The E-Commerce Systems Modelling Based on Petri Networks

Lyubomyr Chyrun

Ivan Franko National University of Lviv, University Street, 1, Lviv, 79000, Ukraine

Abstract

The paper simulates the process of interaction between virtual enterprises (participants) in the Internet world of e-commerce based on Petri nets. The general formal model of mutual relations using payment systems (processes) between participants of e-commerce is described. Regardless of the type of e-commerce, the primary indicator of the impact on the profit of e-commerce is the involvement of potential audiences to visit the relevant e-commerce sites and their further motivation to make the applicable orders. Appropriate e-commerce referral systems should withstand the worst of motivating the user to place an order. Timely further analysis of the statistics of visits and conversions depends on the formation of new goals to encourage potential e-commerce customers. The e-commerce system model based on the Petri net allows determining the set of financial relations between the participants of e-business. Analysis of labelled systems based on the Petri net enables you to build a tree of achievement of e-commerce goals and finding the relationships between site visits and e-business profits. Further application of the methods of correlation and cluster analysis to the obtained data allows classifying the purposes of conversions depending on the market reaction and the current time for the necessary adjustment.

Keywords 1

E-commerce system, information process marking, e-business, Petri net, time series, reach tree, trend detection, information technology, electronic content commerce system, intelligent system, payment system, average method, financial transaction, virtual enterprise, internet shop, successful conversion, cluster analysis, machine learning

1. Introduction

The twenty-first century has challenged humanity in the form of a worldwide “web” of the Internet and the emergence of an interactive business or virtual economy. An on-line business is a business built on the joint actions of a business process in a businessperson and a computer or other means of communication. A virtual economy is an economy based on interactive business and the primary law of man - saving time. Already in developed countries, there are many cases of implementation of this law, namely: using communication means, without leaving home, to operate technological lines in production or financial and commercial activities; maintain accounting; carry outdistance learning, reading books and periodicals; to buy goods; perform banking, exchange and other financial transactions. Nowadays, the Internet influences both the external relations between companies and their partners or clients and the internal structure of the companies themselves. E-commerce is based on the design of traditional commerce, and the use of electronic networks gives it flexibility.

2. Related works

E-commerce is the product/service sale for commercial electronic media, either via the network or via the Internet. Data may include ordering, payment and delivery of goods/services. Component e-commerce (EC) is presented on Fig. 1 [1-3]. The process that remains the EC cycle [4-6]: available

COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine
EMAIL: Lyubomyr.Chyrun@lnu.edu.ua (L. Chyrun)
ORCID: 0000-0002-9448-1751 (L. Chyrun)

content, formal ordering, payment, order placement, after-sales service and support. EC types [7]: trade in content, trade, services.

EC Categories [1-7]:

- Business-to-Business or B2B (which is a company that uses reasoning to order offers, required options and pay);
- Business-to-Consumer or B2C(electronic retail);
- Business-to-Administration or B2A (operations used between companies and government organizations);
- Consumer-to-Administration or C2A (does not yet exist, but it is pre-existing two categories used in different fields as social visits);
- Consumer-to-Consumer or C2C (the user who uses to reduce commercial content, informs, audits the trade representation between individuals)

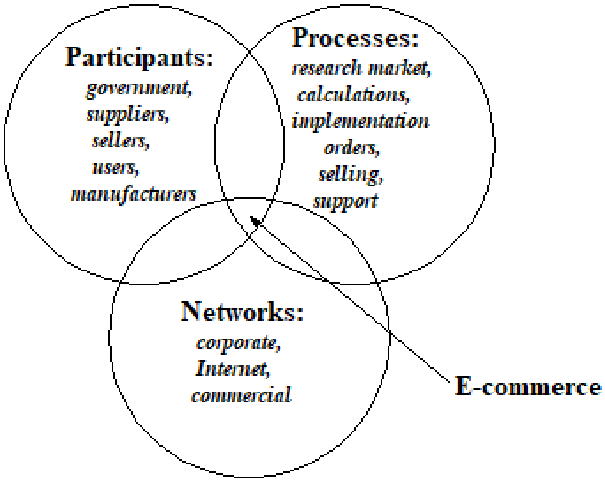

Figure 1: The components of e-commerce

The firm will use the tools to realize EC capabilities can roughly be divided into the following groups [8-11]: business applications; electronic stores; a gateway to the EDI system (electronic data interchange); connection with financial institutions through different payment systems.

The EC dissemination areas [12-15]:

- Marketing, sales and sales promotion;
- Resale, previous arrangements, deliveries;
- Financing and insurance;
- Business operations: ordering, receiving, payment;
- Product maintenance and support;
- Joint product development;
- Distribution joint production;
- Use of public and private services;
- Business administration (concessions, permits, taxes, customs, etc.);
- Transport, transportation and supply;
- General purchases;
- Automatic trading of electronic goods;
- Accounting;
- Resolution of controversial points.

Features of the Internet as the electronic market are as follows [16-21]:

- Openness is access for companies of all sizes and consumers;
- Globalism is access to it from anywhere in the globe;
- The openness of the market is characterized by a low barrier to entry for companies.

The Internet offers the ability to shorten distribution channels and eliminate intermediaries, such as a distributor and wholesaler, who are replaced by a direct buyer-producer relationship. The reason for this reduction is the ability of firms to take on the functions traditionally performed by middleware professionals since the Internet has a more efficient ability to interact with consumers and at the same time allows monitoring consumer information. On the technical side, this is due to developing technologies for building and maintaining databases and automatic processing of incoming requests.

Features of the Internet as a means of retail trade [22-27]:

- The consumer interacts with the Internet and requests for purchase directly through the Internet;
- The consumer controls the content on products and services and receives it at any time of the day;
- Organizing retailing on the Internet requires less investment than traditional retailing.

Compared to television or catalogue trade, consumers may request additional content sufficient to make a purchase, including on-line. Advantages of the Internet market [28-30]:

- Active position of the consumer;
- A comparatively smaller amount of money invested by firms to enter the market;
- The global nature of the market;
- 24/7 market availability;
- Possibility of obtaining the necessary amount of content

The development of the Internet environment as a market requires the implementation of three primary conditions [31-33]:

- The conquest of firms using the Internet as a channel of distribution of products, consumer confidence;
- Ensuring the reliability of participants and their operations;
- Creating security for transmission and further storage of data on the Internet.

Methods of gaining trust [34-37]:

1. Increase in guarantors shifting the risk of buying from buyer to seller;
2. Significant discounts for first-time customers. The first operation significantly improves the trust relationship between the buyer and the firm, which were almost absent before;
3. Details of the firm, including its history, business philosophy and biography, approval marks from other firms certifying the seller, etc.

Not being the only technology, e-commerce is characterized by versatility. It brings together a wide range of business operations, including [38-42]:

- Establishing contact, for example, between a potential customer and a supplier;
- Information exchange;
- Pre- and after-sales support (detailed product and service content, product instructions, customer answers);
- Selling; electronic payment (using electronic transfer of money, credit cards, electronic checks, electronic cash);
- Management of delivery and tracking of physical products, direct delivery of products that can be distributed electronically;
- Virtual enterprises are groups of independent companies working together to provide products and services that are not available to individual companies;
- The company and its trading partners jointly manage shared business processes.

Table 1 presents E-commerce features and benefits [43-45].

Table 1

E-Commerce Features and Benefits

Opportunities for suppliers	Opportunities for customers
Global presence	Global choice
Increasing competitiveness	Quality of service
Fulfilling of needs of the customer	Personalization of goods and services
Shortening of the goods path to the customer	Fast response to demand
Cost savings	Reducing of prices
New business opportunities	New products and services

Examples of specific e-commerce business benefits are as follows [46-51]:

1. Reducing advertising costs;
2. Reduction of shipping costs, mainly for goods that can be obtained electronically;
3. Reduction of design and strategic planning costs;
4. Great opportunities for market research niches in the market;
5. Equal access to the market (for large corporations and small firms);
6. Access to new markets;
7. Involvement of customers in the development and implementation of new products and services

Among the most critical legal issues that require an urgent decision with the participation of the world community are the following [52-57]:

- The order of taxation of transactions in electronic form;
- Tariffs;
- Requirements for the form of agreements and liability;
- Regulation of cryptography;
- Authentication rules;
- Protection of content;
- Consumer protection.

E-commerce now affects the economy and citizens' rights. Legal requirements are required to develop a global and open market by harmonizing legislation and simplifying the rules and procedures applicable in different countries. Developing such norms requires in-depth cooperation between business and government, which would provide significant benefits to manufacturers and consumers worldwide. Today, underdevelopment in general or fragmentation of e-commerce legal norms and significant contradictions between the laws of different countries are obstacles to the online economy's successful functioning. Businesses often exchange commercial content and make payments electronically and with their customers. Traditionally, systems based on EDI (Electronic Data Interchange) and EFT (Electronic Financial Transfer) standards have used to impose stringent conditions on the forms of content transmitted. Mainly large companies, corporations with private VAN networks [58] use such systems. Nowadays, both companies and individuals can conduct business transactions over the Internet without EDI. Commercial content transmitted over the Internet can be divided into two categories: information transactions; financial transactions.

Information provisioning is a fundamental and expensive element of e-commerce. Business information can take several forms [59-63]:

- Statistics (figures, graphs, analyses);
- Corporate information (telephone numbers, addresses, organization structure);
- Product or service information;
- Paid content (news, periodicals, access to databases, etc.)

There are three main classes of financial transactions [59-63]:

- Company-to-company (payments between banks, other financial institutions, transfers of funds from one firm's account to another's account are usually made by electronic transfer or by check);
- Company-to-client (the primary payment method is cash, checks, debit and credit cards);
- Client-to-client (the primary payment method is cash and reviews).

Using the Internet to perform these types of transactions allows you to replace the presentation or display of cash, checks, and credit cards with their electronic equivalents.

Today, there are specially designed electronic versions of payment systems through which you can: order and pay for a purchase, bank payments, investments (insecurities), etc.

The electronic payment system is an authorized information system designed to make payments on the Internet between financial, commercial, manufacturing, governmental organizations and individual users. One of the main requirements for payment systems is [58-63]:

1. Confidentiality;
2. Preserving the integrity of content;
3. Providing payment methods;
4. Providing authentication;
5. Authorization;
6. Minimization of transaction fee.

Payment systems can be classified [64-67]:

1. By payment type:
   a. Credit card-based systems;
   b. Systems based on Internet banking;
   c. Systems based on electronic checks;
   d. Systems using electronic money;
   e. Smart card-based systems;
2. According to the payment scheme:
   a. Credit;
   b. Debit;
3. By prevalence:
   a. International;
   b. National.

Each virtual enterprise p_i independently chooses the type of payment system t_i, which is the best option for conducting financial transactions. However, there is no single *C* system for conducting financial transactions between virtual enterprises with different payment systems in the virtual e-business world. There is no single standard *C* system for controlling financial transactions between participants in the virtual e-business world. Petri networks can represent the system of economic relationships through payment systems (processes) between virtual enterprises (participants) in the Internet world of e-commerce. Such a *C* system will consist of four elements [68]:

1. A finite set of participants $P = \{p_1, p_2,..., p_n\}$ of the electronic market system;
2. Multiple processes of relationships $T = \{t_1, t_2,..., t_m\}$;
3. Input function $I{:}T\rightarrow P$, that is, direct links between processes and participants;
4. Output function $O{:}T\rightarrow P$ (direct links between participants and procedures). (Fig. 2)

3. Material and methods

The system of conducting and controlling financial transactions between virtual enterprises is as follows:

$$C = (P, T, I, O). \quad (1)$$

The power of many virtual enterprises P - $|P| = n$.

The power of multiple payment systems T - $|T| = m$.

A participant (a virtual e-commerce enterprise) p_i is *the input position* of the process t_i, if $p_i \in I(t_i)$ [68-73]. Similarly, p_i is *the output position* of the process t_i, if $p_i \in O(t_i)$. Process inputs and outputs are kits of posts. *The equipment* (the kit)is a generalization of the set. Repeatedly repetitive identical items may be included in the kit. An example of the kit is $O(t_1) = \{p_2, p_3, p_4, p_4\}$.

The multiplicity of the input position p_i for the process t_j is the number of occurrences of the position in the input kit of process $\#(p_i, I(t_j))$ [68-73]. Similarly, *the multiplicity of the output position* p_i for process t_j is the number of position occurrences in the original set of process $\#(p_i, O(t_j))$.

The extended input function $I(p_i)$ is defined as follows [68]:

$$\#(t_j, I(p_i)) = \#(p_i, O(t_j)). \quad (2)$$

The extended output function $O(p_i)$ is defined as follows [68]:

$$\#(t_j, O(p_i)) = \#(p_i, I(t_j)). \quad (3)$$

The graphical representation of the financial relationship system in the virtual world of e-commerce is a bipartite oriented multigraph. The structure of the system consists of a set of participants and processes. Accordingly, the graph of a system consists of two types of nodes – the circle *O* denotes the participant and the bar **|** denotes the process. Therefore, since the vertices of a graph can be divided into two sets - participants and methods, the chart is bipartite.

Oriented arcs (arrows) connect participants and processes, with some arcs directed from participants to methods and some from operations to participants. Such arcs indicate the input and output kits. Multiple arcs represent multiple outputs. Since there can be various arcs between vertices, we have *a multigraph*. The arcs are directional, so this multigraph is oriented.

Thus, graph G of the financial relationship system in the virtual world of e-business is a bipartite oriented multigraph,

$$G = (V,A),$$

where $V = \{v_1, v_2,..., v_s\}$ is the set of vertices, $A = \{a_1, a_2,..., a_r\}$ is set of directional arcs, $a_i = (v_i, v_k)$, where $v_i, v_k \in V$. The set V can be divided into two non-intersecting subsets - P i T, so $V=P\cup T$, $P\cap T=\varnothing$ and for any arc $a_i(v_i, v_k) \in V$ one of two conditions is fulfilled: or $v_i \in T$, $v_k \in P$, or $v_i \in P$, $v_k \in T$. The kit of directional arcs A, corresponding to this system is defined as follows [68]:

$$\#((p_i, t_j), A) = \#(p_i, I(t_j)); \quad (4)$$

$$\#((t_j, p_i,), A) = \#(p_i, O(t_j)). \quad (5)$$

The dual graph of the system $C = (P, T, I, O$ is the graph of the system $\overline{C} = (T, P, I, O)$, that is, the graph in which participants and processes are reversed. Such diagrams are of no practical interest, so we do not consider them.

Marking is assigning a certain number of chips (monetary units) to the members of the relationship system. Marking are user resources, that is, the resources of a virtual e-commerce enterprise that it can use to conduct transactions in the scheme of financial relationships of e-commerce participants.

The marking μ of the system $C = (P, T, I, O)$ is a function that reproduces the participants P set into the nonnegative integers N set. The marking can also be represented as an n-dimensional vector $\mu = (\mu_1, \mu_2,..., \mu_n)$, where $n = |P|$, $\mu_i \in N$. The vector μ defines for each participant p_i the number of chips in this position μ_i. More commonly, marking is used as a function $\mu(p_i) = \mu_i$. The marked system $M = (C, \mu)$ - is the set of the structure of the system $C = (P, T, I, O)$ and the marking μ. The chips on the graph are indicated as a point inside the circle, indicating the given position of the participant.

Running the system C - is a redistribution of chips by *running* processes. When a function is started, the fragments are removed from their input member positions and added to their output. The process can only be activated when allowed. A transition is called allowed if each of its input positions has a chip count of at least the number of arcs from a place in the process. The chips in the entry position that allows the process is called *permits*. The process $t_j \in T$ in the marked Petri net $C = (P, T, I, O)$ with marking μ is allowed if the condition for all $p_i \in P$ is satisfied [68]

$$\mu(\mathrm{pi}) \geq \#(\mathrm{pi}, \mathrm{I}(tj)). \quad (6)$$

The process is started by removing all the allowance chips from their output positions and then adding the chips to each of their initial positions, one chip for each arc. The process of t_j in a marking system with marking μ can be started whenever it is allowed. As a result of triggering the allowed transition t_j a new marking is formed μ', which is determined by this ratio [68]:

$$\mu'(p_i) = \mu(p_i) - \#(p_i, I(t_j)) + \#(p_i, O(t_j)). \quad (7)$$

Its marking determines the status of system C. Running any of the processes changes the state of the system by changing its marking. The states space of a system having n participants is the set of all markings, i.e. N_n. A change in system state is caused by a process start-up and is determined by a function δ, called *the following state function*. The part of the next state δ for the system $C = (P, T, I, O)$ with marking μ and the process $t_j \in P$ is defined if and only if $\mu(p_i) \geq \#(p_i, I(t_j))$ for all $p_i \in P$. If $\mu(p_i)$ is defined, then $\delta(\mu, t_j) = \mu'$, where $\mu'(p_i) = \mu(p_i) - \#(p_i, \mathrm{I}(t_j)) + \#(p_i, \mathrm{O}(t_j))$ for all $p_i \in P$ [68]. Suppose that we have the system $C = (P, T, I, O)$ with initial marking μ^0. The sequential starting of processes can execute this network. Starting a legal t_j process will result in the formation of a new mark $\mu^1 = \delta(\mu^0, t_j)$. Any other allowed process, such as t_k, can be started in the resulting marked graph. We obtain the marking $\mu^2 = \delta(\mu^1, t_k)$. You can continue this process until at least one method is allowed.

When running the system, we get two sequences. The sequence of markings $(\mu^0, \mu^1, \mu^2,...)$ and the sequence of processes that are started $(t_{j1}, t_{j2}, t_{j3},...)$. These sequences are related by the following relation [68]:

$$\delta(\mu_k, t_{jk}) = \mu_{k+1},\ k = 0,1,2,... \quad (8)$$

With the help of this relation, having the structure of the system $C = (P, T, I, O)$, the initial marking μ^0 and the sequence of states $(t_{j1}, t_{j2}, t_{j3},...)$ we can obtain the series $(\mu^0, \mu^1, \mu^2,...)$. The reverse process from markings to the series of processes is possible in most cases, except some degenerate cases. For a system $C = (P, T, I, O)$ with marking μ marking μ' is called *directly achievable* from μ, if there is a process $t_j \in T$, such that $\delta(\mu, t_j) = \mu'$, i.e. μ' can be obtained from μ running one of the allowed processes.

The set of reach $R(C, \mu)$ for the system $C = (P, T, I, O)$ with marking μ - is the most miniature set of markings defined as follows:

1. $\mu \in R$;

2. if $\mu' \in R$, then for all $\mu'' = \delta(\mu', t_j)$, $t_j \in T$ is fulfilled $\mu'' \in R$, that is, if some marking belongs to the set of reach, then all reachable markings from it also belong to this set.

Often use the concept of the advanced function of the next state. The extended input state function is defined for the marking μ and the sequence of processes $\sigma = (t_{j1}, t_{j2},..., t_{jn})$ by this rule [68-73]:

$$E\mu' = \delta(\mu, \sigma) = \delta(\ldots\delta(\delta(\mu, tj1), tj2)\ldots, tn), \quad (9)$$

that is, we first start the process t_{j1}, then t_{j2} and so on, to t_{jn}.

The reach tree of system C is an illustration of the reach set $R(C, \mu)$. Since the reach is infinite in many cases, specific rules allow it to be reflected by the finite reach tree.

The following rules are used to bring the reaching tree to a complete view [68-73].

Passive markings - markings in which no processes are allowed are *terminal vertices* of the reaching tree (do not have their subtree reach).

If the marking received has already encountered in the reaching tree before, such markings are called *duplicate vertices*. Consideration of the same vertices is unnecessary because the resulting subtree will be similar. If in the reaching tree we get a mark that is greater than one of the previous (on the way to the root) markings, then in all positions we put the symbol ω.

With these three rules, a set of the reach of any system can be represented as a finite tree.

4. Experiments, results and discussion

For example, we will build a system of financial relationships between the following participants (Fig. 2): user - client (p_1), Internet - auction (p_2), customer's bank and on-line auction (p_3), Internet – shop (p_4), bank Internet - shop (p_5). Similarly, we have the following processes: payment for a product/service by a customer (t_1), financial interaction process between an Internet - auction, an Internet - shop and a customer's bank (t_2), financial transactions between an Internet - shop and its bank (t_3), transfer money at the level of banks and an Internet - shop (t_4, t_5).

	$I(t_1) = \{p_1\}$;	$O(t_1) = \{p_2, p_3, p_4, p_4\}$;
$C = (P, T, I, O)$;	$I(t_2) = \{p_2, p_3, p_4\}$;	$O(t_2) = \{p_2\}$;
$P = \{p_1, p_2, p_3, p_4, p_5\}$;	$I(t_3) = \{p_4, p_4\}$;	$O(t_3) = \{p_5\}$;
$T = \{t_1, t_2, t_3, t_4, t_5\}$;	$I(t_4) = \{p_5\}$;	$O(t_4) = \{p_3, p_4\}$;
	$I(t_5) = \{p_3\}$;	$O(t_5) = \{p_5\}$.

Figure 2: An example of System C in the virtual world of e-commerce

For the example of the system above, the extended input and output functions show in Fig. 3.

$I(p_1) = \{\ \}$;	$O(p_1) = \{t_1\}$;
$I(p_2) = \{t_1, t_2\}$;	$O(p_2) = \{t_2\}$;
$I(p_3) = \{t_1, t_4\}$;	$O(p_3) = \{t_2, t_5\}$;
$I(p_4) = \{t_1, t_1, t_4\}$;	$O(p_4) = \{t_2, t_3, t_3\}$;
$I(p_5) = \{t_3, t_5\}$;	$O(p_5) = \{t_4\}$.

Figure 3: An example of System C with extended input and output functions

Fig. 4 shows a graph of the system as an example on Fig. 2 with marking $\mu = (1, 1, 3, 2, 1)$.

In Fig. 4, the chip in position p_1 is the permitting for the process t_1. Examples of system C process start-up are shown in Fig. 5. In Fig. 5, process t_4 cannot be started.

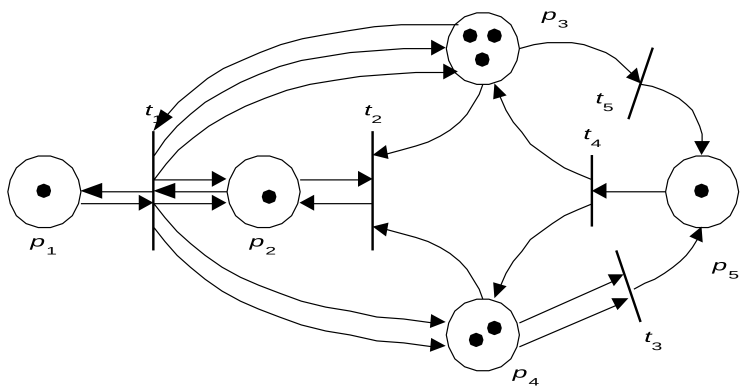

Figure 4: An example of a graph of a labelled C system represented by the Petri net

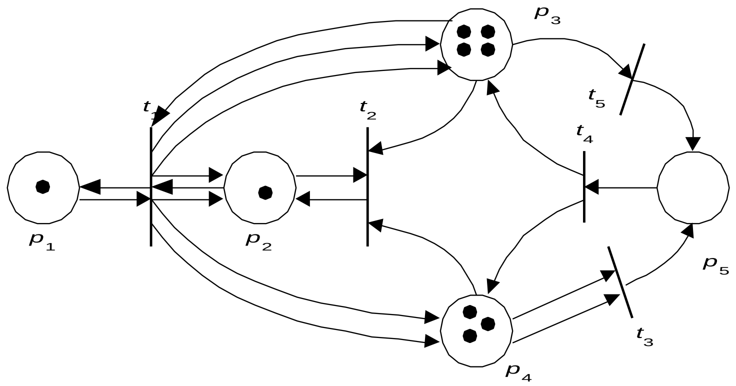

Figure 5: The marking received due to the start of the transition t4 of the graph depicted in Fig. 4.

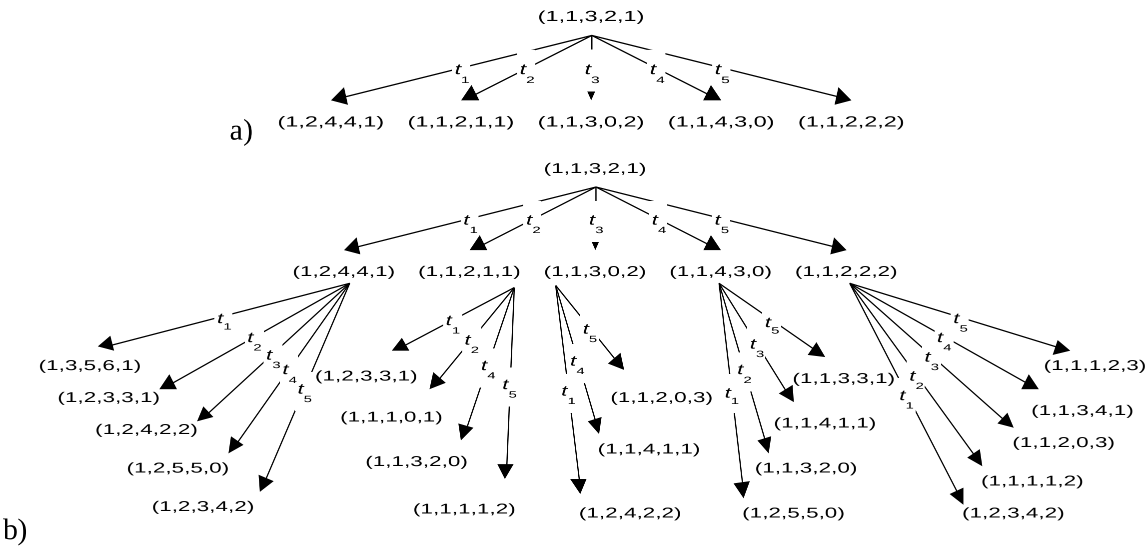

Figure 6: Building a reach tree: a) the first step; b) the second step

For example, consider the system shown in Fig. 4. This initial marking $\mu^0 = (1,1,3,2,1)$ will be the *initial vertex (root)* of the reaching tree (Fig. 6). Two processes - t_1 - t_5 are allowed in μ^0. New tops of the reaching tree are built for all directly accessible from μ^0 markings and arcs are going to them from the initial top marked by the triggering process. As a result of starting t_1 we get $\mu^1_1 = (1,2,4,4,1)$, and as a result of beginning t_2 we get $\mu^1_2 = (1,1,2,1,1)$. Thus, the first step of building a reach tree will be as shown in Fig. 6. Similarly, the following steps are carried out.

Consider the sequence of starts $t_1t_1t_1$. Each time, the marking received differs from the previous one only by the number of chips in p_2 - p_3 (increases by one) and in p_4 (increases by two units). Running the process t_1 many times, you can get any number of chips in it. Since starting the process increases the number of chips in one (several) positions while leaving the number of fragments unchanged in others, the marking of the position (s) in which the number of chips has increased is usually denoted by the symbol ω. The symbol *c* indicates an endless number of chips in the marking position, which does not change in the n step of building the reaching tree. You can extend this notion to a process sequence σ, that increases the number of chips in a position without changing other parts. Thus, if we obtain a mark in the reaching tree that is greater than one of the previous (on the way to the root) markings, then in all larger positions, we affix the symbol ω (Fig. 7).

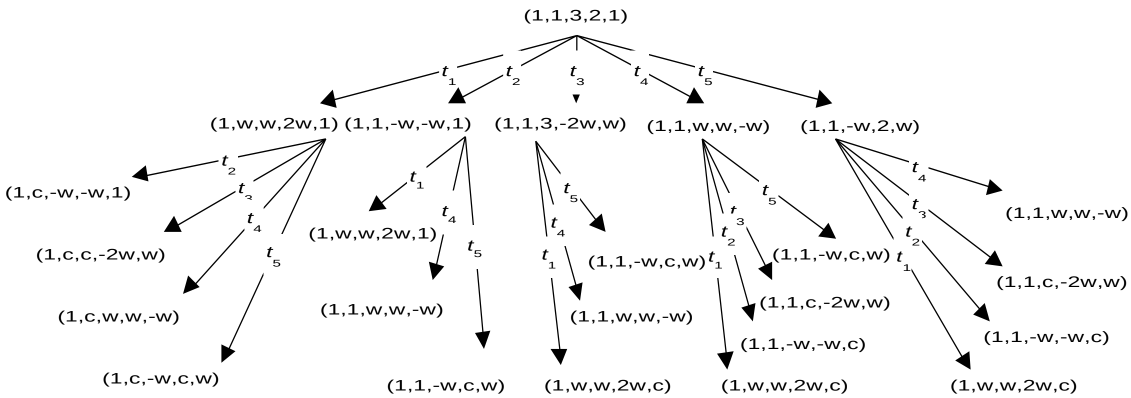

Figure 7: The tree of reach

Regardless of the type of e-commerce, the primary indicator of the impact on the profit of e-commerce is the involvement of potential audiences to visit the relevant e-commerce sites and their further motivation to make the applicable orders. Appropriate e-commerce referral systems should withstand the worst of motivating the user to place an order. Timely further analysis of the statistics of visits and conversions depends on the formation of new goals to encourage potential e-commerce customers. The e-commerce system model based on the Petri net allows determining the set of financial relations between the participants of e-business. Analysis of labelled systems based on the Petri net enables you to build a tree of achievement of e-commerce goals and finding the relationships between site visits and e-business profits. Further application of the methods of correlation and cluster analysis to the obtained data allows classifying the purposes of conversions depending on the market reaction and the current time for the necessary adjustment. The requires:

- Build a correlation field;
- Determine the value of the correlation coefficient;
- Calculate the correlation ratio;
- Plot graphs of autocorrelation functions;
- Break one of the sequences into three equal parts;
- Build a correlation matrix for them; - find the coefficients of multiple correlation.

First, will pre-process the data of the sessions of visiting the e-commerce site when the conversion is achieved and graphic representation of these results (Fig. 8-10):

```
df <- read.csv("D:/V20.csv", header = TRUE) df1 <- read.csv("D:/V20.csv",
header = TRUE) head(df) summary(df)
str(df)
df$Індекс_дня <- as.Date(df$Індекс_дня, format = "%m/%d/%y")
  df$Month  <- as.Date(cut(df$Індекс_дня,  breaks  =  "month"))  df$Week  <-
as.Date(cut(df$Індекс_дня, breaks = "week", start.on.monday = TRUE))
  df$Day <- as.Date(cut(df$Індекс_дня, breaks = "day"))
  ggplot(data = df, aes(x = df$Індекс_дня, y = df$Сеанси)) +
           geom_point(data = df) +
           labs(x = "Month", y = "Total Sessions", title = "Sessions/month")
           + geom_line(data = df)
  r1 <- ggplot(data = df, aes(x = df$Month, y = df$Сеанси)) +
           stat_summary(fun.y=sum, geom="line", colour="mediumvioletred")+
           scale_x_date(labels = date_format("%m"), breaks = "1 month",
                    guide = guide_axis(n.dodge=1)) +
           labs(x="Month", y="Total Sessions", title="Sessions per month")
```

```
> head(df)
  Індекс_дня Сеанси      Month       Week        Day
1 2020-07-18      2 2020-07-01 2020-07-13 2020-07-18
2 2020-07-19      5 2020-07-01 2020-07-13 2020-07-19
3 2020-07-20      5 2020-07-01 2020-07-20 2020-07-20
4 2020-07-21     10 2020-07-01 2020-07-20 2020-07-21
5 2020-07-22      5 2020-07-01 2020-07-20 2020-07-22
6 2020-07-23      6 2020-07-01 2020-07-20 2020-07-23
> summary(df)
   Індекс_дня             Сеанси           Month                 Week                  Day            
 Min.   :2020-01-01   Min.   : 0.000   Min.   :2020-01-01   Min.   :2019-12-30   Min.   :2020-01-01  
 1st Qu.:2020-04-03   1st Qu.: 1.000   1st Qu.:2020-04-01   1st Qu.:2020-03-30   1st Qu.:2020-04-03  
 Median :2020-07-04   Median : 3.000   Median :2020-07-01   Median :2020-06-29   Median :2020-07-04  
 Mean   :2020-07-02   Mean   : 3.647   Mean   :2020-06-17   Mean   :2020-06-29   Mean   :2020-07-02  
 3rd Qu.:2020-09-30   3rd Qu.: 5.000   3rd Qu.:2020-09-01   3rd Qu.:2020-09-28   3rd Qu.:2020-09-30  
 Max.   :2020-12-31   Max.   :36.000   Max.   :2020-12-01   Max.   :2020-12-28   Max.   :2020-12-31  
> 

> str(df)
'data.frame':	1492 obs. of  5 variables:
 $ Індекс_дня: Date, format: "2020-07-18" "2020-07-19" "2020-07-20" ...
 $ Сеанси    : int  2 5 5 10 5 6 1 4 7 3 ...
 $ Month     : Date, format: "2020-07-01" "2020-07-01" "2020-07-01" ...
 $ Week      : Date, format: "2020-07-13" "2020-07-13" "2020-07-20" ...
 $ Day       : Date, format: "2020-07-18" "2020-07-19" "2020-07-20" ...
> 
```

Figure 8: The pre-process the data of the sessions of visiting the e-commerce site when the conversion

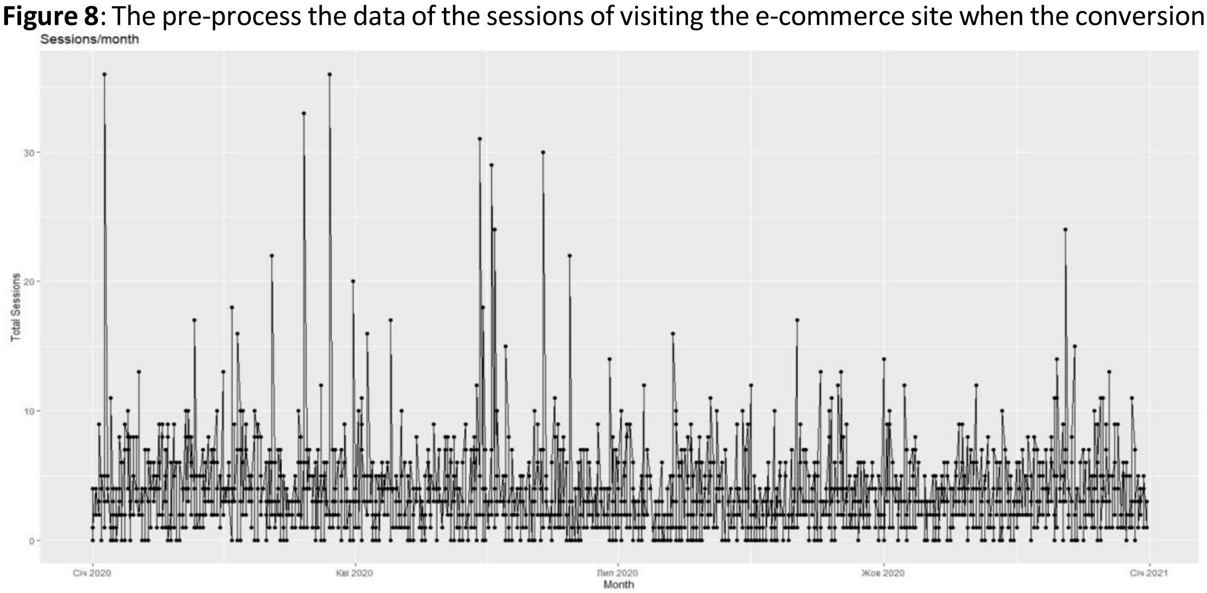

Figure 9: The graphic representation of obtained results (sessions/month)

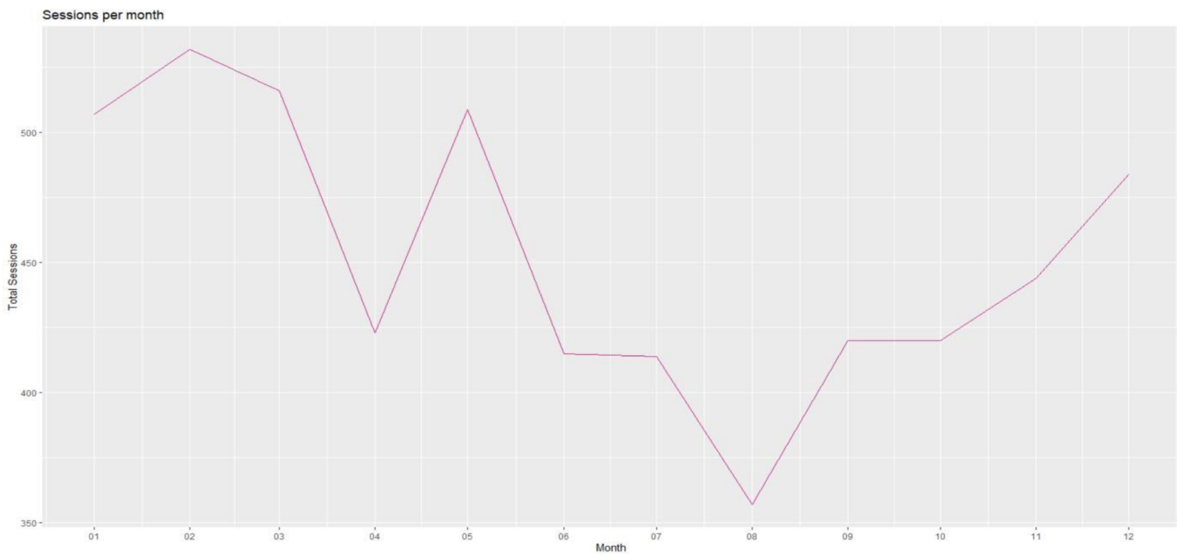

Figure 10: The graphic representation of obtained results (sessions per month)

Detailed analysis of sessions per weeks is more informative than sessions per month (Fig 11-12).

```
r2 <- ggplot(data = df, aes(x = df$Week, y = df$Сеанси)) +
         stat_summary(fun.y = sum,  geom = "line") +
         scale_x_date(labels = date_format("%m-%d"), breaks = "1 week",
                   guide = guide_axis(n.dodge=3)) +
          labs(x = "Day", y = "Total Sessions",  title = "Sessions per week")
```

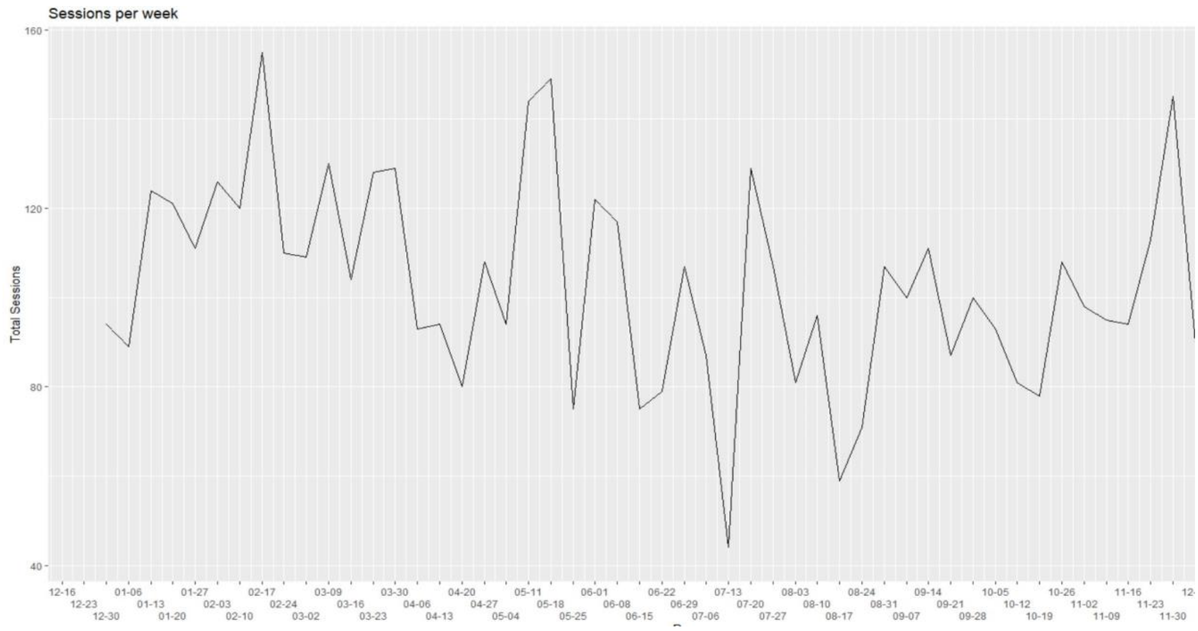

Figure 11: The graphic representation of obtained results (sessions per week)

```
data <- sessions_sum ggplot(data) + geom_point(aes(x=data$Week,
y=data$sum_precip)) +     geom_line(aes(x=data$Week, y=data$sum_precip),
colour = "gold") +   xlab("Month") + ylab("Num") +
   ggtitle("Sessions sum by month")
```

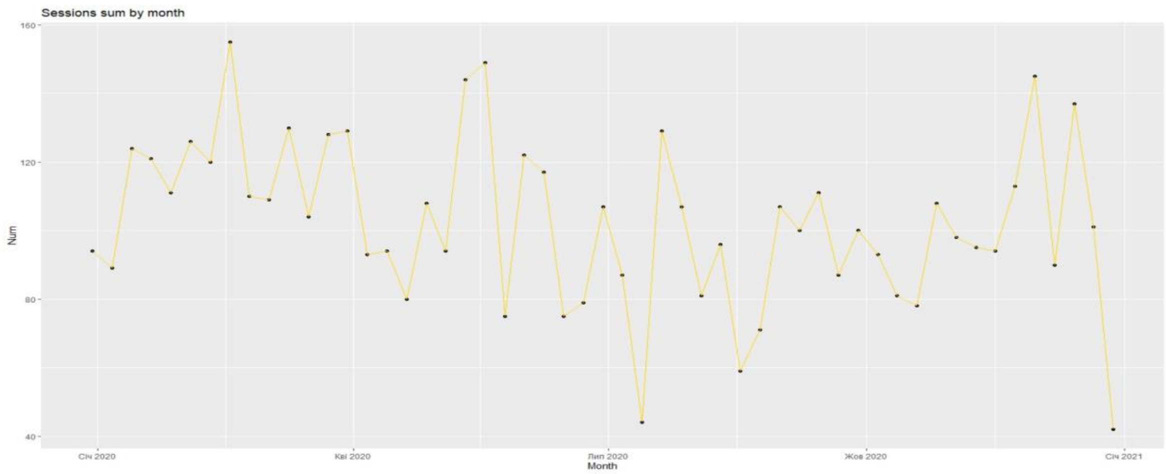

Figure 12: The graphic representation of obtained results (sessions sum by month)

Fig. 13-14 show descriptive statistics of visit sessions with successful conversion - quantitative characteristics of the data. Fig. 15 shows histogram of attendance sessions by successful conversion.

```
stat.desc(df$Індекс_дня)
describe(df$Сеанси)
hist <- r1 + geom_histogram(stat = "identity", colour = "purple")
```

```
> stat.desc(df$Индекс_дня)
                          x
nbr.val      1.492000e+03
nbr.null     0.000000e+00
nbr.na       0.000000e+00
min          1.826200e+04
max          1.862700e+04
range        3.650000e+02
sum          2.752001e+07
median       1.844700e+04
mean         1.844505e+04
SE.mean      2.719117e+00
CI.mean.0.95 5.333701e+00
var          1.103125e+04
std.dev      1.050297e+02
coef.var     5.694197e-03
```

Figure 13: The Descriptive statistics of visit sessions with successful conversion (by day index)

```
> describe(df$Сеанси)
   vars    n mean   sd median trimmed  mad min max range skew kurtosis   se
X1    1 1492 3.65 3.57      3    3.16 2.97   0  36    36 3.28    19.99 0.09
>
```

Figure 14: The Descriptive statistics of visit sessions with successful conversion (by day sessions)

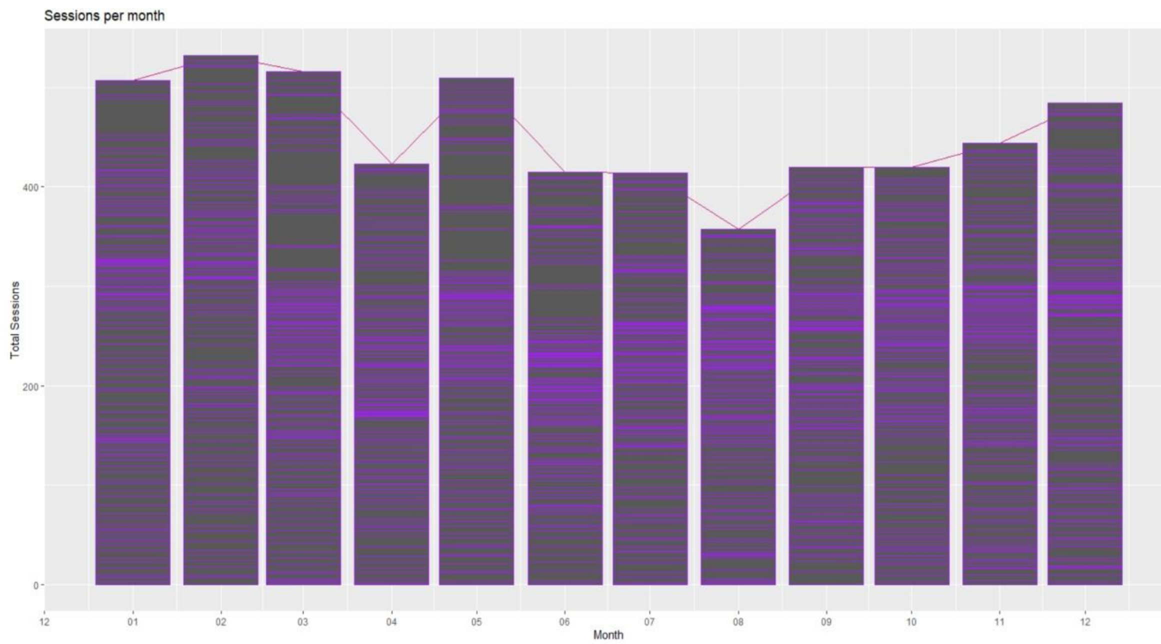

Figure 15: The histogram of attendance sessions by successful conversion (by sessions per month)

Fig. 16-18 show the graphic representation of cumulative attendance sessions

```
sessions_sum  <-  df  %>%      group_by(Month)  %>%      group_by(Day)  %>%
summarise(sum_precip = sum(Сеанси)) %>%   print(sum(df1$Сеанси))
 plot(ecdf(df[,"Сеанси"]),      col = "green")
 num.of.sessions  =  df1$Сеанси  breaks  =  seq(0,  36,  by=1)   num.cut  =
cut(num.of.sessions, breaks, right=FALSE)  num.freq = table(num.cut) freq =
seq(0, cumsum(num.freq)/max(cumsum(num.freq)), by=0.1)
#  кумулята   plot(breaks,  freq,  main="Sessions  frequency",  xlab="Num  of
sessions", ylab="Cumulative frequencies", xlim=c(0,36), ylim=c(0.0,1.0))
lines(breaks, freq)
  polar <- r2 + coord_polar()
```

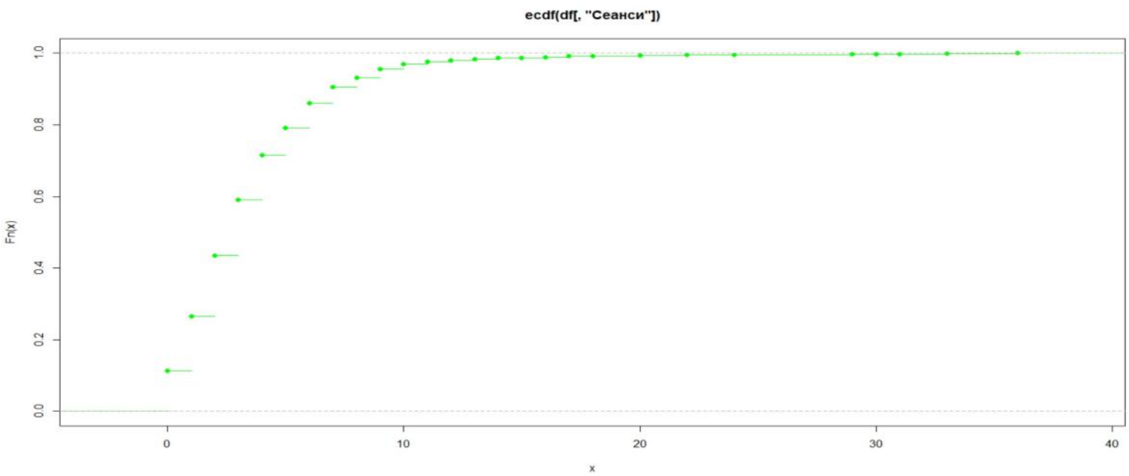

Figure 16: The graphic representation of cumulative attendance sessions

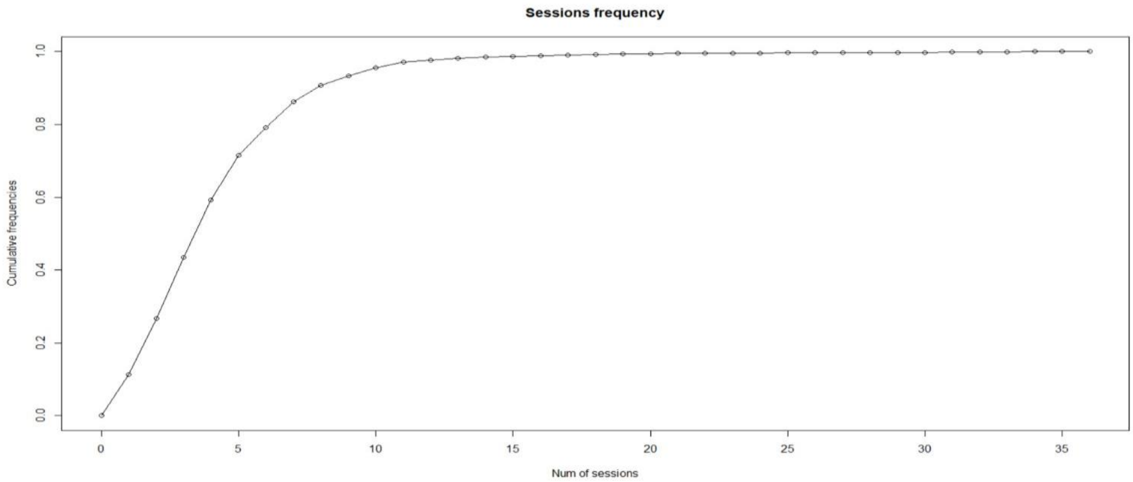

Figure 17: The graphic representation of cumulative attendance sessions frequency

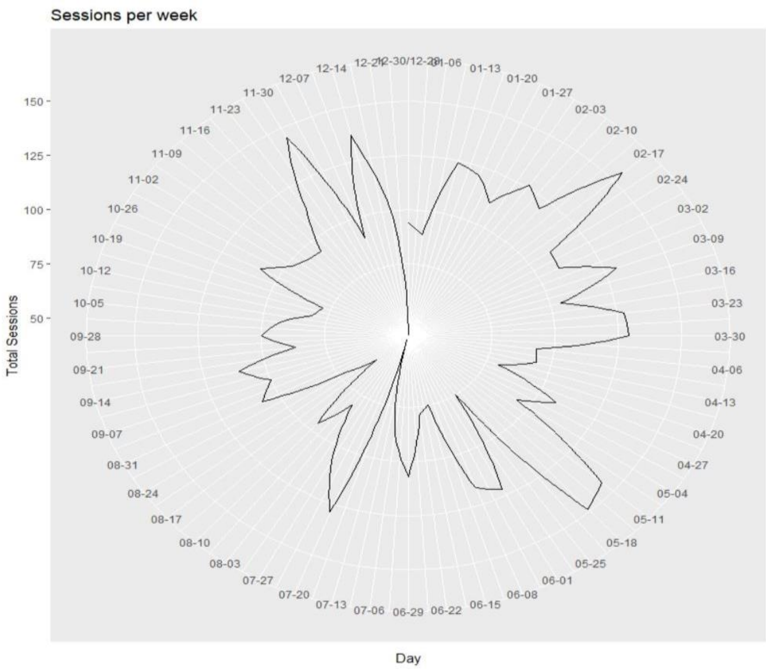

Figure 18: The graphic representation of cumulative visit sessions in polar coordinates (per week)

Fig. 19-27 show detection of the trend of the time series by smoothing methods.

```
data.ts <- ts(sessions_sum$sum_precip, frequency=12, start=c(01,1), end = 31) data.SMA3 <- SMA(data.ts,n=3) data.SMA5 <- SMA(data.ts,n=5) data.SMA7 <- SMA(data.ts,n=7) data.SMA9 <- SMA(data.ts,n=9) data.SMA11 <- SMA(data.ts,n=11) data.SMA13 <- SMA(data.ts,n=13)
data.SMA15 <- SMA(data.ts,n=15) plot.ts(data.ts,col='grey',xlab='Day',ylab = 'Sum of sessions') lines(data.SMA3, col='red') lines(data.SMA5, col='blue') lines(data.SMA7, col='yellow') lines(data.SMA9, col='purple') lines(data.SMA11, col='green') lines(data.SMA13, col='darkgoldenrod4') lines(data.SMA15, col='darkturquoise', lwd=3.5)
 legend(x = "topright",col=c("red","blue","yellow","purple","green",
                        "darkgoldenrod4","darkturquoise"),legend=c("SMA 3",
                        "SMA 5", "SMA 7", "SMA 9", "SMA 11", "SMA 13",
                        "SMA 15"), lty=1,cex=0.6,text.width = 2.2,ncol=3)
```

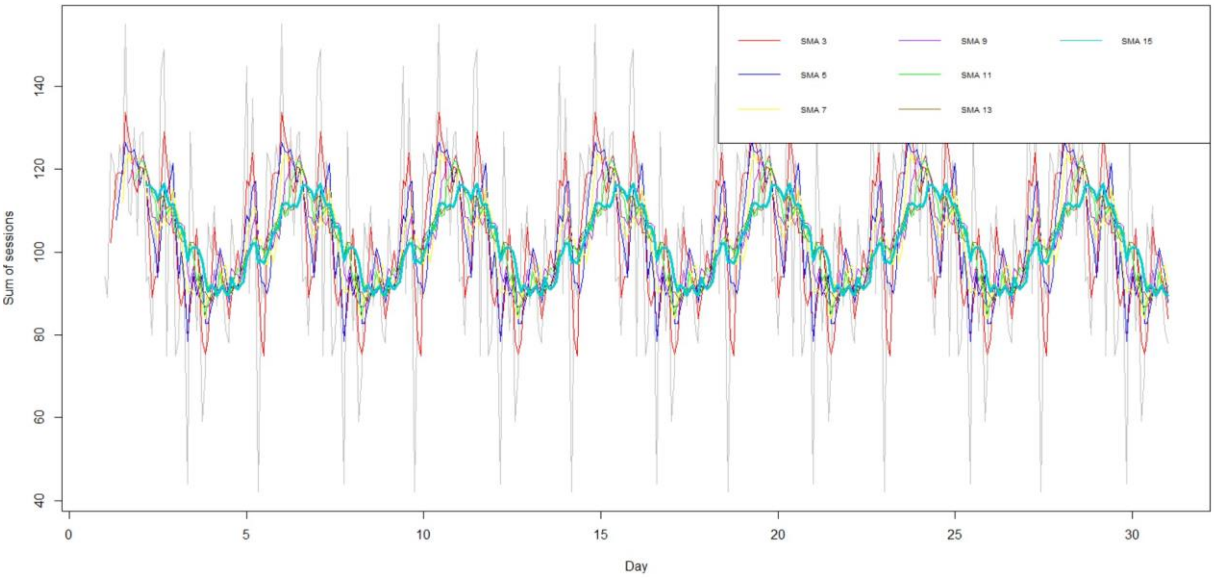

Figure 19: The graphic of detection of the trend of the time series by moving average method (SMA)

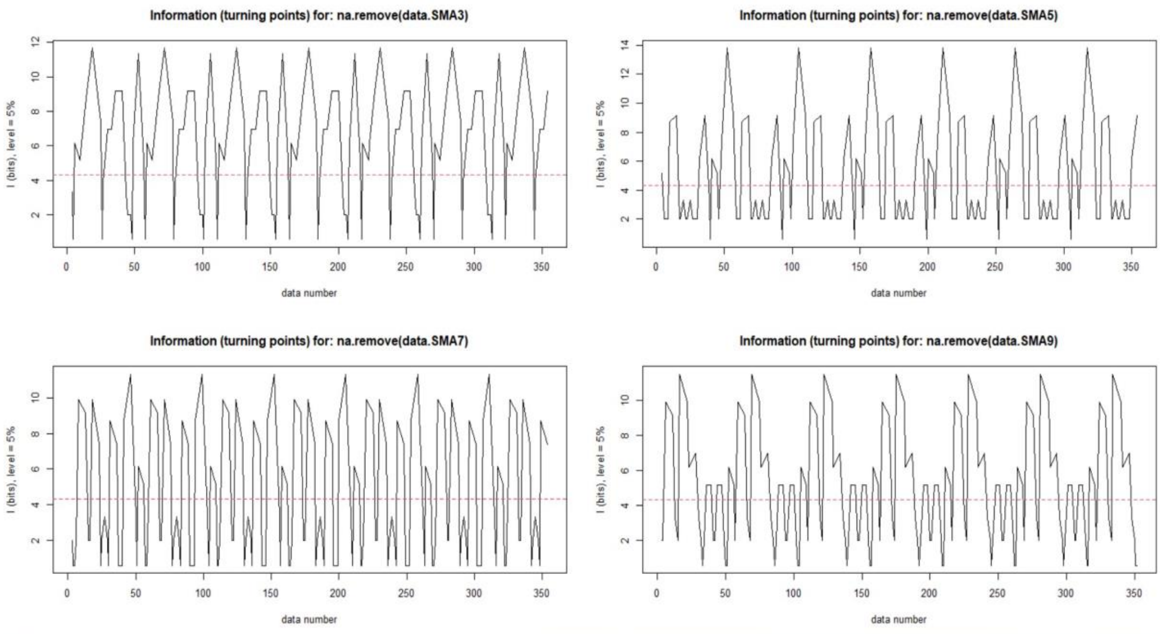

Figure 20: The graphic of trend detection of the time series by moving average method (SMA 3-9)

```
data.SMA3 <- SMA(data.ts,n=3) data.SMA5 <- SMA(data.SMA3,n=5) data.SMA7 <-
SMA(data.SMA5,n=7) data.SMA9 <- SMA(data.SMA7,n=9) data.SMA11 <-
SMA(data.SMA9,n=11) data.SMA13 <- SMA(data.SMA11,n=13) data.SMA15 <-
SMA(data.SMA13,n=15)
  tP1 <- turnpoints(na.remove(data.SMA3)) plot(tP1) tP2 <-
turnpoints(na.remove(data.SMA5)) plot(tP2) tP3 <-
turnpoints(na.remove(data.SMA7)) plot(tP3) tP4 <-
turnpoints(na.remove(data.SMA9)) plot(tP4) tP5 <-
turnpoints(na.remove(data.SMA11)) plot(tP5) tP6 <-
turnpoints(na.remove(data.SMA13)) plot(tP6) tP7 <-
turnpoints(na.remove(data.SMA15)) plot(tP7)
```

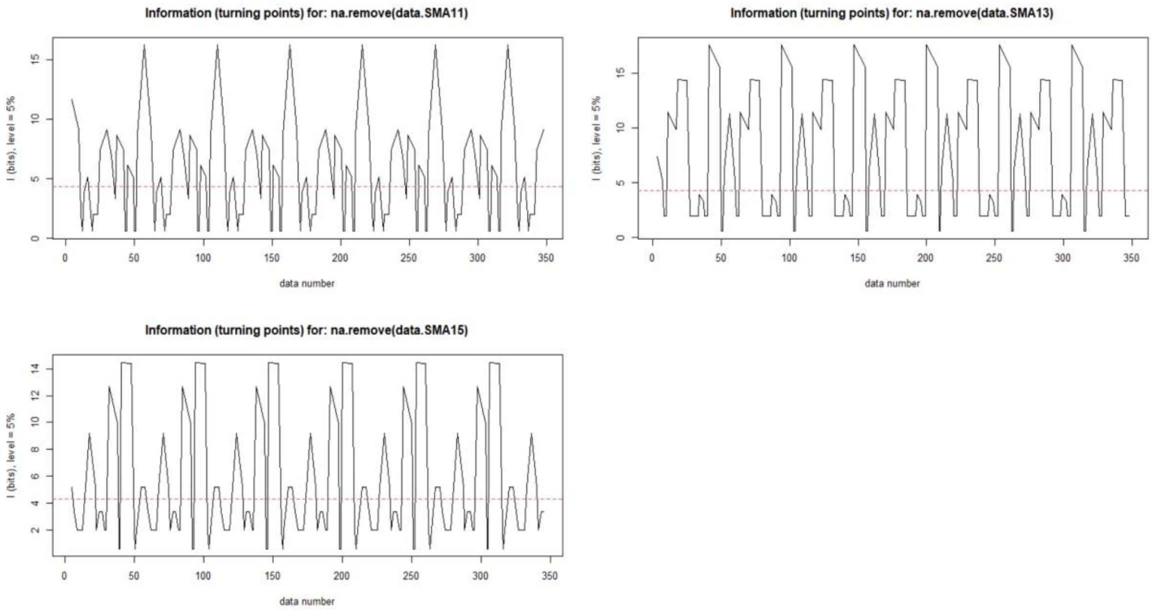

Figure 21: The graphic of trend detection of the time series by moving average method (SMA 11-15)

```
x <- list() corr1 <-
      ccf(data.ts,na.remove(data.SMA3)) x <- list.append(x, max(corr1$acf))
  corr2 <- ccf(data.SMA3,na.remove(data.SMA5)) x <-
      list.append(x, max(corr2$acf))
 par(mfrow=c(1,2)) plot(corr1, main = "data.ts & data.SMA3")
      plot(corr2, main = "data.SMA3 & data.SMA5")
  corr3 <- ccf(data.SMA5,na.remove(data.SMA7)) x
      <- list.append(x, max(corr3$acf))
 corr4 <- ccf(data.SMA7,na.remove(data.SMA9)) x
      <- list.append(x, max(corr4$acf))
  par(mfrow=c(1,2)) plot(corr3, main = "data.SMA5 & data.SMA7")
      plot(corr4, main = "data.SMA7 & data.SMA9")
corr5 <- ccf(data.SMA9,na.remove(data.SMA11)) x <-
      list.append(x, max(corr5$acf))
  corr6 <- ccf(data.SMA11,na.remove(data.SMA13)) x <-
      list.append(x, max(corr6$acf))
par(mfrow=c(1,2)) plot(corr5, main = "data.SMA9 & data.SMA11")
      plot(corr6, main = "data.SMA11 & data.SMA13")
  corr7 <- ccf(data.SMA13,na.remove(data.SMA15)) x <-
      list.append(x, max(corr7$acf))
  corr8 <- ccf(data.ts,na.remove(data.SMA15)) x <-
      list.append(x, max(corr8$acf))
  par(mfrow=c(1,2)) plot(corr7, main = "data.SMA13 & data.SMA15")
      plot(corr8, main = "data.ts & data.SMA15")
 par(mfrow=c(1,1))
plot(x,c(1:15),xlab='Num of SMA',ylab='Corr',main='Corr coeff')
```

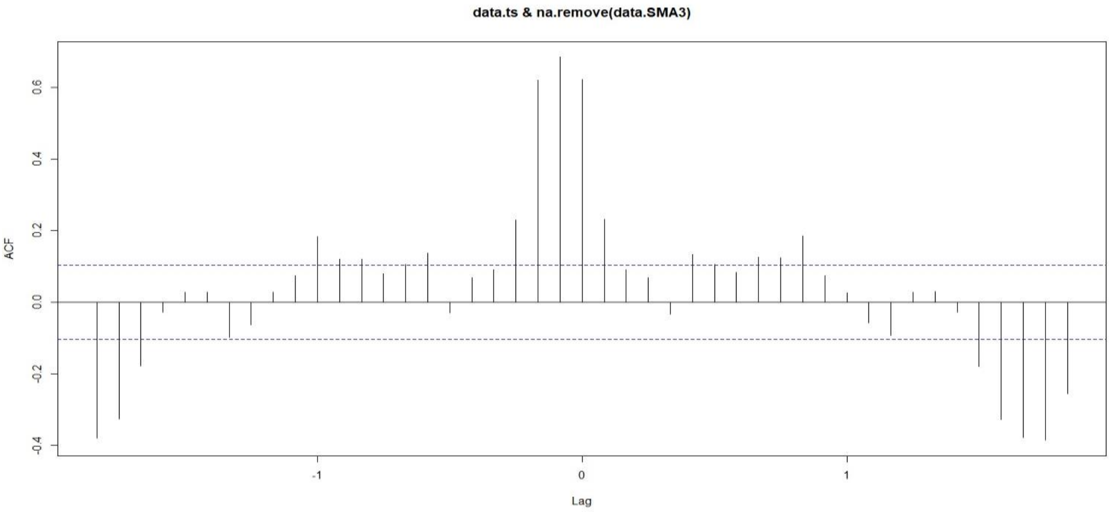

Figure 22: The graphic of trend detection of the time series by moving average method (SMA data/3)

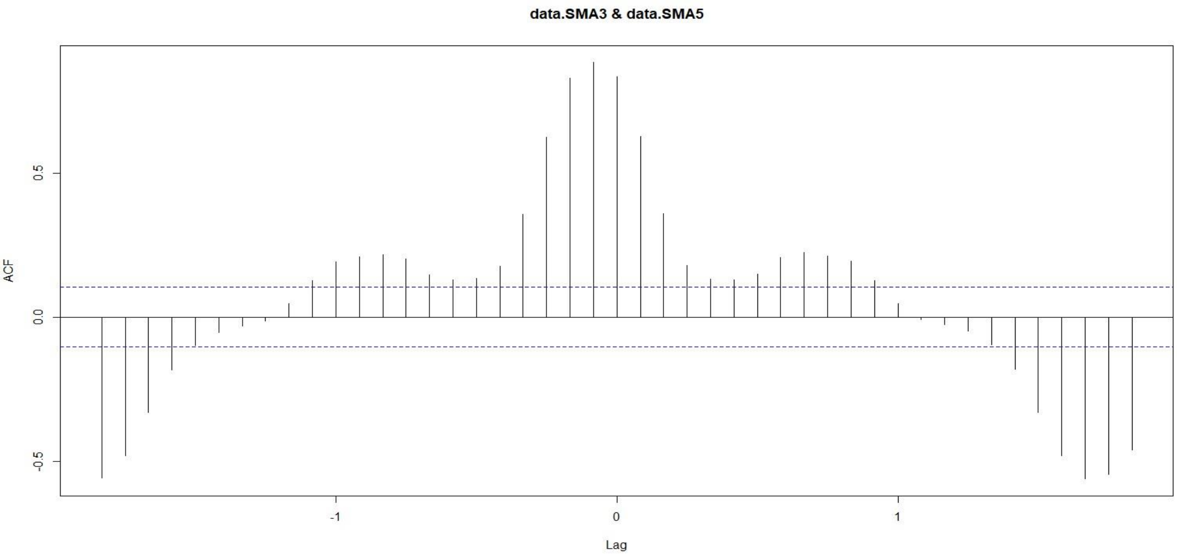

Figure 23: The graphic of trend detection of the time series by moving average method (SMA 3/5)

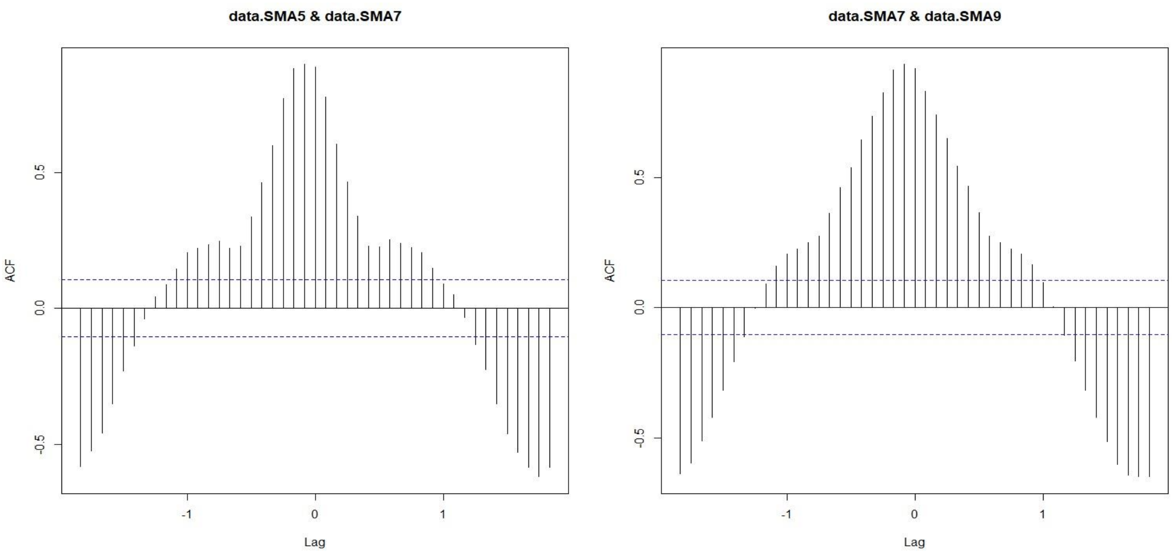

Figure 24: The graphic of trend detection of the time series by moving average method (SMA 5/7/9)

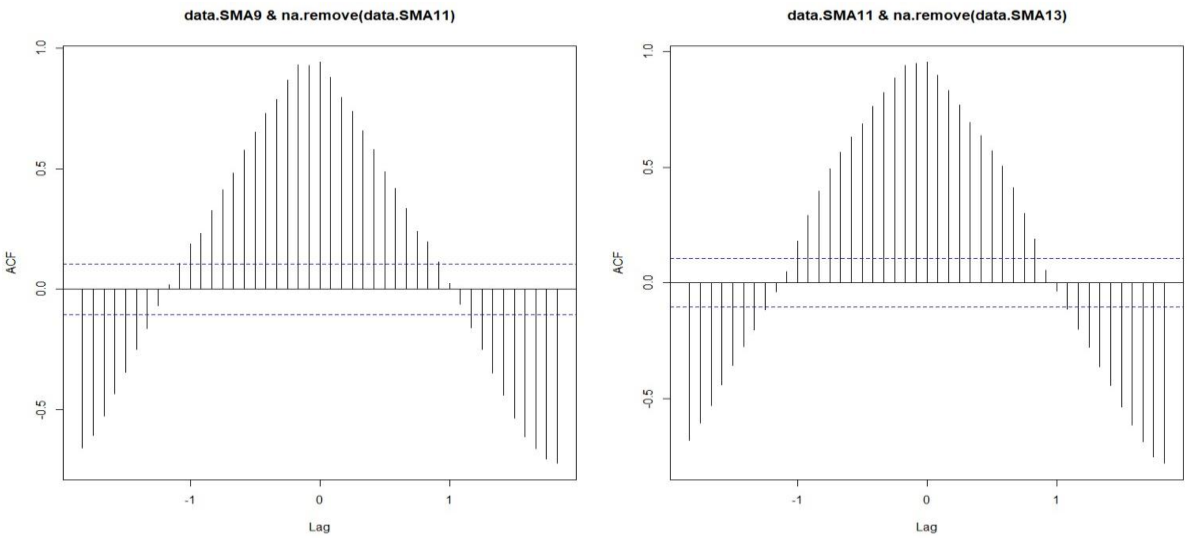

Figure 25: Graphic of trend detection of the time series by moving average method (SMA 9/11/13)

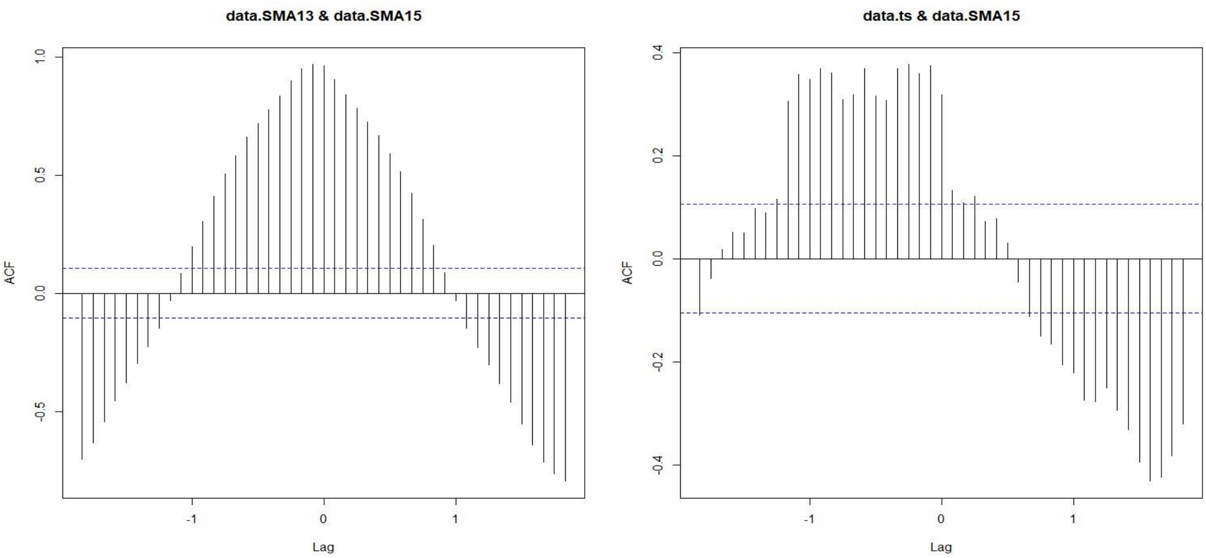

Figure 26: Graphic of trend detection of the time series by moving average method (SMA 13/15/data)

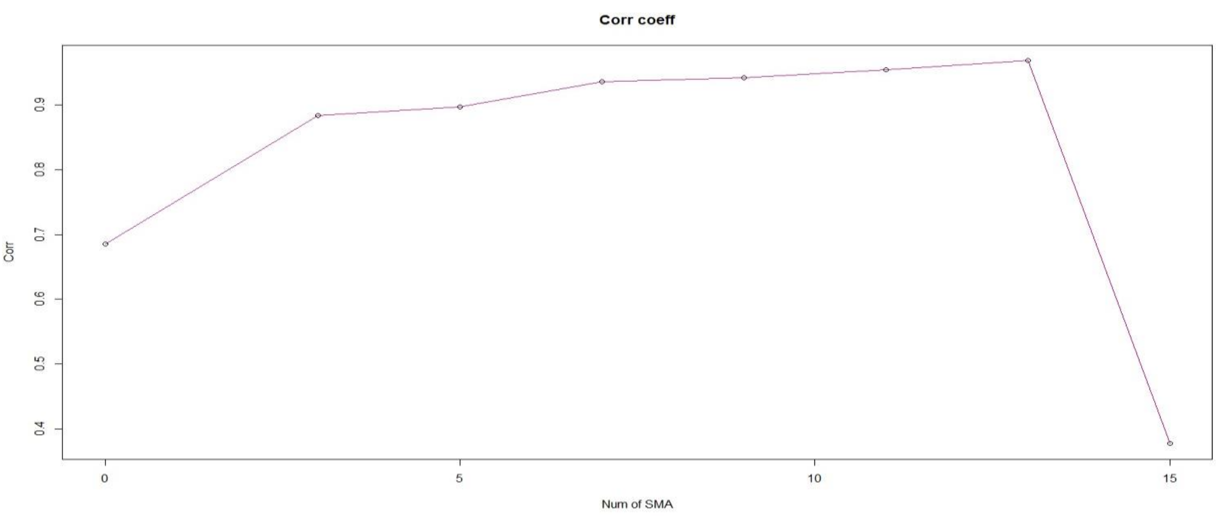

Figure 27: The detection of the trend of the time series by moving the average method of the correction factor

Fig. 28 shows detection of the trend of the time series by the weighted moving average method.

```
ma(data.ts, order = 12, centre = TRUE) autoplot(data.ts, series = "Data") + autolayer(ma(data.ts, order = 12, centre = T), series = "normal") + ggtitle("Daily Sessions") + labs(x = NULL, y = "Sessions")
```

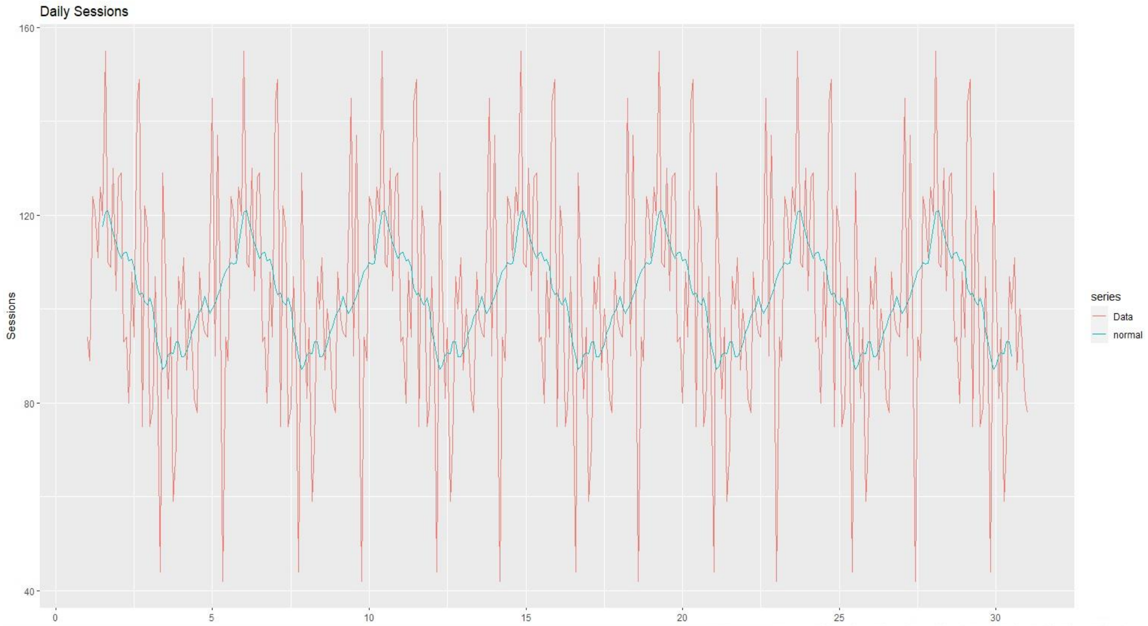

Figure 28: The graphic of trend detection of the time series by weighted moving average method

Fig. 29 shows detection of the time series trend by linear smoothing when w = 3; 5.

```
plot(sma(sessions_sum$sum_precip)) plot(sessions_sum$Day, sessions_sum$sum_precip, main='Simple Moving Average (SMA)') lines(sessions_sum$Day, sessions_sum$sum_precip) lines(sessions_sum$Day, rollmean(sessions_sum$sum_precip,3,fill = NA), col='blue') lines(sessions_sum$Day, rollmean(sessions_sum$sum_precip,5,fill = NA),col='red',lwd=2.5) legend(x = 'topright',col=c('black','blue', 'red'),legend=c('Raw', 'SMA 3', 'SMA 5'),lty=1,cex=0.8)
```

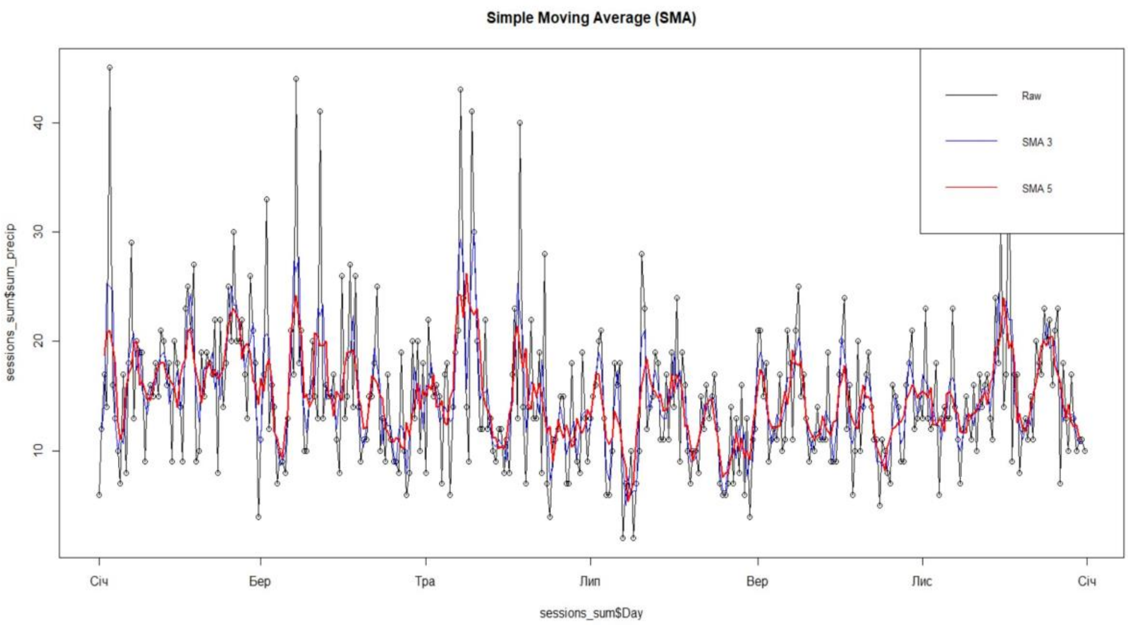

Figure 29: The graphic of detection of the trend of the time series by linear smoothing when w = 3; 5

Fig. 30 shows detection of the trend of the time series by smoothing Kendall formulas.

```
sm <- ksmooth(format(df$Day,'%d'), df$Сеанси, kernel = "normal", bandwidth
= 0.5) plot(sm,xlab='day',ylab='num')
lines(sm,col='blue',lwd=2.5)
```

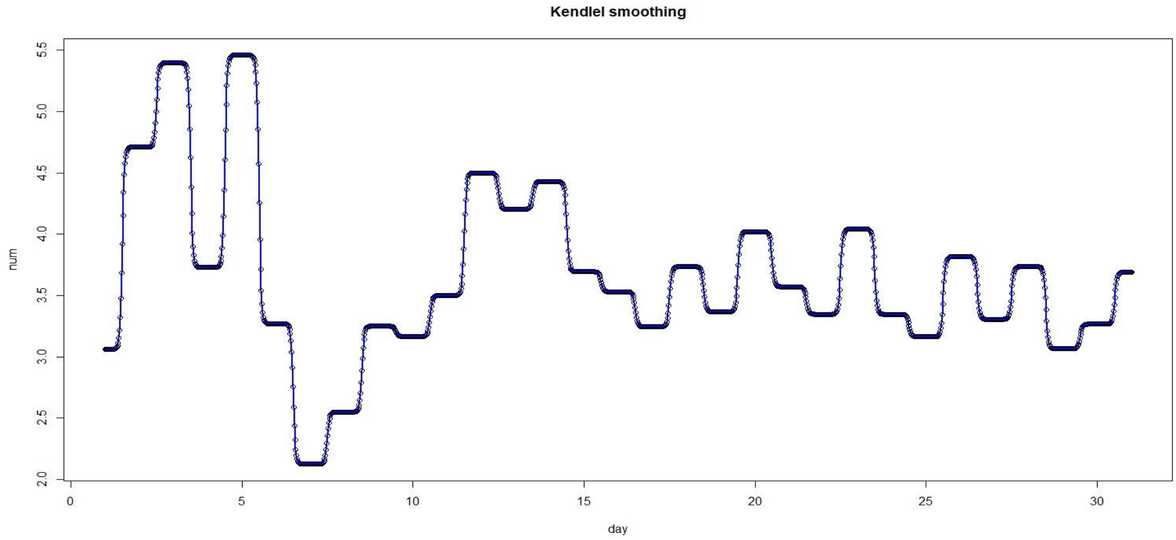

Figure 30: The graphic of detection of the trend of the time series by smoothing Kendall formulas

Fig. 31-32 show detection of the trend of the time series by correlation analysis of time sequences.

```
t1 <- sample(1:nrow(df.cl_an),size = ceiling(0.33*nrow(df.cl_an)),replace =
FALSE)          t2          <-          sample(1:nrow(df.cl_an),size          =
ceiling(0.33*nrow(df.cl_an)),replace = FALSE)
t3 <- sample(1:nrow(df.cl_an),size = ceiling(0.33*nrow(df.cl_an)),replace =
FALSE)   df1  <-  df.cl_an[t1,c(2:7)]  df2  <-  df.cl_an[t2,c(2:7)]  df3  <-
df.cl_an[t3,c(2:7)] res1 <- cor(df1) res2 <- cor(df2) res3 <- cor(df3)
   round(res1, 2)
```

	Сеанси	Відсоток_нових_сеансів	Нові_користувачі	Показник_відмов	Сторінок_за_сеанс	Сер_тривалість_сеансу
Сеанси	1.00	-0.34	0.98	-0.20	0.35	0.29
Відсоток_нових_сеансів	-0.34	1.00	-0.27	0.24	-0.37	-0.35
Нові_користувачі	0.98	-0.27	1.00	-0.20	0.35	0.29
Показник_відмов	-0.20	0.24	-0.20	1.00	-0.36	-0.40
Сторінок_за_сеанс	0.35	-0.37	0.35	-0.36	1.00	0.61
Сер_тривалість_сеансу	0.29	-0.35	0.29	-0.40	0.61	1.00

	Сеанси	Відсоток_нових_сеансів	Нові_користувачі	Показник_відмов	Сторінок_за_сеанс	Сер_тривалість_сеансу
Сеанси	1.0000000	-0.3698150	0.9432800	-0.1918684	0.4112601	0.4653364
Відсоток_нових_сеансів	-0.3698150	1.0000000	-0.2234924	0.3401848	-0.5406602	-0.3675399
Нові_користувачі	0.9432800	-0.2234924	1.0000000	-0.1645100	0.3825443	0.4367804
Показник_відмов	-0.1918684	0.3401848	-0.1645100	1.0000000	-0.3615870	-0.4045469
Сторінок_за_сеанс	0.4112601	-0.5406602	0.3825443	-0.3615870	1.0000000	0.6106418
Сер_тривалість_сеансу	0.4653364	-0.3675399	0.4367804	-0.4045469	0.6106418	1.0000000

Figure 31: Graphic of trend detection by correlation analysis of time sequences (Correlation matrix)

```
coll <- colorRampPalette(c("blue", "white", "red"))(60) heatmap(x = res1,
col = coll, symm = TRUE, margins = c(7,7), scale="column",        cexCol=0.7,
cexRow=0.7)   legend(x="bottomright",   legend=c("min",   "ave",   "max"),
fill=colorRampPalette(c("blue", "white", "red"))(3))
```

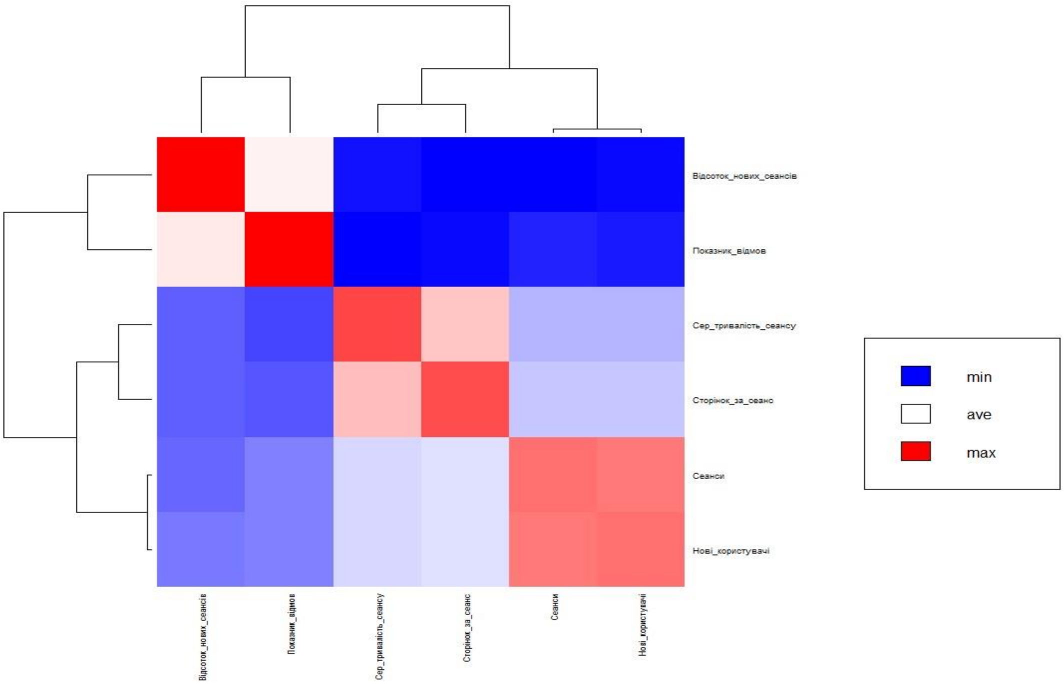

Figure 32: Graphic of trend detection by correlation analysis of time sequences (Correlation matrix)

Fig. 33 shows detection of the trend of the time series by Autocorrelation analysis of time sequences.

```
acf(data.ts, lag.max = 10, plot = FALSE) acf(na.remove(data.SMA15))
```

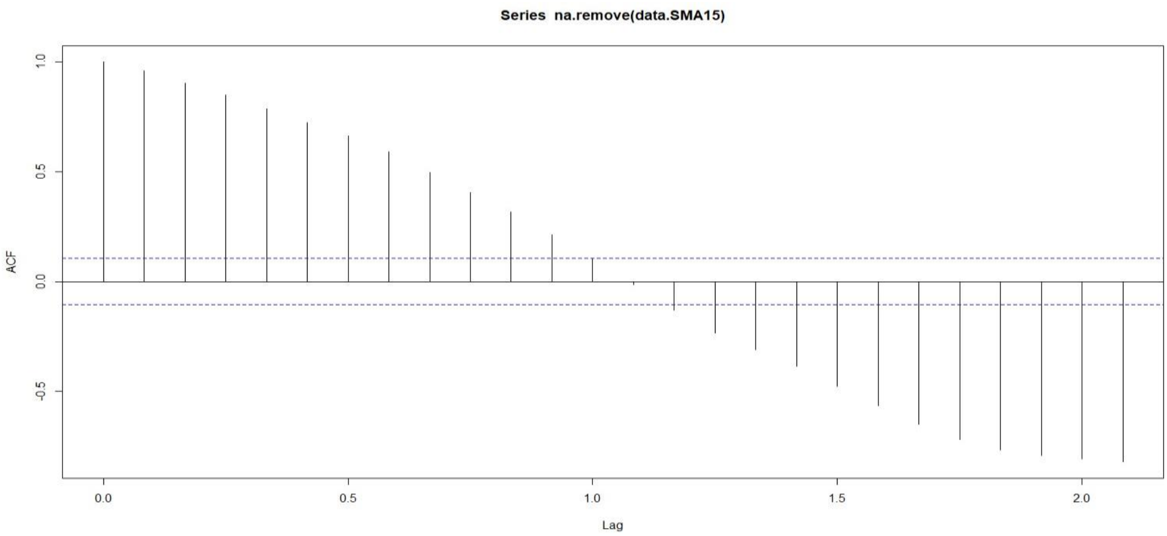

Figure 33: Graphic of trend detection by correlation analysis of time sequences (Autocorrelation)

Fig. 34 shows detection of the trend of the time series by exponential smoothing.

```
data.ts <- ts(sessions_sum$sum_precip, frequency=12, start=c(01,1), end = 31) data.es0.1 <- ses(data.ts, h=11, alpha=0.1, initial="simple") data.es0.2 <- ses(data.ts, h=11, alpha=0.2, initial="simple") data.es0.3 <- ses(data.ts, h=11, alpha=0.3, initial="simple") data.es0.15 <- ses(data.ts, h=11, alpha=0.15, initial="simple") data.es0.25 <- ses(data.ts, h=11, alpha=0.25, initial="simple")
 plot.ts(data.ts,col='grey',xlab='Day',ylab = 'Sum of sessions') plot(data.es0.1, col='darkturquoise', main='e-smoothing, alpha 0.1') plot(data.es0.2, col='blue', main='e-smoothing, alpha 0.2') plot(data.es0.3, col='yellow', main='e-smoothing, alpha 0.3') plot(data.es0.15, col='purple', main='e-smoothing, alpha 0.15') plot(data.es0.25, col='green', lwd=1.5, main='e-smoothing, alpha 0.25')
```

```
 par(mfrow=c(1,2))
plot(data.ts,   ylab='num',main='e-smoothing,   orig   vs   alpha   0.1')
lines(data.es0.1$fitted,                                  col='blue',lwd=2.5)
legend(x="bottomleft",col=c('blue','black'),legend=c('Fitted',
'Original'),lty=1,cex=0.8,bty="o")
 plot(data.es0.3$fitted, ylab='num',main='e-smoothing, alpha 0.3 vs alpha
0.25')           lines(data.es0.25$fitted,           col='green',lwd=2.1)
legend(x="bottomright",col=c('green','grey'),legend=c('Fitted',
'Original'),lty=1,cex=0.8,bty="o")
  corr10   <-   ccf(data.ts,na.remove(data.SMA15))   x   <-   list.append(x,
max(corr8$acf))
  par(mfrow=c(1,2))  plot(corr7,  main  =  "data.SMA13  &  data.SMA15")
plot(corr8, main = "data.ts & data.SMA15")
```

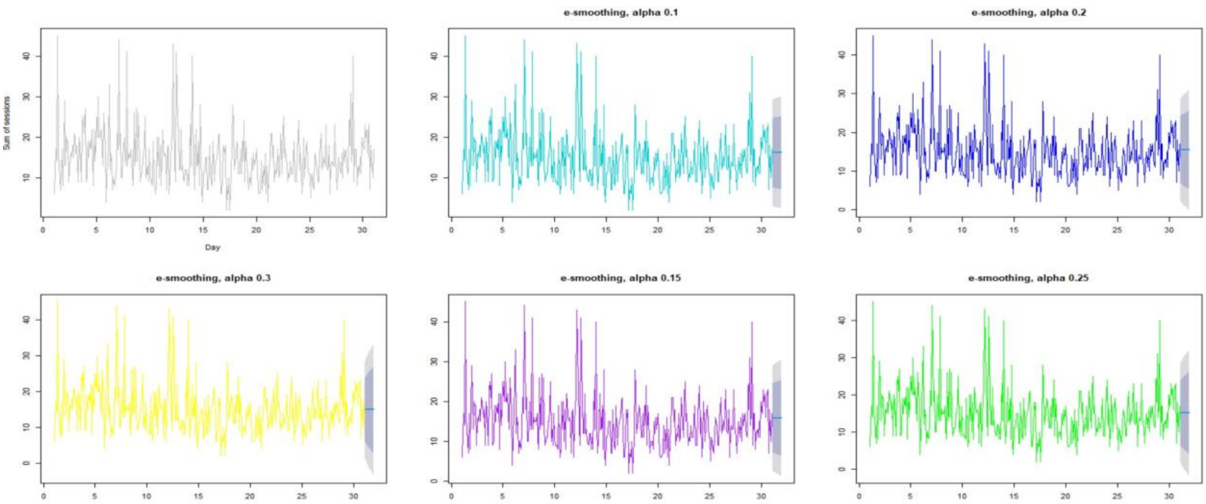

Figure 34: The graphic of detection of the trend of the time series by Exponential smoothing

Fig. 35 shows detection of the trend of the time series by Median smoothing.

```
plot(smooth(data.ts, kind = "3",       twiceit = FALSE, endrule = c("Tukey",
"copy"), do.ends = FALSE), ylab='num')
```

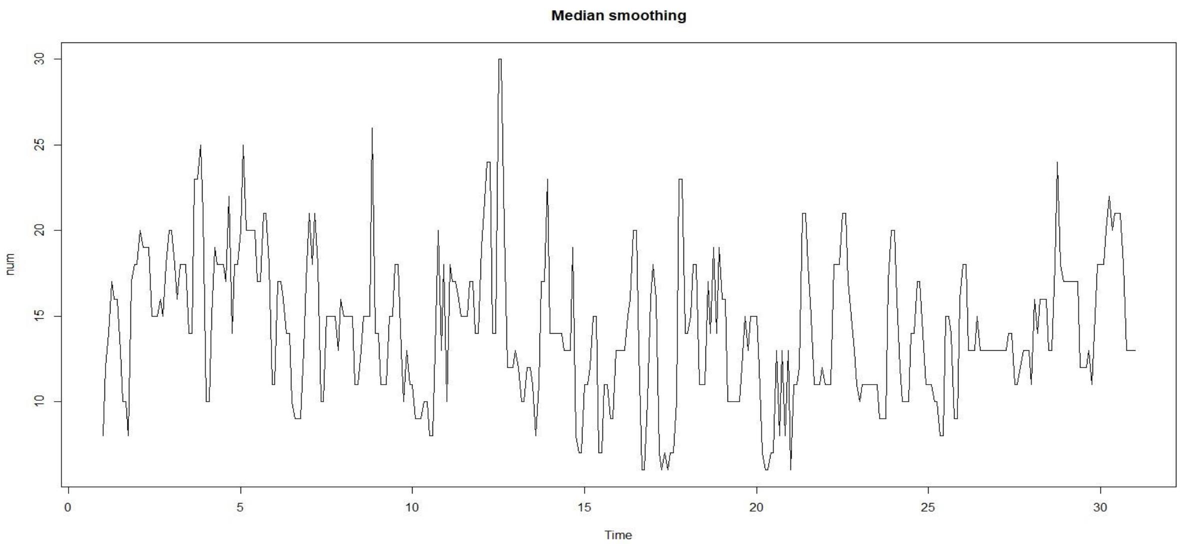

Figure 35: The graphic of detection of the trend of the time series by Median smoothing

Fig. 36 shows detection of the time series trend by hierarchical agglomerative cluster analysis of multidimensional data.

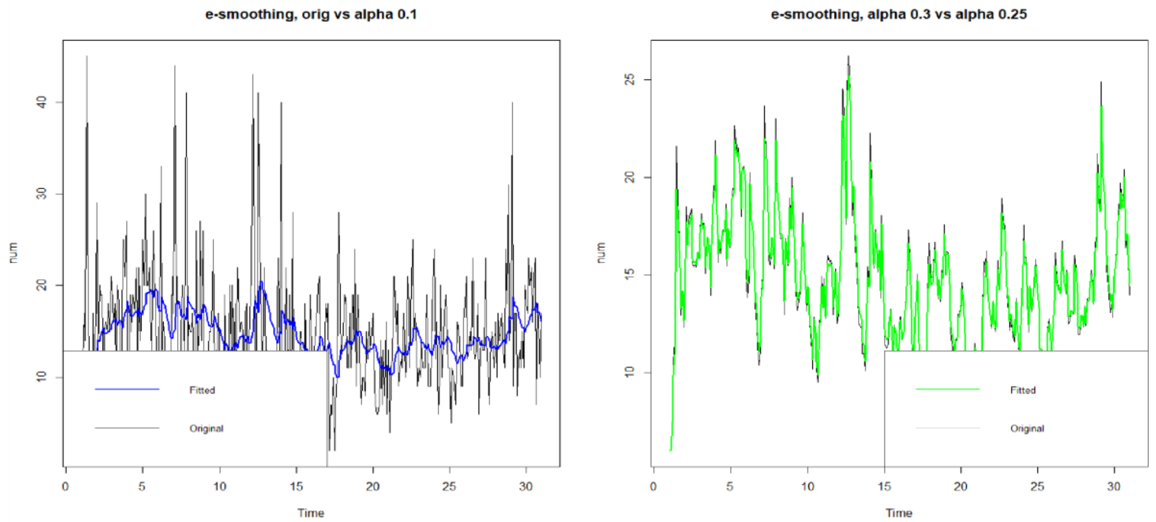

Figure 36: The graphic of detection of the trend of the time series by hierarchical agglomerative cluster analysis of multidimensional data

Fig. 37-38 show construction of the proximity matrix through the rationing of the table "object-property" (first 5 records).

```
array = data.values
# separate array into input and output components
X = array[:,1:7] Y = array[:,0] scaler = Normalizer().fit(X) normalizedX =
scaler.transform(X)        #        summarize        transformed         data
numpy.set_printoptions(precision=3)                         print('Сеанси',
'Відсоток_нових_сеансів', 'Нові_користувачі',
       'Показник_відмов',   'Сторінок_за_сеанс',   'Сер_тривалість_сеансу')
print(normalizedX[:10,:])
```

	Місто	Сеанси	Відсоток_нових_сеансів	Нові_користувачі	Показник_відмов	Сторінок_за_сеанс	Сер_тривалість_сеансу
1	Chengdu	2891	0.3608	1043	0.3023	6.95	270
2	Nuremberg	1051	0.5452	573	0.3368	4.32	315
3	Gloucester	296	0.9966	295	0.8412	1.15	26
4	Nastola	202	0.3911	79	0.2673	4.25	184
5	Santana de Parnaiba	106	0.6509	69	0.4528	4.01	166

Figure 37: The graphic representation of the construction of the proximity matrix through the rationing of the table "object-property" (first five records)

```
Сеанси Відсоток_нових_сеансів Нові_користувачі Показник_відмов Сторінок_за_сеанс Сер_тривалість_сеансу
[[9.370e-01 1.169e-04 3.381e-01 9.798e-05 2.253e-03 8.751e-02]
 [8.491e-01 4.405e-04 4.629e-01 2.721e-04 3.490e-03 2.545e-01]
 [7.069e-01 2.380e-03 7.045e-01 2.009e-03 2.747e-03 6.210e-02]
 [7.101e-01 1.375e-03 2.777e-01 9.397e-04 1.494e-02 6.468e-01]
 [5.078e-01 3.118e-03 3.306e-01 2.169e-03 1.921e-02 7.953e-01]
 [2.708e-01 1.721e-03 1.480e-01 1.648e-03 1.795e-02 9.510e-01]
 [3.481e-01 2.606e-03 1.929e-01 2.288e-03 1.844e-02 9.172e-01]
 [3.439e-01 4.643e-03 2.739e-01 2.865e-03 3.287e-02 8.976e-01]
 [2.068e-01 1.325e-03 7.025e-02 1.178e-03 1.717e-02 9.757e-01]
 [3.994e-01 9.043e-03 3.346e-01 4.960e-03 3.497e-02 8.528e-01]]
```

Figure 38: The graphic representation of the construction of the proximity matrix through the rationing of the table "object-property"

Fig. 39-41 show conducting agglomerative hierarchical cluster analysis.

```
classifiers = [] model1 = xgboost.XGBClassifier() classifiers.append(model1)
model2      =      svm.SVC()      classifiers.append(model2)      model3      =
tree.DecisionTreeClassifier()     classifiers.append(model3)     model4      =
RandomForestClassifier() classifiers.append(model4)
 for clf in classifiers:
```

```
  clf.fit(X_train,  y_train)      y_pred  =  clf.predict(X_test)      acc  =
accuracy_score(y_test, y_pred)    print("Accuracy of {} is {}".format(clf,
acc))   cm = confusion_matrix(y_test, y_pred)
  print("Confusion Matrix of {} is {}".format(clf, cm))
```

```
Accuracy of XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0,
              learning_rate=0.1, max_delta_step=0, max_depth=3,
              min_child_weight=1, missing=None, n_estimators=100, n_jobs=1,
              nthread=None, objective='multi:softprob', random_state=0,
              reg_alpha=0, reg_lambda=1, scale_pos_weight=1, seed=None,
              silent=None, subsample=1, verbosity=1) is 0.03389830508474576
 Accuracy of RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                       criterion='gini', max_depth=None, max_features='auto',
                       max_leaf_nodes=None, max_samples=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, n_estimators=100,
                       n_jobs=None, oob_score=False, random_state=None,
                       verbose=0, warm_start=False) is 0.06779661016949153
 Accuracy of DecisionTreeClassifier(ccp_alpha=0.0, class_weight=None, criterion='gini',
                       max_depth=None, max_features=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort='deprecated',
                       random_state=None, splitter='best') is 0.07627118644067797
```

Figure 39: The graphic representation of choice of integration strategies

Fig. 40 show conducting cluster analysis

```
data  =  pd.read_csv("/content/AnyConv.com__V10.csv", encoding = "ISO-8859-
1")  data.index  =  data.index  +  1  data.columns  =  ["City",  "Sessions",
"New_Session_Percentage", "New_Users", "Bounce_Identity", "Session_Pages",
"Session_Larrow"]   x   =   data   [["Sessions",   "New_Session_Recent"   ,
"Session_Pages", "Session_Series"]] y = data["City "] data[' City '] = data['
City'].astype('category')
X_train,  X_test,  y_train,  y_test  =  train_test_split(x,y,test_size=0.3,
shuffle=True)
  tree.DecisionTreeClassifier().fit(X_train,     y_train)     y_pred     =
clf.predict(X_test)
acc = accuracy_score(y_test, y_pred) * 100 print("\nAccuracy of model is
{}%".format(acc))
X_test.index = range(1,119)
X_test = X_test.dropna(axis=0, how='all') X_test = X_test.iloc[0:30] fig =
ff.create_dendrogram(X_test,      orientation='left',      labels=y_pred)
fig.update_layout(width=800, height=500) fig.show()
confusion_matrix(y_test, y_pred)
```

```
array([[2, 1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [1, 2, 0, 0, 6, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 1, 0, 0, 3, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0],
       [0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 1, 0, 1, 7, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 2, 0, 0, 4, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 1, 0, 0, 2, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 1, 0, 0, 8, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 2, 0, 0, 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 2, 0, 0, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 1, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 3, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0],
       [0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0],
       [0, 0, 0, 0, 4, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0],
       [0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 4, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1]])
```

Figure 40: The graphic representation of conducting cluster analysis

Fig. 41. shows the construction of dendrograms.

```
Accuracy of model is 22.033898305084744%
```

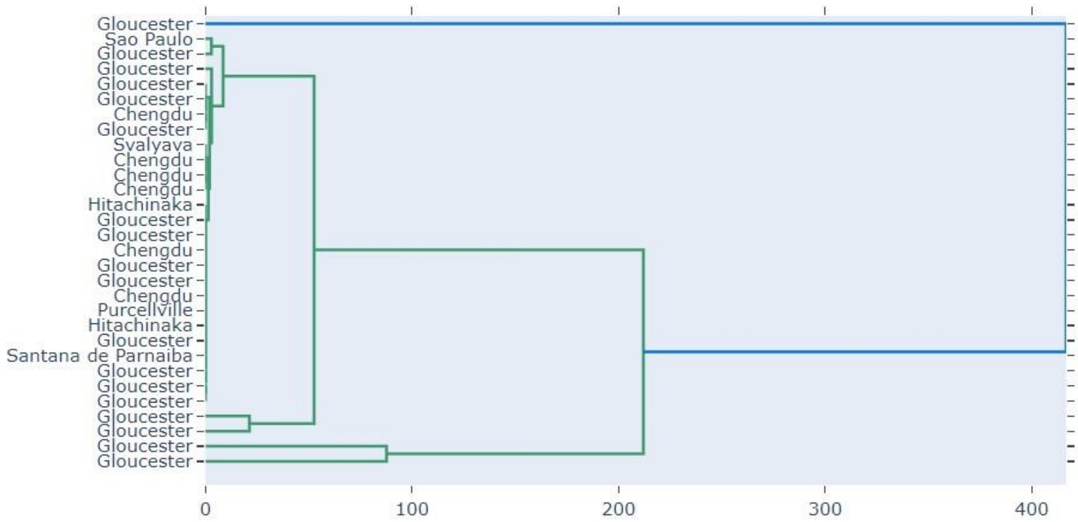

Figure 41: The graphic representation of the construction of dendrograms

5. Conclusions and prospects for further scientific exploration

E-commerce-related legal issues can profoundly affect its development, but they are still waiting to be resolved. They relate to privacy, e-commerce taxation and control over the export of cryptographic technology and legal, financial relationships between participants in the virtual e-commerce world. No matter how the company works, electronic transformation is no longer a matter of choice. It is already a necessity for successful business development.

6. References

[1] V. Vysotska, R. Hasko, V. Kuchkovskiy, Process analysis in electronic content commerce system, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2015, pp. 120-123.

[2] V. Lytvyn, V. Vysotska, Designing architecture of electronic content commerce system, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2015, pp. 115-119.

[3] V. Vysotska, L. Chyrun, L. Chyrun, Information Technology of Processing Information Resources in Electronic Content Commerce Systems, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2016, pp. 212-222.

[4] V. Vysotska, I. Rishnyak, L. Chyrun, Analysis and evaluation of risks in electronic commerce, in: Proceedings of the 9th International Conference of CAD Systems in Microelectronics, 2007, pp. 332-333.

[5] V. Vysotska, L. Chyrun, Methods of information resources processing in electronic content commerce systems, in: Proceedings of 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015-February, 2015.

[6] V. Andrunyk, L. Chyrun, V. Vysotska, Electronic content commerce system development, in: Proceedings of 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015-February, 2015.

[7] K. Alieksieieva, A. Berko, V. Vysotska, Technology of commercial web-resource processing, in: Proceedings of 13th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2015-February, 2015.
[8] V. Vysotska, L. Chyrun, L. Chyrun, The Commercial Content Digest Formation and Distributional Process, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2016, pp. 186-189.
[9] V. Lytvyn, V. Vysotska, V. Shatskykh, I. Kohut, O. Petruchenko, L. Dzyubyk, V. Bobrivetc, V. Panasyuk, S. Sachenko, M. Komar, Design of a recommendation system based on Collaborative Filtering and machine learning considering personal needs of the user, volume 4(2-100) of Eastern-European Journal of Enterprise Technologies, 2019, pp. 6-28.
[10] P. Pukach, K. Shakhovska, The Mathematical Method Development of Decisions Supporting Concerning Products Placement Based on Analysis of Market Basket Content, in: Proceedings of the Experience of Designing and Application of CAD Systems in Microelectronics, CADSM, 2017, pp. 347-350.
[11] V. Vysotska, V.B. Fernandes, M. Emmerich, Web content support method in electronic business systems, volume Vol-2136 of CEUR Workshop Proceedings, 2018, pp. 20-41.
[12] S. Orekhov, H. Malyhon, I. Liutenko, T. Goncharenko, Using Internet News Flows as Marketing Data Component, volume Vol-2604 of CEUR workshop proceedings, 2020, pp. 358-373.
[13] R. Yurynets, Z. Yurynets, I. Myshchyshyn, N. Zhyhaylo, A. Pekhnyk, Optimal Strategy for the Development of Insurance Business Structures in a Competitive Environment, volume Vol-2631 of CEUR Workshop Proceedings, 2020, pp. 79-94.
[14] R. Yurynets, Z. Yurynets, M. Kokhan, Econometric Analysis of the Impact of Expert Assessments on the Business Activity in the Context of Investment and Innovation Development, volume Vol-2604 of CEUR workshop proceedings, 2020, pp. 680-694.
[15] H. Lipyanina, A. Sachenko, T. Lendyuk, S. Nadvynychny, S. Grodskyi, Decision tree based targeting model of customer interaction with business page, volume Vol-2608 of CEUR Workshop Proceedings, 2020, pp. 1001-1012.
[16] I. Oksanych, I. Shevchenko, I. Shcherbak, S. Shcherbak, Development of specialized services for predicting the business activity indicators based on micro-service architecture, volume 2(2-86) of Eastern-European Journal of Enterprise Technologies, 2017, pp. 50-55.
[17] A. Berko, V. Alieksieiev, A Method to Solve Uncertainty Problem for Big Data Sources, in: Proceedings of International Conference on Data Stream Mining and Processing, DSMP, 2018, pp. 32-37.
[18] A.Y. Berko, K.A. Aliekseyeva, Quality evaluation of information resources in web-projects, volume 136(10) of Actual Problems of Economics, 2012, pp. 226-234.
[19] A.Y. Berko, Models of data integration in open information systems, volume 10 of Actual Problems of Economics, 2010, pp. 147-152.
[20] A.Y. Berko, Methods and models of data integration in E-business systems, volume 10 of Actual Problems of Economics, 2008, pp. 17-24.
[21] A. Berko, Consolidated data models for electronic business systems, in: Proceedings of The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM, 2007, pp. 341-342.
[22] O.cKuzmin, M. Bublyk, Economic evaluation and government regulation of technogenic (man-made) damage in the national economy, in: Computer sciences and information technologies (CSIT), 2016, pp. 37–39.
[23] A. Demchuk, V. Lytvyn, V. Vysotska, M. Dilai, Methods and Means of Web Content Personalization for Commercial Information Products Distribution, volume 1020 of Advances in Intelligent Systems and Computing, 2020, pp. 332–347.
[24] A. Gozhyj, I. Kalinina, V. Vysotska, S. Sachenko, R. Kovalchuk, Qualitative and Quantitative Characteristics Analysis for Information Security Risk Assessment in E-Commerce Systems, volume Vol-2762 of CEUR Workshop Proceedings, 2020, pp. 177-190.
[25] V. Vysotska, A. Berko, M. Bublyk, L. Chyrun, A. Vysotsky, K. Doroshkevych, Methods and tools for web resources processing in e-commercial content systems, in: Proceedings of 15th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT, 1, 2020, pp. 114-118.

[26] O. Artemenko, V. Pasichnyk, N. Kunanets, K. Shunevych, Using sentiment text analysis of user reviews in social media for e-tourism mobile recommender systems, volume Vol-2604 of CEUR workshop proceedings, 2020, pp. 259-271.
[27] B. Rusyn, L. Pohreliuk, O. Kapshii, J. Varetskyy, A. Demchuk, I. Karpov, A. Gozhyj, V. Gozhyj, I. Kalinina, An Intelligent System for Commercial of Information Products Distribution Based SEO and Sitecore CMS, volume Vol-2604 of CEUR workshop proceedings, 2020, pp. 760-777.
[28] B.E. Kapustiy, B.P. Rusyn, V.A. Tayanov, Peculiarities of Application of Statistical Detection Criteria for Problem of Pattern Recognition, volume 37(2) of Journal of Automation and Information Science, 2005, pp. 30-36.
[29] B. Rusyn, V. Lytvyn, V. Vysotska, M. Emmerich, L. Pohreliuk, The Virtual Library System Design and Development, volume 871 of Advances in Intelligent Systems and Computing, 2019, pp. 328-349.
[30] B. Rusyn, V. Vysotska, L. Pohreliuk, Model and architecture for virtual library information system, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2018, pp. 37-41.
[31] V. Lytvyn, V. Vysotska, B. Rusyn, L. Pohreliuk, P. Berezin, O. Naum, Textual Content Categorizing Technology Development Based on Ontology, volume Vol-2386 of CEUR Workshop Proceedings, 2019, pp. 234-254.
[32] V. Vysotska, V. Lytvyn, V. Kovalchuk, S. Kubinska, M. Dilai, B. Rusyn, L. Pohreliuk, L. Chyrun, S. Chyrun, O. Brodyak, Method of Similar Textual Content Selection Based on Thematic Information Retrieval, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2019, pp. 1-6.
[33] V. Lytvyn, V. Vysotska, I. Peleshchak, T. Basyuk, V. Kovalchuk, S. Kubinska, L. Chyrun, B. Rusyn, L. Pohreliuk, T. Salo, Identifying Textual Content Based on Thematic Analysis of Similar Texts in Big Data, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2019, pp. 84-91.
[34] O. Veres, B. Rusyn, A. Sachenko, I. Rishniak, Choosing the mathod of finding similar images in the reverse search system, volume Vol-2136 of CEUR Workshop Proceedings, 2018, pp.99-107.
[35] V. Husak, O. Lozynska, I. Karpov, I. Peleshchak, S. Chyrun, A. Vysotskyi, Information System for Recommendation List Formation of Clothes Style Image Selection According to User's Needs Based on NLP and Chatbots, volume Vol-2604 of CEUR workshop proceedings, 2020, pp. 788-818.
[36] S. Makara, L. Chyrun, Y. Burov, Z. Rybchak, I. Peleshchak, R. Peleshchak, R. Holoshchuk, S. Kubinska, A. Dmytriv, An Intelligent System for Generating End-User Symptom Recommendations Based on Machine Learning Technology, volume Vol-2604 of CEUR Workshop Proceedings, 2020, pp. 844-883.
[37] Y. Matseliukh, V. Vysotska, M. Bublyk, Intelligent System of Visual Simulation of Passenger Flows, volume Vol-2604 of CEUR workshop proceedings, 2020, pp. 906-920.
[38] M. Bublyk, Y. Matseliukh, U. Motorniuk, M. Terebukh, Intelligent system of passenger transportation by autopiloted electric buses in Smart City, volume Vol-2604 of CEUR workshop proceedings, 2020, pp. 1280-1294.
[39] I. Bodnar, M. Bublyk, O. Veres, O. Lozynska, I. Karpov, Y. Burov, P. Kravets, I. Peleshchak, O. Vovk, O. Maslak, Forecasting the risk of cervical cancer in women in the human capital development context using machine learning, volume Vol-2631 of CEUR workshop proceedings, 2020, pp. 491-501.
[40] L. Podlesna, M. Bublyk, I. Grybyk, Y. Matseliukh, Y. Burov, P. Kravets, O. Lozynska, I. Karpov, I. Peleshchak, R. Peleshchak, Optimization model of the buses number on the route based on queueing theory in a smart city, volume Vol-2631 of CEUR workshop proceedings, 2020, pp. 502-515.
[41] D. Koshtura, M. Bublyk, Y. Matseliukh, D. Dosyn, L. Chyrun, O. Lozynska, I. Karpov, I. Peleshchak, M. Maslak, O. Sachenko, Analysis of the demand for bicycle use in a smart city based on machine learning, volume Vol-2631 of CEUR workshop proceedings, 2020, pp. 172-183.
[42] O. Soprun, M. Bublyk, Y. Matseliukh, V. Andrunyk, L. Chyrun, I. Dyyak, A. Yakovlev, M. Emmerich, O. Osolinsky, A. Sachenko, Forecasting temperatures of a synchronous motor with

permanent magnets using machine learning, volume Vol-2631 of CEUR workshop proceedings, 2020, pp. 95-120.
[43] A. Berko, I. Pelekh, L. Chyrun, M. Bublyk, I. Bobyk, Y. Matseliukh, L. Chyrun, Application of ontologies and meta-models for dynamic integration of weakly structured data, in: Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP, 2020, pp. 432-437.
[44] M. Bublyk, V. Vysotska, Y. Matseliukh, V. Mayik, M. Nashkerska, Assessing losses of human capital due to man-made pollution caused by emergencies, volume Vol-2805 of CEUR Workshop Proceedings, 2020, pp. 74-86.
[45] I. Rishnyak, O. Veres, V. Lytvyn, M. Bublyk, I. Karpov, V. Vysotska, V. Panasyuk, Implementation models application for IT project risk management, volume Vol-2805 of CEUR Workshop Proceedings, 2020, pp. 102-117.
[46] I. Kalinina, V. Vysotska, P. Bidyuk, A. Gozhyj, M. Vasilev, R. Malets, Forecasting Nonlinear Nonstationary Processes in Machine Learning Task, in: Proceedings of the 2020 IEEE 3rd International Conference on Data Stream Mining and Processing, DSMP, 2020, pp. 28-32.
[47] N. Shpak, N. Seliuchenko, V. Kharchuk, N. Kosar, W. Sroka, Evaluation of Product Competitiveness: A Case Study Analysis, volume 52(2) of Organizacija, 2019, pp. 107-125. DOI: 10.2478/orga-2019-0008.
[48] N. Shpak, N. Podolchak, V. Karkovska, W. Sroka, The Influence of Age Factors on the Reform of the Public Service of Ukraine, volume 13(2) of Central European Journal of Public Policy, 2019, pp. 40-52. DOI: 10.2478/cejpp-2019-0006.
[49] N. Shpak, Z Dvulit, T. Luchnikova, W. Sroka, Strategic development of cargo transit services: a case study analysis, volume 10(4) of Engineering Management in Production and Services, 2018, pp. 76-84. DOI: 10.2478/emj-2018-0024.
[50] N. Shpak, O. Kuzmin, Z. Dvulit, T. Onysenko, W. Sroka, Digitalization of the Marketing Activities of Enterprises: Case Study, volume 11 of Information, 2020, p. 109. DOI: 10.3390/info11020109.
[51] N. Shpak, O. Kuzmin, O. Melnyk, M. Ruda, W. Sroka, Implementation of a Circular Economy in Ukraine: The Context of European Integration, volume 9(8) of Resources, 2020, p. 96. DOI: 10.3390/resources9080096.
[52] M. Sharko, N. Shpak, O. Gonchar, K. Vorobyova, O. Lepokhina, J. Burenko, Methodological basis of causal forecasting of the economic systems development management processes under the uncertainty, volume 1246 of Advances in Intelligent Systems and Computing, AISC, 2021, pp. 423-436. DOI: 10.1007/978-3-030-54215-3_27.
[53] N. Shpak, T. Kyrylych, J. Greblikaite, Diversification Models of Sales Activity for Steady Development of an Enterprise, volume 8(4) of Sustainability, 2016. DOI: 10.3390/su8040393.
[54] O.D. Vytvytska, O.A. Martynyuk, N.O. Shpak, G.T. Karcheva, I.P. Medynsky, L.S. Nodzhak, Modern Structural-functional modeling for the determination of the company's equilibrium conditions in the dynamic business environment, volume 7(1) of Mathematical Modeling and Computing, 2020, pp. 104-111. DOI: 10.23939/mmc2020.01.104.
[55] Y. Tverdokhlib, V. Andrunyk, L. Chyrun, L. Chyrun, N. Antonyuk, I. Dyyak, O. Naum, D. Uhryn, V. Basto-Fernandes, Analysis and Estimation of Popular Places in Online Tourism Based on Machine Learning Technology, volume Vol-2631 of CEUR Workshop Proceedings, 2020, pp P. 457-470.
[56] L. Chyrun, I. Turok, I. Dyyak, Information Model of the Tendering System for Large Projects, volume Vol-2604 of CEUR Workshop Proceedings, 2020, pp. 1224-1236.
[57] V. Lytvyn, A. Gozhyj, I. Kalinina, V. Vysotska, V. Shatskykh, L. Chyrun, Y. Borzov, An intelligent system of the content relevance at the example of films according to user needs, volume Vol-2516 of CEUR Workshop Proceedings, 2019, pp. 1-23.
[58] Y. Kis, L. Chyrun, T. Tsymbaliak, L. Chyrun, Development of System for Managers Relationship Management with Customers, volume 1020 of Lecture Notes in Computational Intelligence and Decision Making, 2020, p. 405-421.
[59] N. Antonyuk, A. Vysotsky, V. Vysotska, V. Lytvyn, Y. Burov, A. Demchuk, I. Lyudkevych, L. Chyrun, S. Chyrun, I. Bobyk, Consolidated Information Web Resource for Online Tourism Based on Data Integration and Geolocation, in: Proceedings of the International Conference on Computer

Sciences and Information Technologies, CSIT, 2019, pp. 15-20. DOI: 10.1109/STC-CSIT.2019.8929790
[60] A. Vysotsky, V. Lytvyn, V. Vysotska, D. Dosyn, I. Lyudkevych, N. Antonyuk, O. Naum, A. Vysotskyi, L. Chyrun, O. Slyusarchuk, Online Tourism System for Proposals Formation to User Based on Data Integration from Various Sources, in: Proceedings of the International Conference on Computer Sciences and Information Technologies, CSIT, 2019, pp. 92-97. DOI: 10.1109/STC-CSIT.2019.8929849
[61] L. Chyrun, A. Kowalska-Styczen, Y. Burov, A. Berko, A. Vasevych, I. Pelekh, Y. Ryshkovets, Heterogeneous Data with Agreed Content Aggregation System Development, volume Vol-2386 of CEUR Workshop Proceedings, 2019, pp. 35-54.
[62] L. Chyrun, Y. Burov, B. Rusyn, L. Pohreliuk, O. Oleshek, A. Gozhyj, I. Bobyk, Web Resource Changes Monitoring System Development, volume Vol-2386 of CEUR Workshop Proceedings, 2019, pp. 255-273.
[63] L. Chyrun, A. Gozhyj, I. Yevseyeva, D. Dosyn, V. Tyhonov, M. Zakharchuk, Web Content Monitoring System Development, volume Vol-2362 of CEUR Workshop Proceedings, 2019, pp. 126-142.
[64] N. Chukhray, N. Shakhovska, O. Mrykhina, M. Bublyk, L. Lisovska, Methodical Approach to Assessing the Readiness Level of Technologies for the Transfer, volume 1080 of Advances in Intelligent Systems and Computing IV, 2020, pp. 259-282.
[65] N. Chukhray, N. Shakhovska, O. Mrykhina, M. Bublyk, L. Lisovska, Consumer aspects in assessing the suitability of technologies for the transfer, in: Computer sciences and information technologies, (CSIT), 2019, pp. 142–147.
[66] M. Bublyk, O. Rybytska, A. Karpiak, Y. Matseliukh, Structuring the fuzzy knowledge base of the IT industry impact factors, in: Computer sciences and information technologies (CSIT), 2018.
[67] M.I. Bublyk, O. M. Rybytska, The model of fuzzy expert system for establishing the pollution impact on the mortality rate in Ukraine, in: Computer sciences and information technologies (CSIT), 1, 2017, pp. 253–256.
[68] J. L. Peterson, Petri nets, volume of 9(3) ACM Computing Surveys (CSUR), 1977, pp. 223-252.
[69] J. L. Peterson, A note on colored Petri nets, volume 11(1) of Process. Lett., 1980, pp. 40-43.
[70] J. Wang, Petri Nets for Dynamic Event-Driven System Modeling, volume 1 of Handbook of Dynamic System Modeling, 2007, p. 24.
[71] M. Wang, G. Liu, C. Yan, C. Jiang, Modeling and vulnerable points analysis for e-commerce transaction system with a known attack, in: International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, 2016, pp. 422-436.
[72] V. Jain, S. Wadhwa, S. G. Deshmukh, e-Commerce and supply chains: Modelling of dynamics through fuzzy enhanced high level petri net, volume 30(2) of Sadhana, 2005, pp. 403-429.
[73] A. Gozhyj, I. Kalinina, V. Gozhyj, V. Vysotska, Web service interaction modeling with colored petri nets, in: Proceedings of International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS, 1, 2019, pp. 319-323.