous years the scope of the term was the focus of many articles. Methods for storing and organizing big data are actually developing at present. Sometimes there is a line between traditional datasets and modern big data. A comparatively big dataset may become a similarly small “big data” (Todoranova and Penchev, 2020). The border between big datasets, data warehouse and big data are obvious (Todoranova *et al.*, 2020) (Nacheva *et al.*, 2019).

In business, processes generate amounts of data, which are normally stored in appropriate databases (Nacheva *et al.*, 2019), but even though sometimes databases could grow larger and larger within duration of several minutes (Nacheva and Sulova, 2020, 2021). These are specific situations, in which companies are migrating from databases to “big data”. Measuring efficiency and hardware storage (Polkowski *et al.*, 2020) uses new opportunities to continuously increase the amount of data. The increased amount of data opens new perspectives for hardware storage (Polkowski *et al.*, 2020) and measuring the disk storage performance (Cristescu, 2019) (Stoyanova, 2020). Digitizing more and more business processes means that data are increasing (Miryanov and Petkov, 2017) (Nikolaev, Milkova and Miryanov, 2018). More energy is needed for their storage and cooling. Sometimes data centres are used. The question concerning the choice “own data centre” or “using a 3PL” is open.

Storing large amount of data needs the application of mathematical methods for creating forecasts with big data (Miryanov and Petkov, 2017) Retrieving data has many approaches (Medvedev and Sergeev, 2020). Creating high reliability forecasts (Miryanov and Petkov, 2017) means accepting, testing and validating all assumptions for the chosen mathematical models (Ana-Maria Ramona, Marian Pompiliu and Stoyanova, 2020) (Polkowski *et al.*, 2020). Sometimes for analyzing big data are used heuristic approaches (Ileanu et al., 2019) (Abdel-Badeeh M. Salem, 2018). Providing web access to big data is another challenge with multiple technological solutions.

At the same time, logistics, which is evolving rapidly and generating a lot of data, needs further digitalization of many business processes. However, there are a number of problems that, even when partially digitized, require additional work. Many of them stem from a lack of control (Kirova, 2009) and other issues. Many of the existing problems are typical for the combined transport or in the transport of goods from point A to point B, when the logistics is performed between different countries and by more than one logistics organization. Some of the problems start from poor or lack of communication between companies, as well as reluctance or inability to cooperate. Some of them could be solved with the help of Apache Kudu.

## 2. Creating methods for analysis of different types of data with the analytical tool Apache Kudu

Big data are evolving as a product that is growing rapidly, but at the same time make a connection between many businesses and other economic units. In this way, new forms of data extraction are emerging. Traditional data sets are also interpreted through the most widely used big data analysis. For the purposes of the study, they are considered articles that describe the methods of big data analysis. In the scientific literature, statistical analysis are known as well as analysis of social media (Baesens, 2019) stating that in the future many benefits are expected from their analysis. By social networks, it is meant not only social media sites such as Facebook, Twitter and LinkedIn, but also any network between customers or companies connected in a certain way (Baesens, 2017). It is made an overview of the considered methods of analysis based on the division into quantitative and qualitative (Fig. 1). Such a separation is present in each of the areas of the economy that involves analysis. The main characteristics of the methods shown in the figure 1 are quantitative analysis with a focus on a large part of the object of study and qualitative analysis - when the data are qualitative, non-numerical specialists turn to qualitative analysis by assigning (coding) numbers. The focus of this analysis is on a smaller part of the studied object.

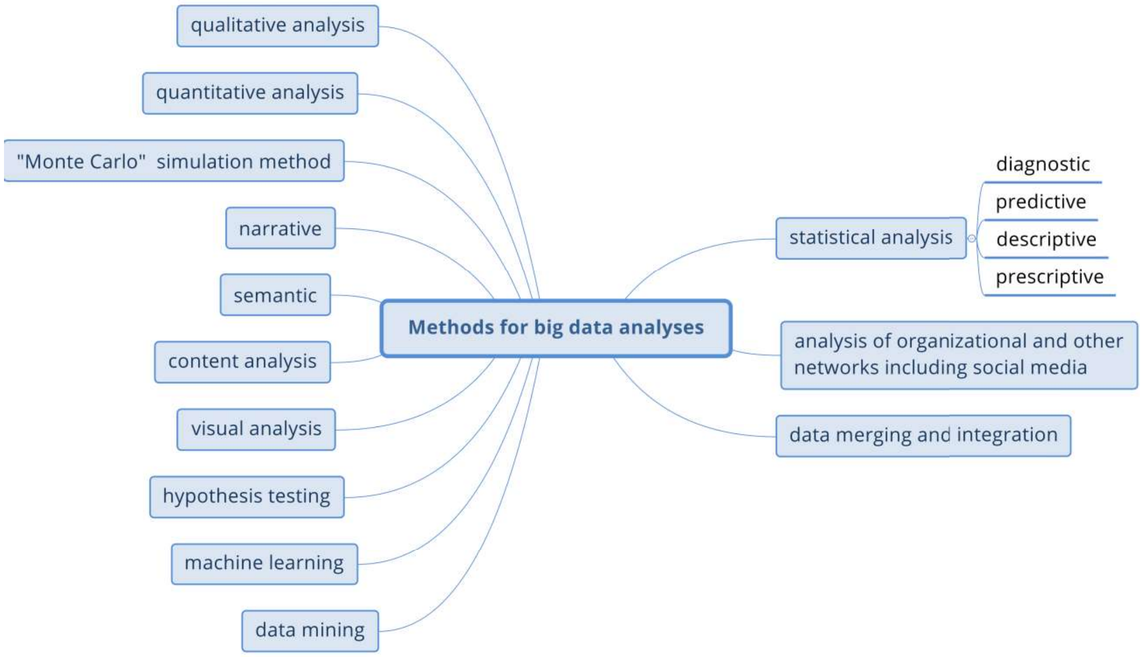


**Figure 1** Methods for big data analysis

Source: The figure is based on a Dr. L. Mileva research on the existing methods in the scientific literature for big data analysis.

The paper presents the possible statistical analyses, which are of several types. 1) Descriptive which analyse the initial data. 2) Diagnostic analysis- establishes the reasons that have caused the changes in the studied phenomenon over a period. It mainly models the development trend and calculates indices of seasonal fluctuations. 3) Predictive analysis- by extrapolation of previously established development trends, different forecasts are developed. The analysis works by developing hypotheses about the expected future behavior of the development of an organization or part of it (T-testing). 4) Prescriptive analysis - a combination of descriptive and predictive, often preferred method, but it also has serious drawbacks, namely: although it is associated with, finding the best course of action for a situation usually requires a lot of labour and large financial costs. 5) Analysis of the connections between the phenomena - regression and correlation analysis, in which it is examined whether there is a connection between certain phenomena, what is the strength of the connection, etc. For a more comprehensive study, specialists focus on performing a combination of statistical or other analysis, because it will get more objectivity and higher quality of the desired result.

Statistical analyses are followed by other types that are widely used in theory and practice: 1) Machine learning, which works with computer algorithms creating assumptions about the studied data in the database. It makes predictions that cannot be made by humans because the amounts of information are too large. This makes machine learning one of the favourites for analysing large data sets. 2) Semantic analysis- collects the semantic information needed for the study from the source code. It is specific and used by information technology specialists. 3) Visual analysis - decision making based on imaged data. It is usually done, so that data users can get a better idea of the prospects for future work, such as rising and falling costs, sales revenue and more. 4) Narrative analysis, which implementation carries out through a story is exploring ideas, attitudes, opinions and stories, reveal the main preferences of the employees of an organization. 5) Content analysis, analysis of texts and performed by extracting data from the content of various formats such as video, text or other extracted from the press, social networks and others. 6) Data mining is similar to content analysis, but with the difference, that it is a search and discovery of useful and valuable information contained in large databases or data warehouses. It combines methods from statistics and machine learning, within the database management. 7) Monte Carlo - a series of computational algorithms that rely on the repetition of random phenomena to achieve numerical results, used for analysis mostly by organizations that want to minimize various risks. 8) Data merging and integration - by combining a set of techniques for integrating and analysing data from multiple sources, better and more accurate predictions are made. The presented list of studied analysis is not exhaustive, as researchers are constantly looking for new ways of analysis and working on new models or combinations of them. The decision usage specific method of analysis depends on the field of work, specific characteristics and purpose of the results that specialists want to achieve. A combination of methods should achieve greater objectivity of the survey, higher quality of the results sought and higher added value for users of the results of surveys conducted using methods of big data analysis.

## 3. Big data analysis using machine learning in the Hadoop ecosystem

Data analysis is a process that relies on methods and techniques to extract information from raw data that are important for achieving business goals. They are represented by statistical methods and machine learning methods. Although statistical methods are derived as a separate branch, they are also contained as part of machine learning. In practice, both types of methods are used in almost all areas of economy and are therefore of great importance and can be used for different purposes, such as comparing different data, end results of different activities, comparing values, studying the presence or absence of mutual connections between individual phenomena. The need for in-depth consideration of machine learning models is due to its growing importance and use in the analysis of large data sets. Machine learning is becoming more popular due to the growing volume of data, which is so diverse that it requires different methods of processing and storage. In addition, we are looking for technology with lower costs and greater processing and storage capacity, something that machine learning provides as opportunities. There are also opportunities for making models that are characterized by the speed of their creation and automation options, while they have the ability to analyse larger, more complex data and accordingly to provide faster, more accurate results- even on a very large scale. Through the usage of machine learning, organizations have the opportunity to build models that help to make optimal decisions and avoid risks. One branch of data analysis is machine learning (ML), which is divided into three parts. These are: classification, clustering and evolutionary algorithms (Raeesi Vanani and Majidian, 2020). After analysing the most popular definitions of Machine learning in science, they are extracted the main features of machine learning (Fig. 2).

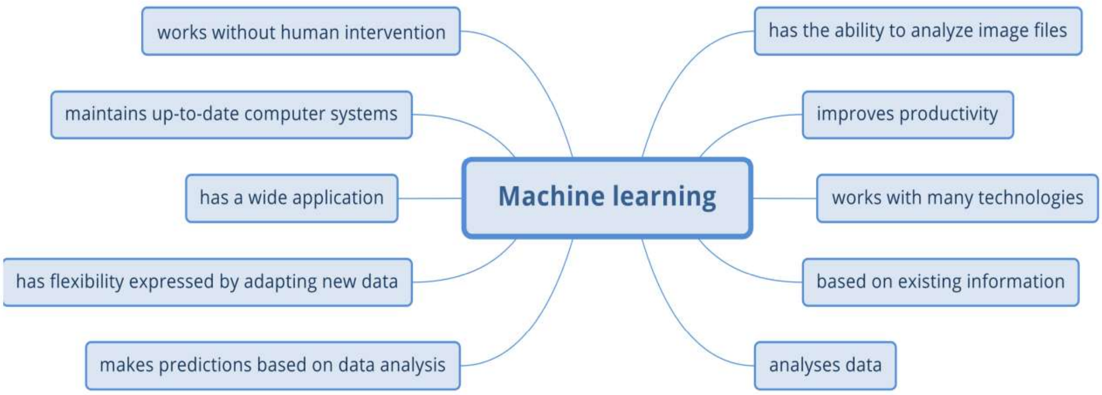


**Figure 2** Characteristics of machine learning

Source: The figure is based on a Dr. L. Mileva research on the existing methods in the scientific literature for big data analysis.

### 3.1 General application of machine learning

Machine learning occupies a central place in the analysis of big data and is widely used in many areas of the economics and its most common usage is in the collection and analysis of data from video surveillance and information collection from social networks, but this is far from use up its application in practice. Machine learning is used to:

- image recognition (most often) video surveillance, customs clearance and airport security check; - in social networks at:
- friends suggestion;
- common acquaintances with other people;
- image recognition and tagging (image tagging on photos);
- markings of location, site, etc.;

- vehicles with a certain autonomy, such as TESLA cars and Google cars;
- personal assistants- Siri, Alexa, etc.- virtual assistants who help mobile phone users to search for objects on the Internet, save data, make schedules, mark and remind important events and more. Real-time translations have recently become especially popular, when the user does not know a certain foreign language and uses the help of a personal virtual assistant;
- analyses of consumer opinion, through surveys- for example, made after buying a product or using a service, after visiting a site, etc.
- whether they are satisfied with the purchase, the service, whether they would use it again, at what price, whether it costs them is that it is too expensive, whether they think there is a need for certain changes and what;
- forecasts based on consumer behavior, for example, a preliminary survey of consumers' opinions on whether they would invest in a new service, previously unknown or product, thus the organization will gain information on whether it is worth investing in a particular product or service or not, should it be soon consider introducing a product and/or service of another kind in order to overcome unforeseen losses and risks;
- analyses of consumer behavior, by collecting input data (usually include information on the number of purchases per month, amount spent, preference for a particular model of shoes, for example - sporty, elegant, etc.);
- real-time traffic forecasts - where passengers will avoid traffic jams and save travel time;
- fraud detection with bank cards and transactions;
- filtering spam and malware messages in e-mail;
- recommendations for visiting tourist sites such as accommodation and meals;
- recommendations for visiting historical and cultural landmarks;
- data collection for project implementation- collection of data for tracking the employees’ behavior within the working day- tracks the time of arrival and departure from the workplace, compliance or non-compliance with the time for breaks, even tracking the activity within the working day;
- collection of data for tracking the behavior of employees within a project- for example, movement of employees, performance of tasks, reporting results, etc.
- video surveillance of residential and construction sites- for security and protection of the site;
- tracking of logistic processes- for tracking of personal and company shipments, logistic processes in sites or from point A to point B during their movement, which helps for undamaged and timely deliveries in real time;

For purposes of the present article, they are derived the applications of Machine learning in logistics area.

Application of machine learning in logistics:

- tracking of logistics processes - for tracking of personal and company shipments, logistics processes in sites or from point A to point B, during their movement, which helps for undamaged and timely deliveries in real time;
- working through a supply chain;
- collecting data to track employee behavior within a project- for example, employee movement, task performance, reporting results, etc. The Hadoop ecosystem handles very large data sets, which requires various tools to retrieve, process, store and distribute them. It also performs multiple analyses using machine learning. In its development, the known basic species complement, modify and compose new subspecies. Each type of machine learning could apply to a different area, goal or task. It is characteristic of Machine learning that depending on the purpose of the work there are opportunities to mix two or more types of methods and techniques. In FIG. 3 and 4 are presented the types of ML, the variables with which they work in the analysis of big data, methods of work, as well as the tools for working in Hadoop.

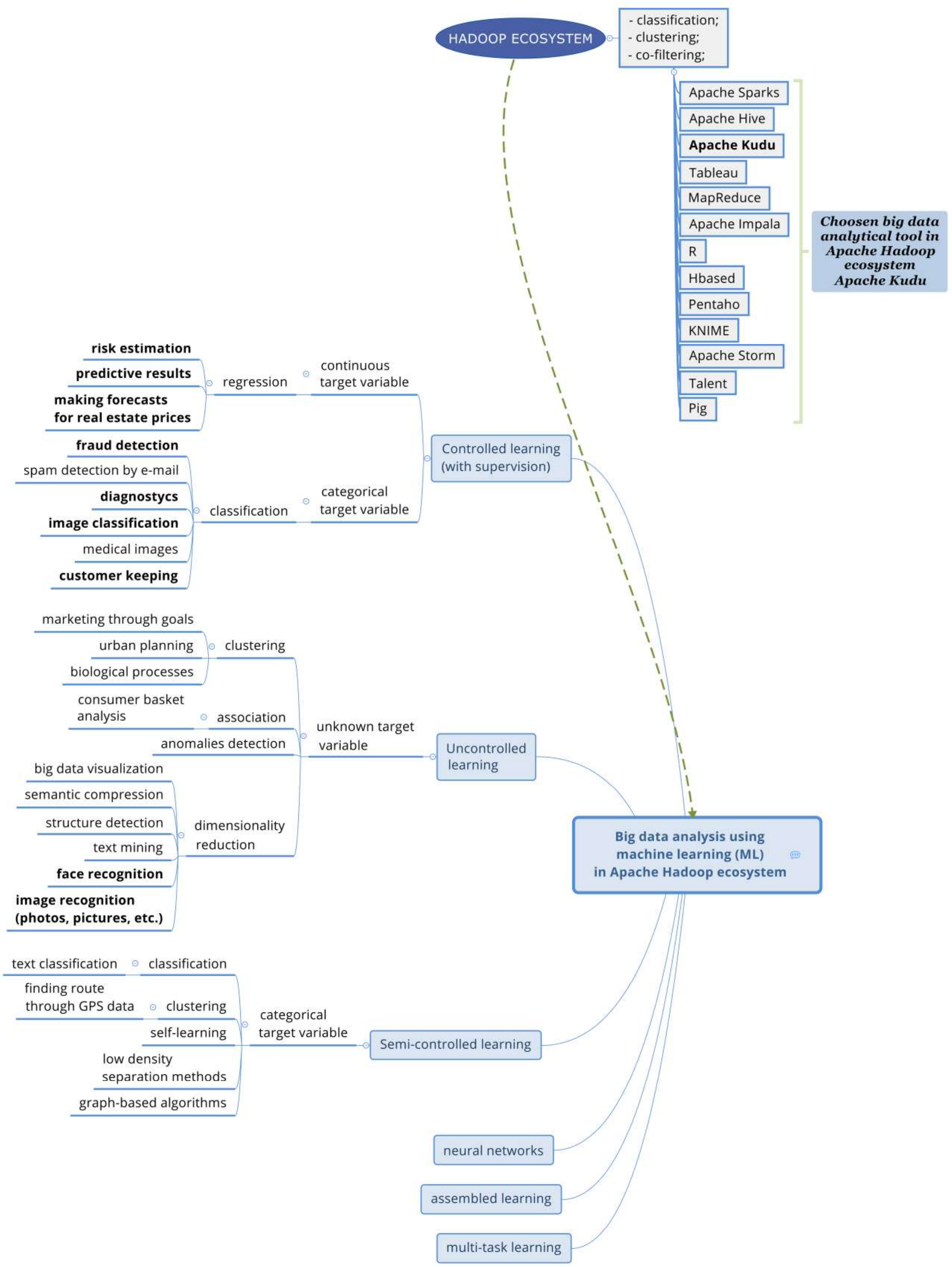


**Figure 3** Big Data Tools

Source: The figure is based on a Dr. L. Mileva's research on existing in the scientific literature methods for big data analysis.

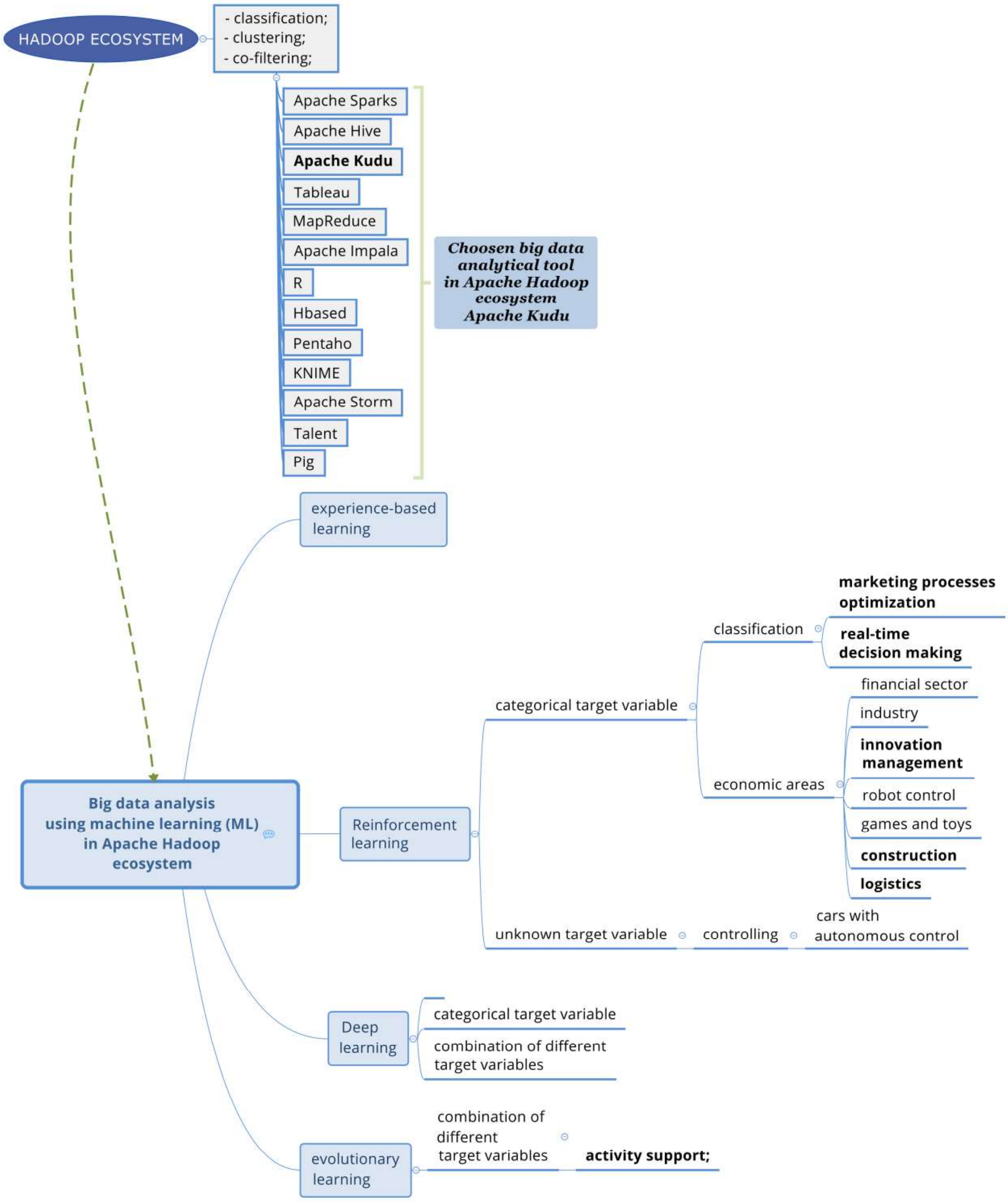


**Figure 4** Big Data Tools
Source: The figure is based on a Dr. L. Mileva's research on existing in the scientific literature methods for big data analysis.

In Fig. 3 and 4 (Heidenreich, 2018) it is also present the tools for working with big data in Hadoop, through the three principles of operation: clustering, classification and co-filtering. Unlike data analysis methods, which aim are the data itself, their aggregation into data clusters and classification, the principles of the tools are based on aggregation, classification and filtering, but no longer on the data itself, but on the computer systems working with them. With Hadoop's big data analysis tools, researchers have the opportunity to design models and use specific methods to make various business decisions not only in the fields of logistics, but also in almost all other economic fields including construction. The analytical tools are presented in fig. 1 and 2:

1. Apache Sparks.
2. Apache Hive.
3. Tableau.
4. Map Reduce.
5. Apache Impala.
6. R.
7. Hbased.
8. Pentaho.
9. KNIME.
10. Apache Storm.
11. Talent.
12. Pig.

This list does not limit known analytical tools, but it is the most applicable and used in practice. Among this tools is the Apache Kudu (*Apache Kudu - Overview*, 2017) , which is characterized by the most advantages among the others and was selected for analysis in Hadoop ecosystem. Apache Kudu is a free, open source tool focused on storing large amounts of data in the Hadoop ecosystem. Its main advantage is that it is compatible with most data processing frameworks in Hadoop. It is also characterized by the speed of data transfer during storage and transfer. The advantages of Apache Kudu are base for creating a model for logistics work in making specific decisions.

## 3.2 Application of the Apache Kudu analytical tool in the field of logistics

Apache Kudu is a column storage manager developed for the Hadoop platform. Kudu shares the general technical features of Hadoop's ecosystem applications, running on stock hardware, being horizontally scalable, and supporting highly accessible operations.

**Advantages**. Kudu is characterized by fast load processing, full integration with Apache Impala, as well as high performance, so it is one of the most preferred analysis tools in Hadoop. The advantages of Kudu include:

- fast processing of OLAP loads (online analytical processing);
- Developed to integrate with Map Reduce, Spark, Flume and other components of the Hadoop ecosystem.
- Full integration with Apache Impala, making it a good, changeable alternative to using HDFS with Apache Parquet.
- strictly defined but nevertheless flexible sequence model that allows on-demand compliance requirements to be selected, including a strict sequence option.
- high productivity at simultaneous execution of consecutive and arbitrary loadings;
- easy to administer and manage via Cloudera Manager;
- high availability. Tablet servers and Masters use the Raft consensus algorithm, which provides availability as long as more replicas are available than not. Readings can be serviced by consecutive read-only tablets, even in the event of a leading tablet failure. Kudu works under certain limitations, and professionals should comply with them.

Apache Kudu could be applied properly in the field of logistics because of its ability to use appropriate solutions (Quinto, 2018). Possible logistics solutions include:

- reporting applications where the new data must be immediately available to end-users; - timeline applications that must support requests for large amounts of historical data while returning detailed requests for an individual site;
- applications that use real-time predictive decision-making models, with periodic updates of the predictive model based on all historical data; These applications are typical in the field of logistics, where it is necessary to track the movement of goods and goods from one point to another to make it possible throughout the process to track goods continuously in real time. For the purposes of the development, they have been identified some well-known problems in logistics, which would be solved with the help of analytical tool Apache Kudu. The advantages of Apache Kudu will be the base for building a model for work in construction and logistics in making specific decisions. Logistics is an area in which digital transformation is present largely.

There are a number of problems that, even when partially digitized, which despite the efforts require additional work. Many of the problems came from a lack of control (Kirova, 2009) lack of flexibility in work, and some are caused by human error. Many of the existing discrepancies are typical for the combined transport (Parvanov and Boeva, 2009) or in the transport of goods from point A to point B, when the logistics is performed between different countries and by more than one logistics organization. Some of the problems came from poor or lack of communication between companies, as well as reluctance or inability to cooperate (sometimes due to lack of sufficient financial and human resources and information, as well as poor management). Some of them could be solved with help of Apache Kudu.

### 3.2.1 Some logistics problems

Some known most common issues are listed below:

- fraud related to theft of customer data;
- fraud involving theft of goods, cargo, fuel (sometimes theft is committed by outsiders, but it is also common for staff members to abuse it to profit from the sale of goods to third parties at higher prices and others.);
- problems with insurance- when there are many damaged and broken shipments, their insurance covers the value, but when there are many such events, the insurer begins to lose from them instead of winning. From there are raised the contributions because the organization becomes a risk client and in turn raises the fees for the clients in order to close the difference for the insurances. Customers become dissatisfied and start using other companies that have lower fees and better service, in which the organization loses market share. Customers are dissatisfied, as their shipments cannot arrive in time. If they are from a store, and if they are their property, they may not recover them at all (in case they are single items). For example, if a picture replicates second time it loses its true value.
- poor infrastructure of some countries on the European road network - incl. Bulgaria, Romania;
  - out-dated fleet of some organizations- leads to delays, technical problems of vehicles, the need for frequent technical maintenance, respectively, and specialists to perform it; - high unemployment;
- unqualified staff; On the one hand, employees complain about lack of work, poor working conditions, and on the other hand, employers complain about the lack of qualified staff (e.g. drivers-suppliers).
- problems with cargo storage;
- inability of the end customer to monitor the movement of his shipment;
- difficult, even impossible traceability of consignments from item A to item B.,
- due to:
- lack of sufficient information - blank data,
- lack of historical data for customers;
- lack of data caused by errors when entering them during registration (as incomplete: address, telephone number, names, etc.), which complicates the process of sending the shipment and requires additional work (call to contact the customer);
- problems with transliteration of the client's names when using more than one alphabet (for example Latin and Cyrillic);
- passing through hard-to-reach terrains;
- crossing limited routes (in connection with hostilities, disasters, accidents, crisis situations (such as the Covid-19 pandemic, the attacks in Turkey, etc.);
- passing through dangerous areas (e.g. areas inhabited by people from problem minority groups);
- remote points along the route;
- Lack of communication IT infrastructure, including scope throughout the supply chain;
  - lack of control;
  - lack of flexibility;
  - unreceived shipments;
  - unsolicited shipments;

The listed problems also require possible solutions to improve the transport and logistics of goods. This can be done by digitizing the processes in logistics organizations.

Suggestions for solutions are:

- better control;
- periodic release of control consignments that require special attention (easily damaged, fragile);
- better flexibility;
- traceability of shipments, cargo, goods in real time;
- real-time vehicle traceability;

- traceability of the personnel performing the transport activities (when the personnel is part of the human resources of the company for the determined period of time);
- predictability of various factors that could cause delays- for example, lack of storage for cargo, delays in transport, staff errors;
- avoiding busy and dangerous routes with the help of a notification system;
- finding unmarked GPS points along the route;
- tracking of consignments, when transported by two or more logistics companies, when the goods are passed from point A to point B, passing through different countries; Highlights:
- tracking staff to avoid fraud;
- cargo tracking;
- better control over shipments;
- better control of staff;
- greater flexibility;

For purposes of the article, was chosen a specific problem, which is largely present in the work of logistics organizations and the solution of which will lead to great benefits for future analysis of big data from the point of view not only of logistics companies but also customers. The main problem observed in the field of logistics is the impossibility of monitoring the transport of goods from more than one logistics organization as well as the transport from one country to another or passing through several foreign territories. Information about it is often lost when it is delivered, which makes it impossible for both the internal system and the end user to track it. This leads to other problems such as:

- inability of the customer to monitor the movement of his shipment;
- inability of the customer to receive notification of his shipment on time;

The inability to receive the shipment on time by the customer leads to:

- delay in receipt;
- charging additional fees for staying in a warehouse for more than three days;
- a fine for the office worker or courier who has failed to deliver the goods within three or more days (depending on the policy of the logistics organization); All this affects both the quality of service and customer satisfaction. It can lead to customer withdrawal, financial losses, and even the organization dropping out of the market. With regard to big data analysis, the current problem leads also to:
- inability to trace;
- loss of information when passing from one courier to another, when passing from one country to another;
- inability to perform analyses with regard to: - tracking the direction of goods and cargo;
- analyses based on time travelled;
- analyses of the movement of goods and cargoes from one country to another;

It can be concluded that all these missed opportunities stop working and make analysis impossible, through big data for business and other purposes. This disrupts work, slows down plans to attract new customers, expand the business and even vice versa can even lead to a contraction of business or dropping out of the market due to losses (Fig. 5).

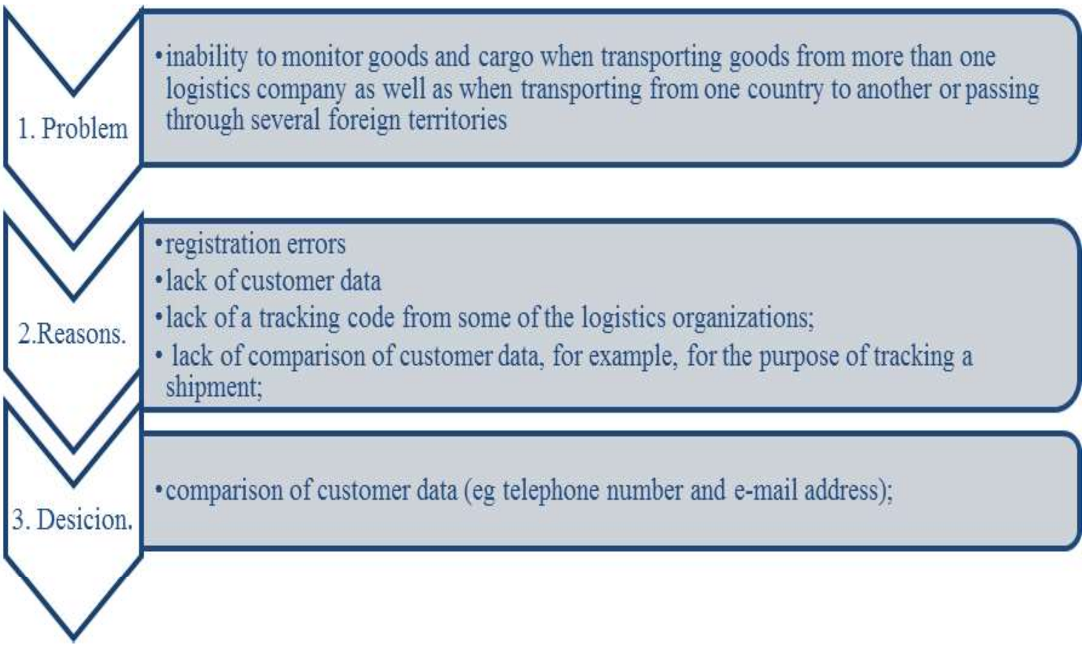


**Figure 5** Decision making process in the field of logistics
Source: based on authors’ research

Solving this problem will achieve many benefits and opportunities in big data analysis using the Apache Kudu tool. With its help, the relevant data first can be stored and then analysed. For this purpose, it is possible to use part of the logistics information system to solve the problem. The whole process of transferring items from one point to another goes through certain stages. The process includes not only the actual physical transport, but also the acceptance of a customer request, data processing, stay in a distribution centre (hub) and delivery. The stages can be systematized as follows (Fig. 6).

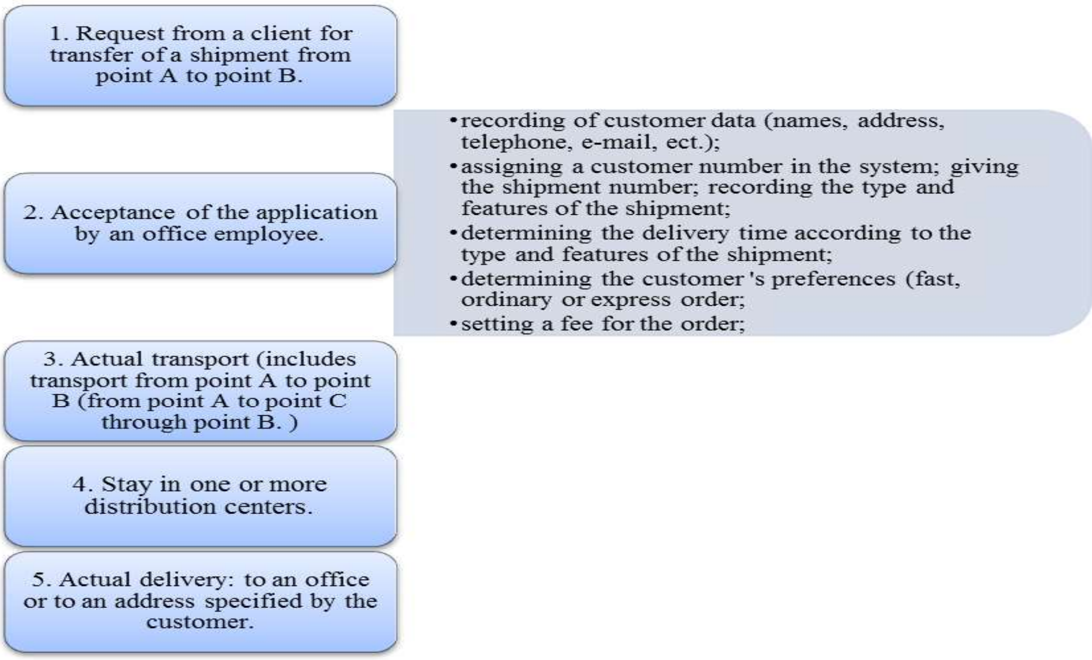


**Figure 6** Shipment delivery process.
Source: based on authors’ research

In order to avoid problems in tracking and to provide an opportunity for continuous monitoring, additional codes should be used in the internal system of the logistics organization to allow constant monitoring, tracking shipments (Fig. 7).

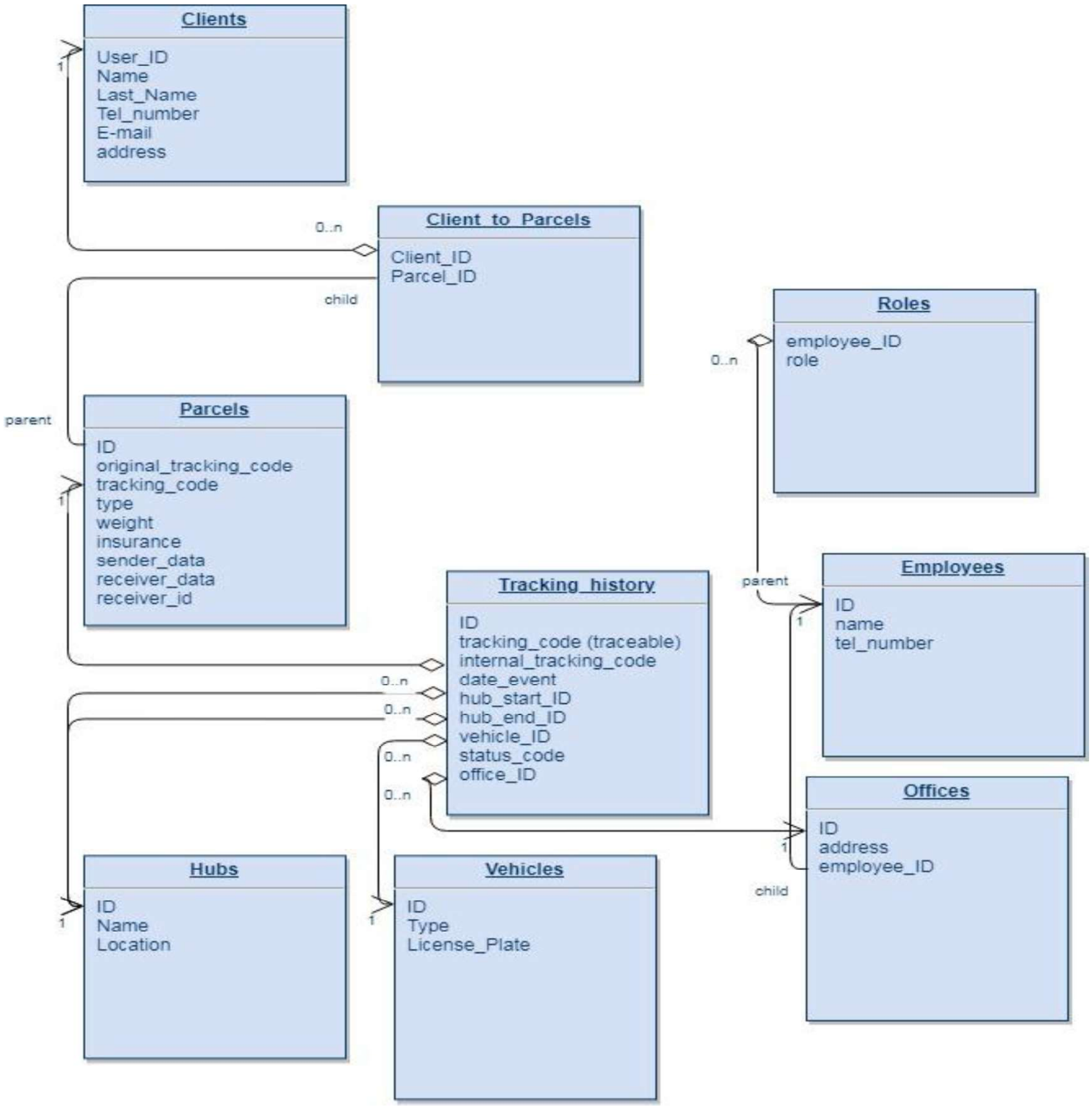


**Figure 7** Model of solution by storing data in Apache Kudu for logistics area
Source figure is based on Dr. L. Mileva's research.

The presented model of solution of the defined problem (Fig. 3) is expressed by:

1) marking in the system of the original code, which is given by the first supplier in the chain;

2) search for matching customer data, data, such as phone number and e-mail address, which will prevent problems caused by transliteration of customer names, as sometimes the transport is carried out between neighbouring countries, but in different languages such as Latin and Cyrillic (within the EU);

3) assigning a new code for tracking the shipment, which would allow the customer to track his shipment, according to the old number, which the company gave him initially upon request to the sender;

The conclusions made so far are about the positive consequences of applying a prototype solution model in logistics in big data analysis in the Hadoop ecosystem using the analytical tool Apache Kudu. They present the possibilities for comprehensive and continuous monitoring of shipments, cargo and goods.

The conclusion is that no data on shipments will be lost during the entire logistics process. This is important for big data analysis, as continuous monitoring will make it possible to analyse:

- direction of transport of goods and cargoes;
- tracking the movement of consignments from one country to another;
- orders of one client;

The possibilities to track orders of a customer could lead to benefits for the organization such as offering various promotional services, concluding contracts at lower prices, a package for loyal customers and more.

- sales revenue (taking into account the movement of consignments from one country to another and vice versa;

It will be possible to assess the risk of losses in transport to destinations that carry a higher risk of damage to resources.

- transport costs;
- profitability of transport;

Except the following features there should be considered limitations when creating the model:

- economic factors;
- geographical factors;
- features of the cargo;
- paid transportation fee. In addition to the loss of data in the tracking of goods, there is another serious problem of transportation- the presence of empty shipments during transportation, which leads to losses of funds.

General problems also arise from the fact that unforeseen events can occur during transport, for which it is difficult to predict how long they would be removed, such as a flat tire, an overturned vehicle, border closures, etc. The proposed model for improving the work aims to reduce the legal burden on land through an optimization task. For the purposes of the task are described, the available criteria and they can be further divided into categories: geographical, economic and infrastructural. Additional restrictions may arise when, as full optimization is not possible in some specific cases, independent of the forecast data, for example a vehicle waiting in line due to a traffic jam of unclear nature (for example: accident, inverted vehicle, etc.), and in some cases human resources will also be used. For the purposes of designing a prototype model for analysis and decision.

## 4. Analysis and interpretation of the results for the field of logistics

Tables 1 to 7 examines the cases of fright transport loaded and empty divided into national and international, represented by Bulgaria and its neighbour countries. The data presenting the performed transportations in vehicle kilometres and thousand journeys. For purposes of the article were made certain calculations. The data about Romania are not available in Eurostat database so that they are not represented in this article. Croatia is presented like country nearest Bulgaria's neighbour Serbia that is not in EU. The data are collected from Eurostat (*Eurostat*, 2021) database.

**NATIONAL**

**Table 1**

Growth rates of Loading status "Empty" (in million vehicle kilometers) by transport coverage "National" transport (in million vehicle kilometers)

| Country | 2019/2018 | 2018/2017 | 2017/2016 |
|---|---|---|---|
| Bulgaria | 0,807 | 0,930 | 1,030 |
| Greece | 0,948 | 1,017 | 0,925 |
| Croatia | 1,000 | 1,005 | 1,053 |

Source: The table is based on author's calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

According to Table 1 results, there is a slight decrease in the share of empty shipments, which marks a positive trend for Bulgaria for 2019/2018 compared to 2018/2017. For neighboring Greece, the results increase as the share of vacancies also decreases, but for the last year compared to the previous 2019/2018 is higher than in Bulgaria. In comparison, Croatia maintains the position of the coefficient around one, which means that there are no major changes in the share of empty shipments, but the values are above one, which means that empty shipments are fact (2018/2017, 2017/2016).

**Table 2**

Growth rates of loading status "Empty" (in thousand journeys) by transport coverage "National" transport (in thousand journeys)

| Country | 2019/2018 | 2018/2017 | 2017/2016 |
|---|---|---|---|
| Bulgaria | 0,823 | 0,932 | 1,023 |
| Greece | 1,031 | 0,942 | 0,971 |
| Croatia | 1,056 | 1,001 | 0,998 |

Source: The table is based on authors' calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

The expectation is that the rates of change in the relative share of empty transportation (calculated on the base of million vehicle kilometres and thousand journeys) will be close in relation to all loads. This rule applies for the three surveyed countries (for national transport) and the three different periods. The exception is for Greece ( 2018/2017), where the rate of change in the relative share of empty loads (calculated on the basis of million vehicle kilometres) is over 1 and the rate of change in the relative the share of empty loads (calculated on the basis of thousand journeys) is below 1. This fact shows Greece's targeted efforts to reduce the relative share of empty loads as thousand journeys at national level.

**Table 3**

Ratio between loading status "Empty" (in million vehicle kilometers) divided by loading status "Total loaded and empty" by transport coverage "National" transport (in million vehicle kilometers)

| Country | 2019 | 2018 | 2017 | 2016 |
|---|---|---|---|---|
| Bulgaria | 0,139 | 0,384 | 0,372 | 0,375 |
| Greece | 0,377 | 0,394 | 0,387 | 0,392 |
| Croatia | 0,110 | 0,381 | 0,375 | 0,375 |

Source: table is based on authors' calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

For the growth rate indicator, empty journeys are compared to the total journeys in national level, Bulgaria shows stable data for 2016-2018 and a decrease in 2019 of the empty journeys compared to the total transport. Greece has shown also decrease in the last year. Croatia shows data similar to those for Bulgaria, which shows a positive trend.

**INTERNATIONAL**

**Table 4**

Growth rates of loading status “Empty” (in million vehicle kilometers) by transport coverage “International” transport in million vehicle kilometers)

| Country | 2019/2018 | 2018/2017 | 2017/2016 |
|---|---|---|---|
| Bulgaria | 0,700 | 0,640 | 1,000 |
| Greece | N/A | N/A | N/A |
| Croatia | 1,084 | 1,021 | 0,979 |

Source: table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

For Bulgaria we have a decrease for 2018/2017 (Table. 4), with a slight increase in 2018/2019. Data for Greece at the international level are missing. For Croatia, we have a slight increase in 2018/2017 and this trend continues in 2019 in terms of empty vehicle in million vehicle kilometres in international traffic. For Croatia, the number one has been exceeded, which means that unloaded shipments have increased. Attention should be focused on international transport in order to reduce empty transportations by road.

**Table 5**

Growth rates of loading status “Empty” (in thousand journeys) by transport coverage “International” transport (in thousand journeys)

| Country | 2019/2018 | 2018/2017 | 2017/2016 |
|---|---|---|---|
| Bulgaria | 0,808 | 0,657 | 1,203 |
| Greece | N/A | N/A | N/A |
| Croatia | 0,994 | 1,056 | 0,973 |

Source: The table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

It is expected the rate of change in the relative share of empty loads relative to all loads (calculated on the basis of million vehicle kilometers and total journeys) to be close. The rule applies to the three countries surveyed for international transport.

**Table 6**

Ratio between loading status “Empty” (in million vehicle kilometers) divided by loading status “Total loaded and empty” (in million vehicle kilometers) by transport coverage “International” transport

| Country | 2019 | 2018 | 2017 | 2016 |
|---|---|---|---|---|
| Bulgaria | 0,098 | 0,107 | 0,117 | 0,109 |
| Greece | 0,087 | N/A | N/A | 0,108 |
| Croatia | 0,174 | 0,161 | 0,171 | 0,176 |

Source: The table is based on authors’ calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

The ratio between Loading status “Empty” (in million vehicle kilometers) divided by loading status “Total loaded and empty” by transport coverage “International” transport (in million vehicle kilometers) (Table 6) shows seriously low values especially for 2019 for international level. Levels for Greece and Croatia are almost the same as Bulgarian.

**Table 7**

Road transport year 2019

| **Country** | **National level** | | **International level** | |
|---|---|---|---|---|
| | **empty VKM/total loaded and empty VKM** | **empty TJ/total loaded and empty TJ** | **empty VKM/total loaded and empty VKM** | **empty TJ/total loaded and empty TJ** |
| Bulgaria | 0,399 | 0,493 | 0,098 | 0,346 |
| Greece | 0,377 | 0,467 | 0,087 | 0,185 |
| Croatia | 0,375 | 0,453 | 0,174 | 0,359 |

VKM – million vehicle kilometers
TJ – thousand journeys
Source: The table is based on authors' calculations through data from: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en

The last table (Table 7) shows road freight transport for 2019. It makes an impression that at national level there is an increase for the three countries in number of journeys without cargo. As for the number of empty shipments to all shipments at international level, there is a high coefficient of empty shipments 0.346. At the same time as the distance traveled by the vehicles is empty- the coefficient is small 0.098, which means that the vehicles travel to very short distances empty. Statistics reports a high number of empty courses, but they are very short distances. This is due to the large number of logistics centers in Western Europe located on the out of the cities. After delivering the cargo to the address/store, the vehicle return empty in the direction of the respective logistics hub and take the next cargo. The values for all remain higher because there are not so many logistics centers in the monitored countries. Therefore, it is important to pay attention to the optimization of transport in Bulgaria and its neighboring countries.

Therefore, a model will be created that helps to overcome the problems of lack of traceability of goods throughout the supply chain and optimize shipments so that they are carried out at the lowest possible cost, with the greatest possible profit.

## 4. Summary and proposal for solving the problem in the field of logistics

From the conclusions made so far, it becomes clear that solving the problem of incomplete tracking of goods along the entire transport route, which leads to data loss, will contribute to solving another serious discrepancy- the problem of available empty traffic, which leads to losses of financial resources for organizations. The solution by implementing an information system in Apache Kudu, by assigning a shipment number for easier and continuous tracking, should also lead to a solution to address the problem of available empty rates for carriers. This can be done by optimizing processes, with the help of questionnaires, which will be used to assess the condition of the road. They will help you choose the right route in real time.

The data will be entered, processed and stored in Apache Kudu (as part of the Hadoop ecosystem), after the process is finished. In this way, they will be avoided possible empty courses of transportation. The proposed solution will implement through a prototype optimization model using the Apache Kudu analytics tool in the Hadoop ecosystem. This will be possible by compiling a model with help of questionnaires. The model presented finding of most profitable route at the corresponding restrictions and weights. When formulating the model, they are important dependencies on the type and type of the vehicle, and additional conditions arise (for example, the direction of the route). Therefore, they a given different weights for the same sections of the route from point A to point B. As an example, a common problem exists in transit traffic when crossing the borders of countries due to driver testing for Covid-19, hence several changes. For example, changing the quarantine period for drivers further blocks the transport of logistics companies. Within 14 (10) days, the vehicle remains blocked, with 20% of the fleet of a small company remaining unusable. For small organizations, this means that they suffer greater losses due to the same external factors, but different internal- for example, in the presence of 5-6 vehicles. For large and medium-sized companies the situation will be different, i.e. when creating an optimization model, one factor will be determining- the size of the organization. As a rule, in case of complete lockdown of a country that is in transit on a given route, the route is extends the journey and the cost of cargo transportation becomes more expensive. This leads to losses for the carrier company. Therefore, the route passes optimization through implementing the questionnaires results in the Apache Kudu analytical tool of the Hadoop ecosystem. However, with such extension of routes, other destinations, which do not lead to extension of the road or go through a longer route with lower costs, are sought. This can be expressed, for example, in lower tolls, cheaper fuel, lower prices of vignettes, reducing the cost of accommodation and day meals for drivers. In the event of a change, it may be possible to transfer additional cargo to the country through which the vehicle passes in order to bypass the closed country. Some of the proposed criteria for modelling the analysis by implementing criteria assessment questionnaires in apache Kudu for the most profitable route are presented in Table 5. Model is based on obtaining and analysing real-time information for an alternative route, at the lowest possible cost. For the purposes of constructing the column of criteria, values (Table 8) from A to D of the criteria are assigned, according to the conditions of the environment at the time of departure and later during the trip (A is the highest and D the highest (lowest)). Estimates of the criteria may change throughout the trip if any of the conditions change. The data are collected in the Apache Kudu storage with the ability to process and analyse in real time. It is important to notice that in this case an important factor is the distance travelled with a load vehicle. According to him, the model is also being built, as it became clear that the transports without cargo in many countries exceed those with cargo. The aim is to offer a solution through which to optimize the transport and the aim is to have them with cargo, so that there can be a greater profit for the logistics organization. Route calculation estimates are given as follows in Table 8.

**Table 8**

Questionnaire for evaluating the strength of each criterion

| | **Criteria** | **Strength of each criterion** |
|---|---|---|
| | Price of transportation- per km (incl. permanent fees- (security, vignette, annual technical inspection, tax) | |
| | Road network condition | |
| | Weather conditions | |
| | Distance with loaded vehicle (in km) | |
| | Distance with empty vehicle (in km) | |

Fill in A, B, C, D or E for each criterion, where "A" is the strongest influence; "E" is the weakest influence

Source: Authors' contribution.

**Table 9**

Aggregated data from a conveyed survey among 5 experts

| **Criteria** | **Aggregated data** | | | | |
|---|---|---|---|---|---|
| | **A** | **B** | **C** | **D** | **E** |
| Price of transportation- per km (incl. permanent fees- (security, vignette, annual technical inspection, tax) | 5 | | | | |
| Road network condition | | | 1 | 4 | |
| Weather conditions | | | 4 | 1 | |
| Distance with loaded vehicle (in km) | | 5 | | | |
| Distance with empty vehicle (in km) | | | | | |

Responses are A, B, C, D or E for each criteria, where "A" is the strongest influence; "E" is the weakest influence.

Source: Authors' contribution

Aggregating data from 5 experts are showed in table 9. According to ranking, each evaluation of criteria could occupy a specific value in the selected range. As some factors are variable, they do not depend on the expert opinion itself, but depend on the specific environmental conditions. For example, when the country borders closed (due to Covid-19 situation), the information should be submitted by the relevant authorities, reported by the organization, entered as data in the information system and taken into account when choosing delivery routes. The same will be the case with factors such as weather road conditions. In order to assess the meteorological situation, the assessment comes from relevant specialists such as meteorological service as an expert statement, which is permanently and its assessment does not change. In advance, only the experts' statements about this factor, re-prioritizing the criteria. This will be repeated for each factor separately. In addition, factors involving the type of vehicle, road surface and others play a major role in making a decision. All evaluated grades at precisely defined point in time will be summarized and the completed assessment will enter the company's information system. Before, entering the information system, the evaluation of the criteria must receive a numerical value. Factors are distributed according to the degree of importance and the points they have received. Experts and specialized commissions should outline the scale for assessing the factors, and appropriate software will extract the data from the various sources in order to determine the most profitable route or the route with the least losses. The recorded data will enter the Apache Hadoop system and will be processed and stored in Apache Kudu so that it can be accessed in real time.

## 6. ICT prototype solution with analytical tool Apache Kudu in logistics area

This chapter discusses the most important moment about working with the model.
**General rules of the model.**

1. Requirements for its implementation.
2. Key points.
3. Used methods.
4. Scope and elements of the model.
5. Methodology for application.
6. Logical sequence of the model evaluation.
7. Conclusions and recommendations.

For the article purposes, on the base of the made studies, will be constructed a prototype model to address specific problems, in particular the problems resolving the lack of traceability along the entire route chain and optimizing routes to minimize unloaded traffic, as a survey has shown that a large proportion of shipments in EU countries are empty. The elements of the model and its use are characterized, which are based on the analysis and assessment of the condition of the transports with the help of road transport by land. The coefficient ratio of the travelled distances with empty transportation (international to national) is calculated. Conclusions are made for the article. It has a comparison between the kilometres travelled in the EU member states. General, national and international transportations are considered. The results of the study introduce several points:

1) main problems in logistics; 2) the factors influencing the work of the logistics organizations in their part for transportation of cargoes, goods and consignments; 3) the points of dependence at which the criteria would save costs and / or bring a profit; 4) the opportunities and areas in which to look for solutions to improve the work.

The model reflects our view that built on the base of a systematic approach outlines the activities and processes that complement each other and are interconnected, through the criteria for optimization of transport to find the most profitable transport route. In addition, when considering the elements of the model, it becomes clear that performing activities of different nature in the process of transporting goods from point A to point B, the system optimization should collect and store data through the Apache storage. Kudu. This process consists of complying with various important criteria. They are located at different levels, in different activities and are performed in different periods of time. The activities related to the optimization of the logistics processes have been established. In a certain period of time, when transporting goods from point A to point B, some factors influence, and in another period - different from the first. At the same time, it is possible in some cases to perform parallel calculations when assessing the situation when deciding which route the vehicle should take in order to avoid the unloaded distance. From the analyses made it can be concluded that the travel criteria are of key importance for optimizing the implementation of logistics processes in their part when transporting goods from point A to point B.

The main principles to be followed in the model are the rules of transparency and objectivity in evaluation; observance of clear rules and systematics in the assessment of the optimization criteria for cargo transportation from item A to item B.

The requirements for the application of the model are outlined: 1) The need for specific knowledge and skills for the processes of the specialists who will deal with the development of a system for optimization of transport through the analytical tool Apache Kudu, in the Hadoop ecosystem. 2) Reliability of the information and observance of the requirements for setting the goals of the model - for clarity, specificity, reachability, measurability. 3) Precisely defined deadline, according to their importance. 4) Logical sequence of execution of the stages. The key points are:

a) The model is valid for the processes in logistics, by continuous tracking of shipments and selection of the most appropriate route, requiring an integrated systematic approach in planning, analysis and management b) The information, used in model needs to be update regularly to meet the needs of the economy and management, including planning and analysis. c) Model should be implemented in the Hadoop ecosystem through data storage in Apache Kudu d) The system approach provides serious opportunities for analogies of processes and phenomena and therefore the model discussed here can be used in various other areas (not only in logistics). Next, the methods used in the model are included: 1) systematic analysis; 2) comparative analysis; 3) content analysis; 4) method of generalization and 5) method of deduction. Sources for collecting information from: analyses of the Ministry of Transport, Information Technology and Communications, the National Statistical Institute, Eurostat; content analysis of the web pages of some logistics companies; documentation of the Kudu analytical tool in Cloudera and other studies by the author. The objectives are: 1) a proposal for the implementation of part of the conceptual model of an information system with the possibility of assigning an additional code for tracking cargo, so as not to interrupt the traceability of the shipment 2) article of data on shipments with cargo, without cargo and total for EU countries. The information is collected through the above methods, the criteria that affect the transport of goods and consignments by land by road are differentiated.

The data processing accomplishes with the help of questionnaires cards, with the development of part of an information logistics system. This is in order to assign a code number to track the cargo along the entire route from point A to point B and rate the route criteria to find on the optimal route (with the least losses) and to implement taken decisions in big data storage system of Apache Kudu. The methodology for application of the model is divided into stages, which include analysis and remodelling (if necessary) which solves the problem of loss of data for loads in tracking from item A. to item B, through which data will be collected for shipments in Apache Kudu, especially in cases where the shipment is carried out by different companies and passes through different countries. Data analysis through the criteria is important for improving the delivery process and calculating the most profitable route with help of a criteria questionnaire assessment. The evaluation of the quality of logistics service includes results such as: 1. Development of a conceptual model of a part of an information system, by assigning a code for tracking shipments along the entire route from item A to item B by storing it in Apache Kudu. 2. Analysis and evaluation of land transport by road for EU countries with and without cargo. 3. Determining the degree of threat from passing routes without cargo, which leads to losses (transportation inside and outside the country) (considering the data of the EU member states). 4. Establishment of a prototype of a model, by answer questionnaire about finding the most optimal way according to certain criteria in order to improve the activity of the logistics organizations from the point of view of performing transportations with the least losses, caused by an empty vehicle. 5. Identify capabilities for big data analysis using the Apache Kudu analytics tool in the Hadoop ecosystem.

Elements- model for analysis of the most profitable route presents a model of a systems approach, where the elements are input, transformation processes and output. The classical model of the systems approach includes several elements; these are input, transformation and management processes, output and feedback. The model is illustrated by fig. 10.

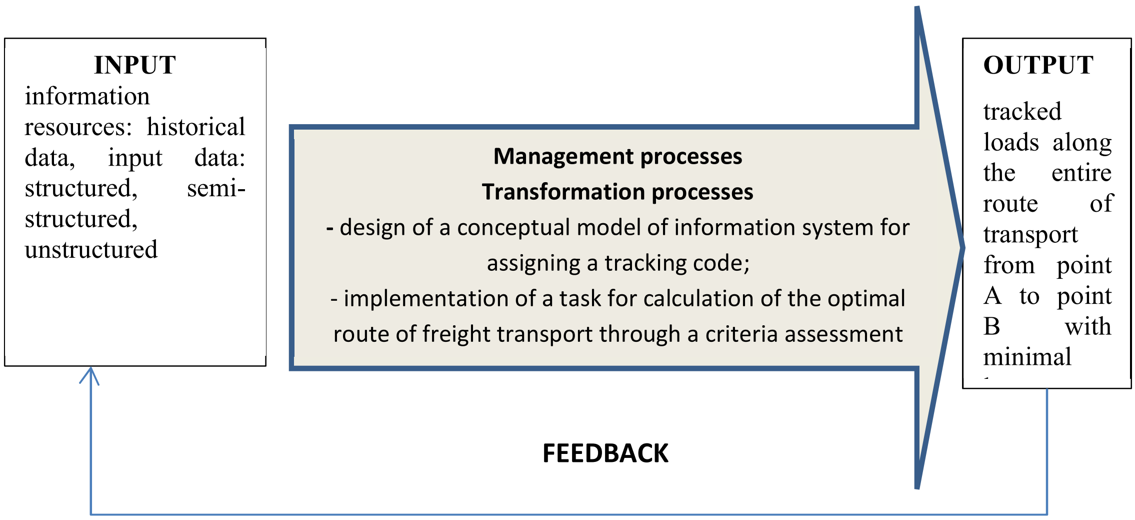


**Figure 8** Prototype of a big data analysis model in Apache Kudu storage
Source: The figure is based on authors' research

Resources that work in the operation of the system represents Model input. In this case, it is better to integrate model based on a systems approach, in order to store and process the data using the Apache Kudu analytics tool as part of the Hadoop ecosystem. Here the important resources for the model are the information resources. They are represented by historical transport data, all delivery data, routes, points through which the cargo pass, travelled distance, travelled time, delivery costs, route changes and all others that are collected and processed with traceability in real time. Then came the management processes and transformation processes, represented by business processes. The management processes are expressed in the degree of management of the tracking information system and the optimization system. By collecting large data in the Apache Kudu storage, as part of the Hadoop ecosystem, the processes of transporting goods, cargo and shipments are tracked. The criteria characterizing the efficiency of the transport are determined and defined on the base of the research of performed transportations in and from the EU member states.

The results of the study lead to the conclusion that one of the most important criteria in transportation is the distance travelled with cargo. The studies also showed that there were too many kilometres travelled in different Member States, leading to losses. Also important for transportation are the cost of transportation and the distance travelled in kilometres. There are cases in which it is better to cover the entire distance without cargo, but they are significantly less than in other cases. Route selection criteria play a significant role in calculating the most profitable route through criteria assessment questionnaires. Calculating the optimal transport route is part of the transformation process. Other factors, such as the condition of the road surface, closure of a road from a certain route due to repairs and accidents and the meteorological situation are also observed here. It should be noticed that there are also conditions in which due to poor road and meteorological conditions there is no possibility to choose an alternative route. This reduces the choices to zero. However, most trips have options available, so optimization can be done in other cases. The result obtained after the calculation of the optimal route is in the output as an element of the model, which has tracked cargo along the entire route of transport from point A to point B with minimal losses.

## Conclusion

The capabilities of big data analysis with Apache Kudu (as part of the Hadoop ecosystem) are expressed with the implementation of the model and data storage with following goals: 1) prevention of data losses due to interruption of traceability of supply codes and 2) optimizing routes, reducing costs and increasing profits for organizations in logistics. Regarding the possibilities for researching additional criteria, in the calculation of the most profitable route outline as follows:

1) the shortest route;
2) route passing on main roads; route according to the type of cargo (for special cargo);
3) route according to the type of vehicle;
4) traffic jams and repairs;
5) closed areas due to Covid 19;
6) closed areas requiring PCR or other tests (during pandemics and other crisis situations);
7) areas with overloaded traffic;
8) areas requiring increased attention due to unsuitable weather conditions.

Solving the listed problems will lead to the following benefits and opportunities for business organizations: 1. Minimization of costs for transportation on route from item A to item B without cargo. 2. Increasing the revenues from transportation along the route. 3. Reducing the risks of misuse of loads by drivers. 4. Reducing the risk of damage to cargo. 5. Reduction of costs for storage of shipments. 6. Reduce the time spent on a vehicle. 7. Greater control over shipments; 8. Greater staff control; 9. Greater flexibility; 10. Reduce the cost of food for a longer stay. The benefits of integrating the model into the Hadoop ecosystem will also be achieved with the help of the Apache Kudu analytical tool, which: 1. Ability to track in real time the cargo at any point along the route from point A to point B. 2. Ability to select compliance requirements on request, including the option of strict consistency of cargo.

At the next stage, the model can use machine learning to make predictions and analyses. To develop the formula for calculating whether the road is good or not, data collected from vehicles already on a route can be used by comparing the estimated time to travel on a route with the actual travelled time. For example, if the estimated time has been set and n serial vehicles passing this route show a longer time than the estimated one, then it follows that in the future, vehicles should not go this way. Another possibility to use machine learning and data storage in Apache Kudu is to track the average speed of vehicles kilometre by kilometre and determine the problematic sections of the road by making calculations with different coefficients the different sections again kilometre by kilometre and thus obtaining more accurate calculations. In this way, logistics organizations will add value to their activities.

## Literature

Abdel-Badeeh M. Salem, S. P. (2018) ‘Developing a web-based ontology for e-business. Int J Electron Commer Stud. 2018;9(2):119–32’.

Ana-Maria Ramona, S., Marian Pompiliu, C. and Stoyanova, M. (2020) ‘Data Mining Algorithms for

Knowledge Extraction’, in *Challenges and Opportunities to Develop Organizations Through Creativity, Technology and Ethics*, pp. 349–357. doi: 10.1007/978-3-030-43449-6_20.

*Apache Kudu - Overview* (2017). Available at: https://kudu.apache.org/overview.html.

Baesens, B. (2017) ‘The big effects of big data’.

Baesens, B. (2019) ‘Real-world techniques for analyzing big data | SAS’, *Sas*. Available at: http://www.sas.com/en_us/insights/articles/big-data/real-world-big-data-tips.html.

Calzon, B. (2021) ‘Your modern business guide to data analysis methods and techiques’, *Datapine*. Available at: https://www.datapine.com/blog/data-analysis-methods-and-techniques/.

Cristescu, M. P. (2019) ‘Specific aspects of the optimization of the reengineering processes of the distributed information applications’, in *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, pp. 627–636. doi: 10.5593/sgem2019/2.1/s07.082.

*Eurostat* (2021). Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=road_go_ta_vm&lang=en.

Heidenreich, H. (2018) ‘What are the types of machine learning?’ Available at: https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f.

Ileanu, B. V. *et al.* (2019) ‘Intriguing behavior when testing the impact of quotation marks usage in Google search results’, *Quality and Quantity*, 53(5), pp. 2507–2519. doi: 10.1007/s11135-018-0771-0.

Kirova, A. (2009) ‘Innovations in the transport services market’, *(in Bulgarian) Mechanics ISSN 1312-3823 Transport issue 3, 2009 Communications*, 3(3). Available at: https://mtc-aj.com/library/388.pdf.

Kuyumdzhiev, I. (2020) ‘A model for timely delivery of it solutions for Bulgarian universities’, in *20th International Multidisciplinary Scientific GeoConference Proceedings SGEM 2020, Informatics, Geoinformatics and Remote Sensing*, pp. 3–10. doi: 10.5593/sgem2020/2.1/s07.001.

Medvedev, A. and Sergeev, A. (2020) ‘No The forecast of the methane concentration changes for the different time periods on the arctic island bely. In: AIP Conference Proceedings. 2020. p. 120019.’

Miryanov, R. and Petkov, J. (2017) ‘An application of the relation between arithmetic mean and geometric mean for rational proving some inequalities’, *Mathematics and informatics*, 60(4), pp. 363–369.

Nacheva, R. *et al.* (2019) ‘Concept map mining approach based on the mental models retrieval’, *TEM Journal*, 8(4), p. 1484. doi: 10.18421/TEM84-54.

Nacheva, R. and Sulova, S. (2020) ‘Internationalization in Context of Education 4.0: AHP Ranking of Bulgarian Universities’, in *ACM International Conference Proceeding Series*, pp. 278–284. doi: 10.1145/3407982.3408006.

Nacheva, R. and Sulova, S. (2021) ‘Research on the Overall Attitude Towards Mobile Learning in Social Media: Emotions Mining Approach’, in *Digital Transformation, Cyber Security and Resilience of Modern Societies. Studies in Big Data*. Springer, Cham, pp. 429–440. doi: 10.1007/978-3-030-65722-2_27.

Nikolaev, R., Milkova, T. and Miryanov, R. (2018) ‘A new meaning of the notion “expansion of a number”’, *Mathematics and informatics*, 61(6), pp. 596–602.

Parvanov, H. and Boeva, O. (2009) ‘Problems of European transport integration’, *(in Bulgarian)*. Available at: https://www.unwe.bg/uploads/ResearchPapers/Research Papers_vol2_2006_No2_H Parvanov, O Boeva.pdf.

Polkowski, Z. *et al.* (2020) ‘Evaluation of aggregated query plans using heuristic approach’, in *Proceedings of the 12th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2020*, pp. 1–4. doi: 10.1109/ECAI50035.2020.9223222.

Quinto, B. (2018) ‘High Performance Data Analysis with Impala and Kudu’, *Next-Generation Big Data*, pp. 101–111. doi: 10.1007/978-1-4842-3147-0_4.

Raeesi Vanani, I. and Majidian, S. (2020) ‘Literature Review on Big Data Analytics Methods’, *Social Media and Machine Learning*. doi: 10.5772/intechopen.86843.

Stoyanova, M. (2020) ‘Good practices and recommendations for success in construction digitalization’, *TEM Journal*, 9(1), pp. 42–47. doi: 10.18421/TEM91-07.

Todoranova, L. *et al.* (2020) ‘A model for mobile learning integration in higher education based on students’ expectations’, *International Journal of Interactive Mobile Technologies*, 14(11), pp. 171–182. doi: 10.3991/ijim.v14i11.13711.

Todoranova, L. and Penchev, B. (2020) ‘A Conceptual Framework for Mobile Learning Development in Higher Education’, in *ACM International Conference Proceeding Series*, pp. 251–257. doi: 10.1145/3407982.3407996.