

Abstracts of Dissertations

Institute of Information and
Communication Technologies

BULGARIAN ACADEMY OF
SCIENCES



2 / 2021



PERSONALITY AND
DECISION-MAKING
MODELS IN INTERNET

Rumen Ketipov

ИНДИВИДУАЛНОСТ И
МОДЕЛИ ПРИ ВЗЕМАНЕ НА
РЕШЕНИЕ В ИНТЕРНЕТ

Румен Кетипов

Автореферати на дисертации

Институт по информационни и
комуникационни технологии

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ISSN: 1314-6351

Поредицата „Автореферати на дисертации на Института по информационни и комуникационни технологии при Българската академия на науките“ представя в електронен формат автореферати на дисертации за получаване на научната степен „Доктор на науките“ или на образователната и научната степен „Доктор“, защитени в Института по информационни и комуникационни технологии при Българската академия на науките. Представените трудове отразяват нови научни и научно-приложни приноси в редица области на информационните и комуникационните технологии като Компютърни мрежи и архитектури, Паралелни алгоритми, Научни пресмятания, Лингвистично моделиране, Математически методи за обработка на сензорна информация, Информационни технологии в сигурността, Технологии за управление и обработка на знания, Грид-технологии и приложения, Оптимизация и вземане на решения, Обработка на сигнали и разпознаване на образи, Интелигентни системи, Информационни процеси и системи, Вградени интелигентни технологии, Йерархични системи, Комуникационни системи и услуги и др.

Редактори

Геннадий Агре

Институт по информационни и комуникационни технологии, Българска академия на науките
E-mail: agre@iinf.bas.bg

Райна Георгиева

Институт по информационни и комуникационни технологии, Българска академия на науките
E-mail: rayna@parallel.bas.bg

Даниела Борисова

Институт по информационни и комуникационни технологии, Българска академия на науките
E-mail: dborissova@iit.bas.bg

Настоящото издание е обект на авторско право. Всички права са запазени при превод, разпечатване, използване на илюстрации, цитирания, разпространение, възпроизвеждане на микрофилми или по други начини, както и съхранение в бази от данни на всички или част от материалите в настоящето издание. Копирането на изданието или на част от съдържанието му е разрешено само със съгласието на авторите и/или редакторите

*The series **Abstracts of Dissertations of the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences** presents in an electronic format the abstracts of Doctor of Sciences and PhD dissertations defended in the Institute of Information and Communication Technologies at the Bulgarian Academy of Sciences. The studies provide new original results in such areas of Information and Communication Technologies as Computer Networks and Architectures, Parallel Algorithms, Scientific Computations, Linguistic Modelling, Mathematical Methods for Sensor Data Processing, Information Technologies for Security, Technologies for Knowledge management and processing, Grid Technologies and Applications, Optimization and Decision Making, Signal Processing and Pattern Recognition, Information Processing and Systems, Intelligent Systems, Embedded Intelligent Technologies, Hierarchical Systems, Communication Systems and Services, etc.*

Editors

Gennady Agre

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: agre@iinf.bas.bg

Rayna Georgieva

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: rayna@parallel.bas.bg

Daniela Borissova

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences
E-mail: dborissova@iit.bas.bg

This work is subjected to copyright. All rights are reserved, whether the whole or part of the materials is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this work or part thereof is only permitted under the provisions of the authors and/or editor.

e-ISSN: 1314-6351

© ICT-BAS 2012

www.iict.bas.bg/dissertations



BULGARIAN ACADEMY OF SCIENCES

Abstract of PhD Thesis

PERSONALITY AND DECISION-MAKING MODELS IN INTERNET

Rumen Rumenov Ketipov

Supervisor: Assoc. Prof. Vera Angelova

Approved by Supervising Committee:

Acad. Ivan Popchev
Prof. Tania Pencheva
Prof. Maria Nisheva-Pavlova
Prof. Daniela Borisova
Assoc. Prof. Irina Radeva



**INSTITUTE OF INFORMATION AND
COMMUNICATION TECHNOLOGIES**
Department of Intelligent Systems

Introduction

The 21st century has already brought many changes in human lives and certainly, there are yet more changes to come. Such an example is the explosion of e-commerce after that consumers' shopping habits have changed drastically. Consequently, on-line purchasing quickly becomes a preferred way to shop for people around the globe and has the potential not only to increase competition within retail markets but also to greatly enhance consumer requirements.

The growth of customers' expectations is a premise for creating a new business strategy that takes into account the specifics of personality that is one of the main sources of our decisions and significantly impacts the way people think, feel, and especially behave [8].

And as each person is different in terms of his or her personality traits [44], identifying the individuality in terms of consumer behavior becomes an invaluable asset, which helps companies to create a revolutionized marketing strategy and to provide more personalized solutions, services, and experience for their customers. It should pay more attention that personality may be defined as the underlying cause not only of the general shopper behavior and perceptions but also generally of our choices, perceptions, and the way we deal in different situations [49]. According to a study conducted at the University of Basel and the Max Planck Institute [38], the risk preference is a personality characteristic in its own right and remains stable over time, which allows to treat the risk averseness as additional personality determinant.

Thus, combining the personality insights and modern technologies, and web design trends could be beneficial for consumers and businesses, as well. This enables the providing of higher quality services, creates a more seamless shopping experience, and would successfully lead to an increase in customer satisfaction. The finely achieved by Machine Learning (ML) methods make it possible to predict, on the one side, the consumer's behavior in the process of decision-making, and on the other side, the products, content, and functionalities that are following individual preferences and expectations.

But despite all the above-mentioned facts, the studies of the application of personalization in the context of e-commerce, based on the users' personality profile, remain relatively scarce and provide a wide field for future analysis and this arouses

the interest of the business community and researchers as well [33].

Aims and Objectives

The purpose of this dissertation sets the scope of the research aims and objectives. The study aims to investigate the presence of significant relationships between personality and some of the basic e-shops features. Based on the obtained results, it also aims to create models for reliable prediction of consumer preferences and behavior in the purchasing decision-making process based on their personality profile.

Based on this the study objectives are defined as follows:

- To study the existing various theories and concepts for personality measurement and to choose an appropriate psychometric model for the study.
- To choose a set of e-shops functionalities that are typical and applicable to most of them.
- To create a research strategy and design, respecting basic standards of ethics and neutrality; to translate the research into three languages - Bulgarian, English, and German to ensure its wider scope and validity; the aspect of risk averseness to be considered as additional personality determinant; to analyze the study's results and to establish if there is a relationship between chosen independent and dependent variables.
- To propose and implement two (or more) ML models in order to be achieved reliable prediction for the dependent variables in the existing significant correlations and to analyze the achieved results; to choose an appropriate ML model and to propose and implement optimization;
- Based on the achieved results of the conducted study to be developed and determined the consumer behavior models in the process of decision-making in the field of e-commerce.

Dissertation Structure

In order to respond to the research aims and objectives, to obtain all necessary information, to conduct a survey and analysis, to fulfill the research aims and objectives, as well as to achieve quality, valid and optimized results at the end of the study, the study is developed in four logically structured chapters as follows.

The first chapter starts with a review of the relevant literature concerning the relationship between personality and the decision-making process. It is paid particular attention to the most modern psycho-dynamic theory - Theory of Personality Traits, as well as the Big Five model.

After that, the study explores in detail the role of personality as a major determinant influencing consumer behavior on the Internet and especially during the process of online purchasing, as well as some basic e-stores features and elements. And since the current work aims to create and propose models for consumer behavior prediction, this chapter also includes a literature review of various studies in the field of personality research applying ML models.

The purpose of the second chapter is to present the main aspects of the methodology used for conducting empirical research and primary data collection that is a basis for forthcoming analysis.

The third chapter focuses on the results of conducted primary research. Additionally, it is also conducted a bivariate analysis to find existing significant correlations between the basic personal determinants and the risk averseness, on the one hand, and the selected functionalities of the web stores, on the other hand.

In the next stage are implemented three regression models in the field of machine learning in order to be achieved a reliable prediction of online consumer behavior based on their personality. It is also proposed and implemented optimization of the most appropriate algorithm for the aim. Based on the obtained results, models of consumer preferences and behavior during the process of decision making in the field of e-commerce are summarized.

Finally, the fourth chapter presents a summary of achieved results as well as some suggestions concerning a possible future upgrading of current research.

Chapter 1

Personality and Decision-Making Process

1.1 Theories and Researches in the Field of Personality Psychology

Although some authors suggest that there are two basic groups of factors that substantially affect consumers during the process of decision-making – internal (culture, social class, family) and external (motivation, attitude, intention, perception) [30], according to Barkhi et. al. (2007), personality is the fundamental determinant of individual behavior, choices as well the way of person process the information from the environment. Thus, consumers prefer brands and products whose characteristics are congruent with their personality profile (for example, house, furniture, and car), as they express and enhance their self-concept by consuming items that evoke positive product user stereotypes for them [3].

Besides, it should be taken into consideration that human decision-making is not always based on principles of rationality. This is often a result of careful assessment of alternatives and results that are influenced by personality, which in turn becomes the basis of individual behavior in all its aspects [29]. So, by people with similar traits is observed a high tendency to behave in a particular way and to use similar decision-making style under certain situations [19].

1.1.1 Theories and Models of Personality Measurement and Analysis

Personality is generally defined as an individual unique and relatively stable pattern of behavior, thoughts, and emotion that significantly impacts human behavior.

These internal factors make one person's behavior consistent and different from the behavior other people would manifest in comparable situations. This aspect of personality is called individual differences [21] and it is of particular interest to researchers.

There are many theoretical perspectives on personality in psychology, which involve different ideas about the way the human develops and forms. Some of the keysets approaches for capturing human individuality are the psychoanalytical personality theory of Sigmund Freud [16], Rogers' Person-centered Theory [47], the Three-Factor Theory of Hans J. Eysenck [13], as well as Five-Factor Theory of Personality - the most used trait approach often referred to as the "Big Five" [19], [10], which is a state-of-the-art measuring model today [8], [21]. It states and measures human nature as a result of mainly biological-determined basic factors: Openness to experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism/Emotional stability [10], [39], [40], [19].

Each of these five broad domains of personality contains additional aspects, which are explaining in detail the individual's behavior [8]. For instance, Agreeableness contains additional facets such as understanding, warmth, morality, pleasantness, cooperation, and tenderness. Taking this into account, it becomes clear, why individuals with high Agreeableness show more sympathy and empathy to others [19]. Likewise, people with a high value of Conscientiousness are individualistic, detail-oriented, efficient, responsible, highly organized, and self-controlled [10], while the lower level of Neuroticism (opposite of emotional stable) characterizes people who are calm and are not easily upset. These with a higher value of Openness to experience are known to be imaginative, independent-minded, and intellectual. Consequently, these five determinants of personality are focused on conceptual individual features. This frame recognizes that most people are not on the polar ends of the black-and-white spectrum but rather somewhere in between [26].

But unfortunately, the Big Five framework is not always applicable in practice due to its volume (240 elements) and therefore the literature presents several shorter, but validated questionnaires, which also successfully apply the Traits theory of personality [6].

For example, the HEXACO framework retains the original factors of the Five-Factor model but adds the determinant "honesty/ humility", which describes the extent to which one puts other people's interests above one's own [1].

Another example is the RIASEC model which evaluates personality based on another six main traits: realistic, investigative, artistic, social, entrepreneurial, and conventional [23].

The Ten-Item Personality Inventory by Gosling [21] is based also on the Big Five personality dimensions, but it includes only ten questions - so, two descriptors giving information about each of the factors.

1.1.2 Influence of Personality on the Online Purchase Decision-Making Process

The specific characteristics of personality have a significant influence not only on the character, attitudes, and habits of the individual but also on the decision-making process. Taking into account that people with the same personality profile behave similarly could be concluded that they also have similar habits and priorities by selecting the most appropriate e-shop for them. And since different types of personality traits make people distinctive in behavior and preferences, this also means that each individual takes decision differently - some people rely on their intuitions, while others prefer to discuss their choices with friends and carefully consider various alternatives [28]. In this regard, as by traditional shopping, online customers behave differently in terms of decision-making and rely on different store features to make a choice. So, each user relies more heavily on certain features of a store to make decisions and paying less attention to others. Because of this, the adaptation of the sales method to the customer's decision-making style is a useful approach to be improved the user's experience [31].

But instead of mechanically market segmenting, personalization, based on Machine Learning allows the applying of algorithms for more accurate prediction of specific user characteristics, which provides an opportunity to offer recommendations for products or content that are consistent with the individual preferences of the clients. In contrast to the statistical approach, which is focused primarily on making inferences and understanding the characteristics of the variables [25], ML approach treats the data as unknowns. It is mainly focused on prediction rather than inference and aims at forecasting unobserved outcomes or future behavior [24].

In practice, there are two main approaches in which ML methods play and will continue to play in future a crucial role in the field of psychology. On the one hand, analyzing large data sets, which is extremely useful for the development and validation of theories in psychology. For example, a large amount of behavioral data in social media might be used for the prediction of a user's personality profile [52]. On the other hand, the individual specifics can be used as predictors of consumer behavior and decision-making on the Internet. And although the key to personalizing an interface is the accurate prediction of user preferences, the research in the field of personalization today based on the user's personality profile remains relatively scarce [33].

Chapter 2

Methodology of Empirical Research

The choice of an appropriate research methodology refers to the regulatory principles for solving a particular research problem so that to be made valid and reliable conclusions at the end of the inquiry [32].

2.1 Research Philosophy

Regarding the purpose of the present dissertation and specific research questions, the author chooses to adopt positivism as a philosophical stance of this investigation [45], because it states that reality is independent and can be measured and predicted impartially through research based on well-defined and logically structured data, which is not affected by personal authorial understandings and perceptions [22].

2.2 Research Approach

For the aim of the present research, the author preferred to adopt a mixed research approach (deductive and inductive), as while the quantitative approach allows the researcher to test new ideas, the qualitative one provides an opportunity to create new ones and this provides better results at the end of the project [45]. Both approaches are not mutually exclusive and are applicable especially in cases of modern research problems [22].

2.3 Research Strategy

The author adopts conducting a survey as a part of his research strategy. So he could determine in the next stage of the project if there is a relationship between consumers' personality profile and their preferences regarding some of the basic features and elements of online stores.

2.3.1 Secondary Data Collection

To collect the necessary secondary information (books, magazines, and newspapers; reports and publications of various associations; public records and statistics) the author reviews the relevant literature and scholars' information, where various authors observe similar research questions related to the relationship between personality, consumer behavior on the Internet, and online shopping habits.

2.3.2 Questionnaire – Primary Data Research

The applying of the survey as a tool for primary data collection makes it possible to be made qualitative conclusions based on collecting quantitative data, which makes it extremely appropriate in cases of the combined research approach [45].

The author creates a questionnaire including 4 sections. For its electronic version is used a web-based survey application (Google Forms). The questionnaire is distributed via email communication in combination with a personal contact on a social network. It is also translated into 3 languages (Bulgarian, English, and German), as the project requires all respondents to be acquainted in detail with the meaning and content of the questions, which are multiple-choice and rank order [45].

- **Section 1** – Online store features/ User preferences;

Based on the fact that what people define as reliable in the network remains stable over the years despite the changing design trends [41], the participants are asked to answer 19 questions related to some of the main web store's features and elements. Thus, the author collects information about consumer preferences, behavior, needs, and requirements.

All the questions are organized into 3 sub-groups - Content and Appearance, User Interface Tools and Risk Reducers (Table 2.1). The assessment of consumers' attitude to each of the observed elements is according to a five-point position of the Likert type (from 1 = never to 5 = always). The asked questions are defined in such a way that they do not take much time and effort of respondents, but at the same time are suitable to collect the needed information.

1.	Product descriptions give me the necessary information to able to make a decision.
2.	Instead of single score, I prefer detailed product ratings.
3.	I read the expert reviews. They are essential in the decision making process.
4.	I read comments which other users have left for different purchases.
5.	I check the product availability as well delivery time before I make a purchase.
6.	I prefer to be able to see where I am in the product purchasing process.
7.	I prefer to see real-time shipping quote estimates.
8.	I prefer to take a look at detailed product images.
9.	I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).
10.	I prefer to by complementary accessories (like insurance and extended warranty) as a bundle.
11.	In order to be able to choose the right product for me I use product categorization resp. featured product filter.
12.	In order to choose among different products, I compare the product details.
13.	I tend to use different features in the cart like one-click reorder, calculate end price etc.
14.	I normally prefer to check different delivery options.
15.	I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.
16.	I tend to write and comment product reviews. They help to clarify uncertainties about desired products.
17.	I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.
18.	I prefer to check the free return possibility; it is essential for me.
19.	I normally check for alternative (secure) payment methods like PayPal etc.

Table 2.1: Online Shop Features

- **Section 2** – Ten characteristics that determine the personality profile;

As it was mentioned previously, the primary benchmarks in the Trait theory for personality measurement today is the Five-Factor frame (or Big Five model). It describes the personality in terms of five basic factors: Openness to experience,

Conscientiousness, Extraversion, Agreeableness, and Neuroticism [20]. These individual determinants underlie the diversity of human behavior and preferences [10]. This concept is strongly applicable in all major cultural regions of the world, and according to different scholars, the observed minimal cultural differences in personality structure could be even ignored [46]. Thus, taking into account the validity of this model, as well as its intercultural applicability, the author decides initially to use this approach in his survey as a tool for respondents' personality appraisal. However, since the original structure is inapplicable in the context of e-commerce due to its significant length, he considers it more appropriate to apply a brief measure of the Big Five personality domains - Ten Item Personality Inventory (TIPI) by Gosling [21], which reached adequate levels validity and reliability and it also consists of only 10 elements - two descriptors giving information about each of the five personality determinants.

- **Section 3** – Risk averseness;

Risk perception is a critical factor, which is defined as a person's current tendency to take or avoid risks and has a significant role in consumer decisions and behavior. So, risk avoidance determines the extent to which consumers are sure or unsure of what they are buying [12]. Considering these issues, the author applies in the current survey a Risk Propensity Scale by Donthu and Gilliland [12] that includes only three elements, to measure participants' desire to avoid or to take risks. The lower value of result is associated with a higher level of risk awareness, while higher - with a lower level of risk awareness (i.e. higher level of risk avoidance).

- **Section 4** – Demographics analysis;

In its last section, the inquiry provides 5 demographic questions related to age, gender, education, citizenship, place of residence, and the frequency of online shopping as well.

To be revealed the weaknesses of the questionnaire and to be proved its effectiveness, it is made a pilot test with 10 people, part of the survey population, as it is recommended in the academic literature [32].

2.3.2.1 Sampling Techniques

The purpose of the conducted study requires to be reached the average customer of online stores and to be made some general inferences [45] for his or her preferences and behavior. In this regard, the author adopts the approach of random sample selection in the process of respondents' selection. The sample is randomly structured so that all elements of the community have had an equal chance of falling into it.

2.3.2.2 Sampling Frame

The questionnaire is distributed to 250 people, representatives of the surveyed population. 226 of all (90.4%) filled in it accurately and completely online and all of them meet the criteria related to the purpose of the survey. According to the literature, the achieved sampling size meets the needs of the current study, so in the end, it could be reached significant conclusions on the researched problem [45].

2.3.2.3 Time Horizon of Research

Considering how dynamic the business and technologies are changing nowadays, the author decides that the ‘snapshot’ approach (a ‘snapshot’ at a certain moment) is the appropriate choice for the current research [45].

2.3.2.4 Validity and Reliability of Data

The potential threats of the current project could be: if some of the respondents are under 18 years or children; if there are not enough participants with online shopping experience; if too few respondents participate in the study; if there are no representatives of all considered age sub-groups; if the survey is only quantitative because the topic requires qualitative analysis of respondents’ attitude to the research problem.

Despite the above-mentioned threats, this project can be considered valid and reliable, because the applied methodological approaches are appropriate for its type and objectives. Additionally, one important instrument that contributes to the validity and reliability of the project is its pilot test, too.

2.3.2.5 Ethical Issues

To be minimized the risk of compromising the research process the author takes into account the next issues [45]:

- all participants are familiar with the survey’s purpose and complete the questionnaire voluntarily;
- no respondents below 18 years participate in the survey;
- the questionnaire does not contain questions related to religious and political affiliation;
- the inquiry is anonymous and because of this no names and personal information are published;

- the respondents are familiar with the fact that the received data is confidential and will not be shared with the third side;
- the collected data is interpreted only based on the study's objective.

2.3.3 Limitations

One of the limitations of the current survey is the insufficient secondary data and statistics related to consumer requirements and expectations regarding the observed features and elements of online stores.

An essential limitation of the current study is also the set time frame because investigating personality is a very complex process that requires validation of the results with other tools and methods during a long period of time.

Another problem would be related to the questionnaire – if the respondents do not understand the questions. But in the case of the present study, this risk is overcome, because all questions are formulated clear and understandable as well as translated into three languages.

Other limitations of the project concern the technical problems during the analysis of questionnaires – if the share of invalid filled-in questionnaires is too large, if there are respondents under 18 years or if the sample size is too small.

Chapter 3

Research. Prediction of Online Consumer Behavior

3.1 Analysis and Results of Empirical Research

3.1.1 Sample Demographic Profile

226 respondents filled in the questionnaire accurately and completely, of which 43% are men and 56% - women. It is observed that neither men nor women have priority in the sample, so at the end of the study, general conclusions without focusing on a specific sub-group are possible.

Since the author does not aim to conduct his analyses concentrating on a specific geographical area, but mostly he wants to reach general conclusions regardless of the consumers national and cultural background, the questionnaire is randomly distributed to participants from different countries (more than 10 countries), as before the questionnaire is translated into three languages (Bulgarian, English, and German).

So, more than half of the survey's respondents (65%) have a Bulgarian origin and lives in the country at the time of the survey (Figure 3.1). The rest of the sample (35%) includes people, who are foreigners or Bulgarians, who have lived for more than 5 years abroad. Thus, the author considers mostly the place of residence of the participants in the last minimum of 5 years, because the intensive geographical mobility observed in recent years is becoming a significant characteristic of consumers, especially in the process of market segmentation. Changing the place of residence naturally leads also to changes in the preferences of customers based on local lifestyle, cultural understandings, the social structure of society, and so on [11].

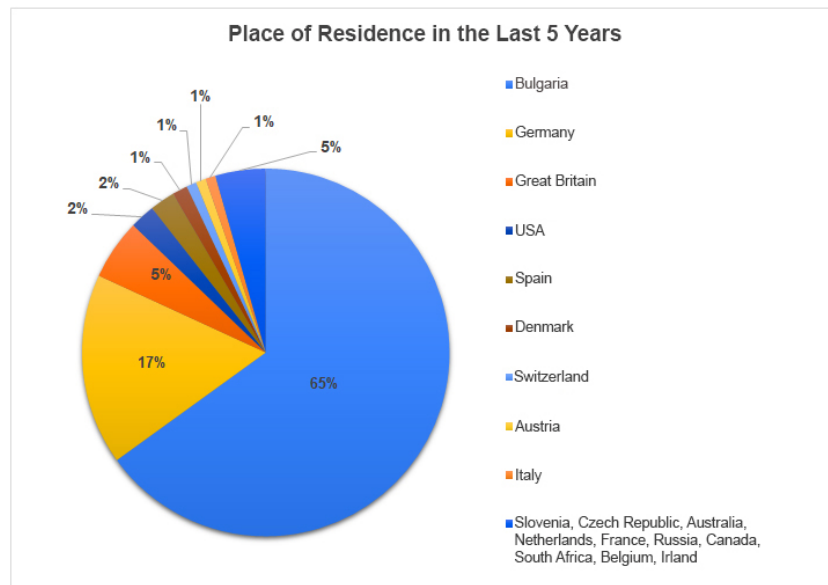


Figure 3.1: Place of residence in the last 5 years

The „Age“ factor strongly influences the likelihood of online buying. In other words, in different age subgroups, it is also observed a different trend to online purchasing. Indeed, according to Farag et al. [14], age is inversely related to the intensity of online shopping. But in the case of this study, all participants are adults and most of them (almost 80%) are in active part of their life, when the Internet occupies a large part of people daily life - 27% of all respondents are between 18 and 30 years old, 60% - between 31 and 45 years old, 11% - between 46 and 60 years old. Only 2% of the sample includes people over 61 years old (Table 3.1).

Another determinant of the literate and skillful Internet use by the average consumer in the 21st century is education. In turn, Bhatnagar and Ghose (2004) [5] add that the lack of education would cause significant barriers in the process of adopting new technologies and working with them, and in the case of e-commerce, the level of Internet use could be considered as one of the main prerequisites for e-shopping decisions. These views are also supported by the data of the current survey - 85% of the sample has higher education and 14% - secondary education. Additionally, as may be observed, 66% of respondents define the intensity of their online shopping experience as high and very high, and 28% of them say that they sometimes order items, but do not define it as rare.

At this point, it could be concluded that the survey sample includes people, who are familiar with both Internet use and the advantages, disadvantages, and risks associated with the process of online purchasing. None of the respondents declared that they had never shopped online (Figure 3.2).

	Criteria	Number	% of Sample
Demographic Sample	226		80%
Age	from 18 to 30	61	27%
	from 31 to 45	136	60%
	from 46 to 60	24	11%
	from 61 to 75	5	2%
Gender	Man	98	43%
	Woman	127	56%
	Other	1	<1%
Education	Secondary	31	14%
	High	193	85%
	No answer	2	1%

Table 3.1: Demographic data

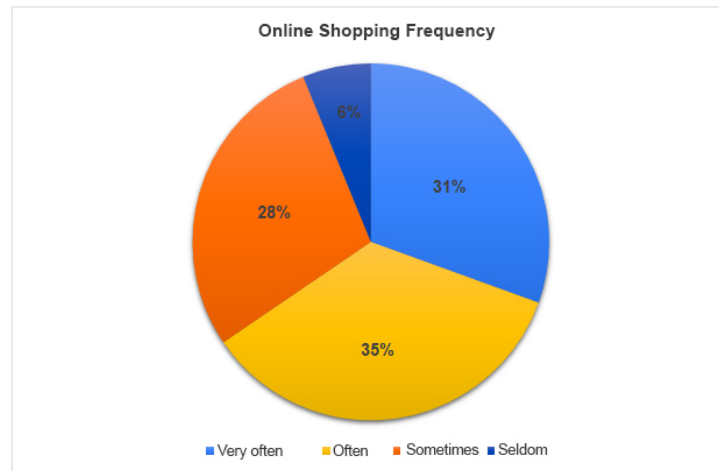


Figure 3.2: Online shopping frequency

3.1.2 Analyzing the Relationship Between Personality and Consumer Preferences

Since personality could be described as a set of specifications influencing human behavior, the availability of data related to personality characteristics provides information that could be used to predict human actions in various situations [48].

In this regard, the author observes the respondents' attitude to 19 main web

store characteristics (Table 3.2), which each e-commerce site is recommended to have to remain competitive. Thus, after establishing the personal profile of participants (applying TIPI test) and their attitude to existing risks (applying Risk Averseness Scale) and after gathering information related to their web site features preferences, the author conducts bivariate analysis to check the existence of a linear relationship between two variables and to analyze later. To construct the equations, the author seeks significant relationships between 6 independent variables (5 personality traits and the respondents' attitude to risk-taking) and each of the observed 19 functionalities of online stores - dependent variables. Thus, he checks a total of 114 items, and the existence of correlations means that it could be formed mathematical equations to assess and predict user preferences. Using the PSPP¹ program for the aim, only significant correlations between variables with correlation levels $p < 0.05$ are considered.

The correlation coefficients in Table 3.2 show that there are 21 significant relationships between the observed dependent and independent variables.

¹GNU PSPP is a free program for statistical analysis - <https://www.gnu.org/software/pspp/>

Dependent variables (each one of the observed 19 functionalities of online stores)	Independent variables (each one of 5 personality traits, as well as the respondents' attitude to risk taking)					
	Extraversion (dynamic and intensive social contacts)	Agreeableness (kindness, resourcefulness, benevolence, easily agreeing)	Conscientiousness (attention and diligence)	Emotional Stability (prone to maintain balance and self-control)	Openness to Experience (frankness and directness)	Risk Propensity (risk perception and awareness)
1 Product descriptions give me the necessary information to able to make a decision.		0.122 ^a				
2 Instead of single score, I prefer detailed product ratings.			0.120 ^a			
3 I read the expert reviews. They are essential in the decision making process.		0.162 ^a				
4 I read comments which other users have left for different purchases.		0.117 ^a				
5 I check the product availability as well delivery time before I make a purchase.			0.194 ^a			
6 I prefer to be able to see where I am in the product purchasing process.			0.176 ^a			0.125 ^a
7 I prefer to see real-time shipping quote estimates.						0.110 ^a
8 I prefer to take a look at detailed product images.			.154 ^a			0.115 ^a
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	0.205 ^a					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			0.164 ^a			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			.133 ^a	.120 ^a		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	0.110 ^a				0.121 ^a	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						0.180 ^a
18 I prefer to check the free return possibility; it is essential for me.				- 0.114 ^a		
19 I normally check for alternative (secure) payment methods like PayPal etc.			0.226 ^a	0.127 ^a		0.157 ^a

^a Significant relationship at the level 0.05

Table 3.2: Checking for significant correlations between 6 independent and 19 dependent variables

3.2 Applying of Machine Learning for Prediction of User Behavior in Internet

The psychological concept of personality accounts for individual differences in the people's enduring emotional, experiential, attitudinal, and motivational styles, whereat personality is supposed to be stable across longer periods of time. Therefore, personality traits are one of the main sources of our decisions [52]. At the same time, the use of machine learning methods allows a reliable estimating of user preferences according to their personality and perception of risk averseness.

In order to respond to the research aims and objectives in this dissertation resp. to be able to make a forecast of user preferences regarding their personality, three regression models are implemented - linear regression, decision tree, and random forest. The equations take the personality traits and risk averseness as input and yield a value showing how important functionality is to the user.

The implementation is done in Python, version 3.8 (64-bit), and all of the regression models are evaluated applying three of the most common metrics for evaluating predictions on regression machine learning problems [43], [35]:

- the Mean Absolute Error (MAE), which is the average of the absolute differences between predictions and actual values; the lower the value, better is the model's performance;
- the Root Mean Squared Error (RMSE) where the errors are squared before they are averaged; in this metric also the lower the value, better is the performance of the model;
- the Mean Absolute Percentage Error (MAPE) can be considered as a loss function to define the error termed by the model evaluation; MAPE estimates the accuracy in terms of the differences in the actual v/s estimated values; the lower the value, the better is the model's performance.

3.2.1 Prediction with Linear Regression

The biggest advantage of linear regression models is the linearity; it makes the estimation procedure simple and, most importantly, these linear equations have an easy-to-understand interpretation on a modular level. This is one of the main reasons why Linear Regression is so widespread in academic fields such as sociology, psychology, medicine, and many other quantitative research fields. At the same time, the linearity is its greatest limitation [36].

In this research, the implementation of Linear Regression starts with importing the necessary libraries and after that randomly splitting the data into training and

testing datasets using the `train_test_split()` function from *scikit-learn* library. In total, 70% of the data is used as training and 30% as test set. The train set is used for fitting the model and the test set for validation.

In this way, using Linear Regression, equations are developed based on the identified significant relationships. The significance of the equations lies in their ability to approximate how the new users, whose personality is known, are likely to behave in an online store. The personality traits and the perception of risk averseness are considered as independent and the preferences of the users as dependent variables.

After the predictions are made, the estimated results are evaluated by applying the metrics for evaluation mentioned above. The obtained results of MAE are presented in Table 3.3, of RMSE in Table 3.4, and of MAPE in Table 3.4. The averaged value of MAE for all significant relationships is 0.77, of RMSE 0.96, and of MAPE 27.55, which means that the accuracy regarding MAPE is 72.45%.

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		0.58				
2 Instead of single score, I prefer detailed product ratings.			0.68			
3 I read the expert reviews. They are essential in the decision making process.		0.81				
4 I read comments which other users have left for different purchases.		0.62				
5 I check the product availability as well delivery time before I make a purchase.			0.84			
6 I prefer to be able to see where I am in the product purchasing process.			0.93			0.90
7 I prefer to see real-time shipping quote estimates.						0.90
8 I prefer to take a look at detailed product images.			0.44			0.43
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	0.76					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			0.76			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			0.74	0.73		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	0.92				0.91	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						1.05
18 I prefer to check the free return possibility; it is essential for me.				1.01		
19 I normally check for alternative (secure) payment methods like PayPal etc.			0.72	0.73		0.72
Average Mean Absolute Error (MAE):		0.77				

Table 3.3: Linear Regression - Mean Absolute Error (MAE)

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		0.73				
2 Instead of single score, I prefer detailed product ratings.			0.87			
3 I read the expert reviews. They are essential in the decision making process.		0.98				
4 I read comments which other users have left for different purchases.		0.77				
5 I check the product availability as well delivery time before I make a purchase.			1.10			
6 I prefer to be able to see where I am in the product purchasing process.			1.15			1.17
7 I prefer to see real-time shipping quote estimates.						1.17
8 I prefer to take a look at detailed product images.			0.50			0.48
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	0.97					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			0.99			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			0.93	0.93		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	1.09				1.08	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						1.20
18 I prefer to check the free return possibility; it is essential for me.				1.16		
19 I normally check for alternative (secure) payment methods like PayPal etc.			0.91	0.94		0.94
Average Root Mean Squared Error (RMSE):	0.96					

Table 3.4: Linear Regression - Root Mean Squared Error (RMSE)

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		15.61				
2 Instead of single score, I prefer detailed product ratings.			20.29			
3 I read the expert reviews. They are essential in the decision making process.		26.34				
4 I read comments which other users have left for different purchases.		18.35				
5 I check the product availability as well delivery time before I make a purchase.			31.37			
6 I prefer to be able to see where I am in the product purchasing process.			35.95			37.69
7 I prefer to see real-time shipping quote estimates.						37.69
8 I prefer to take a look at detailed product images.			9.94			9.77
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	30.57					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			24.39			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			23.30	23.01		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	46.94				45.55	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						37.97
18 I prefer to check the free return possibility; it is essential for me.				37.91		
19 I normally check for alternative (secure) payment methods like PayPal etc.			21.44	22.36		22.05
Average Mean Absolute Percentage Error (MAPE):		27.55				

Table 3.5: Linear Regression - Mean Absolute Percentage Error (MAPE)

3.2.2 Prediction with Decision Trees

Some of the advantages of the Decision Trees method is that it is able to handle both, numerical and categorical, data; it requires little data preparation, it is able to handle multi-output problems, and it is simple to understand and interpret (trees can be visualized). Some of the disadvantages are that decision-tree learners can create over-complex trees that do not generalize the data well (overfitting) or create biased trees if some classes dominate; there are concepts that are hard to learn because Decision Trees do not express them easily, such as XOR or multiplexer problems [43].

The process of implementation of the Decision Trees is similar to the Linear Regression. It starts with importing of necessary libraries and after that, the dataset is randomly split into training (70%) and testing (30%) datasets. After the predictions are made, the estimated results are evaluated using the applied metrics for evaluation.

Table 3.7 represents the values of MAE for all significant relationships between independent and dependent variables, Table 3.4 shows the values of RMSE, and Table 3.8 these of MAPE. The averaged value of MAE for all significant relationships is 0.80, of RMSE 0.98, and of MAPE 27.96.

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		0.61				
2 Instead of single score, I prefer detailed product ratings.			0.74			
3 I read the expert reviews. They are essential in the decision making process.		0.86				
4 I read comments which other users have left for different purchases.		0.69				
5 I check the product availability as well delivery time before I make a purchase.			0.81			
6 I prefer to be able to see where I am in the product purchasing process.			0.93			0.95
7 I prefer to see real-time shipping quote estimates.						0.94
8 I prefer to take a look at detailed product images.			0.44			0.43
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	0.76					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			0.84			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			0.74	0.69		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	0.94				0.95	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						1.08
18 I prefer to check the free return possibility; it is essential for me.				1.01		
19 I normally check for alternative (secure) payment methods like PayPal etc.			0.74	0.83		0.75
Average Mean Absolute Error (MAE):	0.80					

Table 3.6: Decision Trees - Average Mean Absolute Error (MAE)

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		0.77				
2 Instead of single score, I prefer detailed product ratings.			0.92			
3 I read the expert reviews. They are essential in the decision making process.		1.04				
4 I read comments which other users have left for different purchases.		0.84				
5 I check the product availability as well delivery time before I make a purchase.			1.07			
6 I prefer to be able to see where I am in the product purchasing process.			1.16			1.16
7 I prefer to see real-time shipping quote estimates.						1.20
8 I prefer to take a look at detailed product images.			0.53			0.48
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	0.94					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			1.09			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			0.92	0.87		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	1.14				1.11	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						1.24
18 I prefer to check the free return possibility; it is essential for me.				1.19		
19 I normally check for alternative (secure) payment methods like PayPal etc.			0.92	1.05		0.95
Average Root Mean Squared Error (RMSE):	0.98					

Table 3.7: Decision Trees - Root Mean Squared Error (RMSE)

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		16.16				
2 Instead of single score, I prefer detailed product ratings.			21.46			
3 I read the expert reviews. They are essential in the decision making process.		27.14				
4 I read comments which other users have left for different purchases.		19.69				
5 I check the product availability as well delivery time before I make a purchase.			29.29			
6 I prefer to be able to see where I am in the product purchasing process.			35.79			36.34
7 I prefer to see real-time shipping quote estimates.						38.24
8 I prefer to take a look at detailed product images			10.11			9.71
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	29.51					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			26.32			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			22.60	21.16		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	47.72				47.56	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						38.97
18 I prefer to check the free return possibility; it is essential for me.				38.74		
19 I normally check for alternative (secure) payment methods like PayPal etc.			22.36	25.32		22.98
Average Mean Absolute Percentage Error (MAPE):		27.96				

Table 3.8: Decision Trees - Mean Absolute Percentage Error (MAPE)

3.2.3 Prediction with Random Forest

In Random Forests, each tree in the ensemble is built from a sample drawn with replacement (bootstrap sample) from the training set.

The Random Forest algorithm builds multiple decision trees and merges them together to get a more accurate and stable prediction. The forest it builds is an ensemble of decision trees, usually trained with the “bagging” method. Random Forest adds additional randomness to the model while developing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features, which leads to a wide diversity that generally results in a better model. Random Forest reduces overfitting in decision trees and helps to improve the accuracy, it works well with both categorical and continuous values, and it automates missing values present in the data; normalizing of data is not required as it uses a rule-based approach. Some of the disadvantages are that the model requires much computational power as well as resources as it builds numerous trees to combine their outputs and that it requires much time for training as it combines a lot of decision trees to determine the class [42], [43], [35].

Using the library *scikit-learn*, the implementation is similar to the other two ML methods. The dataset is randomly split into training (70%) and testing (30%) datasets, and the number of the trees is set to 150 (*n_estimators* = 150) (default value is 100). After the predictions are made for all significant relationships, the estimated results are evaluated using the applied metrics for evaluation.

Table 3.9 shows the values of MAE for all significant relationships between independent and dependent variables, Table 3.10 represents the values of RMSE and Table 3.11 these of MAPE. The averaged value of MAE for all significant relationships is 0.79, of RMSE 0.98, and of MAPE 27.92.

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		0.56				
2 Instead of single score, I prefer detailed product ratings.			0.73			
3 I read the expert reviews. They are essential in the decision making process.		0.86				
4 I read comments which other users have left for different purchases.		0.70				
5 I check the product availability as well delivery time before I make a purchase.			0.81			
6 I prefer to be able to see where I am in the product purchasing process.			0.93			0.95
7 I prefer to see real-time shipping quote estimates.						0.93
8 I prefer to take a look at detailed product images.			0.42			0.43
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	0.76					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			0.84			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			0.73	0.69		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	0.94				0.95	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						1.08
18 I prefer to check the free return possibility; it is essential for me.				1.00		
19 I normally check for alternative (secure) payment methods like PayPal etc.			0.74	0.84		0.74
Average Mean Absolute Error (MAE):		0.79				

Table 3.9: Random Forest - Mean Absolute Error (MAE)

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		0.76				
2 Instead of single score, I prefer detailed product ratings.			0.91			
3 I read the expert reviews. They are essential in the decision making process.		1.04				
4 I read comments which other users have left for different purchases.		0.85				
5 I check the product availability as well delivery time before I make a purchase.			1.07			
6 I prefer to be able to see where I am in the product purchasing process.			1.16			1.16
7 I prefer to see real-time shipping quote estimates.						1.19
8 I prefer to take a look at detailed product images.			0.48			0.48
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	0.94					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			1.08			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			0.91	0.87		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	1.13				1.11	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						1.24
18 I prefer to check the free return possibility; it is essential for me.				1.19		
19 I normally check for alternative (secure) payment methods like PayPal etc.			0.92	1.05		0.95
Average Root Mean Squared Error (RMSE):	0.98					

Table 3.10: Random Forest - Average Root Mean Squared Error (RMSE)

	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness to Experience	Risk Averseness
1 Product descriptions give me the necessary information to able to make a decision.		16.03				
2 Instead of single score, I prefer detailed product ratings.			21.28			
3 I read the expert reviews. They are essential in the decision making process.		27.34				
4 I read comments which other users have left for different purchases.		19.89				
5 I check the product availability as well delivery time before I make a purchase.			29.23			
6 I prefer to be able to see where I am in the product purchasing process.			35.77			36.38
7 I prefer to see real-time shipping quote estimates.						38.14
8 I prefer to take a look at detailed product images.			10.15			9.67
9 I seek to buy accessories that go along with the product I purchase (accessories and complementary offers, which complement the chosen product).	29.37					
10 I prefer to buy complementary accessories (like insurance and extended warranty) as a bundle.						
11 In order to be able to choose the right product for me I use product categorization resp. featured product filter.						
12 In order to choose among different products, I compare the product details.			26.22			
13 I tend to use different features in the cart like one-click reorder, calculate end price etc.						
14 I normally prefer to check different delivery options.			22.54	21.20		
15 I tend to use different contact/ support possibilities, in order to ensure myself about certain product features resp. to continue the buying process.						
16 I tend to write and comment product reviews. They help to clarify uncertainties about desired products.	47.70				47.41	
17 I avoid saving my personal data in web stores, so I usually prefer to buy as a guest.						38.94
18 I prefer to check the free return possibility; it is essential for me.				38.71		
19 I normally check for alternative (secure) payment methods like PayPal etc.			22.21	25.35		22.80
Average Mean Absolute Percentage Error (MAPE):	27.92					

Table 3.11: Random Forest - Average Mean Absolute Percentage Error (MAPE)

3.2.4 Results Comparison

Based on the presented results about the user preferences depending on their personality, it could be summarized that all three ML methods have achieved quite similar average values of the evaluation metrics (Table 3.12), although there are many techniques that allow optimization of the results. Some of these techniques are for example learning with more data, cross-validation, genetic algorithm, and others [4].

Evaluation metric	ML model	Average value
MAE	Linear Regression	0.77
	Decision Trees	0.80
	Random Forest	0.79
RMSE	Linear Regression	0.96
	Decision Trees	0.98
	Random Forest	0.98
<i>Accuracy</i> reg. MAPE	Linear Regression	72.45 %
	Decision Trees	72.04 %
	Random Forest	72.08 %

Table 3.12: Average values of the evaluation metrics

Before making a choice which ML model is the most appropriate for the purpose it is essential to evaluate all candidates with applicable evaluation metrics but it is also very important to visualize the distribution of the actual and predicted data [53]. From this point of view, Figures 3.3, 3.4, and 3.5 illustrate the distribution and the density of the distribution of the prediction for the dependent variable "19. I normally check for alternative (secure) payment methods like PayPal etc." depending on the risk averseness for all three ML models.

Another example of the significant relationships where the algorithms have achieved higher values regarding the evaluation metrics resp. where they don't make a good prediction is between the extroversion and "16. I tend to write and comment on product reviews. They help to clarify uncertainties about desired products." Regarding MAPE, the Linear Regression reaches an accuracy of the forecast of 54.50%, the Decision Trees and the Random Forest of 52.30%. But according to Lewis [37] these values could still be interpreted as an appropriate prediction.

Figures 3.6, 3.7, and 3.8 represent the distribution of the data of all three ML models. Of course, the distribution of the actual data is similar, but the prediction of Linear Regression is more symmetrically distributed around the median.

As a brief conclusion, it could be summarized that all of the three ML models have achieved quite a similar prediction according to the applied evaluation metrics. And although the results aren't very accurate, they could be categorized as quite

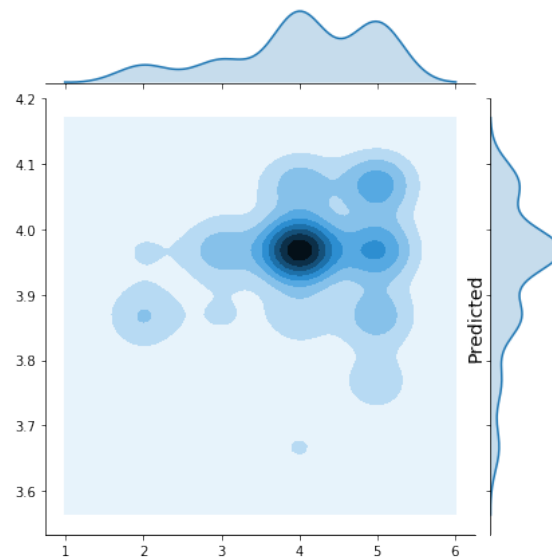


Figure 3.3: Linear Regression - prediction of checking for alternative (secure) payment methods and risk averseness, data distribution

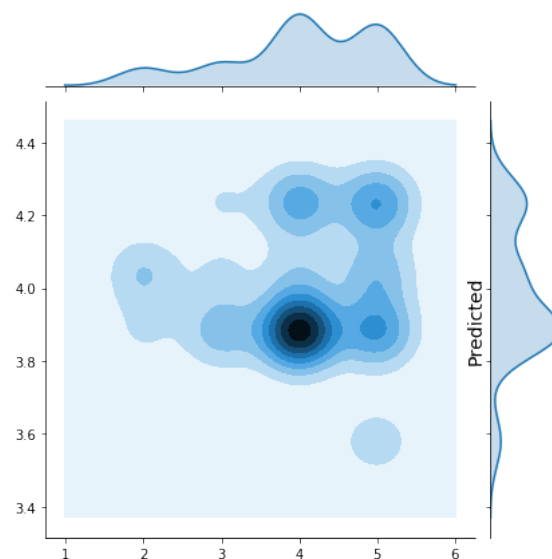


Figure 3.4: Decision Trees - prediction of checking for alternative (secure) payment methods and risk averseness, data distribution

appropriate for the aim [37]. Actually, they aren't very accurate only by a few significant relationships.

In spite of the fact that regarding the literature review and according to the obtained results all three ML models would be appropriate for the aim of this work, in the next step, an optimization of the Random Forest is proposed and implemented.

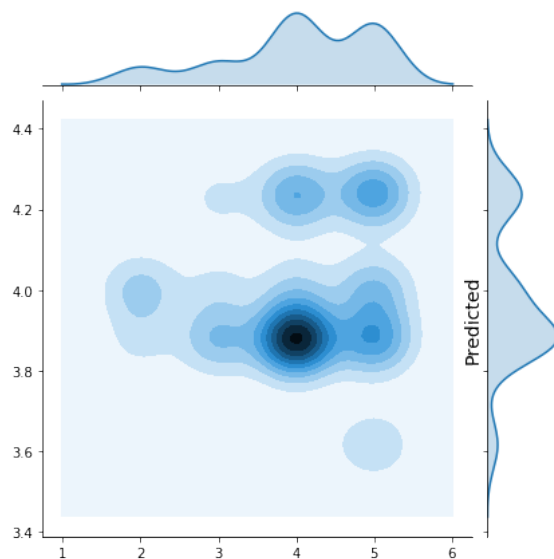


Figure 3.5: Random Forest - prediction of checking for alternative (secure) payment methods and risk averseness, data distribution

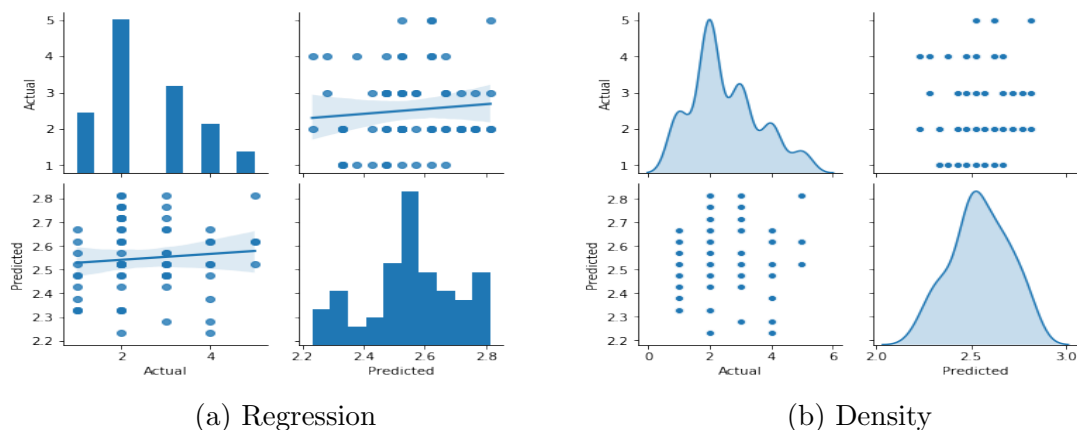


Figure 3.6: Linear Regression - extroversion and comment on product reviews, data distribution

Although the customer decision-making process is often viewed as a linear proceeding, actual researches have shown that decision-making has a non-linear nature resp. it is a dynamic process containing loops [27]. The relationship between human personality and user preferences is very complex, therefore flexible ML algorithms, capable of modeling non-linear effects and interactions, might even allow researchers to use the peculiarities of psychological measurements to increase predictive performance. Random Forest uses an ensemble of decision trees as a basis and therefore has all advantages of decision trees, such as high accuracy and no necessity of scaling data. Moreover, it also has a very important additional benefit, namely perseverance

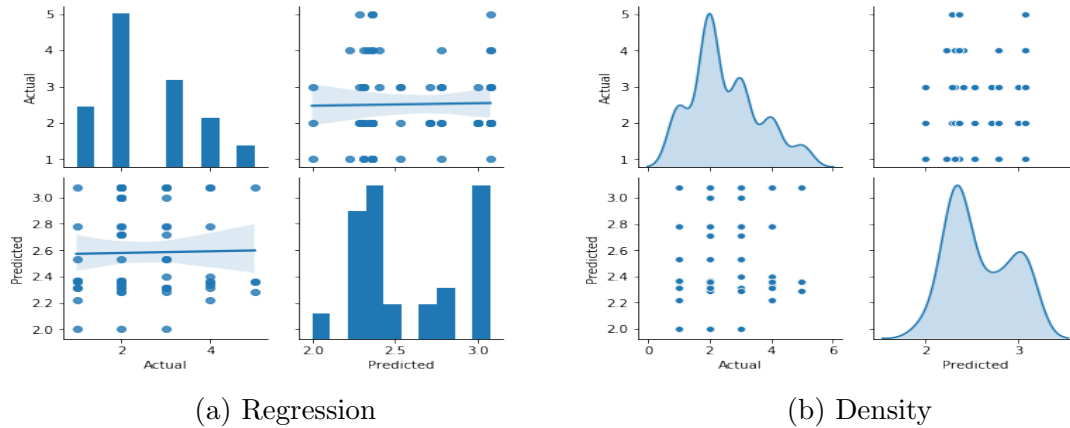


Figure 3.7: Decision Trees - extroversion and comment on product reviews, data distribution

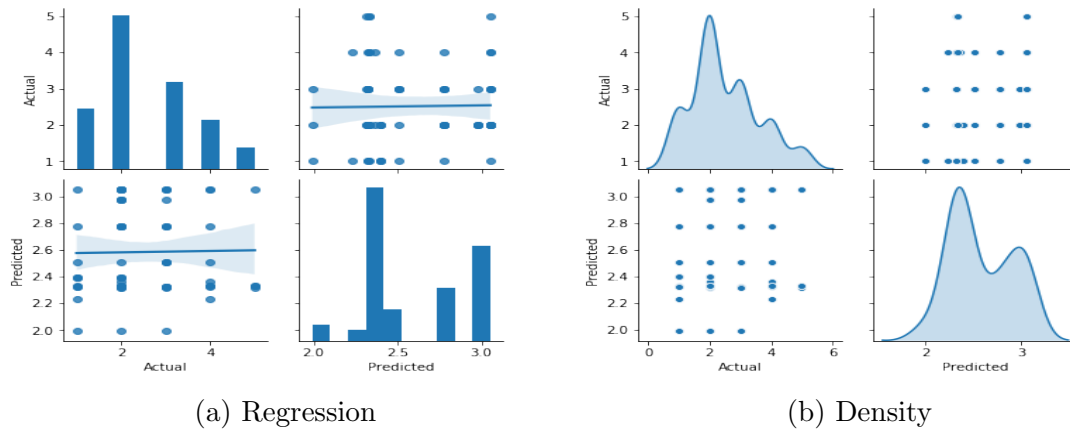


Figure 3.8: Random Forest - extroversion and comment on product reviews, data distribution

to overfitting, which occurs when a model incorporates random variation in a given dataset, that is not caused by the underlying, true relationship between predictors and criterion variables. Random Forest algorithm doesn't require data scaling and has higher prediction accuracy, and it is easier for hyperparameters tuning [34], [7], which makes the algorithm very appropriate for research in the field of personality.

3.2.5 Optimization of Random Forest

The optimization of any ML model is a very important step in the process of solving the global problem, whereat the different models have various hyperparameters which could be tuned in order to optimize the results.

For the aim of this dissertation, it is proposed and implemented an optimization with cross-validation, applying the class *GridSearchCV* of the library *scikit-learn*, as well optimization with TPOT (Tree-based Pipeline Optimization Tool), which uses genetic programming (GP) to explore different pipelines and recommend one with an optimal cross-validated score after a specified number of generations.

In the proposed optimization, *GridSearchCV* goes through all the combinations 10 times because the value of the cross-validation generator is set to 10 ($cv=10$). In this case, there is a total of 120 fits.

In this configuration, the method does not lead to an improvement of the results according to MAPE in 5 of all 21 significant relationships between personality traits and consumer preferences in online shopping. In the other 16 significant relationships, the accuracy regarding MAPE has been improved differently. The highest improvement is in the propensity to check for alternative (and safer) payment methods depending on the emotional stability of the user (2.58%). The improvement in the average accuracy for all 21 significant relationships regarding MAPE is 0.53 % or from 72.08 % to 72.46 %. According to the MAE and RMSE metrics, there is also a slight improvement, which is varied in the different relationships.

TPOT is built on the *scikit-learn* library and follows the *scikit-learn* API closely, it is open source, well documented, and under active development. It can be used for regression and classification tasks and it has special implementations for medical research. TPOT uses a genetic search algorithm to find the best parameters and model ensembles; it tries a pipeline, evaluates its performance, and randomly changes parts of the pipeline in search of better-performing algorithms. By default (100 generations and 100 populations), TPOT would have to evaluate 10 000 configurations before finishing [43].

In the proposed optimization, TPOT has to evaluate 1 100 configurations as the population size is set to 100 and the number of iterations to the run pipeline optimization process is set to 10 ($population_size + (generations \times offspring_size)$). By default, the number of offspring to produce in each genetic programming generation is equal to the number of population size. In this configuration, the algorithm has improved the results regarding MAPE in 19 of all 21 significant relationships. The improvement of the average accuracy for all 21 significant relationships regarding MAPE is 0.69% or from 72.08% to 72.58% accuracy. According to the MAE and RMSE metrics, there is also a slight improvement, which is varied in the different relationships.

As an example, Figure 3.9 illustrates the actual and the predicted values of Random Forest for the dependent variable "I avoid saving my personal data in web stores, so I usually prefer to buy as a guest" depending on the risk averseness, as well as the optimization with *GridSearchCV*. Figure 3.10 represents the optimization with TPOT. Regarding MAPE, Random Forest has achieved an accuracy of the

prediction of 77.20%; after optimization with cross-validation using *GridSearchCV*, the achieved accuracy is of 77.71%, and with TPOT of 78.23%.

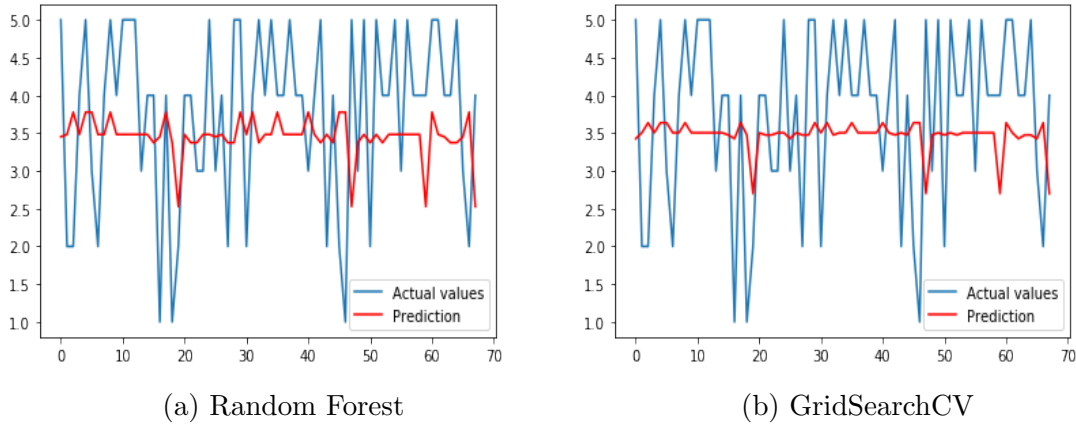


Figure 3.9: Saving personal data and risk averseness - actual and predicted data of Random Forest and optimization with GridSearchCV

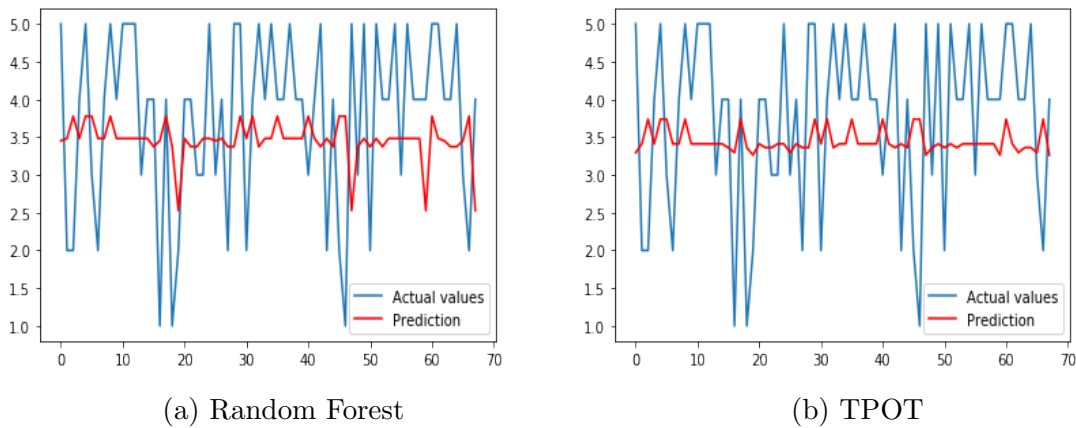


Figure 3.10: Saving personal data and risk averseness - actual and predicted data of Random Forest and optimization with TPOT

Based on the presented findings, it could be summarized that both of the methods lead to improvement of the results of Random Forest according to the achieved values of the evaluation metrics, wherent TPOT has scored slightly better results. At the same time, because of their nature, both algorithms are time-consuming procedures. But with a proper configuration respecting the aim of the research, both of them could achieve satisfactory results within acceptable estimation time.

3.3 Decision Making in E-Commerce Based on User's Personality

Considering that personality could be conceptualized as a set of stable individual differences influencing our specific behavior, attitude, and reactions to environmental stimuli and based on the results of the conducted study, the author's assumption about the existence of empirical relationships between personality and human's behavior is confirmed. It could be concluded that personality has a significant role in the process of online purchasing because each consumer adds his unique characteristics influencing his behavior on a subconscious level.

According to the study results, **more extroverted** individuals would react positively if they have an opportunity to purchase additional articles and accessories related to the already chosen product, whereby Random Forests optimization achieves 71% accuracy of the mean absolute percentage error (MAPE). These individuals are also actively involved in both writing and reading comments, and they believe this plays a significant role in making a purchase decision. More extroverted personalities are more active, enthusiastic, talkative, energetic, and dominant. They tend to have a higher frequency and intensity of social relationships and desire to express their leadership [51], while introverts tend to be more reserved and often require a period of solitude and silence.

More agreeable users prefer to read comments left by other customers before purchasing the item they want. According to the current study, applying the Random Forest model could also predict their preferences with 81% MAPE forecast accuracy. For them, this is essential because they are more resourceful, need more information and shared experience with others not only on the internet [15]. These users also pay particular attention to the products' description from the point of view of its informativeness (84% MAPE forecast accuracy), as well as to the expert evaluation of the considered article (74% MAPE forecast accuracy).

More conscientious people pay special attention to details, prefer to be able to choose between alternative products and to compare their parameters. This is because conscientious people are well organized and purposeful [9]. The Random Forests method here achieves a forecast accuracy of 76% according to MAPE. The purchasing process could become a better experience for these customers if they can see more detailed product photos (90% accuracy of the MAPE forecast), as well as an item evaluation based on different sub-criteria (80% accuracy). These people desire to complete the set tasks to the end [2] and this fact explains why it is so important for them to be able to check the current availability and delivery time of items (70% accuracy of the forecast), as well as the various options for this (78 % accuracy) before they make an online purchase. They also desire to have the opportunity to track their order status and to use alternative and more secure payment methods. By this criteria, the Random Forest algorithm achieves 79%

MAPE forecast accuracy.

The **emotionally stable** people tend to behave confidently and calmly, as well as to use a rational approach by problem-solving [2]. It is clear that they also prefer to have different delivery options and more secure payment methods, whereby the Random Forest algorithm achieves over 75% accuracy of the MAPE forecast. The opposite of emotional stability is neuroticism - neurotic people control difficulty their emotions and their state of stress [49]. This also confirms the current study showing that the neurotic consumers need to have the opportunity for a free return.

Openness to new experiences is defined as a tendency towards active imagination, intellectual curiosity, a willingness to consider new ideas and try new things [9]. The more open online users are usually more creative, and the experience sharing with others is especially important for them to clarify issues related to the selected products and services before their purchase decision [10], [49]. According to the results, people with high levels of openness prefer to comment and ask questions about the considered products to ensure its quality. In this case, the Random Forest method achieves one of the lowest prediction regarding MAPE (53%). Nevertheless, according to Lewis [37], this can still be an acceptable forecast.

In accordance with the degree of digitalization in today's life and the growing risk of fraud on internet, the current study also confirms the existence of a significant link between users' **risk perception** and their willingness to share personal data and to use more secure payment methods on the internet (62% and 78% accuracy of the forecast with Random Forests according to MAPE). Moreover, the risk perception is positively related to consumers' preference to track the status of their order and to see timely the delivery price. It is also particularly important for online customers to see detailed product photos to reduce the likelihood of disappointment in the items after the delivery, whereby the Random Forest algorithm achieves 90% MAPE forecast accuracy.

Conclusion

Knowledge of the user's personality, as well as the techniques allowing prediction of his or her needs and preferences, opens further new horizons. Considering that the human factor plays a crucial role in social and economic processes, the topic of personality is applicable in various fields of science and the contemporary world. For example, Project Management, Strategic Management, Human Resource Management, and Recruitment, Customer Relationship Management (CRM) and also CRM in Social Media (CRM 2.0), Risk Management and Assessment, Marketing and Advertising, Knowledge Management, Expert Systems, Social Commerce, Improving Customer Loyalty and E-commerce. Taking into account that today's economic and social processes are user-oriented, all listed here current scientific trends could find application in the field of E-commerce, as it is a market segment with sustainable development in recent years [50].

The results of the study show that certain e-shops' functionalities are more preferred by certain groups of users. Thus, knowing the consumers' personality and applying the methods of Machine Learning to predict what users' preferences would be, makes it possible to create models of behavior and decision making in e-commerce. For instance, this would make the personalization of the users' interface possible, and that could better meet their expectations and needs.

In conclusion, but not least, it should be emphasized as extremely important that research in this not yet well-known field must respect all legal and ethical norms, and the results have to be used solely for human benefits.

Limitations

As it is already mentioned above, each investigation has its **limitations**, especially when it is human-centered. Although the survey sample has a multicultural background and consists of 226 respondents living in more than 10 countries over the world, most of them are representatives of the European culture. In turn, this limits the application of the achieved results to some extent within these cultural sub-groups. Although according to the scholars, the Big Five concept is considered

consistent and stable in different languages and cultures, Friedman [17] states that studies related to personality are applicable only to the relevant cultural environments with similar political, social, and ideological structure and norms, and they would not be valid in other cultural societies. However, other authors claim that these differences are mainly related to the translation of the measurement tool, as well as to the genetic difference between the participants [18].

Other limitations are related to the reliability of the participants' answers according to the assessment of their personality profile, as well as their preferences for particular functionalities.

Guidelines for Future Research

The study of personality and its application in various fields of modern life, especially in the field of modern technologies, arouse the interest of the business community and researchers as well. But at the same time, it is extremely important to observe ethical norms and not to abuse the achieved results, which should be used only for the users' benefit.

Regarding future research on the dissertation topic, it is recommended to analyze to what extent the personalization of e-shops increases consumer satisfaction and also if the users are enough supported in the decision making process, meaning whether the e-shop usability is significantly improved.

As it is stated above, the TIPI test proves to be a useful tool in identifying user's personalities on the internet, because it is a measure of the Big Five dimensions that is appropriate in cases when a very short instrument with optimized validity is needed. Although most implicit personality assessment methods have the potential for wide application, at the same time, more of them generally suffer from insufficient measurement accuracy. In this aspect, in the future, it is recommendable such approaches to be additionally verified by direct inquiry, as the TIPI test is appropriate for this aim.

Moreover, before widely applying the results in practice, it is recommendable to lead studies like this again and, if possible, with a larger sample. Another way to control the results could be a supplementary eye-tracking test with the participants.

The individual specifics can be used as predictors of consumer behavior and decision-making on the internet, but the key to personalizing an interface is the accurate prediction of user preferences, therefore such studies have to be carefully conducted in accordance with the scientific guidelines and recommendations.

Publications

1. Ketipov, R., Kolev, K., Sevova, Zh., Blagoev, I., Petrov, P., Kostadinov, G., Zakinski, I. Time series trend and seasonality preprocessing with genetic algorithms. // *XXVII International Symposium "Management of Energy, Industrial and Ecological Systems"*, 16-17 May 2019, Bankya, Bulgaria, 65-68. [In Bulgarian]
2. Ketipov, R., Kolev, K., Sevova, Zh., Blagoev, I., Petrov, P., Kostadinov, G., Zakinski, I. Removal of linear component and sinusoidal harmonics from time series with differential evolution. // *Proceedings of the Annual University Scientific Conference of NMU "V. Levski"*, 27-28 June 2019, V. Tarnovo, Vol. 10, Scientific direction "Technical sciences", 1586-1594, ISSN: 2367-7481. [In Bulgarian]
3. Balabanov, T., Zankinski, I., Ketipov, R. Weights Permutation in Multilayer Perceptron. // *Proceedings of International Conference on Big Data, Knowledge and Control Systems Engineering, "John Atanasoff" Union on Automatics and Informatics*, Bulgaria, 2018, 23-27, ISSN: 2367-6450.
4. Balabanov, T., Ketipov, R., Atanassova, Z. MLP with Stochastic Manipulated Hidden Layer. // *Proc. of the International Scientific Conference - UNITECH 2018*, 2, University Publishing House Vasil Aprilov - Gabrovo, 2018, 324-328, ISSN: 1313-230X.
5. Balabanov, T., Ivanov, S., Ketipov, R. Solving Combinatorial Puzzles with Parallel Evolutionary Algorithms. // *Lecture Notes in Computer Science, including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, 2020, 11958 LNCS, 493-500, DOI: 10.1007/978-3-030-41032-2_56, SJR 2019 – 0.427.
6. Ketipov, R., Kostadinov, G., Petrov, P., Zankinski, I., Balabanov, T. Human-Computer Mobile Distributed Computing for Time Series Forecasting. // *Communications in Computer and Information Science (CCIS)*, Springer, 2019, 1141 CCIS, 508-509, DOI: 10.1007/978-3-030-36625-4_40, SJR 2019 – 0.188.

7. Ketipov, R., Kostadinov, G., Petrov, P., Zankinski, I. Genetic Algorithm Based Formula Generation for Curve Fitting in Time Series Forecasting Implemented as Mobile Distributed Computing. // *High-Performance Computing (HPC)*, Borovets, Bulgaria, 2019. *Studies in Computational Intelligence*, 2021, 902 SCI, 40-47, DOI: 10.1007/978-3-030-55347-0_4, SJR 2019 - 0.215.
8. Ketipov, R., Kolev, K., Sevova, J., Blagoev, I., Petrov, P., Kostadinov, G., Zankinski, I. Trend and Seasonality Removal with Differential Evolution. // *Information technologies and control*, 4, 2018, 17-22, ISSN: 1312-2262.
9. Zankinski, I., Keremedchiev, D., Blagoev, I., Ketipov, R., Kolev, K., Kostadinov, G., Petrov, P. Recursive brute-force selection operator in genetic algorithms. // *International Scientific Conference UNITECH*, 15 - 16.11.2019, Gabrovo, 227-232, ISSN:11313-230X.

Results

Base on the conducted secondary and primary research, as well as on the obtained results, it may be concluded that the dissertation aims and objectives are achieved. As a consequence, the following **thesis contributions** to the current state of knowledge may be formulated:

1. It is chosen an appropriate psychometric model to measure the user's personality based on the Big Five concept, which according to the scholars is a state-of-the-art assessment tool today, and besides, it is stable in different languages and cultures. But due to its length, another brief measure is approved which allows exploitation of the user's profile based on the same five determinants. Ten Item Personality Inventory (so-called TIPI test) by Gosling [21] reaches adequate levels for validity and reliability, measuring personality in just about a minute and this makes it perfectly suitable for cases when it is required almost instantaneous personality identification.
2. It is set a list of 19 main functionalities which are categorized into 3 subgroups - content and appearance, user interface tools, and factors influencing risk averseness. They are applicable for the most current e-shops and could be used as a starting point in studies of the relationship between personality and user preferences.
3. An appropriate research strategy and design are created respecting basic standards of ethics and neutrality, whereat the aspect of risk avoidance is considered as a additional personality determinant. The questionnaire is translated into three languages - Bulgarian, English, and German; after analysis of the results of the empirical study, the existence of significant correlations between basic personality determinants and user preferences in the process of online shopping has been determined.
4. Three Machine Learning models (linear regression, decision trees, and random forest) are proposed and implemented which experimentally forecast the users' preferences to some e-shop functionalities based on their personality; the calculations include only the significant correlations found previously between personality determinants (independent variables) and observed e-shops

functionalities (dependent variables); it is also proposed an optimization for random forest.

5. Based on the obtained results, models of consumer behavior in the process of decision making in the online purchasing sequence are summarized. It is pointed out that based on users' personality and through applying the Machine Learning methods consumer preferences could be successfully predicted.


The achieved results at the end of the study could be used as a starting point for further personalization in the field of e-commerce. So, knowing the user's personality, his or her preferences could be detailed predicted and so, the need for an additional inquiry in this regard could be eliminated.

Declaration

This dissertation is the result of my own work and investigations, except where otherwise stated. All thoughts taken directly or indirectly from external sources are properly denoted as such, a bibliography is appended.

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Sofia, 07/05/2021

Signature: 
/Rumen Ketipov/

Bibliography

- [1] ASHTON, M. C., LEE, K. The Prediction of Honesty-Humility-Related Criteria by the HEXACO and Five-factor Models of Personality. // *Journal of Research in Personality*, Vol. 42(5), 2008, p. 1216–1228.
- [2] BAKKER, A. B., VAN DER ZEE, K. I., LEWIG, K. A., DOLLARD, M. F. The Relationship Between the Big Five Personality Factors And Burnout: A Study Among Volunteer Counselors. // *The Journal of Social Psychology*, Vol. 146(1), 2006, p. 31–50.
- [3] BARKHI, L., WALLACE, L. The Impact of personality Type on Purchasing Decisions in Virtual Stores. // *Information Technology Management*, Vol. 8, 2007, p. 313–330.
- [4] BAYLEY, S., FALESSI, D. Optimizing Prediction Intervals by Tuning Random Forest via Meta-Validation, 2018.
- [5] BHATNAGAR, A., GHOSE, S. Online Information Search Termination Patterns Across Product Categories and Consumer Demographics. // *Journal of Retailing*, Vol. 80, 2004, p. 221–228.
- [6] BOLOGNA, C., DE ROSA, A. C., DE VIVO, A., GAETA, M., SANSONETTI, G., VISERTA, V. Personality-Based Recommendation in E-Commerce. // *Conference: EMPIRE 2013 workshop, 1st Workshop on Emotions and Personality in Personalized Services*, 2018.
- [7] BREIMAN, L. Random Forests. // *Machine Learning*, Vol. 45, 2001, p. 5–32.
- [8] BUETTNER, R. Predicting User Behavior in Electronic Markets Based on Personality-mining in Large Online Social Networks. // *Electronic Markets*, Vol. 27(3), 2017, p. 247–265.
- [9] CAMPS, J., STOUTEN, J., EUWEMA, M. The Relation Between Supervisors' Big Five Personality Traits and Employees' Experiences of Abusive Supervision. // *Frontiers in Psychology*, Vol. 7(112), 2016, p. 1–11.

- [10] COSTA, P. T., JR., MCCRAE, R. R. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources, 1992.
- [11] DERI, L., ARMENSKI, T., TESANOVIC, D., BRADIĆ, M., VUKOSAV, S. Consumer Behaviour: Influence of Place of Residence on the Decision-making Process When Choosing a Tourist Destination. // *Economic Research-Ekonomiska Istraživanja*, Vol. 27(1), 2014, p. 267–279.
- [12] DONTU, N., GILLILAND, D. The Infomercial Shopper. // *Journal of Advertising Research*, Vol. 36, 1996, p. 69–76.
- [13] EYSENCK, H. J. Dimensions of Personality. London: Methuen, 1947.
- [14] FARAG, S., KRIZEK, K. J., DIJST, M. Exploring the Use of E-shopping and Its Impact on Personal Travel Behaviour in the Netherlands. // *Transportation Research Record*, Vol.18, 2003, p. 47–54.
- [15] FLANAGIN, A. J., METZGER, M. J., PURE, R., MARKOV, A. Mitigating Risk in Ecommerce Transactions: Perceptions of Information Credibility and the Role of User-generated Ratings in Product Quality and Purchase Intention. // *Electronic Commerce Research*, Vol. 14, 2014, p. 1–23.
- [16] FREUD, S. Three Essays on the Theory of Sexuality. 1905. 24.08.2020 // < [https : //www.sas.upenn.edu/ cavitch/pdf - library/FreudsEThreeEssayscomplete.pdf](https://www.sas.upenn.edu/cavitch/pdf-library/FreudsEThreeEssayscomplete.pdf) >
- [17] FRIEDMAN, H., SCHUSTACK, M. Personality: Classic Theories and Modern Research. 6th ed., Pearson Education Inc., 2016. ISBN 978-0-205-99793-0.
- [18] FUNDER, D. The Personality Puzzle. WW Norton & Company, New York, 2010.
- [19] GOLDBERG, L. The Structure of Phenotypic Personality Traits. // *American Psychologist*, Vol. 48(1), 1993, p. 26–34.
- [20] GOLDBERG, L. R. Language and Individual Differences: The Search for Universals in Personality Lexicons, In L. Wheeler (Ed.), Beverly Hills, CA: Sage. // *Review of Personality and Social Psychology*, Vol. 2, 1981, p. 141-165.
- [21] GOSLING, S. D., RENTFROW, P. J., SWANN, W. B. A Very Brief Measure of the Big-five Personality Domains. // *Journal of Research in Personality*, Vol. 37(6), 2003, p. 504–528.
- [22] GRAY, D., Doing Research in the Real World, Thousand Oaks, CA: SAGE Publications, 3rd edition, 2014.

- [23] HOLLAND, J. L. Making Vocational Choices: A theory of Vocational Personalities and Work Environments. Odessa, FL: Psychological Assessment Resources, 3rd edition, 1997.
- [24] HURWITZ, J., KIRSCH, D. Machine Learning For Dummies. IBM Limited edition. Hoboken, John Wiley & Sons, Inc, 2018.
- [25] IOANNIDIS, J. P., TARONE, R., MCCLAUGHLIN, J. K. The False-positive to False-negative Ratio in Epidemiologic Studies. // *Epidemiology*, Vol. 24, 2011, p. 450–456.
- [26] JOHN, O. P., SRIVASTAVA, S. The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. // *Handbook of Personality: Theory and Research*, New York: Guilford Press, Vol. 2, 1999, p. 102–138.
- [27] KARIMI, K., PAPAMICHAIL, N., HOLLAND C. P. The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behaviour. // *Decision Support Systems*, Vol. 77, 2015, p. 137-147.
- [28] KAZEMINIA, A., KAEDI, M., GANJI, B. Personality-based personalization of online store features using genetic programming: Analysis and experiment. // *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 14(1), 2019, p. 16-29.
- [29] KAHNEMAN D., TVERSKY A. Subjective Probability: A Judgment of Representativeness. // *Cognitive Psychology*, Vol. 3, 1972, p. 430–454.
- [30] KOKLIC, M., VIDA, I. A Strategic Household Purchase: Consumer House Buying Behaviour. // *Managing Global Transitions*, Vol. 7(1), 2009, p. 75–96.
- [31] KOSINSKI, M., BACHRACH, Y., KOHLI, P., STILLWELL, D., GRAEPEL, T. Manifestations of User Personality in Website Choice and Behavior on Online Social Networks. // *Machine Learning*, Vol. 95, 2014, p. 357–380.
- [32] KOTHARI, C. Research Methodology. Methods and Techniques. New-Delhi: New Age International Publishers, 2nd edition, 2004.
- [33] KWON, K., KIM, C. How to Design Personalization in a Context of Customer Retention: Who Personalizes What and to What Extent? // *Electronic Commerce Research and Applications*, Vol. 11, 2012, p. 101–116.
- [34] KUBAT, M. An Introduction to Machine Learning. eBook, Springer, 2017.
- [35] KUHN, M., JOHNSON, K. Applied Predictive Modeling. eBook, Springer, 5th edition, 2016.

- [36] LANGTANGEN, H. P. A Primer on Scientific Programming with Python. Springer, 5th edition, 2016.
- [37] LEWIS, C. D. Industrial and Business Forecasting Methods : A Practical Guide to Exponential Smoothing and Curve Fitting. Butterworth Scientific, London, 1982.
- [38] MAX-PLANCK-GESELLSCHAFT. Risikobereitschaft ist ein relativ stabiles Persönlichkeitsmerkmal. 2017. 18.12.2020 <<https://www.mpg.de/11679764/risikoquotient>>
- [39] MCCRAE, R. R., & JOHN, O. P. An introduction to the five-factor model and its applications. // *Journal of Personality*, 60(2), 175–215, 1992.
- [40] MCCRAE, R. R., COSTA, P. T. More Reasons to Adopt the Five-Factor Model. // *American Psychologist*, Vol. 44(2), 1989, p. 451–452.
- [41] NIELSEN, J. Designing Web Usability: The Practice of Simplicity. New Riders Publishing, Indianapolis, 1999.
- [42] ORRU, G., MONARO, M., CONVERSANO, C., GEMIGNANI, A., SERTORI, G. Machine Learning in Psychometrics and Psychological Research. // *Front. Psychol.*, Vol. 10, 2020. DOI: 10.3389/fpsyg.2019.02970
- [43] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O. ET AL. Scikit-learn: Machine Learning in Python. // *Journal of Machine Learning Research*, Vol. 12, 2011, p. 2825–2830.
- [44] SARKER, S., BOSE, T. K., PALIT, M., HAQUE, E. Influence of Personality in Buying Consumer Goods-a Comparative Study Between Neo-Freudian Theories and Trait Theory Based on Khulna Region. // *International Journal of Business and Economics Research*, Vol. 2(3), 2013, p. 41–58.
- [45] SAUNDERS, M., LEWIS, P., THORNHILL, A. Research Methods for Business Students. Harlow: FT/ Prentice Hall, 5th edition, 2009.
- [46] SCHMITT, D., P., ALLIK, J., MCCRAE, R., R., BENET-MARTÍNEZ, V. The Geographic Distribution of Big Five Personality Traits: Patterns and Profiles of Human Self-Description Across 56 Nations. // *Journal of Cross-Cultural Psychology*, Vol. 38(2), 2007, p. 173–212.
- [47] SMIGIEL, D. Theories of Personalities: Everything you need to know, CogniFit. Mental, physical health and neuroscience. 2018. 15.10.2020 < <https://blog.cognifit.com/theories-of-personalities/> >

- [48] SOURI, AL., HOSSEINPOUR, SH., RAHMANI, A., M. Personality Classification based on Profiles of Social Networks Users and the Five Factor Model of Personality. // *Human Centric Computing and Information Sciences*, 2018.
- [49] TSAO, W.-C., CHANG, H.-R. Exploring the Impact of Personality Traits on Online Shopping Behavior. // *African Journal of Business Management*, Vol. 4(9), 2010, p. 1800–1812.
- [50] TURBAN, E, OUTLAND, J., KING, D. ET AL. Electronic Commerce 2018. A Managerial and Social Networks Perspective. eBook, Springer, 2018.
- [51] USAKLI, H. A Study on Personality Traits and Shopping Online. // *Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach*, Vol. 121, Seria: Administracja i Zarzadzanie (48), 2019.
- [52] VINCIARELLI, A., MOHAMMADI, G. A Survey of Personality Computing. // *IEEE Transactions on Affective Computing*, Vol. 5(3), 2014, p. 273–291.
- [53] WASKOM, M. Seaborn: statistical data visualization. 2015. 11.12.2020 < [https : //seaborn.pydata.org/index.html](https://seaborn.pydata.org/index.html) >



БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

АВТОРЕФЕРАТ НА ДИСЕРТАЦИЯ

за присъждане на образователна и научна степен “доктор” по научна
специалност “Информатика“

ИНДИВИДУАЛНОСТ И МОДЕЛИ ПРИ ВЗЕМАНЕ НА РЕШЕНИЕ В ИНТЕРНЕТ

Румен Руменов Кетипов

Ръководител: Доц. Вера Ангелова

Научно жури:

Акад. Иван Попчев
Проф. Таня Пенчева
Проф. Мария Нишева-Павлова
Проф. Даниела Борисова
Доц. Ирина Радева



**Институт по информационни и
комуникационни технологии**
Секция „Интелигентни системи“

Увод

Както хората са различни, така разнообразни са и техните специфики [9] и затова възможността да бъдат направени заключения по отношение на взаимовръзката между личността и моделите на човешко поведение в интернет се крие във факта, че хората с приблизително подобни относителни черти се държат по подобен начин както офлайн, така и в онлайн средата. Те имат подобни навици, предпочитания и поведение независимо дали става въпрос за социалните медии, по време на онлайн пазаруване или просто при търсене на конкретна информация [50].

Личностните характеристики разграничават хората един от друг и отразяват техните индивидуални различия, като според Йорданова [1] и Кехайова-Стойчева [2] в литературата липсва единна концепция относно разбирането на понятието „личност“. Съгласно съвременната психология личността на човека се формира въз основа на неговото взаимодействие със света и затова характеристиките личностни черти у всеки от нас дават основно обяснение за нашето поведение и действия свързани със заобикалящата ни среда. Ето защо изучаването на детерминантите на характера, известно като „психология на личността“, цели да опише набора от емоционални и поведенчески черти на дадената личност [25], [41], [42].

Според проучване на Университета Базел и Института Макс Планк (Universität Basel & Max-Planck-Institut), проведено през 2017 г., готовността за поемана на риск трябва да бъде разглеждана като стабилна във времето допълнителна детерминанта на личността [40]. Така, съчетавайки оценяването на личността и индивидуалната склонност към риска чрез валидни в науката инструменти, се дава възможност индивидуалността на човека да бъде по-детайлно изучена.

С динамичното развитие на технологиите интернет става неотделна част от нашето ежедневие, в тази глобална мрежа ние се информираме и добиваме нови знания, комуникираме, изразяваме емоции и впечатления, избираме и пазаруваме. В отговор на това, изучаването на взаимовръзките между личността и предпочитанията на отделния индивид в интернет, в това число и в сферата на електронната търговия, буди все повече интерес в научните и бизнес среди.

Цел и задачи на дисертацията

Тази дисертация има за **цел** да анализира и установи дали и как различните детерминанти на личността на потребителите оказват влияние върху тяхното поведение в онлайн търговията. В случай на установяване на съществуващи значими връзки между личностните детерминанти на потребителите, включително тяхното индивидуално отношение към риска и предварително подготвен набор от основни функционалности на електронните магазини, авторът си поставя за цел въз основа на тези връзки да създаде поведенчески модели за прогнозиране на склонността и предпочитанията на потребителите при вземане на решение в процеса на онлайн пазаруване съгласно тяхната индивидуалност.

За постигане на целта, авторът си поставя следните **задачи**:

- да се изучат различните теории и модели за анализ и измерване на личността и да бъде избран подходящ психометричен модел за целта на изследването;
- да бъде избран набор от функционалности на електронните магазини, които са характерни и приложими за повечето от тях;
- да се създаде стратегия и подготви дизайн на емпиричното проучване, като бъдат спазени основни стандарти за етичност и неутралност; изследването да бъде подготвено на три езика - български, английски и немски, с цел по-широк обхват и валидност на изследването; да бъде добавен аспект на индивидуалната склонност на потребителите към риска; да се анализират резултатите от изследването и да се провери за корелации между независимите и зависимите променливи;
- да се изберат и реализират модели от областта на машинното обучение, с цел прогноза на зависимите променливи в установените значими корелации и да се направи анализ на резултатите; да се избере подходящ модел, за който да се предложи и имплементира оптимизация;
- въз основа на получените резултати от цялостното изследване да се съставят модели за поведение на потребителите при вземане на решение в онлайн търговията.

Структура на съдържанието

За събиране на цялата необходима информация, свързана със събирането на вторични и първични данни за провеждане на необходимите анализи и тестове, за реализирането на изследователските цели и задачи, както и за постигането

на качествени, валидни и оптимизирани резултати, настоящият дисертационен труд разработен и организиран в три логически структурирани глави.

В първа глава е направен преглед на голяма част от наличната научна литература относно влиянието на личността върху поведението и решенията на потребителите в интернет. Тук са изброени още и някои от основните теории в личностната психология, като се обръща особено внимание на най-съвременната от тях - Теорията за личностните черти, както и на Петфакторния модел, който оценява личностните детерминанти, лежащи в основата на човешкото поведение. В тази глава подробно е изучено и въздействието на личността върху потребителското поведение в интернет и някои от основните функционалности на електронните магазините. Изучено е също така и индивидуалното отношение на потребителите към риска в процеса на онлайн пазаруване. В края се извършен преглед на възможностите за приложение на методите за машинно обучение с цел надеждното прогнозиране на потребителското поведение в онлайн средата, както и на голяма част от проведените през последните години изследвания в тази сфера.

Във втората глава на дисертацията са представени основните аспекти на методологията, използвана за организирането и осъществяването на емпиричното изследване, което е необходимо за събирането на първични данни за предстоящите на следващ етап анализи. Това предоставя условия да се постигнат максимално качествени резултати, които да послужат при развитието на модели на поведение при вземане на решения в интернет на следващ етап.

Третата част от настоящия труд е съсредоточена върху приложната част на проучването. Тук са представени резултатите от емпиричното изследване и е проведен бивариантен анализ за установяване на съществуващи значими връзки между петте личностни детерминанти и отношението на потребителите към риска, от една страна и разглежданите функционалности на уеб магазините, от друга. В продължение са имплементирани три регресионни модела от областта на машинното обучение, с цел да бъде осъществена надеждна прогноза за вземане на решение на потребителите в електронната търговия въз основа на тяхната индивидуалност. Предложена и имплементирана е също така оптимизация за модела случайни гори, който се оказва особено подходящ за тази цел. С оглед на резултатите от цялостното изследване, в края на трета глава, са резюмирани модели на поведение при вземане на решение в онлайн търговията.

В последната, четвърта, глава е направено обобщение на резултатите от проведения научен труд и са предложени възможности за надграждане на изследването в бъдещ план.

Глава 1

Индивидуалността като фактор при вземане на решение в интернет

1.1 Теории и проучвания на личностната психология

Според науката личността е уникална система, чиито черти, нагласи и склонности са организирани по специфичен начин и имат различна пропорционална конфигурация за всеки индивид, вследствие на което всеки един потребител е уникален по отношение на своите предпочитания, очаквания и нужди [52], [12].

Според наличната в литературата информация, някои автори предполагат, че има две основни групи фактори, оказващи съществено влияние върху потребителите по време на процеса на вземане на решения - вътрешни (култура, социална класа, семейство) и външни (мотивация, отношение, намерение, възприятие) [32].

От друга страна, според други автори, личността е първата детерминанта, която трябва да бъде взета предвид при анализа на потребителското поведение, тъй като тя определя процеса на оценка и избора, както и начина, по който човек обработва информацията от околната среда [6], [24].

1.1.1 Теории и модели за анализ и измерване на личността

Личността е полето в психологията, което изучава мислите, чувствата, поведението, целите и интересите на индивидите. Следователно тя обхваща много широк спектър от важни психологически характеристики. Ето защо съществуват различни теоретични модели за измерване на тези детерминанти. Например според някои хората имат ясни, добре дефинирани цели и активно се стремят

да ги постигнат [45], докато други предполагат, че ние нямаме представа за своите чувства и мотиви, така че поведението ни се влияе от процеси, които работят извън нашето съзнание [44].

Най-съвременният подход за оценка на индивидуалността е Теорията за чертите на личността, чрез която се измерват специфични универсални детерминанти, които формират индивидуалността. Изследователите разглеждат нашето поведение като конструкция от относително стабилни във времето характеристики, които предразполагат човек да действа по определен начин, независимо от ситуацията [24].

В перспективата на личностната теория има няколко основни фигури, които я развиват до нивото, в което ние днес я разбираме и прилагаме. Сред тях са Гордън Олпорт и Хенри Одбърт - Gordon Allport & Henry Odbert (англ.), в основата, на чийто подход стои лексикална хипотеза, според която всички важни характеристики на личността са отразени в езика, който използваме за описване на други хора [3].

Теорията за личността на Айзенк – Eysenk (англ.) [20] се счита за истинска парадигма в психологията, твърдейки, че във всеки от нас има три основни черти - психотизъм, екстровертност и невротизъм, чиято сила варира при отделните индивиди. Но по-късно Реймънд Кател - Raymond Cattell (англ.) [46], заявява, че съществуват много по-голям брой специфики, за да се получи пълна представа за нечия личност. Така той идентифицира 16 личностни фактора, които са общи за всички хора и създава метод за многоизмерен анализ на личностна (16PF), включващ 192 елемента.

По-късно психолози изследват списъка на Кател и установяват, че той може допълнително да бъде намален. Днес много изследователи, сред които Голдберг - Goldberg (англ.) [25], [24] и Маккрей и Коста - McCrae & Costa (англ.) [41] вярват, че основните личностни характеристики се свеждат до пет [48].

1.1.1.1 Големите пет - петфакторен модел на личността

Може би най-прилаганият стандарт в съвременната психология е моделът Големите пет (Big Five), измерващ личността, основавайки се на пет основни характеристики - екстровертност, сговорчивост, добросъвестност, отвореност и емоционална стабилност [48], [41], [12], [16].

Първоначално Маккрей и Коста – McCrae & Costa (англ.) [42], [41], [16] представят триизмерен модел на личността на основата на характеристиките невротизъм, екстровертност и откритост, т. нар. NEO рамка и по-късно нейната преработена версия (NEO-PI-R), която има 240 елемента. По-късно, през 90-те години, те представят петфакторен модел, NEO-Five-Factor-Inventory (NEO-FFI), който се състои от 60 елемента, описващ петте централни специфики на

личността, всяка от които идва със списък от допълнителни аспекти, даващи обяснение на това как даден индивид може да се справя в определени ситуации:

- Невротизъм, Екстроверсия, Отвореност към нови изживявания, Сговорчивост и Добросъвестност.

Всеки един от петте елемента обобщава в себе си широк набор от аспекти, носещи ни информация за много допълнителни измерения, свързани с личността и според изследователите в тази област всяка от големите пет личностни черти може да бъде съпоставена към някой специфичен поведенчески модел на вземане на решения [51].

1.1.1.2 Други модели за анализ на личността

Съществуват и други подобни подходи за идентифициране на личностните особености на потребителя, които биха могли дори да улавят дори и други допълнителни аспекти на нашето поведение, като едни най-приложимите в литературата са изброени по-долу.

Моделът HEXACO - запазва оригиналните черти от Големите пет, но съдържа една допълнителна черта - честност/ смирение, която авторите описват като степента, до която човек поставя чуждите интереси над своите [4].

Моделът RIASEC - създаден е от американския психолог Джон Холанд - John Holland (англ.), като рамката се основава на шест основни личностни черти: реалистичен, разследващ, артистичен, социален, предприемчив и конвенционален и за разлика от останалите, описва силната връзка между средата и личността [28].

Тестът TIPI (Ten-Item Personality Inventory) - поради необходимостта от по-кратка версия на Големите пет, Гослинг и други - Gosling et al. (англ.) разработват Ten-Item Personality Inventory (TIPI тест), състоящ се само от десет елемента, като два дескриптора дават информация за всеки елемент [26]. Оценката се дават по 7-бална скала, която варира от 1 (напълно несъгласен) до 7 (напълно съгласен). В този си вид моделът е създаден, за да бъде завършен за около една минута.

Моделът ESV

(Eigenschaften - Situationen - Verhaltensweisen) (нем.) - моделът "Черти на характера - Ситуации - Поведения" е подход за самооценка на индивидуалните особености, който също така отчита предпочитанията на личността в

определени ситуации и нейното специфично поведение [13]. Чрез този подход в процеса на самооценка, в допълнение към определени качества, се търсят не само предпочитания към определени ситуации, но също така и към конкретни модели на поведение [54]. Според Щангл - Stangl (нем.) [55] наличието на отношение между поведението и личността доказва, че личните предпочитания към ресурсите, независимо дали са икономически или психологически, зависят от специфичните характеристики на човека.

1.2 Влияние на индивидуалността при вземането на решение за онлайн покупка

Възможността да бъдат направени научни заключения по отношение на влиянието на личностните особености върху поведението на потребителя в интернет, се крие във факта, че хората с приблизително подобни относителни черти се държат по подобен начин както в реалния, така и във виртуалния свят. Те имат подобни навици, предпочитания и поведение независимо дали в социалните медии или в процеса на онлайн пазаруване [43], [31]. Боравейки с интерфейса на избран онлайн магазин, потребителят въвежда свой собствен и уникален за себе си културен, психологически и физиологичен контекст, което може да бъде използвано като инструмент за подобряване на ползваемостта [33].

Един инструмент, който позволява надеждно да бъдат изследвани взаимовръзките между личността и поведението на потребителите в интернет са моделите за машинно обучение. Те могат да обработват множество променливи едновременно, като всеки елемент може да бъде използван като отделен предиктор [29]. В практиката се наблюдават два основни подхода, при които методите за машинно обучение се прилагат в изследването и оценката на личността. При единия, анализирайки големи набори от данни, прилагането на методите за машинно обучение помага за анализ на поведенчески показатели на потребителите в интернет (например в социалните медии), които се използват за предсказване на личностните черти.

От друга страна, имайки предвид, че личността е психологическа конструкция, целяща да обясни голямото разнообразие в човешко поведение от гледна точка на няколко стабилни и измерими индивидуални характеристики [59], прилагането на машинно самообучение може да позволи по-нататъшно проникване в прозренията на личностната психология по надежден път. Например индивидуалните специфики на отделния човек и оценките на характеристиките от психометричните модели могат да бъдат използвани като предиктори на поведението [35], с цел да бъдат генерирани модели на поведение при вземането на решения на потребителите в интернет.

Обобщение

Глава първа разглежда личността като доминиращ фактор в процеса на вземане на решение в интернет, както и по-конкретно в електронната търговия, като за тази цел са разгледани различни проучвания и теории на личностната психология. В тази връзка са изучени отделните фактори, които оказват влияние в процеса на вземане на решение, както и различни модели за измерване на личността. Специално внимание е обърнато на най-съвременния подход за оценка на човека, а именно на Теорията за чертите на личността, чрез която се измерват специфични универсални детерминанти, които формират личността. Оказва се, че моделът Големите пет е един от най-често прилаганите инструменти за измерване на личността на човека, но поради обема си, той невинаги е удобен и приложим в контекста на изследване на персоналността в интернет [10]. Поради този факт са изучени и други модели от тази област, които биха могли да отговорят на целта на дисертационния труд. Един подходящ за целта инструмент е ТІРІ тестът, който оценява измеренията на личността от Големите пет, но се състои само от 10 елемента, което позволява той да бъде завършен за около една минута.

Разгледани са конкретни научни изследвания, които изучават връзките между индивидуалността на потребителите и техните предпочитания към отделни функционалности на електронните магазини, както и влиянието на личностните детерминанти на човека при вземане на решение в интернет. След анализ на литературата в тази област са обобщени някои основни елементи и функционалности, които са характерни и необходима част от съдържанието на всеки съвременен онлайн магазин и които на по-късен етап от научното изследване ще бъдат използвани за анализ на взаимовръзката им с индивидуалността на участниците в емпиричното изследване.

Изследвано е също така и приложението на методите от машинното обучение за прогноза на личността, вследствие на поведението на потребителите в интернет, както и на предпочитанията им, позовавайки се на тяхната индивидуалност.

Изучаването на научната литература в тези области прави възможно провеждането на изследването на следващ етап в рамките на тази дисертация.

Глава 2

Методология на емпиричното проучване

Изборът на подходяща методология за провеждане на проучване, основана на систематизирани и структурирани правила, е ключова стъпка на всеки изследователски проект [53]. В тази връзка Котари - Kothari (англ.) [34] твърди, че позоваването на регулативните принципи за решаване на даден изследователски проблем е от основна важност, така че да могат да бъдат направени валидни и надеждни заключения в края на проучването.

2.1 Философия на проучването

За да бъде постигната целта на настоящото проучване, позитивизма бива възприет като философска парадигма с „обективна“ перспектива, тъй като тя се основава на твърдението, че възприятията и позициите на хората са верни или неверни, правилни или грешни и поради убеждението, че знанието е трудно, реално и постижимо [53]. Действително този подход е избран, тъй като той гласи, че реалността е независима и може да се измерва и предсказва безпристрастно чрез научно изследване въз основа на добре дефинирани и логически структурирани данни, които не са засегнати от личните авторски разбирания и възприятия [27].

2.2 Изследователски подход

В настояща дисертация се използва смесен подход за провеждане на изследването, което позволява използването на силните страни както на индуктивното, така и на дедуктивното разсъждение, за да се изучат различните гледни точки

на едно и също явление. Докато количественият подход позволява на изследователя да тества нови идеи, качественият, от своя страна, предоставя възможност за създаване на нови идеи, което на практика осигурява и по-добри резултати [17].

Така в началото на изследването е приложен качествен подход за интервюиране, което има за цел събирането на информация относно предпочитанията и желанията на участниците. Въз основа на техните отговори в последствие се извършва количествен анализ за проследяване на получените резултатите, основан на събраните статистически данни.

2.3 Стратегия на проведеното проучване

Въпреки че провеждането на емпиричното изследване е част от изследователската стратегия в рамките на тази дисертация, то няма за цел да анализира предпочитанията на отделния потребител относно набора от основни функционалности на онлайн магазините. Провеждането на настоящето проучване има за цел да послужи като средство за установяване на зависимости между типа индивидуалност и предпочитанията на потребителя в процеса на вземане на решение в онлайн пазаруването.

В проучването има въпроси с количествен, както и с качествен характер, събиращи информация за отношението на участниците към онлайн пазаруването, както и към различните характеристики на онлайн магазините. Предвидени са още въпроси за оценка на тяхната личностна характеристика, които са основани на валидирана в науката рамка (TIPI тест), а също и такива, които оценяват склонност на участниците към поемане на риск.

2.3.1 Събиране на статистически данни

Настоящият проект е разработен в съответствие с проучванията, използващи данни, събрани както от първични, така и от вторични източници.

Вземайки предвид целите на проекта, анкетата е избрана като подходящ инструмент за първично събиране на данни, тъй като тя изследва явления и контекст на разглеждания проблем и в резултат на това може да обобщи някои отличителни възгледи. Освен това тя е лесно разбираема от хората, позволява събиране на голямо количество данни по изключително икономичен начин и дава възможност чрез събиране на количествени данни да се правят качествени заключения, което я прави изключително подходяща в случаите на смесен изследователски подход.

2.3.1.1 Събиране на вторични данни

За да бъде събрана необходимата справочна информация преди провеждането на първичното изследване, е направен преглед на съответната литература, която разглежда подобни изследователски въпроси, свързани с поведението на потребителя в интернет и по-конкретно в процеса на онлайн пазаруване. Вторичните данни, използвани за целта на проекта, са официални и са налични в различни публикации на някои организации и бизнес асоциации, книги и научни списания, доклади и публикации на различни организации, публични регистри и статистика.

2.3.1.2 Анкета – събиране на първични данни

Използван е самоуправляващ се въпросник като основен изследователски инструмент, чиято електронна версия е създадена с уеб-базирано приложение за проучване (Google Forms) и се разпространява до подходящи участници чрез комуникация по електронната поща в комбинация с личен контакт в социалните мрежи. Анкетата се състои от 4 раздела и е реализирана на 3 езика – български, английски и немски, за да може обхватът на изследването да бъде по-обширен и валиден в интернационален аспект.

- **Първи раздел** – функционалности на електронните магазини/ предпочитания на потребителите;

Първият раздел от проучването се състои от 19 въпроса. Въпросите имат за цел да анализират отношението и мнението на респондентите относно техните предпочитания към съответния онлайн магазин при пазаруване в интернет (Таблица 2.1). Избраният набор от функционалности е характерен и приложим за повечето съвременни електронни магазини. Участниците имат възможността да отговарят на представените им въпроси, основавайки се на скала от 1 до 5 (никога, рядко, понякога, често, винаги).

- **Втори раздел** – десет характеристики, определящи личността на отделния индивид;

За измерване на личността е приложен валидираният в науката ТІРІ тест (Ten Item Personality Inventory), разработен от Гослинг и други - Gosling et al. (англ.) [26], който достига адекватни нива за конвергентна валидност относно Големите пет. ТІРІ тестът се състои само от 10 елемента, като два дескриптора дават информация за всеки елемент. Оценката се дава по 7-степенна скала, която варира от 1 (напълно несъгласен) до 7 (напълно съгласен). В този си вид моделът е създаден, за да бъде завършен за около една минута, което го прави напълно подходящ за прилагане в процеса на персонализация в интернет.

1.	Описанието на продукта ми дава нужната информация, за да мога да взема решение.
2.	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.
3.	Чета експертните оценки, те помагат в процеса на вземане на решение.
4.	Чета коментарите на други потребители за съответните артикули.
5.	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.
6.	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.
7.	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.
8.	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.
9.	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.
10.	Предпочитам да купувам допълнителни продукти като застраховка и/ или удължена гаранция в пакет с избрания продукт.
11.	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.
12.	За да избира между различните продукти, обикновено сравнявам детайлите им.
13.	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.
14.	Проверявам различните опции за доставка.
15.	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.
16.	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.
17.	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.
18.	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.
19.	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.

Таблица 2.1: Функционалности на електронните магазини

- **Трети раздел** – избягване на риска;

Донт и Джилленд - Donthu & Gilliland (англ.) [19] разработват т. нар. скала за избягване на риска, която измерва желанието на човек да избягва или поема риска, която се прилага и в настоящото проучване като кратък и практичен вариант за оценка на нагласите у потребителите. Тя се състои само от 3 елемента, които се оценяват от 1 до 5 (абсолютно несъгласен – несъгласен – не съм сигурен, колебае се – съгласен – абсолютно съгласен). Според нея хората, които имат висока оценка, тоест не са склонни към приемане на известен риск, предпочитат да бъдат много сигурни в това, което планират да купят. Докато потребителите с по-ниска оценка могат да толерират известен риск и несигурност в своите действия. Инструментът е използван за измерване на индивидуалната склонност към риск, която се свързва с индивидуалните различия на субектите по отношение на реакциите им спрямо риска [1].

- **Четвърти раздел** – демографско проучване;

В последната си секция запитването предвижда допълнително анализ на 5 демографски променливи и 1 въпрос относно честотата на онлайн пазаруването (никога, рядко, понякога, често, много често). Социално-демографските фактори, свързани с възраст, пол, образование, гражданство и място на пребиваване, оказват съществен ефект върху намерението и решението за покупка, както и върху това колко често потребителят поръчва артикули и услуги онлайн [15]. Особено внимание е обърнато на местоживеенето през последните минимум 5 години, поради факта, че хората, обитаващи определен регион, обикновено започват да споделят подобни ценности, нагласи и предпочитания [18].

2.3.1.3 Техника при подбор на статистическата извадка

Целта на изследването изисква да се достигне среднестатистическия интернет потребител и да се направят някои общи изводи [53] относно неговото предпочитание и поведение. Извадката е структурирана на случаен принцип, така че всички елементи на общността имат равни шансове да попаднат в нея. Всички участници са помолени да подкрепят изследването като предадат нататък въпросника към свои контакти, с които авторът няма реална връзка. По такъв начин той не оказва влияние, пряко или косвено, върху получените резултати.

2.3.1.4 Рамка на статистическата извадка

Въпросникът е изпратен на 250 души, представители на анкетираната популация. 226 от всички (90.4%) са попълнили точно и изцяло въпросника онлайн.

Въз основа на литературата, достигнатият обем на статистическата извадка отговаря на нуждите на проучването, за да представи то в края си значителни резултати относно изследвания проблем [53].

Всички легитимни участници в анкетата са над 18 години. Извадката обхваща представители на различните възрастови групи, от най-младите пълнолетни интернет потребители до хората в активна и зряла възраст до 75 години. Всички участници са с различен пол, социален статус, образование, разнообразно гражданство и местоживеене в последните минимум 5 години. Всички те съобщават за различна интензивност и честота на реализираните от тях онлайн покупки, което е условие за получаване на безпристрастни и валидни резултати. Изброените факти са предпоставка за постигане на обективен поглед по темата на изследването.

2.3.2 Ограничения при провеждането на проучването

Всяко проучване има ограничения, свързани обикновено с неговата емпирична страна, основани на избраната рамка на подбор на участници и подход.

Едно от ограниченията на настоящото проучване е свързано с недостатъчните вторични данни и статистика относно потребителските очаквания и изисквания към различните функционалности на онлайн магазините. Въпреки това е направен широк набор от функционалности, които са характерни за повечето съвременни електронни магазини.

Друго ограничение е зададената времева рамка за провеждане на първични изследвания, тъй като темите, свързани с индивидуалността и потребителското поведение, изискват да се изследват в по-дълъг период от време и евентуално проверят чрез други методи и инструменти.

Други потенциални ограничения на проекта биха били свързани с анкетата - ако респондентите не разбират зададените им въпроси. Но в случая може да се смята, че този риск е преодолян, тъй като всички въпроси са формулирани ясно, кратко и разбираемо. Освен това въпросникът е преведен на три езика и по такъв начин всички анкетирани имат възможност детайлно да вникнат в смисъла на зададените въпроси и възможностите за отговор.

Обобщение

Във втора глава от настоящата дисертация е представена избраната методология на емпиричното проучване, което е ключова стъпка на всеки изследователски проект, където човекът е обект на проучване. Изложени са избраните изследователски подход и стратегия на проведеното проучване, както и струк-

турата на анкетата. ТІРІ тестът е кратък и изборът му за оценка на личността се оказва особено подходящ за тази цел, като това открива нови възможности за бъдещи изследвания в тази област. Изготвеният набор от 19 функционалности е достатъчно голям, за да може да изследва обстойно преференциите на участниците в изследването и е характерен за по-голямата част от съвременните електронни магазини. Въпросите в анкетата са формулирани кратко и ясно, така че да бъдат максимално разбираеми за всички участници.

Обърнато е внимание също така на валидността и надеждността на данните, както и на етичната страна на проведеното изследване.

Въпреки съществуването на някои ограничения относно избрания изследователски подход, може да се приеме, че проучването е подходящо за целта на дисертационния труд.

Глава 3

Изследване. Прогнозиране на потребителското поведение в интернет

3.1 Анализ на резултатите от емпиричното проучване

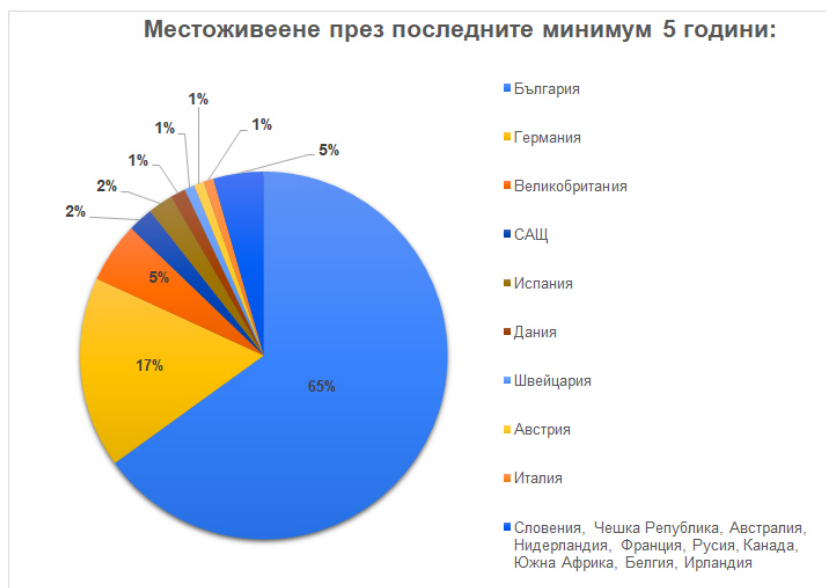
Емпиричното проучване е проведено онлайн, ползвайки приложението Google Forms, във времевата рамка от 30 август до 15 септември 2020 година, като извадката се състои от 226 участника.

3.1.1 Демографски преглед на извадката

От 226 души, попълнили изцяло въпросника, 43% са мъже, а 56% от всички са жени (Таблица 3.1). Възрастта оказва влияние върху онлайн потреблението [21], но в това изследване всички участници са в активна възраст между 18 и 75 години.

Макар повече от половината участници в проучването да са с български произход и да живеят към момента на изследването в страната (65%), останалата част от извадката (35%) обхваща хора, които са чужденци или българи, които вече повече от 5 години живеят в чужбина (Фигура 3.1).

Промяната на местоживеенето води до голяма степен и до промени в предпочитанията на потребителите в зависимост от специфичните характеристики на мястото на пребиваване. Разликите в начина на живот на хората вследствие на продължителното пребиваване в друга държава и съответно сред други културни особености и ценности са причинени най-вече от новата заобикаляща ги



Фиг. 3.1: Местоживеене през последните минимум 5 години

среда, както и от различната социална структура на обществото [18].

Въпреки че участниците в проучването са подбрани на случаен принцип, от получените данни става ясно, че всеки един от тях преди това е реализирал поне няколко онлайн покупки и е достатъчно запознат с процеса на пазаруване в интернет. Никой не е заявил, че няма никакъв опит с онлайн пазаруването (Фигура 3.2).



Фиг. 3.2: Опит в онлайн пазаруването

	Критерии	Брой	% от извадката
Демографска извадка	226		100%
Възраст	от 18 до 30	61	27%
	от 31 до 45	136	60%
	от 46 до 60	24	11%
	от 61 до 75	5	2%
Пол	Мъж	98	43%
	Жена	127	56%
	Друг	1	<1%
Образование	Средно	31	14%
	Висше	193	85%
	Не е посочено	2	1%

Таблица 3.1: Демографски данни

3.1.2 Проследяване на взаимовръзките между индивидуалността и предпочитанията на потребителите

След установяването на личностния тип на участниците в запитването и тяхното отношение към риска, както и след събирането на информация относно техните предпочитания, свързани с основните функционалности на онлайн магазините, е направен бивариантен корелационен анализ. За да се конструират уравненията, променливите, свързани с личностните черти и склонността към поемане на риск, са възприети като независими променливи, докато разглежданите функционалности на онлайн магазините се считат за зависими. Целта е да бъде установено съществуването на значими връзки между тези 6 независими променливи и всяка една от 19-те зависими променливи. Наличието на значими корелации предполага, че могат да бъдат формирани математически уравнения за оценка и прогноза на предпочитанията на потребителите (Таблица 3.2). Вземат се предвид само променливите със значими връзки, които са идентифицирани според техните коефициенти на корелация на нива $p < 0.05$, използвайки програмата PSPP, която е с всеобщ публичен лиценз на ГНУ.

Коефициентите на корелация в Таблица 3.2 показват, че са открити 21 значими връзки между анализираниите зависими и независими променливи.

Зависими променливи (всяка една от 19-те разглеждани основни функционалности на онлайн магазините)	Независими променливи (всяка една от 5-те личностни черти, както и склонността към поемане на риск)					
	Екстремност (проявяват стремеж към внимателни, интензивни социални контакти и динамика)	Сторичност (проявяват любовност, съобразителност, благосклонен, десно съгласяващ се)	Добросъвестност (проявяват внимание, старание и детайлност)	Емоционално стабилност (склонен да запазва баланс и самообладание)	Отвореност към нов опит (проявяват откритост и прямота)	Склонност към поемане на риск (възприемане и осъзнаване на риска)
1. Описанието на продукта ми дава нужната информация, за да мога да взема решение.		0.122 ^a				
2. Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.			0.120 ^a			
3. Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		0.162 ^a				
4. Чета коментарите на други потребители за съответните артикули.		0.117 ^a				
5. Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.			0.194 ^a			
6. Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.			0.176 ^a			0.125 ^a
7. Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.						0.110 ^a
8. Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.			0.154 ^a			0.115 ^a
9. Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	0.205 ^a					
10. Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.						
11. За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.						
12. За да избира между различните продукти, обикновено сравнявам детайлите им.			0.164 ^a			
13. Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.						
14. Проверявам различните опции за доставка.			0.133 ^a	0.120 ^a		
15. Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.						
16. Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	0.110 ^a				0.121 ^a	
17. Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.						0.180 ^a
18. Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.				- 0.114 ^a		
19. Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.			0.226 ^a	0.127 ^a		0.157 ^a

^a Наличието на значима връзка на ниво 0.05

Таблица 3.2: Проверка за значима корелация между 6 независими и 19 зависими променливи

3.2 Приложение на методи от машинното обучение за прогноза на потребителското поведение в интернет

Психологическата концепция за личността отчита индивидуалните различия в трайните емоционални, междуличностни, преживяващи и мотивационни стилове на хората, като личността на човека остава стабилна и налична за по-дълъг период от време. Следователно наличието на информация за личностния тип на индивида предоставя сведения за това как той възприема представената информация, какви са неговите предпочитания и за това как той би реагирал в различните ситуации [59]. В същото време методите на машинното обучение предоставят възможност това поведение на потребителите, вследствие на тяхната индивидуалност, да бъде надеждно прогнозирано [49].

С цел да бъдат прогнозираны предпочитанията на потребителя към различните функционалности на електронните магазините в зависимост от неговата индивидуалност, в рамките на тази дисертация са реализирани три регресионни модела - линейна регресия, дърво на решенията и случайни гори. Алгоритмите вземат 5-те личностни детерминанти на потребителя и склонността му към поемане на риск като вход и дават стойност, показваща колко вероятно е съответната функционалност да бъде предпочетена от дадения потребител. Значението на алгоритмите се крие в способността им да предвиждат как нови потребители, чийто личност и индивидуална склонност за поемане на риск е известна, вероятно ще се държат и вземат решение в един онлайн магазин. За да се построят алгоритмите, личностните детерминанти действат като независими променливи, докато предпочитанията на потребителите към отделните функционалности се считат за зависими.

Програмирането е извършено на Python, версия 3.8 (64-bit), а точността на всеки регресионен модел е проверена чрез приложими метрики за проверка на точността [37], а именно посредством средната абсолютна процентна грешка (Mean Absolute Percentage Error (MAPE)), общата абсолютна грешка (Mean Absolute Error (MAE) и средната квадратична грешка (Root Mean Squared Error (RMSE)).

3.2.1 Прогноза с линейна регресия

В рамките на дисертацията реализирането на линейната регресия стартира с импортиране на нужните библиотеки, визуализиране на заредените данни от цялата извадка с цел проверка, след което множеството се разделя на две части - 70% тренировъчно и 30% тестово подмножество. Разделянето на наблюденията от даденото множество в двете подмножества е на случаен принцип, като

тренировъчното подмножество служи, за да се построи прогнозният модел, т.е. да се намерят коефициентите на регресията, а тестовото подмножество, за да се тества построената вече линейна регресия [38]. След като е направена прогнозата следват визуализацията на резултатите, както и пресмятане на стойностите на метрики за оценка на алгоритъма.

По този начин посредством линейната регресия са направени прогнозни изчисления за всички открити значими връзки между личностните детерминанти на потребителите и предпоченциите им относно отделните функционалности на електронните магазини. Както беше споменато, точността на прогнозите е проверена чрез метриците обща абсолютна грешка (MAE) (Таблица 3.3), корен от средната квадратична грешка (RMSE) (Таблица 3.4) и средната абсолютна процентна грешка (MAPE) (Таблица 3.4). Всички те са отрицателно ориентирани резултати, т.е. по-ниските стойности са по-добри. Средната стойност на MAE при линейната регресия е 0.77, на RMSE 0.96, а стойността на MAPE е 27.55, т.е. средната стойност за точността (*accuracy*) по тази метрика е 72.45%.

	Експертизност	Споможливост	Добросъвестност	Емоционална стабилност	Отвореност към нов опит	Склонност към поемане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	0.58				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		0.68			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	0.81				
4	Чета коментарите на други потребители за съответните артикули.	0.62				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		0.84			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		0.93			0.90
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					0.90
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		0.44			0.43
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	0.76				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		0.76			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		0.74	0.73		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	0.92			0.91	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					1.05
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			1.01		
19	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.		0.72	0.73		0.72
Средна стойност на общата абсолютна грешка (MAE):		0.77				

Таблица 3.3: Линейна регресия - обща абсолютна грешка (MAE)

	Екстремност	Сторитивност	Добростъжност	Емоционална стабилност	Отвореност към нов опит	Склонност към поемане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	0.73				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		0.87			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	0.98				
4	Чета коментарите на други потребители за съответните артикули.	0.77				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		1.10			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		1.15			1.17
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					1.17
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		0.50			0.48
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	0.97				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		0.99			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		0.93	0.93		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	1.09			1.08	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					1.20
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			1.16		
19	Предпочитам плащане с алтернативни и по-сигурни методи на плащане (напр. PayPal и др.) и проверявам за наличието на такива.		0.91	0.94		0.94
Средна стойност на корена от средната квадратична грешка (RMSE):		0.96				

Таблица 3.4: Линейна регресия - корен от средната квадратична грешка (RMSE)

	Експертност	Сговорливост	Добросъвестност	Емоционална стабилност	Отвореност към нов опит	Склонност към поемане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.		15.61			
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		20.29			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		26.34			
4	Чета коментарите на други потребители за съответните артикули.		18.35			
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		31.37			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		35.95			37.69
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					37.69
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		9.94			9.77
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	30.57				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		24.39			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		23.30	23.01		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	46.94			45.55	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					37.97
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			37.91		
19	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.		21.44	22.36		22.05
Средна стойност на средната абсолютна процентна грешка (MAPE):		27.55				

Таблица 3.5: Линеинна регресия - средна абсолютна процентна грешка (MAPE)

3.2.2 Прогноза с метода дърво на решенията

Реализирането на метода дърво на решенията е подобен както при линейната регресия, като множеството също така на случаен принцип се разделя на две части - 70% тренировъчно и 30% тестово подмножество. За изпълнение е използвана библиотеката *scikit-learn* [47].

В Таблица 3.6 са представени отделните стойности на общата абсолютна грешка (MAE) за дървото на решения, в Таблица 3.7 на средната стойност на корена от средната квадратична грешка (RMSE), а в Таблица 3.8 на средната абсолютна процентна грешка (MAPE) за всички открити значими връзки. Средната стойност на MAE при дървото на решения е 0.80, на RMSE 0.98, а на MAPE е 27.96, т.е. средната стойност за точността по тази метрика за всички значими връзки е 72.04%.

	Експертизност	Споморивност	Добросъвестност	Емоционална стабилност	Отвореност към нов опит	Склонност към поемане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	0.61				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		0.74			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	0.86				
4	Чета коментарите на други потребители за съответните артикули.	0.69				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		0.81			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		0.93			0.95
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					0.94
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		0.44			0.43
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	0.76				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		0.84			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		0.74	0.69		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	0.94			0.95	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					1.08
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			1.01		
19	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.		0.74	0.83		0.75
Средна стойност на общата абсолютна грешка (MAE):		0.80				

Таблица 3.6: Дърво на решенията - обща абсолютна грешка (MAE)

	Екстремност	Сговорливост	Добресъвестност	Емоционална стабилност	Отпорност към нов опит	Склонност към посегане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	0.77				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		0.92			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	1.04				
4	Чета коментарите на други потребители за съответните артикули.	0.84				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		1.07			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		1.16			1.16
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					1.20
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		0.53			0.48
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	0.94				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		1.09			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		0.92	0.87		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	1.14			1.11	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					1.24
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			1.19		
19	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.		0.92	1.05		0.95
Средна стойност на корена от средната квадратична грешка (RMSE):		0.98				

Таблица 3.7: Дърво на решенията - корен от средната квадратична грешка (RMSE)

	Екстремност	Сговорчивост	Добротелност	Емоционална стабилност	Отвореност към нов онлиг	Склонност към посегане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	16.16				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		21.46			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	27.14				
4	Чета коментарите на други потребители за съответните артикули.	19.69				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		29.29			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		35.79			36.34
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					38.24
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		10.11			9.71
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	29.51				
10	Предпочитам да купувам допълнителни продукти като застраховка и/ или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		26.32			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		22.60	21.16		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	47.72			47.56	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					38.97
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			38.74		
19	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.		22.36	25.32		22.98
Средна стойност на средната абсолютна процентна грешка (МАРЕ):		27.96				

Таблица 3.8: Дърво на решенията - средна абсолютна процентна грешка (МАРЕ)

3.2.3 Прогноза с метода случайни гори

Процесът на изпълнение на случайните гори в рамките на тази дисертация е подобен както при другите два метода. Стартира се с импортиране на нужните библиотеки, визуализиране на заредените данни от цялата извадка с цел проверка, след което множеството на случаен принцип се разделя на две части - 70% тренировъчно и 30% тестово подмножество. За реализация е използвана библиотеката *scikit-learn*, като зададения брой на дърветата в гората е 150 ($n_estimators = 150$), а по подразбиране това число е 100.

В Таблица 3.9 са презентирани стойностите на общата абсолютна грешка (MAE), в Таблица 3.10 стойностите на корена от средната квадратична грешка (RMSE), а в Таблица 3.11 са презентирани стойностите на средната абсолютна процентна грешка (MAPE). Средната стойност на MAE при случайните гори е 0.79, на RMSE 0.98, а на MAPE 27.92. т.е. средната стойност на точността за всички значими връзки по MAPE е 72.08%.

	Експертизност	Споморивност	Добросъвестност	Емоционална стабилност	Отвореност към нов опит	Склонност към поемане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	0.56				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		0.73			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	0.86				
4	Чета коментарите на други потребители за съответните артикули.	0.70				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		0.81			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		0.93			0.95
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					0.93
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		0.42			0.43
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	0.76				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		0.84			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		0.73	0.69		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	0.94			0.95	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					1.08
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			1.00		
19	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.		0.74	0.84		0.74
Средна стойност на общата абсолютна грешка (MAE):		0.79				

Таблица 3.9: Случайни гори - обща абсолютна грешка (MAE)

	Екстремност	Сторитивност	Добростъжност	Емоционална стабилност	Отвореност към нов опит	Склонност към поемане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	0.76				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		0.91			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	1.04				
4	Чета коментарите на други потребители за съответните артикули.	0.85				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		1.07			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		1.16			1.16
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					1.19
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		0.48			0.48
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	0.94				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		1.08			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		0.91	0.87		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	1.13			1.11	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					1.24
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			1.19		
19	Предпочитам плащане с алтернативни и по-сигурни методи на плащане (напр. PayPal и др.) и проверявам за наличието на такива.		0.92	1.05		0.95
Средна стойност на корена от средната квадратична грешка (RMSE):		0.98				

Таблица 3.10: Случайни гори - корен от средната квадратична грешка (RMSE)

	Експертизност	Сторартизност	Добросъвестност	Емоционална стабилност	Отвореност към нов опит	Склонност към поемане на риск
1	Описанието на продукта ми дава нужната информация, за да мога да взема решение.	16.03				
2	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.		21.28			
3	Вместо крайна оценка на продукта, предпочитам подробна такава за отделните критерии.	27.34				
4	Чета коментарите на други потребители за съответните артикули.	19.89				
5	Проверявам актуалната наличност на продукта, както и времето за доставка, преди да направя покупка.		29.23			
6	Предпочитам да мога да проследя статуса на поръчката в процеса на закупуване и доставка.		35.77			36.38
7	Проверявам калкулираната цена за доставка спрямо конкретно избраните от мен артикули.					38.14
8	Предпочитам да разгледам уголемени (детайлни) снимки на продуктите.		10.15			9.67
9	Предпочитам да купувам и други подходящи продукти и аксесоари заедно с избрания продукт.	29.37				
10	Предпочитам да купувам допълнителни продукти като застраховка и/или удължена гаранция в пакет с избрания продукт.					
11	За да избира най-подходящия за мен продукт, използвам подробен филтър за търсене и категоризация на продуктите.					
12	За да избира между различните продукти, обикновено сравнявам детайлите им.		26.22			
13	Използвам различни функции в количката като бързо добавяне на продукт (с един клик) и изчисляване на крайната цена.					
14	Проверявам различните опции за доставка.		22.54	21.20		
15	Използвам различни възможности за контакт и консултация, за да се уверя в определени характеристики на продукта и продължа процеса на закупуване.					
16	Пиша и коментирам отзиви за продуктите. Те помагат за изясняването на въпроси относно продуктите.	47.70			47.41	
17	Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация.					38.94
18	Предпочитам и проверявам за налично безплатно връщане на стоката. За мен това е от съществено значение.			38.71		
19	Предпочитам плащане с алтернативни и по-сигурни методи на заплащане (напр. PayPal и др.) и проверявам за наличието на такива.		22.21	25.35		22.80
Средна стойност на средната абсолютна процентна грешка (MAPE):		27.92				

Таблица 3.11: Случайни гори - средна абсолютна процентна грешка (MAPE)

3.2.4 Сравнение на резултатите от различните методи

От презентиранияте прогнозни резултати за предпочитанията на потребителите в онлайн пазаруването в зависимост от техните личностни детерминанти може да се заключи, че трите алгоритъма постигат много близки средни стойности, измерени от приложените метрики за оценка на точността на прогнозата. Прави впечатление, че стойностите на метриците за оценка при дървото на решения и случайните гори са много идентични при отделните прогнози, което може да се обясни с тяхната идентична природа, а именно с дървовидната структура на алгоритмите. В Таблица 3.12 в обобщен вид са представени средните стойности на метриците за оценка за трите модела.

Средната стойност на общата абсолютна грешка (MAE) е най-добра при линейната регресия (0.77), въпреки че разликата с другите два алгоритъма е незначителна предвид средните стойности на зависимите променливи. Средната стойност за корена от средната квадратична грешка RMSE, която показва стандартното отклонение на грешките при прогнозирането, т.е. стойността отразява отдалечеността на данните от регресионната линия, също е почти еднаква за трите модела. А получената стойност за точността по средната абсолютна процентна грешка (MAPE) ($Accuracy = 100\% - MAPE$) е най-висока при линейната регресия (72.45%), като разликата със стойностите на другите два модела отново е незначителна.

Метрики за оценка	Модел	Средна стойност
MAE	Линейна регресия	0.77
	Дърво на решенията	0.80
	Случайни гори	0.79
RMSE	Линейна регресия	0.96
	Дърво на решенията	0.98
	Случайни гори	0.98
Accuracy	Линейна регресия	72.45 %
	Дърво на решенията	72.04 %
	Случайни гори	72.08 %

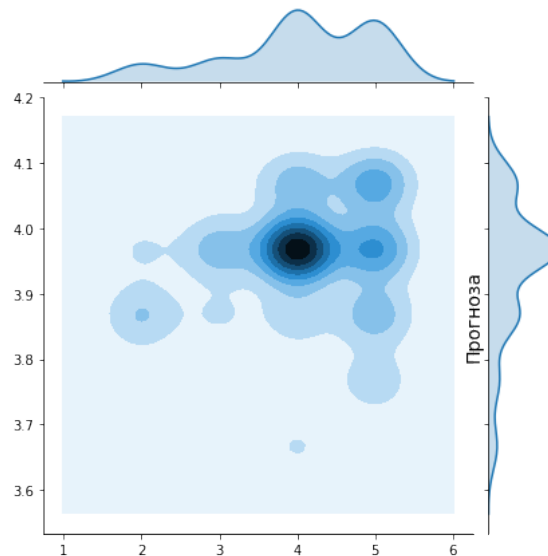
Таблица 3.12: Средни стойности на метриците за оценка

В рамките на тази дисертация трите модела постигат сравнително еднакви резултати според метриците за оценка, но съществуват редица техники за оптимизиране на стойностите на тези метрики, с които те могат да бъдат подобрени, като например обучение с по-голям обем от данни, оптимизация с кръстосано валидиране, с генетично програмиране и други [7].

Преди да бъде избран определен модел е важно той да бъде оценен с различни метрики за оценка, но изборът на правилния модел зависи и от целта,

която е поставена, от актуалния и очаквания обем на данните, както и от техния генезис. При извършване на сравнение, т.е. при избор на модел за машинно обучение, има смисъл, освен да се направи оценка на точността чрез метриците за оценка, да се провери чрез визуализация например дали определени измервания се различават съществено от други, къде и каква е плътността на разпределението в прогнозата, както и на актуалните стойности [60].

Като един пример, с цел добиване на по-добра представа за качеството на прогнозите, на фигурите 3.3, 3.4 и 3.5 са изобразени разпределението и плътността на разпределението на данните в прогнозите на трите алгоритъма за използване на алтернативни методи за плащане преди потребителят да реши да пазарува онлайн, в зависимост от това, до колко той е склонен да поема риск.

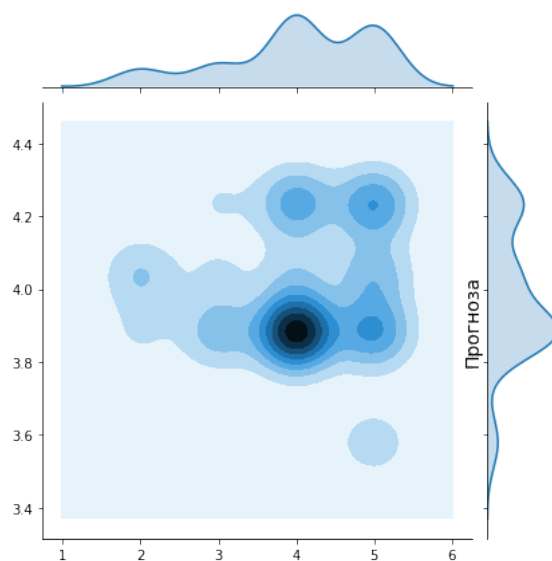


Фиг. 3.3: Линейна регресия - прогноза за проверка на алтернативни методи за плащане и склонност за поемане на риск, плътност на разпределението

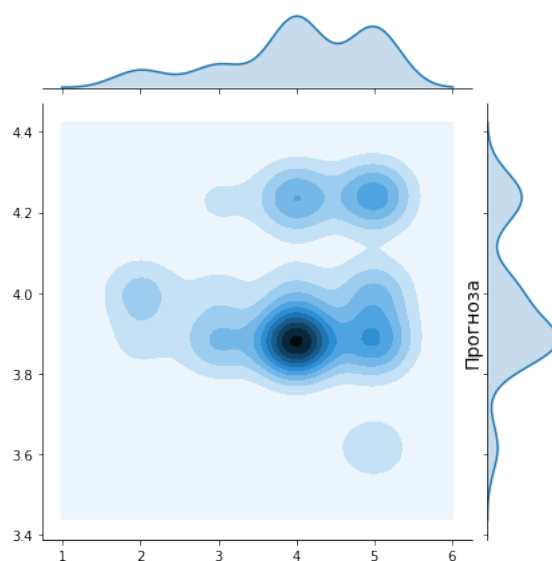
Друг пример от значимите взаимовръзки между личностните детерминанти и предпочитанията на потребителите в онлайн пазаруването, където алгоритмите постигат най-високи стойности на метриците за оценка, т.е. не се справят особено добре, е между екстровертността и ”16. Пиша и коментирам отзиви за продуктите, те помагат за изясняването на въпроси относно продуктите”. Линеината регресия постига точност на прогнозата според MAPE от 54.50%, а дървото на решения и случайните гори едва 52.30%.

На фигурите 3.6, 3.7 и 3.8 е представено разпределението на данните за трите модела.

Разпределението на актуалните стойности, разбира се, е еднакво и при трите



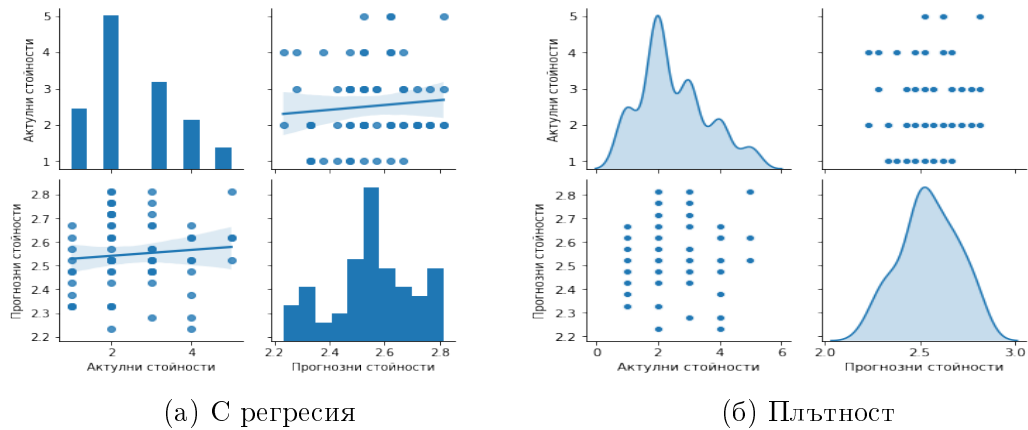
Фиг. 3.4: Дърво на решенията - прогноза за проверка на алтернативни методи за плащане и склонност за поемане на риск, плътност на разпределението



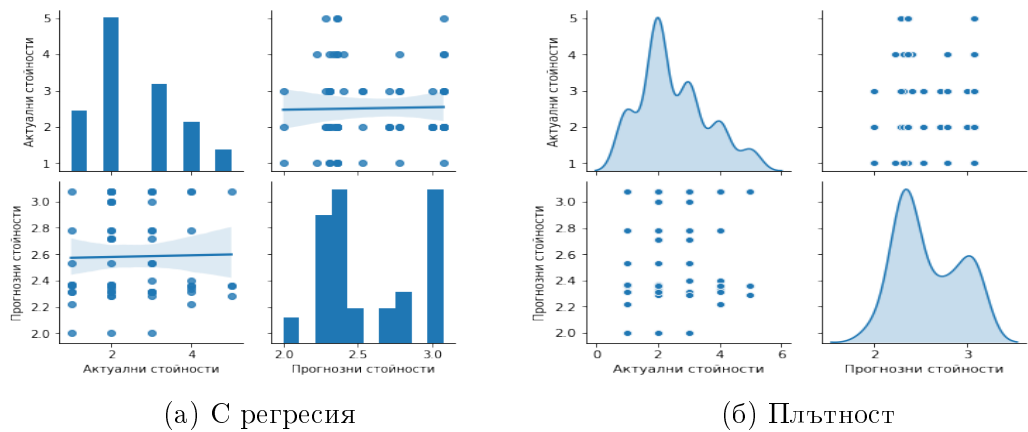
Фиг. 3.5: Случайни гори - прогноза за проверка на алтернативни методи за плащане и склонност за поемане на риск, плътност на разпределението

модела, но разпределението на прогнозните стойности при линейната регресия е по-равномерно около медиана, в сравнение с другите два модела.

Въпреки че при някои от откритите значими връзки отделните алгоритми постигат малко по-добра прогноза от другите, може да се заключи, че прогнозите им са еднакво точни според средните им стойности на метриците за оценка.



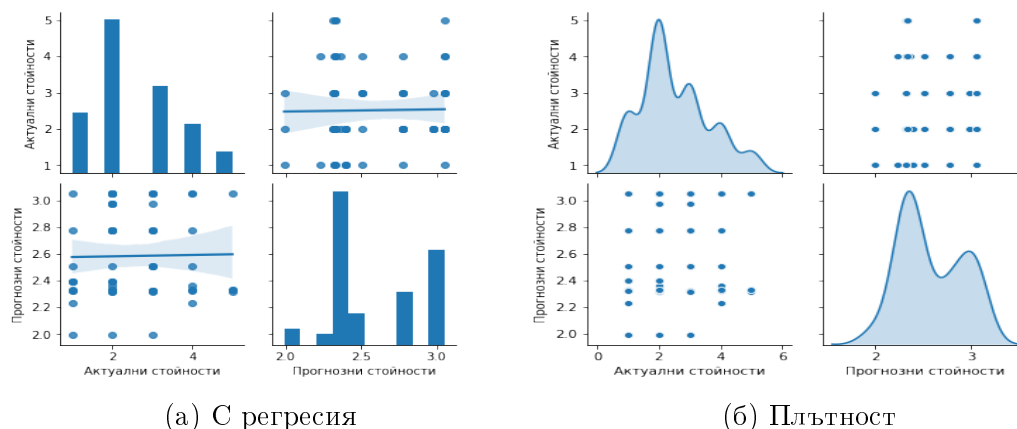
Фиг. 3.6: Линейна регресия - екстровеитност и коментиране на продуктите, разпределение на данните



Фиг. 3.7: Дърво на решенията - екстровеитност и коментиране на продуктите, разпределение на данните

И въпреки че според тези стойности прогнозите им не могат да бъдат определени като изключително точни, те биха могли да бъдат приемливо добри за целта [39]. Всъщност само при някои значими връзки прогнозните стойности не са толкова точни и се отличават значително от точността на прогнозите за другите значими връзки.

Човешката индивидуалност и произтичащите от това предпочитания са сложна взаимовръзка. На процеса на вземане на решение в интернет и по-конкретно в онлайн търговията често се гледа като на последователен процес, който включва необходимост от разпознаване, търсене на информация, проверка на алтернативи, покупка и евентуални активности след покупката, но актуални научни трудове показват, че този процес съдържа цикли и природата му не е напълно последователна в този ред (non-linear nature) [30]. И въпреки че и трите модела



Фиг. 3.8: Случайни гори - екстровеерност и коментирание на продуктите, разпределение на данните

в това изследване се оказват подходящи за тази цел, което се потвърждава и в разгледаната литература, и че стойностите на метриците за оценка за прогнозиране на всички три модела биха могли да бъдат подобрили чрез различни техники, в следващата глава е предложена оптимизация на модела случайни гори. Алгоритъмът е устойчив на прекомерно нагаждане (*overfitting*) и поради природата му при него няма нужда от трансформация на данните, той работи добре с цифрови и категорични характеристики на данните. Дори ако се въведат нови параметри в набора от данни, общият алгоритъм не е засегнат много, тъй като новите данни могат да повлияят на едно дърво, но е много трудно за него да повлияе на всички дървета [36], [11], което го прави много подходящ за изследване на индивидуалността в интернет, предвид динамиката на активността на потребителите в мрежата.

3.2.5 Оптимизация на метода случайни гори

Оптимизирането на избрания модел е една от важните стъпки при решаване на цялостната задача чрез средствата на машинното обучение, като отделните алгоритми имат различни параметри, които могат да се настроят така, че да се подобри работата на модела.

За целите в тази дисертация е предложена оптимизация на случайни гори с кръстосано валидиране, като се използва решетъчно търсене, прилагайки класа *GridSearchCV* на библиотеката *scikit-learn*, както и оптимизиране чрез ТРОТ (Tree-based Pipeline Optimization Tool), който използва генетично програмиране.

В предложената оптимизация *GridSearchCV* преминава през всяка една комбинация 10 пъти, тъй като на параметъра за кръстосана валидация е за-

дадена стойност 10 ($cv=10$). В този случай се получават общо 120 комбинации.

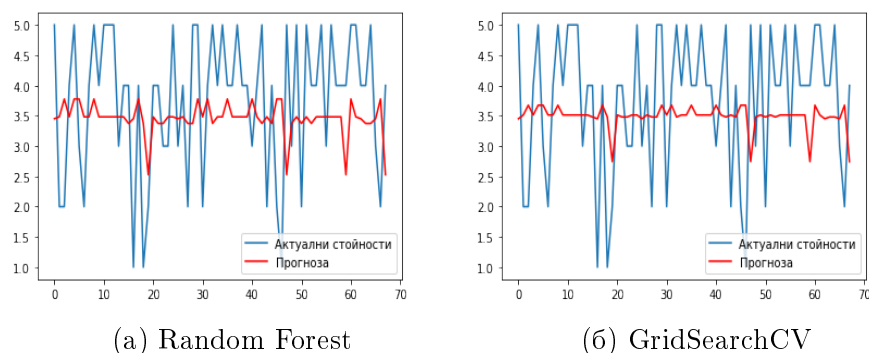
При 5 от всички 21 значими връзки между личностните детерминанти и предпочитанията на потребителите в онлайн пазаруването методът не води до подобрение в точността според MAPE с тези параметри. В останалите 16 значими връзки стойностите на точността според MAPE са различно подобрени, като най-високото подобрение е 2.58% при склонността за проверка за алтернативни и по-сигурни начини за заплащане в зависимост от емоционалната стабилност на потребителя. Подобрението на средната стойност на точността за всички 21 значими връзки според MAPE е 0.53% или от 72.08% на 72.46%. Според метриците MAE и RMSE също се наблюдава незначително подобрение, което при отделните взаимовръзки е различно.

ТРОТ е изграден върху библиотеката *scikit-learn* и с добре документиран отворен код, който е в активна разработка. Проектът е бил под ръководството на учени от Университета Пенсилвания (University of Pennsylvania) и има специални разработки за медицински изследвания. ТРОТ базира на алгоритъм за генетично търсене, за да намери най-добрите параметри и ансамбли на модела. С настройките по подразбиране (100 генерации и 100 популации) алгоритъмът ще трябва да оцени 10 000 конфигурации на конвейера, преди да завърши, т.е. поради природата си алгоритъмът е сравнително бавен [47].

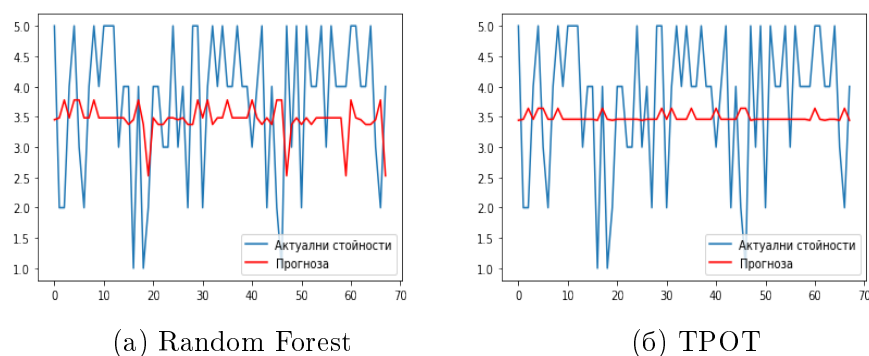
В предложената оптимизация с ТРОТ броят на генерациите е 10, а броя на индивидите за всяка генерация е 100, като така алгоритъмът трябва да оцени 1100 конфигурации ($population_size + (generations \times offspring_size)$). Така конфигуриран, алгоритъмът води до подобрение на стойностите на точността според MAPE при 19 от всичките 21 значими взаимовръзки. Подобрението на средната стойност на точността за всички 21 значими връзки според MAPE е 0.69% или от 72.08% на 72.58% точност.

Като един пример на Фигура 3.9 са изобразени актуалните стойности и прогнозата на случайни гори за зависимата променлива "17. Избягвам да оставям (запазвам) лични данни в онлайн магазините и предпочитам да купувам като гост, без регистрация" вследствие на избягването на риск в онлайн пазаруването, както и оптимизацията с *GridSearchCV*, а на Фигура 3.10 оптимизацията с ТРОТ. За тази взаимовръзка между независимата и зависимата променлива точността на случайни гори според MAPE 77.20%, оптимизацията с кръстосано валидиране постига точност от 77.71%, а оптимизацията с ТРОТ подобрява точността на 78.23%.

Във формата на кратко обобщение може да се каже, че и двата алгоритъма водят до леко подобрение на резултатите на случайни гори според метриците за оценка на точността на прогнозата, като резултатите постигнати с ТРОТ са незначително по-добри. В приложената конфигурация времето за изчисление и на двата алгоритъма е приемливо, въпреки че поради естеството си и двата



Фиг. 3.9: Склонност за запазване на лични данни вследствие на избягване на риска - актуални стойности и прогноза с метода случайни гори и оптимизация с GridSearchCV



Фиг. 3.10: Склонност за запазване на лични данни вследствие на избягване на риска - актуални стойности и прогноза с метода случайни гори и оптимизация с TROT

алгоритъма имат голяма консумация на изчислително време.

TROT е създаден за оптимизация на задачи от сложно естество [47], но все пак при по-малко комплексни задачи с по-малък обем от данни, алгоритъмът може да постигне добри резултати за сравнително малко време. *GridSearchCV* преминава през всяка една комбинация от предварително зададените параметри, което е причината той да се нуждае от по-голям времеви изчислителен ресурс. Едно алтернативно решение например би било приложена комбинация от решетъчно и рандомизирано търсене при по-голяма база от данни, като се направят няколко итерации с *RandomizedSearchCV*, след което върху получените резултати да се използва *GridSearchCV*. Недостатък в този случай би било, че времето за изпълнение също няма да бъде много по-кратко и че предишни итерации няма да бъдат взети под внимание при определяне на стойностите на параметрите.

3.3 Модели на поведение при вземане на решение в онлайн търговията в зависимост от индивидуалността на потребителя

Въз основа на проведеното проучване в литературата и получените резултати от настоящето изследване, може да бъде обобщено, че индивидуалността играе важна роля в процеса на предприемане на онлайн покупки, тъй като отделните потребители притежават различни специфични особености, които се отразяват върху тяхното поведение на подсъзнателно ниво. Резултатите от това проучване показват, че с помощта на методите за машинно обучение е възможно това поведение на потребителите да бъде надеждно прогнозирано. Така след оптимизация с ТРОТ, който базира на алгоритъм за генетично търсене, методът случайни гори постига средна точност на прогнозата от 73% за всички открити значими взаимовръзки. Така с оглед на тези резултати могат да бъдат установени модели на поведение при вземане на решение в онлайн търговията, които са резюмирани по-долу.

Според резултатите от проведеното изследване **по-екстровеъртните** личности биха реагирали положително при наличие на предложение за покупка на други допълнителни продукти и аксесоари, свързани с разглеждания от тях продукт, където чрез оптимизация случайни гори постигат 71% точност на прогнозата по средната абсолютна процентна грешка (МАРЕ). Тези личности също така активно участват както в писането, така и в четенето на коментари и това според тях играе значима роля при вземането на решение за покупка. По-екстровеъртните личности са по-активни, ентузиазирани, приказливи, енергични и доминиращи. При тях е установена тенденция към по-висока честота и интензивност на взаимоотношения и желание за изява и лидерство [58], докато интровертите са склонни да бъдат по-резервирани и често изискват период на уединение и тишина.

По-сговорчивите потребители предпочитат да четат коментари, оставени от други клиенти, преди да закупят желаниа от тях продукт. Чрез прилагане на метода случайни гори според настоящето изследване това предпочитание на по-сговорчивите потребители може да бъде добре прогнозирано (81% точност по МАРЕ). Тези потребители отделят също особено внимание на описанието на разглежданите продукти, на неговата детайлност и информативност (84% точност на прогнозата по МАРЕ), както и на експертните оценки относно разглежданите продукти (74% точност на прогнозата). По-сговорчивите личности имат нужда от повече информираност и споделен опит с другите, когато са в процес на вземане на решение в интернет [22].

По-добросъвестните предпочитат да имат възможност за избор между различни, но подобни продукти и да могат да сравнят техните детайли. Това

е така, защото добросъвестните хора са добре организирани, подредени и целенасочени [14]. При тези потребители процесът на вземане на решение бива улеснен и чрез предоставяне на детайлни снимки на продуктите (90% точност на прогнозата по МАРЕ), както и чрез налична подробна оценка по отделните критерии на артикулите (80% точност на прогнозата). Добросъвестните хора също се характеризират със стремежа си към постижения, те се стремят към завършване на поставените задачи до край [5], което обяснява защо за тях е от важно значение предоставянето на възможност за проверка на актуалната наличност на артикулите и времето за доставка (70% точност на прогнозата), както и различните опции за това (78% точност) преди да се решат да направят покупка онлайн. Тези потребители също предпочитат да могат ясно да проследят статуса на поръчката в процеса на закупуване и доставка, както и да имат възможност за заплащане с алтернативни и по-сигурни методи на заплащане, където оптимизиран алгоритъмът случайно гори постига 79% прогнозна точност по МАРЕ.

За **емоционално стабилните** потребители е важно предлагането на различни опции за доставка, както и наличието на алтернативни и по-сигурни методи за заплащане, където алгоритъмът случайно гори постига над 75% точност на прогнозата по МАРЕ. Полярната противоположност на емоционалната стабилност е невротизмът, като невротичните хора се характеризират с трудност да контролират своите емоции и да овладяват състоянието си на стрес [56]. Това се потвърждава и от резултатите на изследването, според които за по-невротичните потребители е изключително важно предоставянето на възможност за безплатно връщане на стоката.

Според резултатите на изследването хората с високи стойности на **отвореност към нов опит** предпочитат да коментират и задават въпроси относно продуктите, защото за тях това е начин да се уверят в качеството на дадения продукт. Методът случайно гори тук постига една от най-ниските прогнозни стойности по МАРЕ (53%), като според Люис - Lewis (англ.) тази стойност все пак може да бъде приета като една приемлива за целта прогноза [39].

В проведеното проучване се потвърждава и силната зависимост между **склонността към поемане на риск** на потребителите като личностна детерминанта и споделянето на лични данни в интернет, както и предоставянето на алтернативни и по-сигурни методи за заплащане (62% и 78% точност на прогнозата със случайно гори по МАРЕ). Освен това възприемането на риска е положително свързан с предпочитанията на потребителите да имат възможността да проследят статуса на закупените от тях продукти, както и да имат предоставена своевременно калкулираната цена за доставка. За тези потребители също така се оказва особено важно и предоставянето на детайлни снимки на разглежданите продукти, където алгоритъмът случайно гори постига 90% точност на прогнозата по МАРЕ.

Обобщение

В глава трета са представени резултатите от изследването, като в началото са изложени тези от емпиричното проучване, което служи за събиране на данни. Презентирани са откритите значими взаимовръзки между личността на потребителите и индивидуалната им склонност за поемане на риск в онлайн пазаруването и предпоченциите им към отделните функционалности на електронните магазини, като същите са подкрепени с удостоверени в научната литература аргументи.

На следващ етап от изследването са презентирани резултатите от приложението на избраните три модела за машинно обучение за прогноза на потребителските предпоченции като резултат от тяхната индивидуалност. Трите модела постигат приблизително еднакви средни стойности на прогнозите според избраните метрики за оценка, като според тях и трите модела могат да бъдат подходящи за подобни изследвания. Моделът случайни гори е избран за оптимизация на следващ етап, тъй като той позволява гъвкаво нагласяване на параметрите и е устойчив на прекомерно нагаждане. При него няма нужда от трансформация на данните, той работи добре с количествени и категорийни характеристики на данните, което го прави подходящ за изследване на индивидуалността в интернет, където процесите и активността на потребителите са динамични. Предложената оптимизация с кръстосано валидиране, прилагайки класа *GridSearchCV* на библиотеката *scikit – learn*, както и с ТРОТ, който базира на алгоритъм за генетично търсене, демонстрира, че и двата метода водят до подобряване на прогнозните стойности според метриците за оценка.

Базирайки се на резултатите от цялостното изследване, в края на трета глава са резюмирани модели на поведение при вземане на решение в онлайн търговията.

Заклучение - резюме на получените резултати

Познаването на личността на индивида, както и техниките за прогнозиране на предпоченциите му като следствие на неговата индивидуалност открива много хоризонти. Предвид, че човешкият фактор играе решаваща роля в социалните и икономически процеси, темата за личността е приложима в различни научни области на съвременния свят. Само някои от тези сфери и дисциплини са например стратегическото управление, управлението на проекти, управлението на човешки ресурси и подборът на кадри, управлението на взаимоотношенията с клиентите, управлението и оценка на риска, маркетингът и рекламата, социалната търговия, подобряването на лоялността на клиентите и електронната търговия. Превдид също факта, че съвременният свят и съответно икономическите модели са предимно потребителски ориентирани, всички тези области могат да се допълват и/или да намират приложение в онлайн търговията, един сегмент, който бележи устойчиво развитие през последните години [57].

Резултатите от научния труд показват, че определени функционалности в електронните магазини са по-предпочитани от определени групи потребители. Така познавайки индивидуалността на потребителите е възможно чрез методите за машинно обучение да се прогнозира какви биха били техните предпоченции в онлайн търговията и така да бъдат компилирани модели на поведение при вземане на решение. Това би направило например персонализирането на потребителския интерфейс възможно, така че той да отговаря по-добре на очакванията на потребителите и така те да бъдат улеснени в процеса на вземане на решение.

Ограничения на изследването

Всяко научно изследване има своите ограничения, особено, когато в центъра на проучването стои човекът. Така освен ограниченията относно емпиричното изследване, може да бъде допълнено, че резултатите на научното проучване не могат да бъдат общовалидни в географски и културен аспект [23]. Въпреки че

извадка на проучването се състои от 226 участника от над 10 държави, те са главно представители на европейската култура, което ограничава приложението на резултатите в рамките на този културен кръг.

Друго ограничение остава въпросът до колко отговорите на участниците в изследването отговарят на истинността, както относно оценката на техните личностни детерминанти, така и до техните предпочитания към отделните функционалности. И въпреки че извадката е достатъчно голяма, за да могат резултатите да бъдат сметнати за валидни, преди широко приложение им в практиката е препоръчително да бъде направена проверка на същите чрез други методи и изследвания.

Насоки за бъдещи изследвания

Като една от основните насоки за бъдещи изследвания върху тематиката на дисертацията се препоръчва да бъде проучено дали чрез персонализация, изхождайки от индивидуалността на потребителя и неговите предпочитания, ползваемостта на електронните магазини бива значително подобрена.

Препоръчва се също така имплицитните методи за оценка на личностните детерминанти на потребителите в интернет да бъдат допълвани и контролирани чрез експлицитни, като ТРІ тестът, поради своя обем, предоставя една такава възможност.

Преди широко прилагане на резултатите в практиката е препоръчително подобни изследвания да бъдат проверявани повторно или при възможност с по-голяма извадка. Един друг начин, позволяващ например получените резултати да бъдат допълнително проверявани, е проследяването на погледа на участниците чрез Eye Tracking. Поради факта, че човешкото око (fovea centralis) поема информация само по време на фиксация върху екрана, която обикновено трае между 200 и 600 милисекунди [8], тази технология позволява да бъде проверено дали дадена област, например определена функционалност, в това число опция или меню на даден уеб сайт, е наистина от интерес за даден потребител.

Темата за изучаване на индивидуалността и приложението ѝ в различни области на съвременния живот, особено чрез приложението на модерните технологии, се радва на все по-засилен интерес в научните среди и в бизнеса. Но е изключително важно в това приложение да бъдат спазвани етичните норми и да не се злоупотребява, а постигнатите резултати да бъдат използвани в услуга на човека.

Публикации

1. Кетипов, Р., Колев, К., Севова, Ж., Благоев, И., Петров, П., Костадинов, Г., Занкински, И. Предварителна обработка на тренда и сезонността във времеви редове с генетични алгоритми. // *XXVII Международен симпозиум „Управление на енергийни, индустриални и екологични системи“*, 16-17 май 2019 г., Банкя, България, 65-68.
2. Кетипов, Р., Колев, К., Севова, Ж., Благоев, И., Петров, П., Костадинов, Г., Занкински, И. Премахване на линеен компонент и синусоидални хармоници от времеви редове с еволюция на разликите. // *В: Сборник доклади от Годишна университетска научна конференция на НВУ „В- Левски“*, 27 – 28 юни 2019, В. Търново, том 10, Научно направление „Технически науки“, 1586-1594, ISSN: 2367-7481.
3. Balabanov, T., Zankinski, I., Ketipov, R. Weights Permutation in Multilayer Perceptron. // *Proceedings of International Conference on Big Data, Knowledge and Control Systems Engineering, “John Atanasoff” Union on Automatics and Informatics*, Bulgaria, 2018, 23-27, ISSN: 2367-6450.
4. Balabanov, T., Ketipov, R., Atanassova, Z. MLP with Stochastic Manipulated Hidden Layer. // *Proc. of the International Scientific Conference - UNITECH 2018*, 2, University Publishing House Vasil Aprilov - Gabrovo, 2018, 324-328, ISSN: 1313-230X.
5. Balabanov, T., Ivanov, S., Ketipov, R. Solving Combinatorial Puzzles with Parallel Evolutionary Algorithms. // *Lecture Notes in Computer Science, including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*, 2020, 11958 LNCS, 493-500, DOI: 10.1007/978-3-030-41032-2_56, SJR 2019 – 0.427 (Q3).
6. Ketipov, R., Kostadinov, G., Petrov, P., Zankinski, I., Balabanov, T. Human-Computer Mobile Distributed Computing for Time Series Forecasting. // *Communications in Computer and Information Science (CCIS)*, Springer, 2019, 1141 CCIS, 508-509, DOI: 10.1007/978-3-030-36625-4_40, SJR 2019 – 0.188 (Q2).

7. Ketipov, R., Kostadinov, G., Petrov, P., Zankinski, I. Genetic Algorithm Based Formula Generation for Curve Fitting in Time Series Forecasting Implemented as Mobile Distributed Computing. // *High-Performance Computing (HPC)*, Borovets, Bulgaria, 2019. *Studies in Computational Intelligence*, 2021, 902 SCI, 40-47, DOI: 10.1007/978-3-030-55347-0_4, SJR 2019 - 0.215 (Q4).
8. Ketipov, R., Kolev, K., Sevova, J., Blagoev, I., Petrov, P., Kostadinov, G., Zankinski, I. Trend and Seasonality Removal with Differential Evolution. // *Information technologies and control*, 4, 2018, 17-22, ISSN: 1312-2262.
9. Zankinski, I., Keremedchiev, D., Blagoev, I., Ketipov, R., Kolev, K., Kostadinov, G., Petrov, P. Recursive brute-force selection operator in genetic algorithms. // *International Scientific Conference UNITECH*, 15 – 16.11.2019, Gabrovo, 227-232, ISSN:1313-230X.

Научно-приложни резултати

С оглед на извършената работа в този дисертационен труд и на резултатите, получени в хода на проведените изследвания, авторът счита, че поставените задачи са изпълнени и целта на дисертационния труд е постигната. Като следствие на това могат да бъдат формулирани следните **научно-приложни резултати**:

1. Избран е ТІРІ тестът като подходящ психометричен модел за оценка на личността на потребителите според петфакторния модел Големите пет (Big Five).

След направения обстоен литературен обзор за изучаване на различните теории и модели за измерване на личността и тяхното прилагане, петфакторният модел Големите пет се оказва особено подходящ за подобни изследвания. ТІРІ тестът е доста кратък, което прави неговото приложение за директно запитване на потребителите в интернет възможно. В комбинация той би бил мощен инструмент за ревизиране или допълване на резултатите от имплицитните методи за оценка на личността на потребителите. В рамките на изучената литература този психометричен инструмент не е използван до сега за оценка на личността на потребителите с цел прогноза на техните преференции в електронната търговия.

2. Конкретизиран е набор от 19 основни функционалности на електронните магазини.

Наборът от 19 основни функционалности на електронните магазини са категоризирани в 3 подгрупи – съдържание и външен вид, инструменти от потребителския интерфейс и фактори, влияещи върху избягването на риска, така че да обхванат ключови функционалности, които са характерни за повечето съвременни онлайн магазини. Този набор от функционалности е подходящ инструмент за изследване на предпочитанията на потребителите в онлайн търговията в зависимост от техните личностни детерминанти.

3. Създадена е стратегия и дизайн на емпиричното изследване на три езика с добавена "6-та" личностна детерминанта "избягване на риска".

Създадената стратегия и дизайн на емпиричното изследване на три езика е съобразено с основните стандарти за етичност и неутралност. След

провеждане на проучването е направен анализ на резултатите като са открити значими корелации между отделните личностни детерминанти и предпочитанията на потребителите в онлайн пазаруването. Получените резултати от проучването съвпадат в по-голямата си част с тези в изучената литература, но добавянето на индивидуалната склонност към риска като "б-та" личностна детерминанта (б-та независима променлива) е новост.

4. Разработени са три модела за машинно обучение: линейна регресия, дърво на решенията и случайни гори. Реализирана е оптимизация за метода случайни гори.

С разработените три модела експериментално е направена прогноза за предпочитанията на потребителите към определени функционалности на електронните магазини в зависимост от тяхната индивидуалност, като в изчисленията влизат само откритите значими връзки между отделните личностни детерминанти и функционалности. След направения сравнителен анализ на резултатите според приложими метрики за оценка на прогнозата на отделните модели за машинно обучение е предложена и реализирана оптимизация за един от тях, а именно за метода случайни гори, който се оказва подходящ в изследваната област.

5. Компилирани са модели за поведение на потребителите при вземане на решения в онлайн търговията според тяхната индивидуалност.

Вследствие на получените резултати от цялостното изследване са компилирани модели за поведение на потребителите при вземане на решения в онлайн търговията според тяхната индивидуалност, където предпочитанията на потребителите могат успешно да бъдат прогнозирани чрез приложение на методите за машинно обучение. Поради факта, че компилираните модели са следствие на резултатите от емпиричното изследване и получените резултати от експеримента с моделите за машинно обучение, те нямат аналог в този вид в изучената литература.

Постигнатите резултати може да бъдат използвани като изходна позиция при персонализация в онлайн търговията, като резултатите могат да бъдат приложени към нови потребители, с цел да бъдат прогнозирани техните предпочитания, като се знае тяхната индивидуалност и по този начин необходимостта от изрично запитване за предпочитание да отпадне.

Декларация за оригиналност на резултатите

Декларирам, че настоящата дисертация съдържа оригинални резултати, получени при проведени от мен научни изследвания (с подкрепата и съдействието на научния ми ръководител). Резултатите, които са получени, описани и/или публикувани от други учени, са надлежно и подробно цитирани в библиографията.

Настоящата дисертация не е прилагана за придобиване на научна степен в друго висше училище, университет или научен институт.

София, 07.05.2021 г.

Подпис: 
/Румен Кетипов/

Библиография

- [1] ЙОРДАНОВА, Т. Личностни детерминанти на иновационното потребителско поведение. Автореферат на дисертационен труд. Софийски университет „Св. Климент Охридски”, 2018.
- [2] КЕХАЙОВА-СТОЙЧЕВА, М. Социално-психологически аспекти на потребителското поведение. Библиотека “Проф. Ц. Калянджиев”, Икономически университет – Варна, 2008 г.
- [3] ALLPORT, G. W., ODBERT, H. S. A Psycholexical Study. // *Psychological Monographs*, Vol. 47, 1936.
- [4] ASHTON, M. C., LEE, K. The Prediction of Honesty-Humility-Related Criteria by the HEXACO and Five-factor Models of Personality. // *Journal of Research in Personality*, Vol. 42(5), 2008, p. 1216–1228.
- [5] BAKKER, A. B., VAN DER ZEE, K. I., LEWIG, K. A., DOLLARD, M. F. The Relationship Between the Big Five Personality Factors And Burnout: A Study Among Volunteer Counselors. // *The Journal of Social Psychology*, Vol. 146(1), 2006, p. 31–50.
- [6] BARKHI, L., WALLACE, L. The Impact of personality Type on Purchasing Decisions in Virtual Stores. // *Information Technology Management*, Vol. 8, 2007, p. 313–330.
- [7] BAYLEY, S., FALESSI, D. Optimizing Prediction Intervals by Tuning Random Forest via Meta-Validation, 2018.
- [8] BENTE, B. Erfassung und Analyse des Blickverhaltens. Lehrbuch der Medienpsychologie. Göttingen, 2004, 298 - 321.
- [9] BLYTHE, J. Consumer Behaviour. London: Sage Publications Inc., 2nd edition, 2013.
- [10] BOLOGNA, C., DE ROSA, A. C., DE VIVO, A., GAETA, M., SANSONETTI, G., VISERTA, V. Personality-Based Recommendation in E-Commerce. //

- Conference: EMPIRE 2013 workshop, 1st Workshop on Emotions and Personality in Personalized Services*, 2018.
- [11] BREIMAN, L. Random Forests. // *Machine Learning*, Vol. 45, 2001, p. 5–32.
- [12] BUETTNER, R. Predicting User Behavior in Electronic Markets Based on Personality-mining in Large Online Social Networks. // *Electronic Markets*, Vol. 27(3), 2017, p. 247–265.
- [13] BURISCH, M. Approaches to Inventory Construction. // *American Psychologist*, Vol. 39, 1984, p. 214–227.
- [14] CAMPS, J., STOUTEN, J., EUWEMA, M. The Relation Between Supervisors' Big Five Personality Traits and Employees' Experiences of Abusive Supervision. // *Frontiers in Psychology*, Vol. 7(112), 2016, p. 1–11.
- [15] CHEUNG, C. M. K., CHAN, G. W. W., LIMAYEM, M. A., CHERRY, K. Critical Review of Online Consumer Behavior: Empirical research. // *Journal of Electronic Commerce in Organizations*, Vol. 3(1), 2005, p. 1–19.
- [16] COSTA, P. T., JR., MCCRAE, R. R. Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual. Odessa, FL: Psychological Assessment Resources, 1992.
- [17] DECARLO, M. Scientific Inquiry in Social Work. Open Social Work Education. 2018. 20.09.2020 <<https://openlibrary-repo.ecampusontario.ca/jspui/bitstream/123456789/550/8/Scientific-Inquiry-in-Social-Work-1549383798.html#chapter-62-section-1>>
- [18] DERI, L., ARMENSKI, T., TESANOVIC, D., BRADIĆ, M., VUKOSAV, S. Consumer Behaviour: Influence of Place of Residence on the Decision-making Process When Choosing a Tourist Destination. // *Economic Research-Ekonomska Istraživanja*, Vol. 27(1), 2014, p. 267–279.
- [19] DONTU, N., GILLILAND, D. The Infomercial Shopper. // *Journal of Advertising Research*, Vol. 36, 1996, p. 69–76.
- [20] EYSENCK, H. J. Dimensions of Personality. London: Methuen, 1947.
- [21] FARAG, S., KRIZEK, K. J., DIJST, M. Exploring the Use of E-shopping and Its Impact on Personal Travel Behaviour in the Netherlands. // *Transportation Research Record*, Vol.18, 2003, p. 47–54.
- [22] FLANAGIN, A. J., METZGER, M. J., PURE, R., MARKOV, A. Mitigating Risk in Ecommerce Transactions: Perceptions of Information Credibility and the Role of User-generated Ratings in Product Quality and Purchase Intention. // *Electronic Commerce Research*, Vol. 14, 2014, p. 1–23.

- [23] FRIEDMAN, H., SCHUSTACK, M. *Personality: Classic Theories and Modern Research*. 6th ed., Pearson Education Inc., 2016. ISBN 978-0-205-99793-0.
- [24] GOLDBERG, L. The Structure of Phenotypic Personality Traits. // *American Psychologist*, Vol. 48(1), 1993, p. 26–34.
- [25] GOLDBERG, L. R. Language and Individual Differences: The Search for Universals in Personality Lexicons, In L. Wheeler (Ed.), Beverly Hills, CA: Sage. // *Review of Personality and Social Psychology*, Vol. 2, 1981, p. 141–165.
- [26] GOSLING, S. D., RENTFROW, P. J., SWANN, W. B. A Very Brief Measure of the Big-five Personality Domains. // *Journal of Research in Personality*, Vol. 37(6), 2003, p. 504–528.
- [27] GRAY, D., *Doing Research in the Real World*, Thousand Oaks, CA: SAGE Publications, 3rd edition, 2014.
- [28] HOLLAND, J. L. *Making Vocational Choices: A theory of Vocational Personalities and Work Environments*. Odessa, FL: Psychological Assessment Resources, 3rd edition, 1997.
- [29] HURWITZ, J., KIRSCH, D. *Machine Learning For Dummies*. IBM Limited edition. Hoboken, John Wiley & Sons, Inc, 2018.
- [30] KARIMI, K., PAPAMICHAIL, N., HOLLAND C. P. The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behaviour. // *Decision Support Systems*, Vol. 77, 2015, p. 137-147.
- [31] KAZEMINIA, A., KAEDI, M., GANJI, B. Personality-based personalization of online store features using genetic programming: Analysis and experiment. // *Journal of Theoretical and Applied Electronic Commerce Research*, Vol. 14(1), 2019, p. 16-29.
- [32] KOKLIC, M., VIDA, I. A Strategic Household Purchase: Consumer House Buying Behaviour. // *Managing Global Transitions*, Vol. 7(1), 2009, p. 75–96.
- [33] KOSINSKI, M., BACHRACH, Y., KOHLI, P., STILLWELL, D., GRAEPEL, T. Manifestations of User Personality in Website Choice and Behavior on Online Social Networks. // *Machine Learning*, Vol. 95, 2014, p. 357–380.
- [34] KOTHARI, C. *Research Methodology. Methods and Techniques*. New-Delhi, New Age International Publishers, 2nd edition, 2004.
- [35] KWON, K., KIM, C. How to Design Personalization in a Context of Customer Retention: Who Personalizes What and to What Extent? // *Electronic Commerce Research and Applications*, Vol. 11, 2012, p. 101–116.

- [36] KUBAT, M. An Introduction to Machine Learning. eBook, Springer, 2017.
- [37] KUHN, M., JOHNSON, K. Applied Predictive Modeling. eBook, Springer, 5th edition, 2016.
- [38] LANGTANGEN, H. P. A Primer on Scientific Programming with Python. Springer, 5th edition, 2016.
- [39] LEWIS, C. D. Industrial and Business Forecasting Methods : A Practical Guide to Exponential Smoothing and Curve Fitting. Butterworth Scientific, London, 1982.
- [40] MAX-PLANCK-GESELLSCHAFT. Risikobereitschaft ist ein relativ stabiles Persönlichkeitsmerkmal. 2017. 18.12.2020 <<https://www.mpg.de/11679764/risikoquotient>>
- [41] MCCRAE, R. R., & JOHN, O. P. An introduction to the five-factor model and its applications. // *Journal of Personality*, 60(2), 175–215, 1992.
- [42] MCCRAE, R. R., & JOHN, O. P. Validation of the Five-Factor Model of Personality Across Instruments and Observers. // *Journal of Personality and Social Psychology*, Vol. 52(1), 1987, p. 81–90.
- [43] MCCRAE, R. R., COSTA, P. T. More Reasons to Adopt the Five-Factor Model. // *American Psychologist*, Vol. 44(2), 1989, p. 451–452.
- [44] MCCLELLAND, D. C., KOESTNER, R., WEINBERGER, J. How do Self-attributed and Implicit Motives Differ? // *Psychological Review*, Vol. 96, 1989, p. 690–702.
- [45] MCGREGOR, I., MCADAMS, D. P., LITTLE, B. R. Personal Projects, Life Stories, and Happiness: On Being True to Traits. // *Journal of Research in Personality*, Vol. 40, 2006, p. 551–572.
- [46] MULLINS, J. L. Management and Organisational Behaviour. Prentice Hall Financial Times, New York, 9th edition, 2010.
- [47] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O. ET AL. Scikit-learn: Machine Learning in Python. // *Journal of Machine Learning Research*, Vol. 12, 2011, p. 2825–2830.
- [48] POWER, R. A., PLUESS, M. Heritability Estimates of the Big Five Personality Traits Based on Common Genetic Variants. // *Translational Psychiatry*, Vol. 5(7), 2015.
- [49] ORRU, G., MONARO, M., CONVERSANO, C., GEMIGNANI, A., SERTORI, G. Machine Learning in Psychometrics and Psychological Research. // *Front. Psychol.*, Vol. 10, 2020. DOI: 10.3389/fpsyg.2019.02970

- [50] RENNER, K. -H., SCHÜTZ, A., MACHILEK, F. Internet und Persönlichkeit: Differentiell-psychologische und diagnostische Aspekte der Internetnutzung. Göttingen, Hogrefe, 2005.
- [51] RIAZ, M. N., RIAZ, M. A., BATOOL, N. Personality Types as Predictors of Decision Making Styles. // *Journal of Behavioural Sciences*, Vol. 22(2), 2012, p. 100–114.
- [52] SARKER, S., BOSE, T. K., PALIT, M., HAQUE, E. Influence of Personality in Buying Consumer Goods-a Comparative Study Between Neo-Freudian Theories and Trait Theory Based on Khulna Region. // *International Journal of Business and Economics Research*, Vol. 2(3), 2013, p. 41–58.
- [53] SAUNDERS, M., LEWIS, P., THORNHILL, A. Research Methods for Business Students. Harlow: FT/ Prentice Hall, 5th edition, 2009.
- [54] STANGL, W. Eigenschaften - Situationen - Verhaltensweisen - ESV: Eine Ökonomische Ratingsform des 16PF. // *Zeitschrift für Experimentelle und Angewandte Psychologie*, Vol. 36, 1989, p. 665–671.
- [55] STANGL, W. Personality and the Structure of Resource Preferences. // *Journal of Economic Psychology*, Vol. 14(1), 1993, p. 1–15.
- [56] TSAO, W.-C., CHANG, H.-R. Exploring the Impact of Personality Traits on Online Shopping Behavior. // *African Journal of Business Management*, Vol. 4(9), 2010, p. 1800–1812.
- [57] TURBAN, E, OUTLAND, J., KING, D. ET AL. Electronic Commerce 2018. A Managerial and Social Networks Perspective. eBook, Springer, 2018.
- [58] USAKLI, H. A Study on Personality Traits and Shopping Online. // *Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach*, Vol. 121, Seria: Administracja i Zarządzanie, Vol. 48, 2019.
- [59] VINCIARELLI, A., MOHAMMADI, G. A Survey of Personality Computing. // *IEEE Transactions on Affective Computing*, Vol. 5(3), 2014, p. 273–291.
- [60] WASKOM, M. Seaborn: statistical data visualization. 2015. 11.12.2020 <<https://seaborn.pydata.org/index.html>>

Abstracts of Dissertations

Number 2, 2021

INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES
BULGARIAN ACADEMY OF SCIENCES

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ

ИНСТИТУТ ПО ИНФОРМАЦИОННИ И КОМУНИКАЦИОННИ ТЕХНОЛОГИИ

Брой 2, 2021

Автореферати на дисертации