Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Statistics

Diploma Thesis

Predictive analysis of rate of happiness

Rodovolskiy Mikhail

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

Bc. Mikhail Rodovolskiy

Systems Engineering and Informatics
Informatics

Thesis title

Predictive analysis of rate of happiness

Objectives of thesis

The goal of the thesis is to identify important factors affecting level of happiness. The analysis will be based on well-being statistics provides by European Social Survey.

Methodology

Assessment will be based on exploration of several classification models for quantitative data. To reach the aim there will be employed statistical procedures, such as exploratory data analysis, regression analysis or multivariate statistical methods.

The proposed extent of the thesis

60 – 80 pages

Keywords

Big Data, Analysis, Survey, Questionnaire, European Social Survey, predictive modeling

---

Recommended information sources

Abbott, D. Applied Predictive Analytics : Principles and Techniques for the Professional Data Analyst. Praha: John Wiley & Sons, Incorporated, 2014. ISBN 9781118727935.

Field, A., 2018. Discovering Statistics Using IBM SPSS Statistics. London: SAGE.

Hintz, A., Dencik, L. and Wahl-Jorgensen, K., n.d. Digital Citizenship In A Datafied Society.

Mayer-Schönberger, V. and Cukier, K., n.d. Big Data.

Paul C. Zikopoulos., 2013. Harness The Power Of Big Data: The IBM Big Data Platform. McGraw-Hill/Osborne.

Tufféry, Stéphane. Data Mining and Statistics for Decision Making. UK, West Sussex: Wiley, 2011. ISBN 978-0-470-68829-8.

White, T., 2012. Hadoop. Sebastopol, Calif.: O'Reilly

Witten, I., Frank, E. and Hall, M., 2011. Data Mining. Amsterdam etc.: Morgan Kaufmann.

Expected date of thesis defence

2020/21 SS – FEM

The Diploma Thesis Supervisor

Ing. Tomáš Hlavsa, Ph.D.

Supervising department

Department of Statistics

Electronic approval: 23. 11. 2020

prof. Ing. Libuše Svatošová, CSc.

Head of department

Electronic approval: 24. 11. 2020

Ing. Martin Pelikán, Ph.D.

Dean

Prague on 21. 02. 2021

Declaration

I declare that I have worked on my diploma thesis titled "Predictive analysis of rate of happiness" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the diploma thesis, I declare that the thesis does not break copyrights of any their person.

In Prague on 31.03.2021 ______________________________

Acknowledgement

I would like to thank my supervisor Ing. Tomáš Hlavsa, Ph.D., for his time and very helpful consultations, my parents, my close friends, Anna Fomichova and A.Soldatova for their advice and their support during my work on this thesis.

Predictive analysis of rate of happiness

Abstract

This paper examines the key factors affecting the level of happiness of an individual. It consists from two parts: theoretical and practical. The first part is dedicated to the determination of basic concepts of big data, classification of the methods of big data analysis. It also contains a literature review of articles related to the studies of happiness. The determinants of happiness are described in this part.

The second part of thesis is dedicated to the practical research and analysis based on European Social Survey Dataset in order to approve or reject the stated hypotheses as well as to highlight the main factors that affect the level of happiness of a person.

The results obtained in the course of the study reveal the positive effects of GDP per capita, income, and trust in society, health and being religious. Within the positive ones, some factors affecting the happiness level negatively were discovered as well. They are: unemployment status, felling of safety, level of satisfaction.

Keywords: Big Data, analysis, statistics, predictive modelling, analytics, happiness, level of happiness, European Social survey, survey

Prediktivní analýza míry štěstí

Abstrakt

Tento článek zkoumá klíčové faktory ovlivňující úroveň štěstí jednotlivce. Skládá se ze dvou částí: teoretické a praktické. První část je věnována stanovení základních pojmů velkých dat, klasifikaci metod analýzy velkých dat. Obsahuje také literární přehled článků souvisejících se studiem štěstí. Determinanty štěstí jsou popsány v této části.

Druhá část práce je věnována praktickému výzkumu a analýze na základě datového souboru Evropského sociálního průzkumu za účelem schválení nebo odmítnutí uvedených hypotéz a také zdůraznění hlavních faktorů, které ovlivňují úroveň štěstí jednotlivce.

Výsledky získané v průběhu studie odhalují pozitivní účinky HDP na obyvatele, příjem, a důvěra ve Společnost, Zdraví a být náboženský. V těch pozitivních byly také objeveny některé faktory ovlivňující úroveň štěstí negativně. Jsou to: stav nezaměstnanosti, kácení bezpečnosti, úroveň spokojenosti.

Klíčová slova: Big Data, analýza, statistika, prediktivní modelování, analytika, štěstí, úroveň štěstí, Evropský sociální průzkum, průzkum

Table of contents

1 Introduction

Recently, there is a huge increase in the volume of data, which is a key factor in the Big Data scenario. Big data requires new high-performance processing. Now it is important to study the origin of the term "Big Data", the main approaches of scientists to its interpretation and examples from the real practice of developed countries on its implementation. It is believed that the concept of Big Data is a promising area of research for various sectors of the world's economy, which opens up new opportunities for doing business and managing the economy. Big data is a socio-economic phenomenon associated with the emergence of new technological opportunities for analyzing a huge amount of data. The implementation of Big Data technologies allows efficiently and quickly gain benefits from a huge array of information. With their help, government agencies and business representatives optimize various processes, and end users receive high-quality services. Active involvement of scientists in the methodology of implementing the information technology concept of Big Data is a strategic direction for the further development of Big Data. Further studies of Big Data will allow to use them in decision-making processes and risk management processes, and will inevitably bring the area of study to a qualitatively new, higher level.

As it is seen, scientists cover various problems of human development at the micro and macro levels: education, health, investment in these components, public financing, and so on. However, despite the presence of a significant number of publications devoted to this topic and their depth, the problem of determining the level of realization of human potential, its level of happiness on the basis of both objective and subjective indicators remains insufficiently studied.

The lack of research in the scientific literature on information technologies for analyzing, processing and storing data using Big Data methods has led to the relevance and choice of the research topic.

2 Objectives and Methodology

2.1 Objectives

The purpose of the study is to find out what has an effect on people's feeling of happiness, its calculation features, and its significance for improving the standard of living of the population.

To achieve this goal, there is a need to solve the following tasks:

- To analyze the existing approaches to the definition of the subject of the level of happiness, to give a reasonable definition, to identify the role and place of the level of happiness in the system of socio-economic categories;
- Highlight the factors that could have impact on the level of happiness;
- Build a predictive model based on these factors;
- State the proper hypotheses base on previous studies of the subject of this thesis;
- Analyze these factors in order to understand the how they affect the level of happiness: positively or negatively.

The object of the study is the dataset of European Social Survey taken in 23 European countries in 2016. It requires a comprehensive analysis to extract useful information. The subject of the study is the indicators of the level of happiness and their measurement.

2.2 Methodology

The main data source for this study will be the European Social Survey database. ESS is a large-scale academic social research project designed to track and explain the dynamics of interactions between changing European institutions and attitudes and behavioral patterns of the European population.

The main principle of constructing an ESS sample is to observe the probabilistic, random nature. The basis for its construction is internal statistical estimates. The actual ways to achieve the probabilistic nature of the sample vary from country to country, depending on their individual capabilities. The minimum sample size for each country is 1,500 people (with the exception of countries with a population of less than 2 million people – for them, the lower threshold is 800 people).

The ESS uses the following data weighting scheme.

1) w = 1/(PROB1*...* PROBk) is the basis for constructing the data weighting vector; k depends on the number of sampling stages;

2) All weights are recalculated taking into account that their sum must be equal to n, for example, rescaled weights = n*w/sum (w).

3) Overweighting of population data from different countries occurs in the following cases. The variable responsible for the weight of the population data of a particular country (PWEIGHT) is used to adjust the indicators when using a combination of data sets of two or more countries. The formula for calculating it is as follows: PWEIGHT=[Population over 15 years old]/[(The volume of the selection in the array)*10 000]. The source of data on the size of the population is Eurostat.

The type of data used in this study is cross-sectional, and, accordingly, regression analysis of cross-sectional data will be used in the work.

It is assumed that the variables may influence the dependent variable linearly. Thus, the multiple linear regression model will be estimated. The multiple linear regression formula is presented below:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon, \tag{1}$$

where y is the dependent variable;

β_0 is the intercept of the model;

$\beta_1, \dots, \beta_{\mathrm{k}}$ are the regression coefficients;

$x_1, \dots, x_{\mathrm{k}}$ are the independent variables;

ε is a random error of the model.

Based on the studies of the related articles in theoretical part some independent variables possible predictors of happiness were highlighted. They are: ***income, satisfaction, discrimination, safety, trust, GDP per capita, unemployment, age, gender, marital status, health, education and religion.*** These variables will be explained in more detailed way in the practical part. Including them, the model will look like this:

$$\begin{gathered} Happiness_i = \beta_0 + \beta_1 \cdot Income_i + \beta_2 \cdot Satisfaction + \beta_3 \cdot Discrimination_i + \\ + \beta_4 \cdot Safety_i + \beta_5 \cdot Trust_i + \beta_6 \cdot GDP\ per\ capita_i + \beta_7 \cdot Unemployment_i + \\ \beta_8 \cdot Age_i + \beta_9 \cdot Gender_i + \beta_{10} \cdot Marital\ status_i + \beta_{11} \cdot Health_i + \\ \beta_{12} \cdot Education_i + \beta_{13} \cdot Religion_i \end{gathered} \tag{2}$$

The model will be estimated with the help of the OLS method. The quality of the results of the model estimation will be analyzed with the help of R-squared, number of significant variables and verification of the necessary assumptions.

Thus, we can proceed to the analysis. During this stage, the preprocessing of the data will be conducted as well as the exploratory data analysis. After that the regression model is built that estimates socio-demographic and economic characteristics of respondents and their impact on the happiness level. The main method of modeling is the construction of multiple linear regression equation.

3 Literature Review

3.1 Theoretical basis of Big Data studies

3.1.1 Big Data: basic concepts, its nature, and attributes

Nowadays we live in a world where the amount of data generated every day is constantly growing. A proportion of the data coming from digital media, social networks and the "Internet of things", etc. is increasing.

According to the scientific research made at the University of San Diego and the University of Southern California[1]:

- During one hour of sitting in social networks, we see as many people as we would not have met in a lifetime if we had lived 100 years ago;

- We get as much information as a person of the Middle Ages received in a lifetime only just in a week;

- Modern people consume 34 GB of media content per day;

- Every day, each of us receives so much information, which can be placed in 174 printed publications;

- Scientists estimated a 100 KB/s of brain traffic during surfing the Internet;

- The average American consumes "information food" 11.8 hours a day, "digesting" 100500 words and 34 GB of audio-visual information.

In other words, we can say that modern society is experiencing another information technology boom, but this time it is associated with a rapid, exponential growth in the volume of information. It is evidenced by the given estimate of IBS analysts: "the entire volume of all world data" in 2020 is forecasted to be 40-44 ZB (1 Zetabyte = 2^{70} bytes), and in 2025 this amount will increase by 10 times[2].

As we can see, the growth rate of data volume is impressive, and these data arrive at a high, but different speed, different type of structure, and contain large amount of information that can be a key to achieving the different goals of different people, competing companies, and even countries. The ability to analyze this huge amount of data represents a new era of productivity growth, innovation, and security.

---

[1] The 2018 Digital Future Report. Center for the Digital Future at USC Annenberg. / https://www.digitalcenter.org/reports/

[2] Technology and Innovations. Big Data. - https://it-enterprise.com/knowledge-base/technology-innovation/big-data-bolshie-dannye

Due to the fact that the main part of the increase in the volume of information is unstructured or partially structured data, due to significant amount, it is necessary to process and store such data with special hardware and software. On the other hand, classical data processing algorithms are ineffective when applied to unstructured data problems.

To solve these problems, some of the world's largest companies in the IT industry have started developing completely new approaches to the problem of processing and storing information in order to obtain useful knowledge. As a result, there is a system of new tools and methods for analyzing large volumes of data and weak structuring. This system is called "Big data".

The Big Data term was introduced into scientific environment by Berkeley University computer science doctor Clifford Lynch (2008) with the publication of a special issue of the journal NATURE "How can technologies that open up the possibility of working with large amounts of data affect the future of science?

Currently, the most common interpretation of big data is understanding it as a huge amount of information generated by the Internet in the process of using modern digital technologies by people. In other words, big data is identified with the aggregation of digital footprints that people leave when they surf the Internet, download mobile apps or music, communicate with friends on social networks, use GPS, buy goods in online stores, and so on.

In the Oxford English dictionary, the term "Big Data" is defined as follows: "The largest data sets that can be calculated to reveal patterns, trends, and associations, especially in relation to human behavior and interaction"[3]. The technological component in this interpretation is closely related to the social component and focuses on the practical possibilities of analyzing human activity.

The interpretation of the Cambridge dictionary emphasizes the importance of methods and tools used to work with big data, but it also postulates the role of man in relation to technologies: "Big data - very large data sets produced by people in the process of using the Internet, which can be stored, understood and used only with the help of special tools and methods"[4].

Big data is understood in a complex way, that is, as "information assets characterized by such large volumes, speed and diversity, requiring special technologies and analytical methods to convert them into value" (De Mauro et al, 2015). David Boyd and Kate Crawford

[3] The Oxford English dictionary. URL: https://en.oxforddictionaries.com/definition/big_data
[4] Cambridge Dictionary URL: http://dictionary.cambridge.org/dictionary/english/big - data

(2013) in their article "Critical questions of big data" propose to define big data as a cultural, technological, and scientific phenomenon based on the interaction of:

1. Technology;
2. Methods of analysis;
3. Myths and human representations.

Thus, the last (at the moment) stage in the evolution of understanding big data is connected with moving away from focusing on the volume of data and the complexity of processing it. There is a shift in focus from technological issues to socio-humanic aspects, especially the ethical use of big data.

Big data is online data, but it contains information not only about the Internet environment. Big data is still understood as social signals of the Internet that reflect the state of society and can measure its trends (Rodgers, 2009; Rodgers, 2019), as a "manifestation of social behavior", as well as "manifestations of social processes and human activities" (Resnyansky, 2019), as a "manifestation of social and cultural reality" (Shaw, 2015). This approach to big data "... it challenges sociologists by encouraging them to generalize the concept of social knowledge taking into account its new contexts caused by the formation of the digital environment of human habitat" (Dencik, 2018).

Big data is a concept that is used to characterize a set of data arrays that are so large and complex that it is difficult to process using traditional database management tools or data processing applications (Mayer-Schönberger, 2013).

In information technologies, big data is a set of methods and tools for processing structured and unstructured multi-type dynamic data of large volumes in order to analyze it and use it to support decision-making processes (Uskenbayeva, 2013).

The defining characteristics of Big Data are volume (in the sense of physical volume), velocity (in the sense of growth rate and the need for high-speed processing and obtaining results), and variety (in the sense of the possibility of simultaneous processing of different types of structured and partially structured data).

Two other features of big data are often used as well: cost (veracity, in the sense of the economic effect that the technology provides to users) and reliability (value, in the sense of the quality of the collected data that can differ significantly) (Zikopoulos, 2012).

Usually, there are three attributes that used to describe various aspects of big data (figure 1):

- Volume,

- Speed,

- Structure.

These three attributes allow us to determine the nature of the data available for analysis (Sharma, 2013).

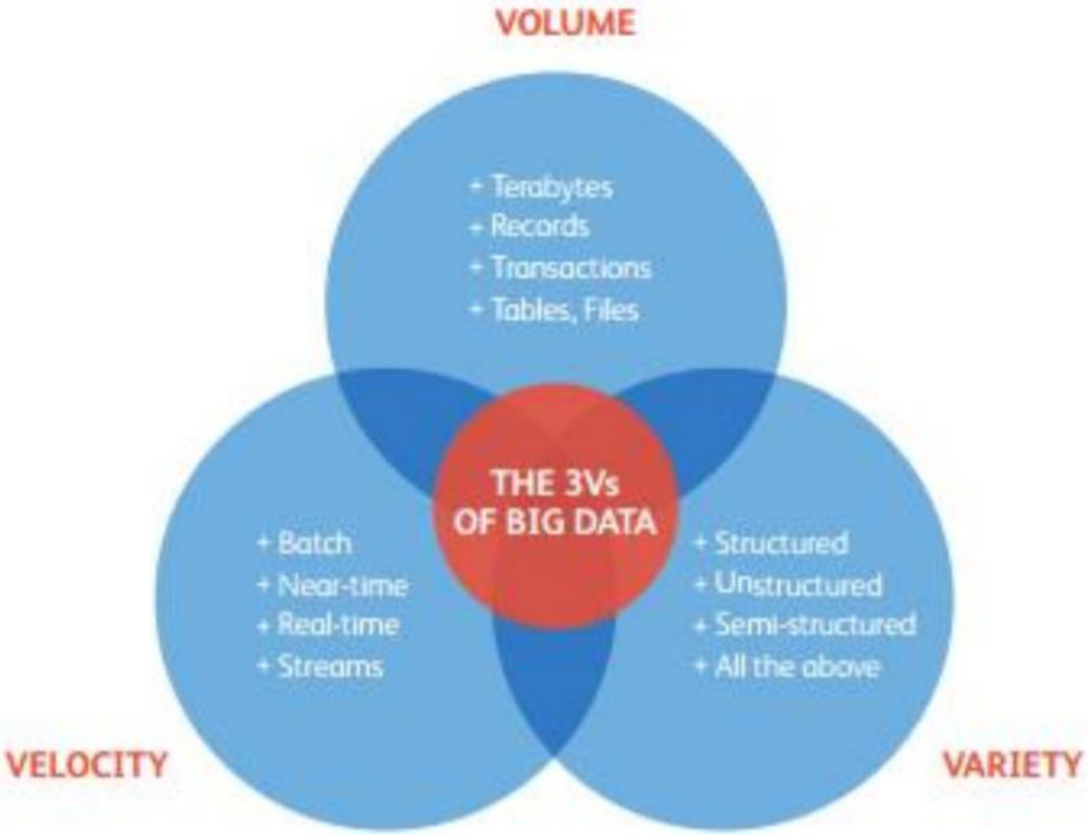

Figure 1. 3Vs of Big Data. Source: https://rightpeoplegroup.com/us/what-is-business-intelligence-and-big-data/

- The volume of data.

Volume is the most complex aspect of big data, as it requires scalable storage and a distributed approach to queries.

Large companies already have a large amount of data accumulated over many years. They can be presented in the form of system logs, accounting, and so on. This data reaches such volumes that relational database management systems (DBMS) can't handle it. The "Data Warehouse" solution which is a data warehouse (a subject-oriented information database specifically designed for reporting and business analysis to support decision-making processes in the organization) may not necessarily be able to process and analyze this data due to the lack of a parallel computing architecture.

A lot of things can be obtained from text data, geolocation data, or log files. For example, email communication patterns, consumer preferences and trends, and security research. Big data technologies offer solutions for obtaining useful information from these huge amounts of data (Sharma, 2013).

- The velocity of data.

Data is sent to users at high speed. Internet and mobile technologies have allowed companies to establish feedback loop with their customers. As an example, online stores have revolutionized the way consumers and suppliers interact. Online stores can now keep logs of all user transactions, store the history of these transactions, and quickly use this data to provide recommendations, increasing the company's competitiveness. Internet marketing organizations get many advantages by being able to quickly get predictive information. With the invention of smartphones, the world generates a lot of geolocation data, and the ability to use advantages of large amounts of data has become important (Sharma, 2013).

- The structure of the data.

Data generated by social networks and digital technologies is usually not structured. Unstructured text documents, video, audio data, images, financial transactions, and social media interactions are examples of unstructured data.

Relational databases support the ability to store large objects, but they have their own limitations. This leads to data loss. Big data technologies, on the other hand, contribute to the complete preservation of all data for later analysis. The main principle of big data technologies is that important and valuable information can be hidden in every bit of data.

To work effectively with big data, there is a need in comprehensive monitoring solutions, structuring, filtering, and searching for hierarchical relationships.

Using big data allows you to observe a significantly huge sets of variables and, based on the information provided, identify global trends and conclusions on the strategy of various situations.

3.1.2 Classification of methods for big data analysis

Taking into account the attributes of big data described in the previous paragraph, we should remember that the volume of data is large, it arrives rapidly and in a much unstructured form that does not correspond to the usual structures of relational databases.

The amount of information containing big data is quickly expanding due to the spreading of digitization and datafication processes. Digitization and datification are two interrelated parts of the same process - translating the entire heritage of humanity into a machine-readable format, that is, in Big Data arrays. Danification is now seen as the concept of converting everything around us into a quantitative online data format, even what at first glance is not perceived as information, such as a person's location, engine vibrations, or bridge loads. At the same time, the possibility of quantitative analysis is a necessary component of the resulting data array.

The differences between digitization and classification can be clearly demonstrated by the well-known Google Books project regarding the complete digitization of printed heritage enabling people around the world to search for and view books for free via the Internet (Mayer, 2013). The project has started in 2004 and today we already use its abilities.

The first text digitized by Google was presented as scanned pages that were converted into digital copies, accessible to anyone via the Internet from anywhere in the world. However, the text could not be found by keywords or analyzed. Only images of text were available, which people could reread and turn into useful information.

The second step is identification, which was carried out by Google thanks to an optical character recognition program that could recognize letters, words, sentences and paragraphs in a digital image. As a result, we received digitized texts instead of digitized page images. Now information from pages is available not only for people to read, but also for processing on computers and for analysis using algorithms. In addition, the texts have become indexed, which means that they are searchable. The NgramViewer service by Google Books provided text analysis capabilities, which allowed us to take a new look at the process of spreading ideas and developing human thinking over the centuries (Chen and Yan, 2016).

Danification transforms our daily activities and interactions into a data format that can be analyzed and provide information about individuals, communities, and society as a whole. Facebook, Instagram, Etc., for example, collect and store information about us and our preferences. The presence of this data, accumulated by social media sites and search engines, provokes the desire to use it for various analysis purposes, since traditional data collection is a difficult process that requires a lot of time and resources.

However, not everyone can get access to such data, social media companies restrict access (especially in transactional data), and this create a new type of inequality - information inequality. "An anthropologist working for Facebook or a sociologist working for Google may have access, but other parts of the scientific community do not" (Boyd and Crawford, 2013). For example, a survey of more than 9,000 social scientists conducted in 2016 to study the degree of availability of big data to academic researchers found that for 32% of respondents, access to commercial or proprietary data was a "big problem" (Metzler et al, 2016). Some companies restrict access to data in its entirety, others sell access privileges for very high fees; and others offer small data sets for University researchers.

Thus, not all "big data" is equal. In this connection, Boyd and Crawford (2013) suggest classifying them by the openness criterion:

- "Rich big data" that contains complete information, but is available only to a limited number of researchers;
- "Poor big data" that cannot pretend to full information coverage about the phenomenon under study, but is available to a wide audience of researchers.

So, we see that big data, on one hand, has a great potential to meet the information needs of researchers in modern society, but on the other hand, it creates new problems.

With so much hidden information in this data, an alternative way to analyze the data is needed. Large corporations may have enough resources to handle this task, but the total amount of data generated on a daily basis will easily outweigh the company's ability to process such volumes. Thanks to cheaper hardware, cloud computing, and open source software, processing large amounts of data has become much cheaper.

A lot of data means a lot of hidden valuable information. The ability to quickly analyze large amounts of data means you can learn more about customers, market trends, marketing and advertising drivers, hardware monitoring and performance analysis, and more. And this is an important reason why many large companies need reliable technologies and data analysis tools.

A characteristic feature of big data technology is the processing of different types of information from different sources of information: structured, partially structured, and unstructured.

A **data model** is a collection of tools that describe data structures for a specific application or class of applications. This concept combines data types and structures, a system of operations, and tools for describing constraints (Lipinson, 2008). The features of the structured data model are, on one hand, the placing of pre-known restrictions on the data by the type and length of attributes, and, on the other, the data structure is known and defined with the of data scheme. This can lead to certain difficulties when working with data, for example, in terms of modifying the model in order to meet requirements that have changed over time. An example of a structured data model is a relational database management system (DBMS).

Unstructured data, unlike structured data, lacks a specific structure, which makes it more difficult to build a model for it. Unstructured information includes text files of various documents, emails, SMS messages, video clips, digital images, audio files, etc. With the help of certain analysis methods, this data can be very informative due to the presence of so-called "Hidden knowledge" in it. The latter can be obtained using Data Mining. Data Mining is a decision support process based on the search for hidden patterns in "raw" data that are previously unknown, non - trivial, practically useful and accessible to interpretation of knowledge necessary for decision-making in various areas of human activity (Barseghyan, 2009).

When building a partially structured data model, you should take into account the features of structured and unstructured data. The main problems that occur when working with partially structured data include:

1) there is a degree of their correctness that needs to be taken into account when building a model of tools for assessing their reliability;
2) the data scheme may not fully meet the processed data or its absence makes it impossible to interpret them;
3) Some attributes may be missing or incomplete in order to meet the conditions of correctness for them.

Previously, relational databases were used to store large amounts of data. Big data technologies outperform relational databases in a number of criteria, including the need for more complex backups, recovery, and faster search algorithms (Pierre, 2011). It should be noted that the benefits of using big data technologies can lead to loss of data privacy.

The vast majority of information is generated by a giant network of devices that interact with other data networks, such as sensors and smart devices. Such growth rates in the amount of data in the world will lead to a tenfold increase in the number of virtual and physical servers due to the expansion and creation of new Data centers. This increases the need for efficient use and monetization of this data. And since the use of Big Data in business requires a lot of investment, business efficiency can be improved by reducing costs and / or increasing sales.

Today, the scope of use of big data is extremely wide, and it will only grow over time. Therefore, it is necessary to take into account that the great importance of Big Data is due to the results of processing and analysis, and not to the data itself or its volumes.

The use of Big Data has great potential in various areas of the economy. Trends in the rapid development of information and communication technologies and the involvement of all forms of economic activity in the development of the information society have also affected our country.

The information technology sector is one of the few areas that shows significant growth in the economy. Our country retains the status of one of the largest IT outsourcing centers in the world and continues to be among the top ten leaders in software development.

Big Data technologies can be useful for solving the following main tasks:

- Marketing and increasing sales;
- Forecasting the market situation;
- Effective customer segmentation;
- Improvement of products and services;
- Making deeper management and operational decisions based on Big Data analysis;
- Increasing the level of labor productivity;
- Efficient logistics;
- Monitoring the state of fixed assets;
- Optimization of the investment portfolio.

There are well known companies such Oracle, Microsoft, SAP, and IBM that develop various tools for Big Data processing and analysis.

There are plenty of technologies that are used to collect and process big data:

- Mass parallel processing (MPP)
- MapReduce - a computing paradigm proposed by Google;

- Complex event processing – processing information online from different sources; data processing depends on time;
- Hadoop is a project of the Apache Software Foundation that implements the MapReduce paradigm;
- RDBMS (Relational Database Management System) – a database management system based on a relational model;
- Cassandra – alternative for Hadoop HDFS, a database which is like NoSQL;
- Hive - file storage, which is created by Facebook;
- NoSQL database management systems based on this paradigm are quite different from relational database management systems, since they do not use the SQL language.

We describe groups of methods and technologies for big data Analytics that are often classified in the formal model of information technology, taking into account functional relationships, namely:

- Data mining methods,
- Text mining technologies,
- Map Reduce technology,
- Data visualization,
- Other technologies and methods of analysis.

Methods of intellectual data analysis (Data Mining). The use of data mining methods and technologies allows solving the following tasks (Barseghyan, 2009; Chubukova, 2006; Dyuk, 2001; Paklin, 2009; Witten, 2011; Zhuravlev, 2006; Zinoviev, 2000):

- Classification,
- Clustering,
- Association,
- Sequence, or sequential association,
- Forecasting,
- Deviation Detection, deviation or outlier analysis,
- Estimation,
- Link Analysis,
- Visualization ,
- Graph Mining,
- Summarization - description of specific groups of objects using the data set under consideration.

Data mining methods are divided into two groups: Supervised Learning and Unsupervised Learning (Barseghyan, 2009; Dyuk, 2001; Paklin, 2009). Another classification divides the variety of data mining methods into two groups: statistical and cybernetic methods. This separation scheme is based on different approaches to teaching mathematical models (Barseghyan, 2009; Chubukova, 2006; Paklin, 2009; Witten, 2011; Zinoviev, 2000).

Let's describe the most suitable ones for big data analysis (Barseghyan, 2009; Chubukova, 2006; Paklin, 2009; Zinoviev, 2000).

- **Association rules (Association Rule Learning).** A set of techniques for identifying relationships, i.e. Association rules, between variables in large data sets. Hidden patterns analysis (Association Analysis) is used to analyze the market basket.
- **Classification.** A set of techniques that allows predicting consumer behavior in a particular market segment (purchasing decisions, outflow, consumption volume, etc.).
- **The Decision Trees method.** One of the most popular methods for solving classification and forecasting problems. In its simplest form, a decision tree is a way to represent rules in a hierarchical, sequential structure. The decision tree method is usually referred to as the "naive" approach.
- **Cluster analysis.** A statistical method for classifying objects into groups by detecting previously unknown common features. An example is market segmentation. To solve the clustering problem on graphs, the Girvanand Newman algorithm of the MLP method (Markov Cluster Algorithm) is used.
- **Dynamic Quantum Clustering (DQC).** It was developed for the analysis of large multidimensional data and implements the search paradigm as "let the data speak for itself" (Weinstein, 2013). The DQC method (like many other methods of Big data Analytics) "works" without prior information about the properties of the "structure", their type and topology, which can be "hidden" in the data and discovered as a result of its application. The

method works well with multidimensional data and depends linearly on the dimension in the analysis.

- **Regression.** A set of statistical methods for identifying patterns between changes in a dependent variable and one or more independent variables.
- **Time Series Analysis.** A set of methods borrowed from statistics and digital signal processing for analyzing data sequences that repeat over time.
- **Outlier analysis** is used for fraud detection, personal marketing, and medical analysis.
- **Machine learning.** The field in computer science (historically it was called "artificial intelligence"), which aims to create self-learning algorithms based on the analysis of empirical data. Machine learning is now used: to recognize spam or non-spam emails to gain knowledge about user preferences and recommendations based on this information; to determine the best content to attract potential customers; to determine the probability of winning a case and compliance with legal standards of invoices.
- **Supervised and Unsupervised Learning.** A set of techniques that are based on machine learning technologies that allow identifying functional relationships in the data sets. Unmanaged learning has common features with cluster analysis.
- **Ensemble Learning.** This method uses a variety of predicative models, which improves the quality of forecasts.
- **Evolution Analysis, Genetic Algorithms.** Genetic algorithms are inspired by the nature of evolutionary processes – that is, mechanisms such as inheritance, mutation, and natural selection. These mechanisms are used to "evolve" useful solutions to problems that require optimization. In this technique, possible solutions are presented as "chromosomes" that can be combined and mutated. As in the process of natural evolution, the fittest individual survives.
- **Neural Networks** are a class of models based on the analogy of the human brain and are designed to solve various problems of data analysis after passing the data training stage. Neural networks can be used, for example, to predict sales volumes, market indicators, recognize signals, and develop self-learning systems.
- **Visualization.** Methods for graphical representation of big data analysis results in the form of diagrams or animations to simplify interpretation and facilitate understanding of the results obtained. Visualization of analytical data – representation of information in the form of drawings, graphs, diagram using interactive features and animation for results, as well as source data for further analysis (Paklin, 2013).

Visual representation of the results of big data analysis is essential for their interpretation. Human perception is limited, and scientists continue to conduct research to improve modern methods of presenting data in the form of images, diagrams, or animation.

New advanced visualization techniques are: the tag cloud of cluster graph; historical flow; spatial stream.

- **Technology of Text Mining.** The basis pf this technology is statistical and linguistic analysis, and artificial intelligence methods. This technology is used for analysis, navigation and search in unstructured texts[5]. The use of Text Mining information systems allows users to acquire new knowledge.

Text Mining technologies are a set of methods that are designed to extract information in texts based on modern ICT, which allows identifying patterns that provide users with useful data and new knowledge. The main goal of Text Mining is to provide analysts with the ability to work with large volumes of source data by automating the process of obtaining the necessary data.

Let's also describe several technologies and disciplines of data research from the point of view of big data technology.

- **A / B testing, Split testing.** A marketing research technique in which a control sample is compared in turn with others. This method is used to optimize Web pages in accordance with the specified goal.
- **Natural Language Processing (NLP).** A set of techniques for recognizing human natural language borrowed from computer science and linguistics.
- **Sentiment analysis.** Methods for assessing consumer mood that are based on human natural language recognition technologies. Sentiment analysis helps researchers determine the mood of speakers or authors on a topic.
- **Network analysis**. A set of techniques for analyzing connections between nodes in networks. Social network analysis allows analyzing the relationships between individual users, companies, communities, and etc.
- **Optimization.** A set of numerical methods for redesigning complex systems and processes to improve one or more metrics. It helps in making strategic decisions, such as the composition of the product line to be put on the market, in conducting investment analysis. etc.

---

[5] Text Mining [Electronic resource]. - Access mode: http:/ /statsoft.ru/home/textbook/modules/sttextmin. html#index

- **Pattern recognition.** A set of techniques with self-learning elements for predicting consumer behavior.
- **Predictive modeling**. A set of techniques that allow creating a mathematical model of a pre-defined probable scenario.
- **Signal processing.** A set of techniques borrowed from radio engineering that aims to recognize a signal against a background of noise and then analyze it.
- **Spatial analysis.** Spatial analysis is the use of topological, geometric, and geographical information in data. A set of data analysis techniques partially borrowed from statistics. The source of big data in this case is geographic information systems (GIS).
- **Statistics.** The science of collecting, organizing, and interpreting data, in particular developing questionnaires and conducting experiments. Statistical methods are often used to make value judgments about the relationship between certain events.
- **Simulation (Simulation).** Modeling the behavior of complex systems is often used to forecast, predict, and process various scenarios in planning.
- **Crowdsourcing.** Methodology for collecting data from a large number of sources.

Crowdsourcing is the categorization and enrichment of data by a wide, indeterminate circle of people, in order to use their creative abilities, knowledge and experience using information and communication technologies.

- **Data Fusion and Data Integration.** A set of techniques that allow integrating heterogeneous data from different information sources for in-depth analysis. This set of techniques allows analyzing comments from users of social networks and compare them with sales results in real time.
- **The Map Reduce Framework.** Creation and support of data warehouses in terabytes, petabytes and more is possible thanks to distributed file system technologies (Stonebrake, 2010). Distributed data processing systems, instead of storing data in a single file system, store, and index data on several (even thousands) hard drives and servers. A "map" is also created, which contains information about the location of certain data. One of the most well-known systems using this approach is Hadoop.

Despite significant changes in the field of machine data processing so far and, in the future, it is impossible to process data without Data Science specialists who are able to work with data professionally and formulate tasks that are understandable in accordance with analysis algorithms. Such specialists should possess the following groups of competencies: IT literacy, mathematical and statistical knowledge, knowledge, know how to apply previous groups of skills, in particular, knowledge of mathematics, mathematical analysis, mathematical statistics, probability theory; knowledge of English; knowledge of the main programming languages that have components for working with large data sets (Java (Hadoop), C ++ (BigARTM, Vowpel Wabbit, XGBoost), Python (Matplotlib, Numpy, Scikit, Skipy)) knowledge of statistical tools - SPSS, MATLAB, SAS Data Miner, Tableau; knowledge of business development laws; economic knowledge.

Data Scientist, as a scientist, not only collects and analyzes data, but also studies it in different contexts and from different angles. An important and special quality of Big data specialists is the vision of logical connections in a complex system of collected information, and the development of effective business solutions based on quantitative analysis. In today's competitive and extremely dynamic world, in a rapidly growing flow of information, Data Scientist is indispensable for management in terms of making the right business decisions and decision-making processes.

3.1.3 Big data in the context of research of the modern world problems

As we can see, big data has increasingly begun to attract the attention of the whole scientific community (Boyd and Crawford, 2013; Burrows and Savage, 2014). Scientists sought to understand the phenomenon of big data in its entirety and predict its impact not only on science, but also on the development of society. The understanding of big data has expanded and has undergone certain transformations associated with the inclusion of the human component. For example, Hopkinson and Evelson proposed the definition of big data as a combination of methods and technologies that extract meaning from data on the extreme edge of practicality, as well as people with appropriate analytical skills who are able to make information that is hidden in the diversity of "capacious and heterogeneous data" available (Hopkins and Evelson, 2012).

At this time, calls for a "new literacy" are becoming more active, radical ideas arise that statistics and sociology are losing their priority positions in the study of society, while computer scientists are taking on purely sociological tasks. Indeed, the trend of "privatization" of sociology by representatives of computer science was actualized by the appearance of big data. Here are the statements of the authors of the book "Big Data and Social Science: A Practical Guide to Methods and Tools": "The world has changed for empirical researchers. New types of "big data" have created a whole new scientific field - data science. This world is dominated by computer scientists who have created new ways to create and collect data, developed new analytical and statistical methods, and proposed new ways to visualize and present information. These new sources of data and methods are transforming applied social sciences" (Foster et al, 2016). Such statements cannot but worry the scientific community, the real threat exists and requires both articulation and the search for ways to overcome it.

Access to big data, as well as its processing, is impossible without the use of special tools and methods. This is probably why the terms "Big Data" and "Big Data Analytics" are often identified. All big data methods are based on the concept of mining, that is, on automatic search for patterns in the available data (Press, 2013). The difference between data mining and Big Data Analytics is that Big Data Analytics is the application of Data Mining technologies to a new type of data. It is possible to say that Big Data Analytics is a combination of online data (digital traces), access technologies and data mining. The first attempts to study big data for statistical analysis led to the need to address a number of methodological issues, as well as issues of "repurposing" the methods of processing online data used by Internet platforms in order to solve statistical problems.

Digital sociology has emerged in response to a certain hype about how "new data" transforms the ways of knowing the society. It is understood as "computational social science". Some critical authors identify it with a particular form of data analysis. But digital sociology offers an alternative to the narrow definition of digital social research. In contrast, digital sociologists seek to explore a much broader set of interactions between data, people, and technology that overwhelm, exceed, and do not "fit" into the simple story of new forms of data analysis taking the place of older social research methods such as surveys or fieldwork (Marres, 2017). Digital sociology, according to its apologists, is designed to understand the modern digital world and requires new ways of thinking about the social. Its followers develop concepts, tools, and practices for analyzing the intersection of social and digital.

The goal of digital sociology is to study the patterns of modern social life integrated into the digital Internet space. The object is digital society as a new socio-cultural reality. The subject is social relations that arise in the digital environment, digital social life, which includes various social phenomena that arise in the digital environment, as well as their relationship with the material social reality. However, from the very beginning of its formation, digital sociology has experienced the greatest criticism due to the lack of a theoretical basis. Today it can be stated that it has not found ways to transform it into a special sociological theory.

On the other hand, sociology gradually accumulates all the methods developed by statisticians, promotes broad interest in the direction called "digital methods" and in fact is a detailed methodology for conducting sociological research based on big data, which is now already considered "not as an innovation, but as a mainstream" (Snee et al, 2015).

The concept of "digital methods" was introduced into scientific circulation by the Dutch scientist R. Rogers in 2007 in order to differentiate new methods that have appeared in the last two decades and significantly expanded the Arsenal of statistical research methods (Rogers, 2009). The need to distinguish digital methods as a separate type of analysis methods was due to two factors:

- Social reality has become to be understood as "augmented", that is, one where "real" (material) and "virtual" (online) events are interconnected (Jurgenson, 2012), where the digital environment is the natural environment for the deployment of social processes.
- The need to study augmented social reality, especially its "virtual" component, caused the migration of sociological research from the "real world" to the "virtual world of the Internet", from offline to online (Rogers, 2009).

This migration provoked, firstly, the need to adapt traditional methods of obtaining sociological information (surveys, interviews, observations, etc.) to the digital environment, and secondly, the need to develop new methods based on using the capabilities of the digital environment to obtain data on various social phenomena. Under "virtual" (or "digitized») we began to understand research methods, moved from the usual reality (Offline) to the virtual reality of the Internet (i.e. online) and adapted to the online specifics of Internet research.

In contrast, digital methods are methods that were originally developed for the digital environment, fully utilize the capabilities of the Internet and cannot be used outside of it. Digital methods are unique in way that they are embedded in online devices (Google, Facebook, Twitter, etc.) And are able to use both data and computing capabilities of online platforms. At the same time, the Internet is not only a source of data, but also a tool for its research (Rogers, 2009; Rogers, 2019; Venturini and Latour, 2009; Venturini and Rogers, 2019). The main problem that arises when using digital methods is determined by the difficulties in separating the studied social phenomena from the specifics of the media in which they are observed (as we know, the media is a notification).

The use of digital methods is based on the strategy of "following the environment" (Rogers, 2019; Venturini and Bounegru, 2018), the development of which is due to the specifics of digital data, which is inextricably linked both with the medium where they were created and with the methods by which they can be studied.

The first stage is data selection. Its goal is to select from the vast amount of available data what can help to achieve the research goal. Firstly, it is necessary to analyze the adequacy of the source that will be used to obtain data regarding the research object, it is necessary to identify the extent to which the observed phenomenon is widespread in the environment which used to study it. For example, if you are investigating a community, you should find out whether it is online, whether its members use a specific platform, space, or device from which data can be collected (Twitter, Facebook, a website, or a mobile app).Secondly, it is necessary to find out the availability of data containing information about the social phenomenon being studied (search queries, hyperlinks, tags, retweets, URLs, user IP addresses, timestamps, events, promotions, etc.), and determine the level of completeness of the information contained in them.

The second stage of the "following the environment" strategy is the selection of analysis tools. At this stage, there is need to find out which tools (i.e. Internet devices) process the selected objects, and select those that can serve as analysis.

The third stage is re - profiling of certain devices, taking into account the specifics of the research object. It should be taken into account that research based on digital methods is a kind of secondary research, since the data (digital traces) that serve as an empirical base were not created for the purpose of analysis, so their application requires non-trivial operationalization and coordination of research issues with digital media.

In the fourth stage the question is to the validity of conclusions made on the basis of online data. In other words, there is a need to check them using traditional offline methods.

The leader in digital methods development is The leading European research group "The Digital Methods Initiative" (DMI), which consists of new media researchers developing methods and tools for repurposing Internet devices and platforms (such as Twitter, Facebook, Google) for research on social and political issues.

The emergence of digital methods is the result of the evolution of online methods of environmental research, that is, they fully use the capabilities of online tools for the purposes of sociological research due to their repurposing. Repurposing (secondary use) is a very important concept in digital methods. Its essence can be explained by the example of such a common device as Google Analytics, created to help webmasters analyze Internet sites, for example, to view the presence of links to their pages from other resources and optimize their visibility (primary use). However, search queries can be considered much more broadly than webmasters do.

Search queries contain signs of social problems and are markers of social dynamics. Their study makes it possible to move from the study of the digital environment to the use of Internet devices for studying society (Marres, 2017; Rogers, 2019). So, today digital methods are the most developed and tested methodology for studying modern society based on big data, which is confirmed by numerous publications, for example (Van Es and Schäfer, 2017).

The impact of big Data on society is usually described through the success stories of startups that implement Big Data technologies. However, the impact of big data is not limited to this. Big data pioneer Alex Petland wrote: "with Big Data, we can start to really look at the details of social interaction and how they play out and are no longer limited to averages like market indexes or election results. This is an amazing shift. The ability to see the details of events, political revolutions and be able to anticipate and control them – this, of course, can be compared to the gift of Prometheus, which can be used for both good and evil" (Pentland, 2014).

It is clear that without proper transparency, accountability, and control, big data systems can be used in ways that violate civil rights. The key here is to recognize that with big data, the adverse consequences may not even be deliberate abuse, but simply a "system glitch." At the moment, citizens do not have the information or resources necessary for meaningful interaction with these changes that are already taking place.

The example of China, which has been introducing a social rating system since 2014, which is also called the social credit system, is simply appalling. According to the "program for creating a social credit system", by 2020 the country should launch a system for evaluating each citizen, it will work in real time, and the results of its work should be published on the Internet. The program's stated goal is to build a society in which "being honest and trustworthy becomes prestigious and desirable."

The evaluation criteria are chosen by the state authority, which has the right to determine whether you should be trusted. In addition, the rating will be public information that will determine whether you get a loan, a new position, the ability to travel and even go on a date with you or not. The algorithm for calculating the rating is not fully disclosed, but 5 main factors affecting the rating are published:

1) Credit history;
2) Ability to fulfill obligations;
3) Verified personal data, such as address or mobile phone number;
4) Personal preferences and behavior;
5) Relations between citizens

3.1.4 The essence of predictive analytics

To be successful, companies need to forecast and analyze the market, but the abundance of data makes it harder to identify patterns. This is what artificial intelligence helps: thanks to its ability to learn, it can analyze large amounts of data and make high – quality forecasts.

Predictive Analytics allows you to use statistical methods of data mining, game theory, and analyze current and historical facts to make predictions about future events.

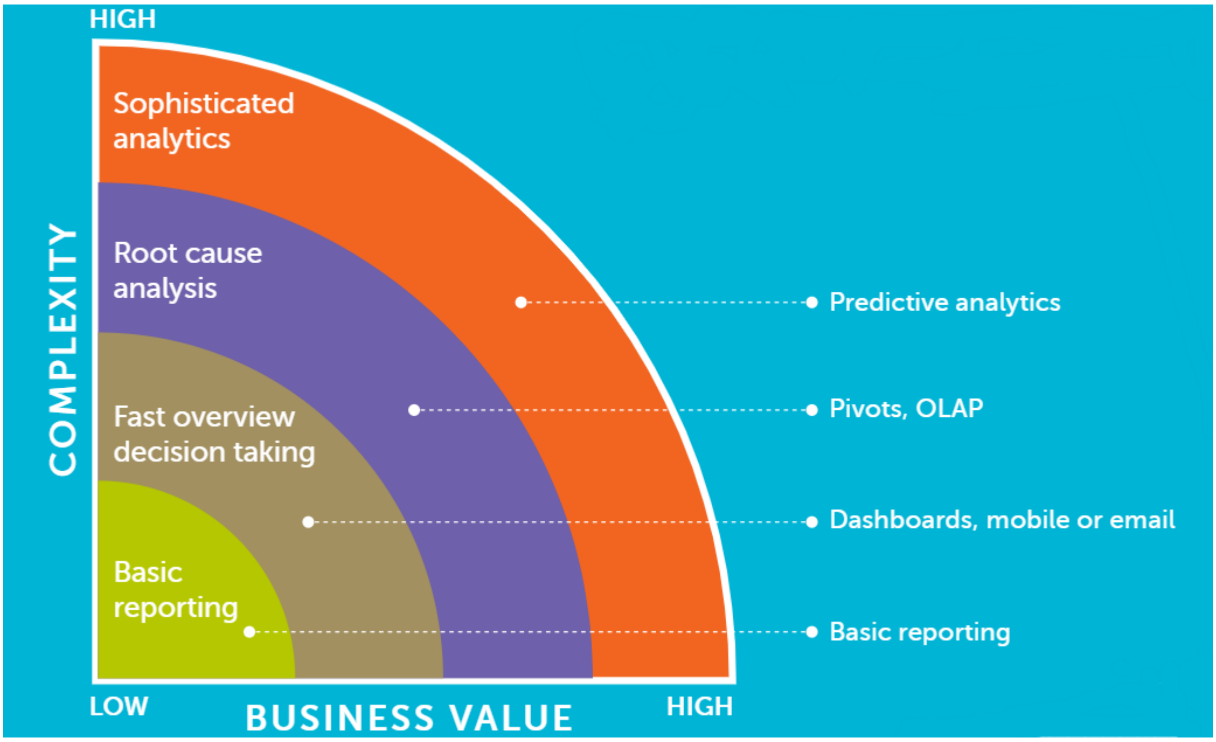

Figure 2. Predictive analysis in business analysis. Source: https://trends.rbc.ru/trends/futurology/5ecd6a7e9a79475b696df28c

Predictive Analytics service. Predictive Analytics is based on automatic search for patterns, relationships, and anomalies between various factors.

How predictive modeling works[6]:

- The model is built based on the identified patterns in the analysis of historical events;
- Based on the identified patterns, the forecast of future processes is made;
- Based on the forecast made, measures are taken to optimize business processes.

Companies often use advanced technologies to get the most accurate forecasts. The accuracy of the results guarantees a significant amount of data used, regardless of the source. The process can use the SQL, Excel, SalesWorks, CSV, etc. platforms. This way, the highest level of Analytics is achieved with the most accurate forecast.

Main features of predictive Analytics. The ability to make fairly accurate forecasts provides virtually unlimited business opportunities:

- Identification of implicit patterns;
- Carrying out of ranging of clients;
- Conducting product segmentation;
- Building a multi-factor sales forecast in the long and short-term periods;
- Ability to build multiple models of the development of events;
- Search for the target audience;
- Building a forecast of changes in the customer base;
- Building a set of "what if" / "what if" forecast models;
- Calculating the impact of competitors on pricing;
- Analysis of the impact of various factors, mainly key ones, on sales with subsequent forecasting of future impact; forecasting price elasticity;
- Analysis of promotional sales;
- Search for manipulations;
- Forecasting the appearance of new product items accounting for seasonality, temperature conditions, weather conditions;
- Forecasting demand in different time frames;
- Accounting for product substitution.

Predictive Analytics has found its application in many areas of business and life:

Direct marketing: the goal is to increase the number of reviews for integrating customer data from various open Internet sources.

[6] Softserve Business Systems. URL: https://www.softservebs.com/uk/solutions-2/prediktivna-analitika/

Predicting the effectiveness of promotional campaigns by segmenting customers. Predicting the effectiveness of advertising content.

Detection of fraud in insurance and loan obtaining, since the client's behavior model is accurately calculated.

Investment risk management and forecasting.

Predictive Analytics is able to assess the potential of any startup and assets with a high degree of probability. This method is effectively used by companies when choosing partners.

Predicting the situation can predict the possible departure of the client and identify the reasons for this behavior. Knowing this information will help correct the situation. Recommendation services: based on users ' preview of product groups, you can recommend other products. In education, the analysis of various teaching methods can reveal the most optimal in each case. In political campaigns, you can model the voting process. Based on the analysis of various factors, the patient's propensity to various diseases, the occurrence of which is provoked by lifestyle, can be predicted. In insurance, lending, you can determine the exact amount of coverage.

Advantages of predictive Analytics predictive Analytics allows you to reduce risks; Optimize resources; Improve the company's efficiency in conditions when other market participants are experiencing a crisis or you can simply stabilize the situation in the company; Increase profits by meeting the needs of customers as much as possible, since they have their own opinion about the product; Increase competitiveness; Optimize operational activities; Improve and simplify the decision-making process. Predictive Analytics has now become an affordable tool for different levels of business, providing the company's needs with an accurate and valuable forecast.

3.2 Literature review of happiness related studies

3.2.1 Happiness level and its definition

Questions of happiness have always worried philosophers since antiquity. Throughout human history, a huge number of works have been created and a number of theories have been derived that try to understand the essence of happiness and find ways that can lead a person to it.

The first paradigm in the framework of the consideration of modern concepts of happiness is the theory of goals and needs (Syrovezhkin A, 2012). These theories say that happiness can be achieved only if the need is satisfied, if a person achieves his goal. The theory of needs speaks about the presence of innate and acquired needs in a person. Some of them he is aware of – some of them he is not. In goal theory, on the contrary, we are talking about a conscious selective attitude to desires, the choice of certain goals to achieve, exactly those that will make a person happy. Here we must not forget about the conflict of small and global desires, as it can make the achievement of happiness fundamentally impossible for a person. Brief periods of happiness from current tasks are often unable to cover the general dissatisfaction from delaying the implementation of global plans. Another obstacle to achieving happiness in this paradigm lies in the lack of human resources to achieve the set goals (for example, if the set threshold was initially too high).

The idea of the second paradigm, which we consider in this review, is to achieve happiness through maintaining a balance of pain and pleasure in a person's life (Guidi M., 2007). Within this group of theories, any human need is directly caused by a lack of something. Accordingly, the higher the overall dissatisfaction with life, the more happiness compensatory satisfaction of needs brings. Moreover, there is also a symmetrical contrasting ability of a person to experience emotions – the more he is able to feel happiness, the more pronounced he can feel unhappy. Finally, this theory is supported by the difference in the level of positive and negative emotions, depending on the personal significance of a particular event in a person's life.

In another approach, activity theories, happiness is defined as a product that accompanies human activity. Here we find a fundamentally different approach to happiness – it is perceived not as a goal or final achievement, but as a process. American Professor of Psychology Csikszentmihaly writes that satisfaction from an activity is achieved only if individual abilities correspond to the abilities necessary to perform this activity (Bukayas

T.,1995). The feeling of happiness arises precisely when the business that a person is engaged in is not too difficult and not too easy for him, when he finds it fascinating and interesting.

Another set of theories about happiness is based on the skill of associative thinking in humans (Sysoeva T., 2004). An important role here is played by the phenomenon of conditioned evocation of emotions. According to this approach, the feeling of happiness is formed from a chain of associations and memories that the person himself causes in connection with a particular event. In the case of the habit of forming positive groups of memories, a person perceives a greater number of events in his life as pleasant, and, accordingly, more often feels happy. Another way to increase the individual level of happiness is to control negative thoughts and consciously try to reduce them. Proponents of this paradigm even talk about a peculiar habit of being happy, which consists in a priori waiting for future events as positive.

Another theory describing the phenomenon of happiness speaks of its relativity. Veenhoven (1991) characterizes its essence as follows. According to this paradigm, the level of happiness of a person depends not on objective well-being, but on a comparative subjective position in relation to other people. Based on this, the concept of happiness from a research point of view seems quite useless, on the one hand, since it does not fully correlate with objective indicators of a successful life, and on the other – too biased and half – hearted-due to the constant change in the standards of happiness for the person himself. That is why, despite the compliance with common sense and logic, in practice, this theory rarely forms the basis of scientific and empirical research. According to Veenhoven, despite the fact that at the individual level, each of us improves our life situation in order to become happier, at the generalized, collective level, people still need the state and expect from it guarantees of legal and social security, economic well-being, in order to maximize their own comfort and make their own lives more satisfactory.

From the above, it follows logically that, within the framework of this approach, the assessment of the level of happiness "in general" consists of two components: the level of subjective feeling of well-being/satisfaction with life and the ratio of oneself with various parameters and accepted assessments of success, well-being and well-being. The affective component (the hedonic level of happiness) is the positive experience of a person – everything that brings him pleasure; as a cognitive one - how much his achievements and achievements are evaluated by others, how he considers them, how they are ranked in the society around him.

Despite the presence of large body of literature describing the happiness, within the socio-economic literature, there is a self-reflective assessment of the level of happiness, also referred to as" life satisfaction "and" subjective well-being" (Shmotkin D., 2005). The latter formulation of the concept is most common. It describes how a person evaluates their own life. This estimated value can reflect both directly judgments about the perception of life, and the personal ratio of positive and negative in the perception of things in general (the notorious pessimists and optimists). In itself, "subjective well-being" is a complex construct that contains both cognitive and affective components (Diener, E., 1994). The terms "contentment" and "happiness" appear in the literature when it comes to "well-being" in general, as a certain level or meaning, without highlighting each individual layer of elements.

A common basic definition of happiness in social and economic research is given by Dutch scientist Ruut Veennhoven, head of the World Database of Happiness and founder of the Journal of Happiness Studies. He describes this phenomenon as "the degree to which an individual evaluates the overall state of their life as positive." In other words – how much a person likes the life he lives. Directly, the term "happiness "is the equivalent of the term" subjective well-being", referring the researcher to the psychological and emotional side of the issue, along with all the others. In other words, happiness characterizes a person's attitude to their own life, its subjective perception.

Another fundamental approach to defining happiness is to distinguish it from subjective well-being. One of the most significant works devoted to the phenomenon of happiness and its relation to the concept of well-being is the book "The pursuit of unhappiness" by David Haybron (2008): The elusive psychology of well-being”. In this work, he insists that happiness cannot be associated with pleasure, because the latter is too illusory and vague in its psychological effects. Life satisfaction also does not fully correspond to the concept of happiness, since it refers to the assessment of life as a whole. In addition, happiness, of course, is a long-term state, and people evaluate their lives most often at this particular moment, and these estimates are extremely influenced by situational factors.

Haybron, on the other hand, puts forward the idea that happiness should be regarded strictly as an emotional phenomenon. In other words, happiness in his opinion is determined by the general emotional state of a person during a certain period of life (p. 109). The author believes that happiness in this sense correlates with well-being from two sides. On the one hand, a high level of happiness seems to be a fairly reliable indicator that a person's life is going well. On the other hand, the true value of happiness is shown in its essential contribution to the self-fulfillment of the emotional part of human life. The importance of happiness is determined by the fact that the nature of each person is undeniably emotional, to a greater or lesser extent. Highborn claims that this emotional part consists of a set of presets about the need to be happy under certain circumstances. However, happiness can act as a form of life fulfillment only if it does not depend on the values instilled in a person in the course of third-party manipulations, false beliefs and affective states. Such inauthentic happiness is not a material for self-fulfillment, since it does not reflect the real essence of a person, his aspirations and feelings, although it brings a certain pleasure.

Denis McManus, in his review of this book (Phylosophy Magazine, October 2009), identifies three controversial, from his point of view, points covered in this work. The first is that the emotional nature of a person does not necessarily consist in achieving happiness under certain circumstances, since this completely takes the rest of the spectrum of emotions out of play and oversimplifies and unifies human nature. The second is addressed to Haybron's rejection of sujectivism, although his approach to the definition of happiness automatically introduces an element of it into his theory. Otherwise, there may be a feeling that the concept of happiness as an emotional state is at odds with the feeling of life, its perception. Finally, the third is related to the self-fulfillment and authenticity of happiness, namely, the fact that inauthentic happiness cannot fill a person's emotional life. McManus raises the quite reasonable question of whether a person distinguishes between the real conditions under which he should be happy and those that have been instilled in him in some way, and where the boundary between them actually lies.

Another serious attempt to draw a conceptual distinction between the concepts of happiness and well-being is made in one of the works of Jason Raibley (2012). In many ways, it draws on the work of Haybron: partly supporting it, partly criticizing it. The author appeals to several main aspects. The first of them refers to the philosophical doctrines of happiness, namely, the division of such into episodic, (or "feeling of happiness", "hedony" in philosophical literature), and happiness as a certain attribute of personality. Episodic happiness can be recorded physiologically - at the level of measuring hormonal and neurological indicators. It is described by the theory of "objective happiness" of Kahneman (1999), the work of Davis (1981), Sumner (1999), and so on. This type of happiness is extremely dependent on time and event fluctuations. Attributive happiness is more stable and much less amenable to operationalization and measurement.

Raibley writes that regarding the concept of subjective well-being, philosophers are much more uniform in their point of view: a high level of subjective well-being is observed precisely when life goes well for a particular person in a certain way. At the same time, there is a fundamental difference between the assessment of life in terms of its well-being and its emotional assessment. The latter is necessary, because no matter how highly the quality of a person's life is evaluated from the outside, for him such a life can be unbearable.

Thus, having considered a wide range of existing interpretations of the phenomenon of happiness and the theories that explain it, we come to the need to define this concept within the framework of a specific research project. We certainly find the concepts of Haybron and Raibley interesting in the light of the fact that they distinguish between happiness and subjective well-being. This division is also observed in the European Social Survey, since there are two separate questions for life satisfaction and happiness. Moreover, the structure of the questionnaire and the very wording of the question allows us to say that in this paper, the term "happiness" will be understood as attributive happiness in the interpretation of James Raibley. Thus, by happiness we mean a person's emotional assessment of their own life over a long period of time and their personal moral perception of the events that happen to them. This concept is complimentary, but not identical to subjective well-being and life satisfaction.

3.2.2 Approaches to measuring happiness in the economic and social studies

The authors have repeatedly discussed not only the definition of happiness, but also its measurement. Researchers ask whether happiness is measurable in principle – and if so, how. In particular, the researchers were interested in whether it is really possible to measure happiness objectively using questionnaires, or whether we can only get subjective indicators. Is the questionnaire the only measurement method? Finally, it was not clear whether they actually had an idea of a certain level of satisfaction/positive perception of their own life, and whether their answers to the question adequately reflected this idea. Ruut Veenhoven (1991) argues that the vast body of empirical research devoted to these questions allows us to answer them in many ways at the moment.

First, he points out that the objective measurement in the social sciences is different from that in the exact sciences, as measurement of the temperature, for example. The reason for this lies in the fact that the real perception of life is only partially reflected in the social behavior of a person. Moreover, such attributes of happiness, such as, say, a joyful appearance, are certainly more common among happy people, but can also be recorded in unhappy people. Even body language is not recognized as the most reliable indicator (Noelle-Neumann, 1977). Therefore, observation, as a method of measuring happiness, is not so reliable.

Another method is the self-assessment of the level of personal happiness of the respondent, which is expressed in various kinds of answers to questions - both direct and indirect, during an anonymous questionnaire or a personal interview. Despite the fact that the validity of this method was doubted by many scientists, empirical studies have shown that it is quite reliable. One of the controversial points described above-namely, whether respondents have their own opinion about the perception of their life as a whole-is leveled by a fairly stable distribution of people's responses in relation to their satisfaction with life, if they are at least partially tied to a certain time period and typical, approved answers, on the contrary, are quite rare. There is also a fairly common stereotype that people imagine themselves to be happier than they actually are, however, it is not confirmed in practice.

Another serious aspect in assessing happiness is its exposure to situational influences – the wording of the question, mood, weather, morning news, and so on. This is one of the few serious shortcomings in the assessment of happiness, identified and confirmed in the course of many empirical studies. However, as practice shows, the measurement bias data is a random error and, in fact, disappears with large sample sizes. Although there is a more systematic measurement error. It is caused directly by the wording of the question, the response formats, the special sequence of topics in the interview – in other words, poorly developed research tools. According to the estimates of Andrews and Witney (1976), this error accounts for up to 50% of the response spread in the happiness studies. There are several reasons for such a high sensitivity of the studied indicator to the research tool. On the one hand, even a person who has a certain idea of their own level of happiness is not always able to correlate it with a ten-point scale, which means that their answers can vary even with a constant level of happiness. On the other hand, the process of analyzing the results already obtained also depends on the subjective perception of the researcher. Thus, we get the risk of double data distortion.

In the economic and social sciences, there are two main traditions of the scale measurement of happiness. Psychologists approach the assessment of happiness in a quantitative way, assuming that the difference in the assessment between the individual values of the scale are equal for different respondents (for example, 4 differs from 5 as much as 8 differs from 9). In the framework of economic research, this approach to assessing happiness seems unreliable (Ng Y, 1997). Much more common is the ordinal assessment of the level of individual happiness, in which we can only track the relative dependence, which does not differ in a single interpretation for all respondents.

Hence, the most important, primary unobservable factors are the individual characteristics of the respondents. The undeniable advantage of quantifying the level of happiness is the leveling of the time effect with a linear specification of the model. On the other hand, most studies in the field of economics resort to models with latent variables, in which a change in the primary variables leads to a bias in the estimated results and an absentee assumption that there is no influence of static factors in the economic analysis of the individual level of happiness.

Ferrer-i-Carbonel and Frijters compared the results of various models based on both "psychological" and "economic" principles. Their conclusions are based on the analysis of the results of the influence of variables found in the vast majority of models: age, income, partner availability, number of children in the household, health status. In order to reduce the influence of other factors included in the models, the authors tested each of them on the data of the German Socio-Economic Panel (GSOEP) and adjusted the compared parameters. The main conclusion of the researchers was that the choice of a particular type of model leads to a small difference, while timeless factors related to variables are extremely important in explaining happiness.

Speaking about the methodology for assessing the level of happiness, it is impossible to ignore the problem of correctly posing the question in the questionnaire. In the sociological and psychological traditions, it is customary to use questions about the subjective assessment of the respondent's own level of happiness. As part of the European Social Survey (wave 2010), this question is as follows:

Considering all aspects of your life, how happy are you? For the answer, please use the card. / MARK ONE NUMBER/

0 1 2 3 4 5 6 7 8 9 10

Very unhappy Very happy

(88) I find it difficult to answer

Questions of this type are usually referred to as questions about General Satisfaction with Life (General Satisfaction questions). Within the framework of the European Social

Survey, the respondent is given a choice of 11 numerical categories with a verbal interpretation of the polar points – "0" as "Very unhappy" and "10" as "Very happy". Other possible options for the palette of answers to this type of question can be 7 or 5 categorical scales with or without verbal indicators, (for example, "very happy-happy-feel neutral-partly unhappy-very unhappy"). Regardless of the form, the result of this measurement is a categorical assessment of the quality of human life.

The basic assumption that allows us to talk about the admissibility of this question and the validity of the data collected in this way is the idea that the respondents share the same idea of the state of happiness. No matter how absurd this assumption may seem from the point of view of the layman, the results of research by psychologists allow us to speak about its validity.

On the one hand, psychological research shows that people are able to recognize and predict a wide range of human emotions (Diener et al, 1999), including how satisfied or happy other people are. The application of this assumption to cross-cultural communications is controversial in this case. There are both theories that suggest the existence of a single language of emotions for people regardless of their culture (for example, in the works of Paul Ekman, 2010), and works that indicate the significant role of culture (the works of Earley Hoschild, and others, 2006). However, in this study, we will stick to Paul Ekman's point of view about the comparative universality of emotions, and therefore happiness.

On the other hand, the study of B. Van Praag (1991) showed that people who belong to the same language environment share a sufficiently uniform measurement of internal sensations with an abstract digital scale. In other words, they evaluate certain states of human emotions in the same way relative to others, which makes communication between people possible. These two assumptions allow for further analysis on data collected in this form. And the construction of regression models based on them.

Finally, the comparability of the distances between different scale divisions is extremely important and ambiguous. That is, for an adequate and competent construction of analytical models, it is necessary that on the abstract scale of coefficients, the distance between two and three is equal to the distance between five and six and between nine and ten. The need for this is also due to the need to smooth out the trends towards polar responses, in the case of which the difference between the groups of average values is not large. There are scientific studies, the results of which indicate in favor of this interpretation of the distances between the divisions of the scale. Thus, Schwartz (1995) states that when filling out the questionnaire, the respondent most often tries to guess what exactly the researcher would have asked them if he had spoken to him personally. The graphical representation of the scale of the answer to the question in this case quite clearly forms the respondent's idea of equal distances. Although it is equally true that the numbers on such a scale are perceived in an ordinal format and are recognized by respondents only at the "more-less" level, however, we will adhere to the point of view about equal distances set out above.

When modeling the level of happiness, we assume to take into account the similar nature of the variable we use in the analysis. Since within our database, the variable can take values from 0 to 10, (11 different gradations, the distance between which we think of as equal), we have the right to use the linear regression model for modeling.

3.2.3 Determinants of happiness

One of the most important aspects of the study of happiness is the search for the most favorable conditions for people to experience this feeling. Overly generalizing, it can be concluded that studies of Western societies create the following portrait of a happy person. They are most likely to live in an economically developed country with a high level of democratization and freedom, most likely belong to an ethnic majority, are married, and get along well with friends and relatives. It is very likely that he is healthy, physically active, open to various ideas, feels control over his life, puts moral and ethical values above material ones (Veenhoven, 1991). It is unlikely that such a general description is suitable for use in real life and solving real problems. That is why more specialized studies describing this or that aspect in detail are of much greater interest.

An increasing number of studies confirm that the individual level of happiness is largely influenced by genetic and biological factors. Neurologists have found that the level of the so - called "happiness hormones"- dopamine and seratonin, (which are closely related to self-referential indicators of happiness), is largely determined by genetic predisposition (Hamer, 1996). This conclusion is confirmed by the results of studies of the level of happiness among genetically identical twins. Regardless of whether they were brought up together or separately, whether their life experience is similar or different, or how different their current life situation is, this type of twin, unlike twins, invariably showed a similar level of happiness. The influence of the genetic factor on the level of happiness, therefore, is extremely high. The importance of this conclusion is largely due to the fact that previous studies of happiness have shown that due to variations in income, fields of activity, differences in gender identity, an extremely small percentage of the variance in its level is explained-only about 4%, while more than half is explained due to the genetic factor (Inglehart and Klingemann, 2000). The conclusion is that the biological determinant is the most powerful and influential.

However, Inglehart and Klingmann in their article prove that cultural and historical factors also have very strong positions. Thus, they note that cultural diversity is a phenomenon that is practically not manifested within one community, but is clearly noticeable between several. At the same time, studies of the relationship between genetics and happiness are most often conducted within the same community/country, where the level of happiness a priori fluctuates within certain limits peculiar to this country. At the moment when there is a transition from intra-country to inter-country analysis, the influence of genetic and socio-cultural factors becomes approximately equal.

One of the main determinants of happiness is considered to be a person's income, which is discussed in scientific circles. Some of the earliest studies have shown that an increase in income can have a positive effect on life satisfaction in the short term (Abramovitz, 1959). But in the long term, this effect disappears quite quickly due to the habituation of a person to a new situation, and the level of happiness of a person returns to the previous equilibrium point. This is confirmed by the fact that the level of happiness in many countries has remained at the same level for several decades, despite their significant economic growth (Dorn, 2007). Moreover, the study conducted by Easterlin, back in 1974, shows a positive relationship between income and happiness at one point in time but not over time. It's known as the Easterlin paradox or happiness paradox. Income aspiration (Easterlin, 2006), relative income (Blanchflower and Oswald, 2004; Clark et al., 2008), or income distinction are all potential causes of the Easterlin Paradox. Some researchers discovered a strong association between happiness and income as well (Blanchflower and Oswald, 2004; Rahayu, 2016). Some authors, on the other hand, believe that income will continue to rise until it reaches a certain point, after which it will begin to decline or stay unchanged. According to Stevenson and Wolfers (2013), the presence of a threshold income is widely believed despite the absence of observational data.

Another aspect that consistently attracts the attention of researchers is the relationship between the level of happiness and age. Most studies are conducted using cross-sectional data, which includes a one-time survey of various people of different ages. Some of them do not show any connection between these variables, others say that there is a weak positive relationship (Hansson et al, 2005), others show a weak inverse relationship (Chen, 2001), and finally, the fourth leads to conclusions about the existence of a nonlinear relationship with a peak of satisfaction in 30-40 years (Easterlin, 2006). Probably the most influential works devoted to this topic over the past decade belong to the authorship of Brandflower and Oswald (2004), who established the presence of a U-shaped relationship between age and the level of happiness. Initially, their findings related to the United Kingdom and the United States, but were later expanded to more countries. According to their findings, all other things being equal, the level of life happiness of a typical person reaches its minimum in middle age. Later, this conclusion was questioned due to the fact that the result obtained could be an artifact caused by a set of included covariates or using cross-sectional data and the intervention of the cohort effect. The study with the same purpose have been conducted by Morgan and Thompson (2015), who look at the happiness-age relationship in 29 European countries (N = 46,301) to see how it is moderated by national income, as measured by GDP per capita. Only the most developed countries' eudemonic and hedonic happiness remained largely constant over time; in poorer countries, there was either a fluctuating or gradual age-related decline. These studies challenge into question the happiness-age relationship's cultural universality and propose that theories of how age affects happiness should incorporate a socioeconomic level of analysis. However, in most of the studies the variable of ages is used as a control variable.

Also, a large number of works were devoted to the search for the connection between gender and fluctuations in the level of happiness. All of them found only a slight difference in overall life satisfaction between men and women. As in the case of age, belonging to a particular gender affects more indirectly than directly. For example, studies conducted in Taiwan have shown that gender in combination with age has an impact on the level of happiness through social support provided to a person (Lu, 1997). Another study that measured happiness using the Oxford Happiness Inventory among teenagers, students, and employees in Kuwait found that in the last two groups, men were significantly more satisfied with life than women (Abdel-Khalek, 2003).

Another study, also conducted in Kuwait by Ahmed Abdel-Khalek, sought to find a link between happiness, religiosity, and health – both mental and physiological (Abdel-Khalek, 2006). The results obtained in the study of a group of students showed that happiness is largely associated with a person's psychological health, while religiosity has a much smaller impact on life satisfaction. The subjective assessment of physical health allows us to explain an extremely small percentage of changes in the level of happiness of a person. These conclusions are true for both boys and girls. Religiosity can be also connected with the discrimination factor. Green and Elliot (2010) discovered a connection between health and happiness as well, using an intermediate variable. They discovered that people who are more religious are healthier and happier. According to Singer et al. (2007), decreasing in health with increasing age does not make anyone sad due to emotional maturity. Unemployment can lead to a decline in health, which is exacerbated for non-volunteers.

Marital status is one more variable of interest among the researchers. One of such studies have been conducted by Hori and Kamo (2018). The research tests the hypotheses about gender, marital status and social support, and the findings suggest that, in East Asian countries, marital status is a good predictor of happiness, especially for men, but not always for women. Regarding other countries, the variable of the marital status is often used as a control variable to control the differences between the respondents.

Education is also considered as one of the determinants in the studies. Typically, there is a positive relationship between education and happiness (Castriota, 2006). According to Michalos (2008) education does not have a direct impact on satisfaction. It has an indirect channel by social capital (network) or self-confidence and self-esteem. Despite all of positive results, scientific research on the link between education and happiness remains inconclusive. Some authors, for example, Oswald (1997) find the opposite: after adjusting for wages, more educated people are less happy. The outcome may be influenced by two major factors. For instance, individuals with a higher degree have higher career expectations, which are more difficult to achieve. Being overqualified for a position causes annoyance. When a person holds a job that does not require the amount of education they have earned, they are referred to as “over-schooled”. While there are other advantages to education that contribute greatly to improving the standard of life of people and society as a whole, this can be considered wasteful and a waste of money. Second, as one becomes more educated, the wealth disparity widens. As opposed to those with the same education level but a higher income, there may be a negative impact.

Satisfaction with the life is another determinant of happiness. Happiness and life satisfaction are not interchangeable in sociology (Veenhoven, 1991). Happiness is described as an overall enjoyment of one's life as a whole, which makes people feel happy. These meanings are consistent with Bentham's description of satisfaction, which is defined as the accumulation of pleasures and pains. However, happiness differs from life satisfaction in psychology. Therefore, happiness can be positively influenced by this factor.

Taking into account the society itself and its influence on one particular person, another determinant can be find, which is a discrimination factor. While systemic stigma (discriminatory legislation, practices, and cultural attitudes) against sexual minorities is correlated with poor health and well-being studied within a single region, it leads to even less level of happiness investigating a number of countries (Pachankis and Bränström, 2018).

The connection between social capital and happiness can be seen in the community's confidence as well. According to Helliwell (2007), having more social capital reduces suicide and improves subjective well-being. Unhappiness was triggered by interpersonal distrust (Tokuda and Inoguchi, 2008). Relational assets that are not traded, such as confidence in people, membership, and trust in institutions, contribute to a higher level of life satisfaction (Sarracino, 2012).

Another social factor is the feeling of safety, which can relate to the happiness. Happiness level is supposed to grow with an increase of safety feeling. It signifies a higher degree of protection in a particular region. People have more safety to do some activities until late at night without doubts and worries (Rahayu, 2016). However, this factor is not widely used in the studies and its effect is still controversial.

So, in this section, we have given a brief overview of research on the studies for common determinants of happiness. In the next section, we will discuss in more detail the methods used in these studies.

3.2.4 Models used in the related articles

One of the latest research identifying the determinants of happiness is the study of Rahayu (2016). Ordinal Generalized Linear Model has been used to check the influence of the factors on happiness. Such factors as absolute income, relative income, perceived health vector, education level, social capital vector and vector of demographic characteristics have been used as the independent variables. Endogeneity has been challenged in this study due to the association between error and the independent variable. Endogeneity can contribute to parameter bias and inconsistency. Powdthavee (2008) uses instrumental variables to discuss income endogeneity. As instrumental factors in this analysis, side jobs and schooling years were included. To approximate parameters in the right hand side model, this analysis uses the Conditional Mixed Process (CMP) technique. Since the CMP approach could fix endogeneity in the estimation model, it was selected. CMP is suitable for two forms of estimation cases, according to Roodman (2011). Second, estimation in which a completely modeled and posited recursive data-generating method is posited. Second, estimation in which there is simultaneity but instruments allow, such as in two-stage least squares estimation (2SLS).

Moreover, cross-sectional data used in the research may cause a heteroskedasticity problem. However, it has been also controlled in this work. The standard errors are wrong and the metrics are skewed because the error variances are not the same in all situations (Williams, 2006). The Ordinal Generalized Linear Model (OGLM) is used in this analysis, which is able to estimate ologit, oprobit, and hetprob. The OGLM method is a general method for estimating a heterogeneous choice as well as a heteroscedastic ordered model. If this is the case, OGLM is an effective approach for dealing with the estimation model's heteroscedasticity problem.

However, the common multiple linear model and parameter estimation approaches such as ordinary least squares and instrumental variable were predominantly used in previous studies on the relationship between different factors and happiness (Gu and Wei, 2018). For example, Rana et al (2014) use simple linear regression to find out the independent contribution of forgiveness and its domain to happiness. To run simple linear regression analysis, the basic assumptions of linear relationship between the predictor (forgiveness) and criterion (happiness) were established. This is one of the main assumptions of the linear model. However, the authors mention that it is necessary to check that other assumptions are met as well. It can be done with the testing after the model estimation.

Using the discrete dependent variable of happiness level, Sekulova et al (2016) mention that an Ordered Probit model seems to be a better theoretical method for assuming that answers are only normally comparable since the indicator of subjective happiness used here is a discrete non-continuous number. Besides, Ozkan et al (2008) estimated the ordered probit and ordered logit models, as the most appropriate econometric techniques. Despite this, Sekulova et al (2016) in their research evaluated the data using a typical OLS multiple regression model, in which variance in a variety of measurable characteristics is used to describe happiness, due to the reason that both methods show the same results in terms of coefficients signs and significance. In happiness research, Ferrer-i-Carbonell and Frijters (2004) suggest that OLS requirements are almost as stable as Ordinary Probit. As a result, only the OLS regression results have been estimated.

Therefore, it has been decided to use multiple regression model for the analysis of the data as a predictive model, since it gives significant results if all the assumptions are met.

4 Practical Part

4.1 Description of dependent and independent variables included in the model

Particular attention should be paid to how the variables included in the model are calculated.

Seligman (2003) categorizes three common theories of happiness, i.e. Hedonism Theory, Desire Theory, and Objective List Theory. The theory of interest is Objective List Theory, which claims that happiness will come from the fulfillment of life goals such as material desires, freedom, health, education, knowledge, and friendship. Besides, there are numerous articles exploring different determinants of happiness that is why these variables have been collected in the table and classified according to the articles, they have been used in (Table 1).

The following variables have been included in the model: level of income of the respondent, level of satisfaction, discrimination existence, safety feeling, trust, GDP per capita, employment status, age and gender of respondent, his/her marital status, health, education level and if he/she religious or not. These variables are able to cover all the socio-economic factors used in other relative works. Two dependent variables are met in the literature to be a measurement of the happiness, which are happiness index (Chu et al, 2005; Lee et al, 2016; Berggren et al, 2017) and the question of the survey “how happy are you?” (Kye & Park, 2014; Rahayu, 2016; Hubert & Soni, 2017). However, as the dataset consists of the responses of particular people, to comply with the proportionality, it has been decided to choose the question “how happy are you?” as the dependent variable.

4.2 Independent variables description

Variable	Description	Articles
Income	The survey question of income level of the respondent, where 1 is the minimum level and 10 is the maximum level of income	Seligman (2003); Tian and Yang (2007); Rahayu (2016)
Satisfaction	Categorical variable based on the survey question “how satisfied are you?”, where 1 is “Not satisfied” and 5 is “Completely satisfied”	Sarracino (2012), Rahayu (2016), Hubert & Soni (2017)
Discrimination	Dummy variable based on the survey question “Are you discriminated?”, where is 0 “Not discriminated” and 1 is “Discriminated”	Shields & Wailoo (2002); Berggren et al (2017); Pachankis & Bränström (2018)
Safety	Categorical variable based on the survey question “Do you feel safe?”, where is 1 “Don’t feel safe” and 4 is “Feel completely safe”	Rahayu (2016), Musa et al (2018), Yoon (2018)
Trust	Categorical variable of trust formed from the section of trust questions	Sarracino (2012); Rahayu (2016)
GDP per capita	Numeric variable of GDP per capita of the country where respondent lives	Dipietro & Anoruo (2006); Morgan & Thompson (2015)
Unemployment	Dummy variable, which takes 1 if respondent is unemployed and 0 otherwise	Böckerman & Ilmakunnas (2006); Winkelmann (2014)
Age	Age of the respondent	Kye & Park (2014), Rahayu (2016); Eren & Aşıcı (2017)
Gender	Dummy variable of gender, where 1 is for male and 0 for female	Mookerjee & Beron (2005); Lobos et al (2016); Hori & Kamo (2018)
Marital status	Dummy variable of marital status, where 1 is for married people and 0 otherwise	Rogers & White (1998); Hori & Kamo (2018)

Health	Categorical variable based on the survey question “How healthy are you?”, where is 1 “Very unhealthy” and 5 is “Very healthy”	Cornelisse-Vermaat et al (2006); Kye & Park (2014); Rahayu (2016)
Education level	Categorical variable based on the survey question “What is the highest education level you have?”, where is 0 “Not completed ISCED level 1” and 6 is “ISCED 6, doctoral degree”	Chan et al (2005); Michalos (2008)
Religion	Dummy variable, which takes 1 if respondent belongs to particular religion and 0 otherwise	Mookerjee & Beron (2005); Hubert & Soni (2017)

Table 1. Independent variables descrition
Source: own elaboration

4.3 Hypothesis

As the main goal of this work is to identify the determinants of happiness index, the following hypothesis have been introduced:

- ***H_1: Happiness and income are positively related.***

  It is believed that with an increase of income a person feels happier, but only in one point of time according to the happiness paradox or Easterlin paradox.

- ***H_2: Happiness and satisfaction have positive relationship.***

  Satisfaction and happiness are considered to be synonyms in some articles. However, satisfaction is considered as a determinant of happiness as well. It is believed that with higher satisfaction with life, a person feels happier.

- ***H_3: With an increase of discrimination, there is a decrease of happiness.***

  Discrimination against different reasons still takes place in the modern world. Based on the results of happiness research of minorities (Berggren et al, 2017), it is assumed that with higher level of discrimination, a person becomes less happy.

- ***H4: Happiness and safety have positive relationship.***

  Higher safety will increase happiness, which may reflect higher security in a person's living area. People have more safety to do some activities until late at night without worry.

- ***H_5: Happiness and trust have positive relationship.***

  Non-market relational products such as individual trust and institutional trust make life satisfaction greater (Sarracino, 2012). Thus, the positive relationship between happiness and trust is considered.

- ***H_6: GDP per capita has a positive influence on happiness.***

  With the better economics, there is a higher level of trust to the government and higher the happiness index. Thus, the happiness of the particular person is influenced positively as well.

- ***H_7: Education and happiness have positive relationship.***

  In other research, it has been shown that education level does not affect happiness directly. It has an indirect channel through network (social capital) or through self-confidence and self-estimation. However, the influence of education should be considered, and it is assumed that it has a positive impact on happiness.

- ***H_8: Health and happiness have positive relationship.***

  Decrease in health leads to many other problems such as higher expenses, lack of activity and others. Thus, the happiness level will decrease as well. That is why health influences happiness level positively.

- ***H_9: Unemployment and happiness are negatively related.***

  Unemployment usually reduces life satisfaction, as it gives uncertainty and reduces income as well. Life satisfaction is related to happiness of the person that is why it is believed that unemployed people are less happy than the employed ones.

Other demographic variables such as age, gender, marital status, and religion are considered as the vector of control variables. However, additional results can be achieved for these variables as well.

4.4 Research method

The type of data used in this study is cross-sectional, and, accordingly, regression analysis of cross-sectional data was used in the work.

It is assumed that the variables may influence the dependent variable linearly. Thus, the multiple linear regression model was estimated.

$$\begin{gathered} Happiness_i = \beta_0 + \beta_1 \cdot Income_i + \beta_2 \cdot Satisfaction + \beta_3 \cdot Discrimination_i + \\ + \beta_4 \cdot Safety_i + \beta_5 \cdot Trust_i + \beta_6 \cdot GDP\ per\ capita_i + \beta_7 \cdot Unemployment_i + \\ \beta_8 \cdot Age_i + \beta_9 \cdot Gender_i + \beta_{10} \cdot Marital\ status_i + \beta_{11} \cdot Health_i + \\ \beta_{12} \cdot Education_i + \beta_{13} \cdot Religion_i \end{gathered} \tag{1}$$

The model was estimated with the help of the OLS method.

4.5 Preliminary data analysis

This research is based on survey data, which has a cross-sectional type. The number of observations is equal to 47086. The source of data is European Social Survey. Besides, the GDP per capita variable has been added from the source of Happy Planet Index. The sample includes 15 variables, which have been calculated from the given dataset and one of which is dependent variable. Model building and data analysis were performed in the SPSS software package.

It is necessary first to prepare the data for the study and clear the dataset of missing values. After the exclusion of missings the number of observations reduced to 30499.
For the better understanding of the data, the following description statistics for the numeric variables have been calculated:

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
GDP per capita	30499	5659,38	101563,70	34494,33	21867,91	,938	1,044
Age	30499	1	77	37,89	18,40	-,059	-,925
Trust	30499	0	30	15,86	5,959	-,433	-,127
Valid N (listwise)	30499						

Table 2. Descriptive Statistics
Source: own elaboration

As it can be seen, the average GDP per capita is equal to 34494 dollars. Besides, it can be seen that the age of the respondents varies from 1 to 77. The responses for the children have been given by their parents. The average respondent is 38 years old with the trust level of 16 from the range of values 0-30.

All of the distributions seem to be different from normal, based on the skewness and kurtosis values.

For the consideration of categorical and dummy variables the frequency tables have been built. Frequency table of income variable looks as follows:

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,0	407	1,3	1,3	1,3
	1,5	1080	3,5	3,5	4,9
	2,0	1612	5,3	5,3	10,2
	2,5	3553	11,6	11,6	21,8
	3,0	4074	13,4	13,4	35,2
	3,5	4231	13,9	13,9	49,0
	4,0	4131	13,5	13,5	62,6
	4,5	3892	12,8	12,8	75,3
	5,0	3219	10,6	10,6	85,9
	5,5	2736	9,0	9,0	94,9
	6,0	793	2,6	2,6	97,5
	6,5	358	1,2	1,2	98,6
	7,0	263	,9	,9	99,5
	7,5	50	,2	,2	99,7
	8,0	43	,1	,1	99,8
	8,5	41	,1	,1	99,9
	9,0	16	,1	,1	100,0
	Total	30499	100,0	100,0	

Table 3. Frequency table of income

Source: own elaboration

When measuring income in ESS, the decile approach have been applied. The categories in the variable income are national and are based on deciles of monthly income range in the given country. The median income is the reference point and the 10 deciles are calculated with the median itself at the top of the fifth decile.

As the missing values have been deleted from the table the Valid Percent column can be omitted. Cumulative Percent column can be also omitted, but only for dummy variables. The classes of income from 1 to 9 are estimated by the respondents and they give the answer to the question: “how do you estimate your level of income based on the scale from 1 to 9”. The most frequent class of income is 3.5 (13.9%), which can be considered as the people with average income. After this class, the number of people with better income reduces. 85.9 % of all observations are covered by the classed from 1 to 5, while people with higher income make up only 14.1 percent. Classes 8, 8.5, 9 have only 0.1% each from the whole dataset.

Frequency table for dummy variables has been created separately:

		Frequency	Percent
Gender	Female	16535	54,2
	Male	13964	45,8
	Total	30499	100,0
Married	No	29798	97,7
	Yes	701	2,3
	Total	30499	100,0
Religious	No	12413	40,7
	Yes	18086	59,3
	Total	30499	100,0
Unemployed	No	28904	94,8
	Yes	1595	5,2
	Total	30499	100,0
Discrimination	No	30467	99,9
	Yes	32	0,1
	Total	30499	100,0

Table 4. Frequency table for dummies
Source: own elaboration

By the table it is seen that there are more females in the dataset (54.2 % against 45.8% of males), most of the respondents are not married (97.7%). Besides, more than a half of respondents belong to any of religions (59.3%). The number of unemployed respondents is much lower than the employed ones (5.2% against 94.8% of employed). Most of the people are not discriminated and only 0.1% of the sample feel discrimination (32 observations).

It is necessary to consider dependent variable of happiness separately as well:

		Frequency	Percent	Cumulative Percent
Valid	0	202	,7	,7
	1	149	,5	1,2
	2	338	1,1	2,3
	3	628	2,1	4,3
	4	880	2,9	7,2
	5	2446	8,0	15,2
	6	2421	7,9	23,2
	7	5488	18,0	41,2
	8	8944	29,3	70,5
	9	5441	17,8	88,3
	10	3562	11,7	100,0
	Total	30499	100,0	

Table 5. Frequency table for happiness level

Source: own elaboration

Classes of happiness level have been estimated by the respondents as well. The respondents have been asked a question: "how happy are you based on the scale from 1 to 10". Value of 0 means that the respondent is extremely unhappy, while value of 10 means extreme happiness. Most of the respondents estimate their happiness level as 8, which is 29.3% and quite a good result. It is necessary to admit that the number of unhappy people in the sample is pretty low (15.2% of people with the happiness level 5 and lower).

The frequency tables of other variables are presented below:

		Frequency	Percent	Cumulative Percent
Valid	1	1708	5,6	5,6
	2	9119	29,9	35,5
	3	7155	23,5	59,0
	4	10197	33,4	92,4
	5	2320	7,6	100,0
	Total	30499	100,0	

Table 6. Frequency table for satisfaction

Source: own elaboration

		Frequency	Percent	Cumulative Percent
Valid	1	9190	30,1	30,1
	2	15480	50,8	80,9
	3	4794	15,7	96,6
	4	1035	3,4	100,0
	Total	30499	100,0	

Table 7. Frequency table for safety

Source: own elaboration

		Frequency	Percent	Cumulative Percent
Valid	Very bad	415	1,4	1,4
	Bad	2055	6,7	8,1
	Fair	8249	27,0	35,1
	Good	12867	42,2	77,3
	Very good	6913	22,7	100,0
	Total	30499	100,0	

Table 8. Frequency table for health

Source: own elaboration

		Frequency	Percent	Cumulative Percent
Valid	Not completed ISCED level 1	342	1,1	1,1
	ISCED 1	2172	7,1	8,2
	ISCED 2	4754	15,6	23,8
	ISCED 3	10840	35,5	59,4
	ISCED 4	2186	7,2	66,5
	ISCED 5	9830	32,2	98,8
	ISCED 6	375	1,2	100,0
	Total	30499	100,0	

Table 9. Frequency table for education

Source: own elaboration

The satisfaction level has been estimated by the respondents and it gives an answer for the question: "how satisfied are you based on the scale from 1 to 5". For the safety level the following question has been asked: "how safe do you feel based on the scale from 1 to 4". Based on these frequency tables, it can be concluded that the greatest percentage of the respondents are almost satisfied with their life (33.4%), but half of them do feel almost unsafe (50.8%). It is interesting to pay attention to the health and education of the respondents. Thus, 42.2% of people have good health and 35.5% have completed ISCED 3 education level.

One of the important steps is the construction of a correlation matrix to eliminate the problem of multicollinearity in the model. After constructing this matrix, it will be possible to draw a conclusion about the possibility of including certain variables in the model. The correlation matrix is shown in Table 10. Pearson correlation coefficient has been calculated.

	GDP	Income	Satisfaction	Safety	Trust	Age	Happy	Health	Education
GDP	1	,066**	-,004	-,183**	,354**	-,048**	,215**	,175**	,124**
Income	,066**	1	,003	-,108**	,133**	,023**	,174**	,112**	,244**
Satisfaction	-,004	,003	1	,021**	-,095**	,007	-,111**	-,047**	,067**
Safety	-,183**	-,108**	,021**	1	-,217**	,109**	-,218**	-,226**	-,140**
Trust	,354**	,133**	-,095**	-,217**	1	-,020**	,303**	,190**	,200**
Age	-,048**	,023**	,007	,109**	-,020**	1	-,124**	-,402**	-,178**
Happy	,215**	,174**	-,111**	-,218**	,303**	-,124**	1	,357**	,141**
Health	,175**	,112**	-,047**	-,226**	,190**	-,402**	,357**	1	,231**
Education	,124**	,244**	,067**	-,140**	,200**	-,178**	,141**	,231**	1

** - significance at 10% level

Table 10. Correlation matrix

Source: own elaboration

The independent variable of happiness does not have strong correlation coefficients with dependent variables. Its highest correlation is with health variable, as the value of correlation coefficient is 0.357. Regarding multicollinearity, there is no strong correlation obtained between the independent variables.

4.6 Empirical results

The multiple linear regression has been estimated. The following results have been obtained:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,470[a]	,221	,221	1,678

a. Predictors: (constant), Unemployment, Discrimination, Gender, Marital Status, Satisfaction, Health, Religious, Income, GDP per capita, Education, Safety, Trust, Age
b. Dependent Variable: Happy

Table 11. Summary model

Source: own elaboration

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig
1	Regression	24380,293	13	1875,407	666,344	,000[b]
	Residuals	85799,203	30485	2,814		
	Total	110179,496	30498			

Table 12. ANOVA test

Source: own elaboration

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound
1	(Constant)	4,616	,083		55,330	,000	4,452	4,780
	GDP per capita	7,134E-6	,000	,082	14,910	,000	,000	,000
	Income	,160	,008	,108	20,279	,000	,145	,176
	Satisfaction	-,137	,009	-,077	-15,095	,000	-,154	-,119
	Safety	-,253	,014	-,102	-18,708	,000	-,280	-,227
	Trust	,057	,002	,178	31,517	,000	,053	,060
	Age	-,001	,001	-,009	-1,631	,103	-,002	,000
	Discrimination	-,069	,099	-,004	-,697	,486	-,263	,125
	Gender	-,185	,020	-,049	-9,231	,000	-,224	-,146
	Married	-,282	,064	-,022	-4,396	,000	-,408	-,157
	Health	,560	,012	,270	46,529	,000	,537	,584
	Education	-,007	,008	-,005	-,871	,384	-,022	,008
	Religious	,081	,020	,021	4,031	,000	,042	,121
	Unemployed	-,412	,044	-,048	-9,423	,000	-,498	-,326

Table 13. Coefficients table

Source: own elaboration

In this specification, the coefficient of determination is 0.470, which means that the regression equation describes 47% of the variance of the endogenous variable; the remaining 53% are not explained. According to the results of the F-test, this model is significant at the significance level of 1%, as the probability value is 0, which means that the regression model predicts the dependent variable significantly well.

Let's interpret the coefficients in the model. Such variables as GDP per capita, Income, Satisfaction, Safety, Trust, Gender, Marital Status, Health, Religious and Unemployed are significant at 1 % significance level, as the probability value is less than 1% significance level. Constant is significant as well. However, Age, Discrimination and Education variables are not significant in this model. ***Thus, the hypotheses H_3 and H_7 can be rejected on this step***.

Each variable is considered separately further:

- *Income variable has a positive impact on the level of happiness based on the results of the model. The interpretation of its coefficient is as follows: with all other variables being equal, with an increase of income on 1 unit, on average the happiness level will increase on 0.160.*

  Hypothesis H_1 can be approved.

- *Satisfaction is negatively related to the level of happiness, which is quite an unpredictable result. Based on the coefficient value, with all other variables being equal, with an increase of satisfaction on 1 unit, on average the happiness level will decrease on 0.137.*

  Hypothesis H_2 should be rejected.

- *Safety has a negative impact on happiness level as well. Coefficient of -0.227 means that with all other variables being equal, with an increase of safety on 1 unit, on average the happiness level will decrease on 0.253.*

  Hypothesis H_4 should be rejected.

- *Trust is positively related to happiness level, as it has been considered before. With all other variables being equal, with an increase of trust on 1 unit, on average the happiness level will increase on 0.057.*
  Hypothesis H_5 can be approved.

- *GDP per capita has a positive impact on the happiness level. Coefficient of 7,134E-6 means that with all other variables being equal, with an increase of GDP per capita on 1 dollar, on average the happiness level will increase on 7,134E-6.*
  Hypothesis H_6 can be approved.

- *Feeling healthier makes a person happier. Without health problems people will be more productive so that they could earn more money or save more money rather than spending them on medication that makes them happier. The interpretation of its coefficient is as follows: with all other variables being equal, with an increase of health on 1 category up, on average the happiness level will increase on 0.553.* ***Hypothesis H_8 can be approved.***

- *The dummy-variable of unemployment influences happiness level negatively. The interpretation of the coefficient is as follows: compared with employed people, with all other things being equal the happiness level of unemployed people is 0.412 less.* ***Hypothesis H_9 can be approved.***

Additional results have been obtained as well. The dummy-variable of marital status influences happiness level negatively. The interpretation of the coefficient is as follows: compared with unmarried people, with all other things being equal the happiness level of married people is 0.282 less. Besides, gender variable has the negative impact as well. With all other variables being equal, compared with females the happiness level of males is 0.185 less. One more variable is significant, which is the relation to any of religions. Religion has a positive impact on the happiness level. With all other variables being equal, compared with people who do not relate to any religion the happiness level of religious people is 0.081 higher.

The constant of the model is significant as well. The intercept (often labeled the constant) is the expected mean value of the dependent variables when all independent variables are equal to zero. It can be interpreted in the following way: with all the variables equal to zero, the average happiness level is equal to 4.616. The interpretation of the intercept makes sense here, as means that a person without influence of any factors will estimate his level of happiness as 4 or 5.

It is necessary to check the results of the model and estimate its quality.

First, collinearity diagnostic should be done, as this problem appeared before in the research. Collinearity Statistics have been obtained:

Model		Collinearity Statistics	
		Tolerance	VIF
1	(Constant)		
	GDP per capita	,843	1,186
	Income	,908	1,102
	Satisfaction	,975	1,026
	Safety	,854	1,171
	Trust	,802	1,247
	Age	,780	1,282
	Discrimination	,998	1,002
	Gender	,925	1,081
	Married	,995	1,005
	Health	,756	1,322
	Education	,850	1,177
	Religious	,936	1,068
	Unemployed	,973	1,027

Table 14. Collinearity statistics

Source: own elaboration

It can be seen that the value of the VIF (Variation Inflation Factor) near each independent variable is less than 10, then the effect of multicollinearity is not observed, and the regression model is acceptable for further interpretation.

An important point is the analysis of the residuals, that is, the deviations of the observed values from the theoretically expected ones. The residuals must appear randomly (i.e., not systematically) and follow a normal distribution. The following histogram of the residuals has been built:

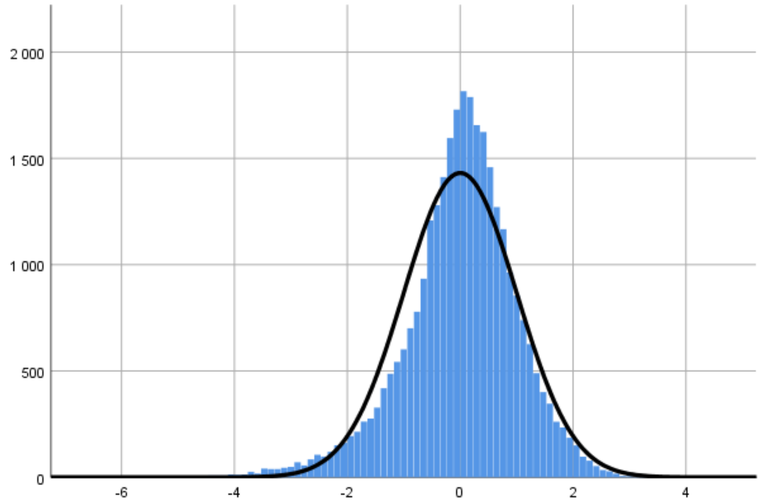

Figure 3: Histogram of the residuals

Source: own elaboration

Based on the histogram, it can be concluded that there is a good coherence between the histogram of the residuals and the normal distribution.

The study highlighted some limitations. Thus, the distributions of some of the variables do not meet the conditions of normality. Also, the number of observations has been reduced almost twice because of the existence of omitted values.

The elimination of these limitations may serve as a future direction of research. Another direction may be the application of a similar methodology used on the data of other countries or used on the country level on a greater scale to check the impact of the same factors on the happiness level.

Despite the above-mentioned limitations, the objective of the work was achieved. The determinants of happiness level have been identified. Four hypotheses have been approved, which are ***H_1, H_5, H_6, H_8, and H_9***. ***GDP per capita, Income, Satisfaction, Safety, Trust, Gender, Marital Status, Health, Religious and Unemployed can be considered as the determinants of happiness level.***

5 Conclusion

Thus, within the framework of this project, the main approaches to the definition and assessment of happiness were considered, a brief review of research on this topic was conducted, and their contribution was supplemented by own empirical research.

Among the results of this work, we can identify an evaluation of the influence of socio-economic factors on the level of human happiness, which has been measured by the respondents of the European Social Survey. Therefore, significant results have been obtained.

Nine hypotheses have been stated, among which five hypotheses have been approved, which are H1, H5, H6, H8, H9. To be more specific, at the individual level, the significant influence of such factors as the respondent's health, employment, and trust in society, income, and GDP per capita was determined. Significant results were also obtained regarding unconfirmed hypotheses, but they contradict the hypotheses put forward. Thus, the effect on a person's happiness is also provided by a person's satisfaction and his sense of safety. Additional and significant results have been obtained for marital status, gender and religion, which have effect on happiness as well. Therefore, GDP per capita, Income, Satisfaction, Safety, Trust, Gender, Marital Status, Health, Religious and Unemployed can be considered as the determinants of happiness level.

Regarding these effects, several interesting results have been revealed. Such variables as GDP per capita, Income, Trust, Health, Religious have positive impact on happiness level, while other variables negatively related to it. The unexpected results have been achieved for the variables of satisfaction and safety, which were assumed to have positive effect on happiness, but the resulting negative coefficient indicates the opposite.

The study highlighted some limitations. Thus, the distributions of some of the variables do not meet the conditions of normality. Also, the number of observations has been reduced almost twice because of the existence of omitted values.

This work raises a large number of questions. First and foremost, how measurable is happiness, and is it possible to apply a ruler with an arbitrarily large number of divisions to it? In many ways, the future development of this area is seen in nonparametric methods and expert assessments, as well as in a deep qualitative approach to this topic. Another question that is faced: how universal is the measurement of happiness possible across cultures? Since the survey includes several countries, this question is quite acute because of the differences in the mentality of people. The third question is what exactly the cultural characteristics are and within what limits they can be taken into account, according to what indicators, and whether such indicators exist.

In any case, the answer to each of these questions requires a lot of work and unbiased scientific interest. That is why each of the works devoted to this topic makes us one step closer to them.

Despite the above-mentioned limitations and questions, the objective of the work was achieved. The determinants of happiness level have been identified, which brings new results to the field of happiness research.

6 References

1. Abdel-Khalek A. M. et al. Happiness in Kuwaiti samples //Derasat Nafsiyah [Psychological Studies]. 2003. Vol. 13. No. 4. P. 581-612.

2. Abdel-Khalek A. M. Happiness, health, and religiosity: Significant relations //Mental Health, Religion & Culture. 2006. Vol. 9. No. 1. P. 85-97.

3. Abramovitz M. The welfare interpretation of secular trends in national income and product //The Allocation of Economic Resources: Essays in Honor of Bernard Francis Haley, Stanford University Press, Stanford. 1959. Vol. 1959. P. 1-22.

4. Akhmetova A. A. Statistical study of the level and quality of life of the population of Kazakhstan. / A. A. Akhmetova // Questions of statistics. - 2011. - No. 4. - pp. 49-53.

5. Andrews F. M., Withey S. B. Social Indicators of Well-Being: Americans Perceptions of Life Quality. – Plenum Publishing Corporation, 1976. P. 678.

6. Antipina, O. N. Economics of happiness as the direction of scientific research // Voprosy ekonomiki. - M., 2012. - No. 2. - p. 94-107

7. Barseghyan A. A. Analysis of data and processes / A. A. Barseghyan, M. S. Kupriyanov, I. I. Kholod, M. D. Tess, S. I. Elizarov – - 3rd ed. pererab. and additional – St. Petersburg: BHV-Petersburg, 2009 – - 512 p.

8. Barseghyan, A. A. Analysis of data and processes / A. A. Barseghyan, M. S. Kupriyanov, I. I. Kholod, M. D. Tess, S. I. Elizarov – - 3rd ed. pererab. and additional – St. Petersburg: BHV-Petersburg, 2009 – - 512 p.

9. Big data sends cybersecurity back to the future [Electronic resource] : - Access mode: http://community.mis.temple.edu/mis520817/2017/03/26/big-data-sends-cybersecurityback-to-the-futur/ (date zvernennya 03.03.2018) - The name of the ekran.

10. Blanchflower D. G., Oswald A. J. Well-being over time in Britain and the USA //Journal of Public Economics. 2004. Vol. 88. No. 7. P. 1359-1386.

11. Boyd D., Crawford K. Critical Quest ions for Big Data: Provocations for a cultural, technological and scholarly phenomenon. Information, Communication & Society. Vol. 15 (5), 2013. p. 662-679.

12. Burrows R., Savage M. After the crisis? Big Data and the methodological challenges of empirica l sociology. Big data & societ. Vol. 1(1), 2014. URL: https://journals.sagepub.com/doi/10.1177/2053951714540280

13. Buyakas T. M. On the phenomenon of pleasure in the process of activity and the conditions of its occurrence (according to the works of M. Chiksentmihaya) / / Vestnik MSU. 1995, pp. 53-61.

14. Cambridge Dictionary URL: http://dictionary.cambridge.org/dictionary/english/big – data.

15. Castriota, S., 2006. Education and happiness: A further explanation to the Easterlin Paradox. Unpublished paper.

16. Chen C. Aging and life satisfaction //Social Indicators Research. 2001. Vol. 54. No. 1. P. 57-79.

17. Chen Y., Yan F. Centuries of sociology in millions of books. The Sociological Review. Vol. 64 (4), 2016 p. 872-893.

18. Chernyak L. Big Data-new theory and practice [Electronic resource] / / Open Systems. DBMS. - M.: Open Systems. - 2011. - No. 10. - Access mode: http://www.osp.ru/os/ 2011/10/13010990/

19. Chubukova I. A. Data Mining: textbook. posob. / I. A. Chubukova. - M.: Internet University of Information Technologies: BINOM: Laboratory of Knowledge, 2006. - 382 p.

20. Clark A. E., Frijters P., Shields M. A. Relative income, happiness, and utility: An explanation for the Easterlin paradox and other puzzles //Journal of Economic Literature. 2008. P. 95-144.

21. Clifford Lynch. Big data: science in the petabyte era. Nature. – 2008, 455. – P. 1-50.

22. Davis W. A theory of happiness //American Philosophical Quarterly. Vol.14. No. 1. 1981. P. 111-120.

23. De Mauro A., Greco M., Grimaldi M. What is Big Data? A Consensual Definition and a Review of Key Research Topics. AIP Conference Proceedings. Vol. 1644 (1), 2015 p. 97-104.

24. Dencik L. Understanding the Datafied Society by Decentring Data. iCS Symp osium. 2018. URL: https://snurb.info/node/2424.

25. Denis McManus, New books review, Philosophy, 2009, Vol. 84. No. 04. P. 624-629.

26. Di Salvo P. (2019). The role of whistleblowing in the datafied society . URL: https://en.ejo.ch/digital - news/technology/the - role - of - whistleblowing - in - the - datafied - society/

27. Diener E. et al. Subjective well-being: Three decades of progress //Psychological bulletin. 1999. Vol. 125. No. 2. P. 276.

28. Diener, E., Subjective well-being. Psychological Bulletin, 1994 No. 95. P. 542–575.

29. Dijcks, Jean-Pierre. Big Data for the Enterprise [Electronic resource] / / Oracle. – October, 2011. - Access mode: http://Big Datawithoracle–521307. pdf

30. Dokuchaev K. Google knows everything about you: 14 things that can Surprise you, 2017 . URL : https://blog.themarfa.name/google - znaiet - o - vas - vsio/

31. Dorn D. et al. Is it culture or democracy? The impact of democracy and culture on happiness //Social Indicators Research. 2007. Vol. 82. No. 3. P. 505-526.

32. Dyuk V. Data Mining: a training course (+CD) / V. Dyuk, A. Samoylenko. - St. Petersburg: Peter, 2001. - 368 p.

33. Easterlin R. A. Does economic growth improve the human lot? Some empirical evidence //Nations and households in economic growth. 1974. Vol. 125. P. 89-125.

34. Easterlin R. A. Life cycle happiness and its sources: Intersections of psychology, economics, and demography //Journal of Economic Psychology. 2006. Vol. 27. No. 4. P. 463-482.

35. Einav L. The Data Revolution and Economic Analysis [Electronic resource] / Liran Einav, Jonathan Levin / / NBER Working PaperNo. 19035, 2013. - Access mode : http://www.nber.org/chapters/c12942.pdf

36. Ericsson E. The Impact of Datafication on Specific Landscapes . 1st ed. [ebook] London: Ericsson, Imperial College Business School & UK's Sustainable Society Network, 2014. URL: https://www.ericsson.com/assets/local/news/2014/4/the - impact - of - datafication - on - strategic - l andscapes.pdf

37. Ferrer-i-Carbonell A., Frijters P. How Important is Methodology for the estimates of the determinants of Happiness? //The Economic Journal. 2004. Vol. 114. No. 497. P. 641-659.

38. Foster I., Ghani R., Jarmin R. S., Kreuter F., Lane J. Big data and social science: A practical guide to methods and tools . London, UK/New York, NY: CRC Press, 2016.

39. Green, M., & Elliott, M. (2010). Religion, Health, and Psychological Well-Being. Journal of Religion and Health, 49, 149-163.

40. Gu, Y. and Wei, H.L., 2018. Significant indicators and determinants of happiness: Evidence from a UK survey and revealed by a data-driven systems modelling approach. Social Sciences, 7(4), p.53.

41. Guba K. Big data in sociology: new data, new sociology? Sociological Review. Vol. 17 (1)., 2018.pp. 213 - 236.

42. Guidi M. E. L. Jeremy Bentham's Quantitative Analysis of Happiness and its Asymmetries //Handbook on the Economics of Happiness, 2007, P. 68-94.

43. Hadoop's adolescence: an analysis of Hadoop usage in scientific workloads / K. Ren, Y. Kwon, M. Balazinska, B. Howe. / / VLDB Endowment. - 2013. - No. 6. - pp. 853-864.

44. Hamer D. H. The heritability of happiness //Nature genetics. 1996. Vol. 14. No. 2. P. 125-126.

45. Hansson A., Hillerås P., Forsell Y. Well-being in an adult Swedish population //Social Indicators Research. 2005. Vol. 74. No. 2. P. 313-325.

46. Happy Planet Index [Electronic resource]. Access mode: http://happyplanetindex.org/

47. Haybron D. M. The pursuit of unhappiness: The elusive psychology of well-being. OUP Oxford, 2008. P. 85 – 90

48. Helliwell, J.F. (2007). Well-Being and Social Capital: Does Suicide Pose a Puzzle?. Social Indicators Research, 81(3), 455-496.

49. Hintz A., Dencik L., Wahl - Jorgensen K. Digital citizenship in a d atafied society . John Wiley & Sons, 2018.

50. Hopkins B., Evelson B. Forrester: Big data – start small, but scale quickly. Computer Weekly, 2012. URL: http: //www.computerweekly.com/opinion/Forrester - Big - Data - Start - Small - But - Scale – Quickly

51. Hori, M. and Kamo, Y., 2018. Gender differences in happiness: The effects of marriage, social roles, and social support in East Asia. Applied Research in Quality of Life, 13(4), pp.839-857.

52. Ian H. Witten. Data Mining: Practical Machine Learning Tools and Techniques / Ian H. Witten, Eibe Frank, Mark A. Hall. – 3rd Edition. – Morgan Kaufmann, 2011. – 664 c.

53. Inglehart R., Klingemann H. D. Genes, culture, democracy, and happiness //Culture and subjective well-being. 2000. P. 162-174

54. Jurgenson N. When atoms meet bits: Social media, the mobile web and augmented revolution . Future Interne. No. 4(1), 2012. p.83 - 91.

55. Kahneman, D. Objective happiness. Kahneman D., Diener E., Schwartz N. Well-being: Fundamentals of Hedonic Psychology. - Publications of the Russell Sage Foundation, 2003. pp. 247-268

56. Kim D. O. Factor, discriminant and cluster analysis: trans. from English / J. O. Kim, C. W. Muller, W. R. Klskka [et al.]. - M.: Finance and Statistics, 1989. -215 p.

57. Kislova O. N. Augmented reality through the prism of Internet practices of modern students. Methodology, theory and practice of sociological analysis of suchasny suspension . Vipusk 20, 2014. p. 351-356.

58. Kislova O. N. Intellectual data analysis: the history of the formation of the term . Ukrainian Social Science journal . No1-2 , 2011. pp. 83-94.

59. Kislova O. N. Sociology in the context of the challenge of "big data": the role of data mining in the development of new approaches to sociological research. Visnik Harkivskogo natsionalnogo universitetu imeni V. N. Karazina seriya " Sotsiologichni doslidzhennya suchasnogo suspilstva: methodology, theory, method and". No. 1148, 2015. pp. 57-62.

60. Kislova O. N. To be or not to be digital sociology? Visnik Harkivskogo natsionalnogo universitetu imeni V. N. Karazina seriya " Sotsiologichni doslidzhennya suchasnogo suspilstva: methodology, theory, methods". No. 1045, 2013 p. 9-15.

61. Lande D. Deep analysis of texts: the technology of effective analysis of text data [Electronic resource] / Dmitry Lande. - Access mode: http://visti.net/~dwl/art/dz/

62. Lipinson, S. Cloud Computing: information and processes [Electronic resource] / S. Lipinson / / Open Systems. - 2008. - No. 11. - Access mode: http://www.osp.ru/os/2008/08/1858076073/_p2.html.

63. Lu L. et al. Personal and environmental correlates of happiness //Personality and Individual Differences. 1997. Vol. 23. No. 3. P. 453-462.

64. Managing data transfers in computer clusters with orchestra / [M. Chowdhury, M. Zaharia, J. Ma ta in.]. / / SIGCOMM

65. Manyika J. Big data: The next frontier for innovation, competition, and productivity / Manyika James. Mc Kinsey Global Institute, June, 2011 – - 156 p.

66. Marr B. Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance / Bernard Marr – - John Wiley&Sons Ltd, 2015. - 256 p.

67. Marres N. Digital Sociology The Reinvention of Social Research . Cambridge: Polity Press, 2017.

68. Mayer - Schönberger V., Cukier K. Big data: A revolution that will transform how we live, work, and think . Houghton Mifflin Harcourt, 2013.

69. Mayer-Schoenberger, V. Big data. The Revolution that will change the way we Live, work and think / V. Mayer-Schoenberger, Kenneth Kukier; translated from English. Inna Gaidyuk. - M.: Mann, Ivanovi Ferber, 2014. - 240 p.

70. Metzler K., Kim D. A., Al lum N., Denman A. Who is doing computational social science? Trends in big data research (White paper). London, UK: SAGE Publishing, 2016.

71. Michalos, A.C., 2008. Education, happiness and wellbeing. Social Indicators Research, 87(3), pp.347-366.

72. Morgan, J., Robinson, O. and Thompson, T., 2015. Happiness and age in European adults: The moderating role of gross domestic product per capita. Psychology and aging, 30(3), p.544.

73. Ng Y. K. A case for happiness, cardinalism, and interpersonal comparability //The Economic Journal. 1997. Vol. 107. No. 445. P. 1848-1858.

74. Noelle-Neumann E. Politik und Glück //Freiheit und Sachzwang. 1977. P. 208-262. (244).

75. November 2017 | TOP500 Supercomputer Sites [Electronic resource] : [Website]. - Electronic tribute. – Access mode to the resource: http://www.top500.org/lists/2017/11/ (date zvernennya 04.03.2018) - The name of the ekran.

76. Oswald, A.J., 1997. Happiness and economic performance. The economic journal, 107(445), pp.1815-1831.

77. Özcan, K.M., Böke, S.S. and Kara, M., 2008, August. The determinants of the perception of happiness about quality of life amongst Turkish people. In Economic Research Forum. Working Paper Series (No. 425).

78. Pachankis, J.E. and Bränström, R., 2018. Hidden from happiness: Structural stigma, sexual orientation concealment, and life satisfaction across 28 countries. Journal of Consulting and Clinical Psychology, 86(5), p.403.

79. Paklin N. B. Business analytics: from data to knowledge (+ CD) / N. B. Paklin, V. I. Oreshkov. - St. Petersburg: Piter, 2009 – - 624 p.

80. Paklin, N. B. Business analytics: from data to knowledge [Text] : textbook. posob. / N. B. Paklin, V. I. Oreshkov. - 2nd ed., ispr. –- St. Petersburg : Piter, 2013. - 702 p.

81. Pentland A. Reinventing society in the wake of big data systems, 2012 URL: http://edge.org/conversation/reinventing -society - in - the - wake - of-big-data .

82. Pentland A. Social Physics: How Good Ideas Spread. The Lessons from a New Science. New York: The Penguin Press, 2014.

83. Powdthavee, N. (2008). How Much Does Money Really Matter? Estimating the Causal Effect of Income on Happiness. Discussion Paper in Economics no. 2009/02, University of York.

84. Press G. A Very Short History of Big Data, 2013 URL: http://www.forbes.com/sites/gilpress/2013/05/09/a - very - short - history - of - big - data .

85. Proceedings of the ACM SIGCOMM 2011 conference. – 2011. – №41 (4). – Pp. 98-109.

86. Pugno M. Scitovsky and income-happiness paradox // The Journal of Socio -Economics 43(2013), P. 1 -10

87. Rahayu, T.P., 2016. The determinants of happiness in Indonesia. Mediterranean Journal of Social Sciences, 7(2), p.393.

88. Raibley J. Happiness is not Well-Being //Journal of Happiness Studies. 2012. Vol. 13. No. 6. P. 1110-1147.

89. Rana, S., Hariharan, M., Nandinee, D. and Vincent, K., 2014. Forgiveness: A determinant of adolescents' happiness. Indian Journal of Health and Wellbeing, 5(9), pp.1119-1123.

90. Redden J. Democratic governance in an age of datafication: Lessons from mapping government discourses and practices. Big Data & Society . Vol. 5 (2), 2018. URL: https://journals.sagepub.com/doi/full/10.1177/2053951718809145 .

91. Resnyansky L. Conceptual frameworks for social and cultural Big Data analytics: Answering the epistemological challenge. Big Data & Society . Vol. 6 (1), 2019. URL: https://journals.sagepub.com/doi/10.1177/2053951718823815

92. Rogers R. A. The End of the Virtual: Digital Methods . Amsterdam: Amsterdam University Press, 2009.

93. Rogers R. Digital methods for web research. Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource, 2015. p. 1-22. Rogers, R. Doing Digital Methods. SAGE Publications Limited, 2019.

94. Roodman, D. (2011). Fitting Fully Observed recursive Mixed-process Models with CMP. The Stata Journal, 11(2), 159-206.

95. Ruppert E., Isin E., Bigo D. Data politics. Big Data & Society . Vol. 4 (2). 2017. URL: https://journals.sagepub.com/doi/full/10.1177/2053951717717749

96. Sarracino, F. (2012). Money, Sociability and Happiness: Social Erosion and Unhappiness? Time Series Analysis of Social Capital and Subjective Well-being in Western Europe, Australia, Canada and Japan. Social Indicator Research, 109, 135-188.

97. ScandiNews – multimedia multilingual informative portal about living in the Nordic countries (2019), "Norway is the happiest country in the world", available at: https://scandinews.fi/society/5109"norvegiya"samaya" schastlivaya"strana"v"mire.

98. Schäfer M. T., Van Es K. F. The datafied society: Studying culture t hrough data. Amsterdam University Press, 2017.

99. Schwarz N. What respondents learn from questionnaires: The survey interview and the logic of conversation //International Statistical Review/Revue Internationale de Statistique. 1995. P. 153-168.

100. Sekulova, F. and Van den Bergh, J.C.J.M., 2016. Floods and happiness: Empirical evidence from Bulgaria. Ecological Economics, 126(C), pp.51-57.

101. Sharda R. Business Intelligence: A Managerial Perspective on Analytics / R. Sharda, D. Delen, E. Turban. - New Jersey: Prentice Hall Press Upper Saddle River, 2013. - 416 p.

102. Shaw R. Big data and reality. Big Data & Society. Vol. 1(4), 2015. p. 1-4.

103. Shmotkin D., Happiness in the Face of Adversity: Reformulating the Dynamic and Modular Bases of Subjective Well-Being Review of General 2005, Vol. 9, No. 4. P. 291–325

104. Shubham Sharma. Big Data Landscape / Shubham Sharma. // International Journal of Scientific and Research Publication. – 2013. – №6.

105. Singer, J., Rexhaj, B. and Baddeley, J., 2007. Older, wiser, and happier? Comparing older adults' and college students' self-defining memories. Memory, 15(8), pp.886-898.

106. Snee H., Hine C., Morey Y., Roberts S., Watson H. Digital Methods for Social Science: An Interdisciplinary Guide to Research I nnovatio . Palgrave Macmillan, London, 2015.

107. Softserve Business Systems. URL: https://www.softservebs.com/uk/solutions-2/prediktivna-analitika/

108. Sonia Buchholts, Maciej Bukowski, Aleksander Sniegoski. Big & Open Data in Europe: A growth engine or a missed opportunity? // Report commissioned by demosEUROPA – Centre for European Strategy Foundation within the «Innovation and entrepreneurship» programme. – Warsaw, Mdruk, 2014. – 116 p.

109. Stevenson, B. and Wolfers, J., 2013. Subjective well-being and income: Is there any evidence of satiation?. American Economic Review, 103(3), pp.598-604.

110. Stonebraker M. MapReduce and parallel DBMS: Friends or enemies? [Electronic Resource] / Michael Stonebraker, Daniel Abadi, Davit Dewitt, Sam Madden, Eric Paulson, Andrew Pavlo, Alexander Razin; translated from English. Sergey Kuznetsov / / Communications of the ACM, vol. 53, No. 1, January 2010. - Access mode: http://citforum.ru/database/articles/mr_vs_dbms-2/

111. Stonebraker M. Object-relational DBMSs, tracking the next great wave / M. Stonebraker, P. Brown, D. Moore. - San Francisco, California: Morgan Kauffman Publishers, Inc, 1998. - 216 p.

112. Sumner L. W. Welfare, happiness, and ethics. – Oxford University Press, USA, 1999. P.625

113. Syrovezhkin A. A. THEORY OF HAPPINESS FOR CONSUMER SOCIETY //Youth and Science: Collection of materials of the VIII All-Russian Scientific and Technical Conference of Students, Postgraduates and Young Scientists, dedicated to. – 2012.

114. Sysoeva T. A. INFLUENCE OF THE EMOTIONAL STATE ON MNEMIC PROCESSES: THE EFFECT OF CONGRUENCE //C 69 Social intelligence: Theory, measurement, research/Ed. DV Lyusina, DV Ushakova. - M.: Publishing House "Institute of Psychology of the Russian Academy of

Sciences", 2004.176 p.(Proceedings of the Institute of Psychology of the Russian Academy of Sciences). 2004, p. 89.

115. Technology and Innovations. Big Data. - https://it-enterprise.com/knowledge-base/technology-innovation/big-data-bolshie-dannye

116. Text Mining [Electronic resource]. - Access mode: http:/ /statsoft.ru/home/textbook/modules/sttextmin. html#index

117. The 2018 Digital Future Report. Center for the Digital Future at USC Annenberg. / https://www.digitalcenter.org/reports/

118. The Oxford English dictionary. URL: https://en.oxforddictionaries.com/definition/big_data

119. Tokuda, Y. and Inoguchi, T., 2008. Interpersonal mistrust and unhappiness among Japanese people. Social Indicators Research, 89(2), pp.349-360.

120. Uskenbaeva, R. K. Tasks of resources provision of distributed computer system's functionality / R. K. Uskenbayeva, A. A. Kuandykov, A. U. Kalizhanova. – Dubai, World Academy of Science, Engineering and Technology. – 2012. – Iss.70. – P. 580-581.

121. Uskenbayeva, R. K. Tasks of creating big data-Big Data / R. K. Uskenbayeva, Zh. B. Kalpeeva, A. Kasymova / / Computing and information technologies in science, technology and education: international conference. - Ust-Kamenogorsk: EKSTU named after D. Serikbayev, 2013. - p. 209-213.

122. Van Praag B. Ordinal and cardinal utility: an integration of the two dimensions of the welfare concept //Journal of Econometrics. 1991. Vol. 50. No. 1. P. 69-89.

123. Veenhoven R. Is happiness relative? //Social Indicators Research. 1991. Vol. 24. No. 1. P. 2

124. Venturini T., Bounegru L., Gray J., Rogers R . A reality check (list) for digital methods. New media & society. Vol. 20 (11), 2018. p. 4195-4217.

125. Venturini T., Latour B. The Social Fabric: Digital footprints and quali - quantitative methods. Proceedings of future en seine, 2009. p. 87-103

126. Venturini T., Rogers R. API - Based Research or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach. Digital Journalism , 2019. p. 1-9

127. Viktor Mayer-Schönberger. Big Data: A Revolution that Will Transform how We Live, Work, and Think / Viktor Mayer-Schönberger, Kenneth Cukier – - New York: Houghton Mifflin Harcourt, 2013. - 242 p.

128. Weinstein M. Analyzing Big Data with Dynamic Quantum Clustering [Electronic resource] / M. Weinstein, F. Meirer, A. Hume, Ph. Sciau, G. Shaked, R. Hofstetter, E. Persi, A. Mehta, D. Horn. arXiv: 1310.2700.– Access mode : https://arxiv.org/ftp/arxiv/papers/1310/1310.2700.pdf

129. White T. Hadoop: The Definite Guide / Tom White. - Boston: O'Reilly Media, Inc., 2012. – 657 p.

130. Williams, R. (2006). Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. The Stata Journal 1, 58-82.

131. Wolfers J. The Economics of Happiness, Part 4: Are Rich People Happier than Poor People? // The New York Times.– 2008.– April 22.

132. World happiness report 2019 [Electronic resource]. Access mode: http://worldhappiness.report/wp " content/uploads/sites/2/2019/03/HR19. pdf

133. Zhuravlev Yu. I. Recognition. Mathematical methods. Software system. Practical applications / Yu. I. Zhuravlev, V. V. Ryazanov, O. V. Senko. - M.: Fazis, 2006. - 176 p.

134. Zikopoulos, P. Harness the power of big data the IBM big data platform / P. Zikopoulos, K. Parasuraman, T. Deutsch, J. Giles, D. Corrigan // McGraw Hill Professional, New York, NY. – 2012. - Mode of access: http://books.google.com/books?id=HhSON 0xOCQoC

135. Zinoviev A. Yu. Visualization of multidimensional data / A. Yu. Zinoviev. - Krasnoyarsk: Ed. Krasnoyarsk State Technical University un-ta, 2000. - 180 p.