BULGARIAN ACADEMY OF SCIENCES
INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGIES

RUMEN RUMENOV KETIPOV

PERSONALITY AND DECISION-MAKING MODELS IN INTERNET

ABSTRACT OF PHD THESIS

Supervisor:

Assoc. Prof. Vera Angelova

Sofia, 2021

The PhD thesis was discussed and allowed to be defended during an extended session of the Department of Intelligent Systems at IICT-BAS, which had been held on 19/04/2021.

The full volume of the dissertation is 181 pages. It consists of an introduction, three chapters, and a conclusion. The list of references contains 273 items. The work includes 2 Appendices. The text of the dissertation includes 27 tables and 36 figures.

The defence of the PhD thesis had been held on .... at .... in Room ………, Block …………, IICT-BAS.

Approved by Supervising Committee:

1. Acad. Ivan Popchev, DSc.
2. Prof. Daniela Borissova, DSc.
3. Prof. Maria Nisheva, PhD
4. Prof. Tania Pencheva, PhD
5. Assoc. Prof. Irina Radeva, PhD

Author: **Rumen Rumenov Ketipov**
Title: PERSONALITY AND DECISION-MAKING MODELS IN INTERNET

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, online 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 **M**achine **L**earning (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 individualtiy 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].

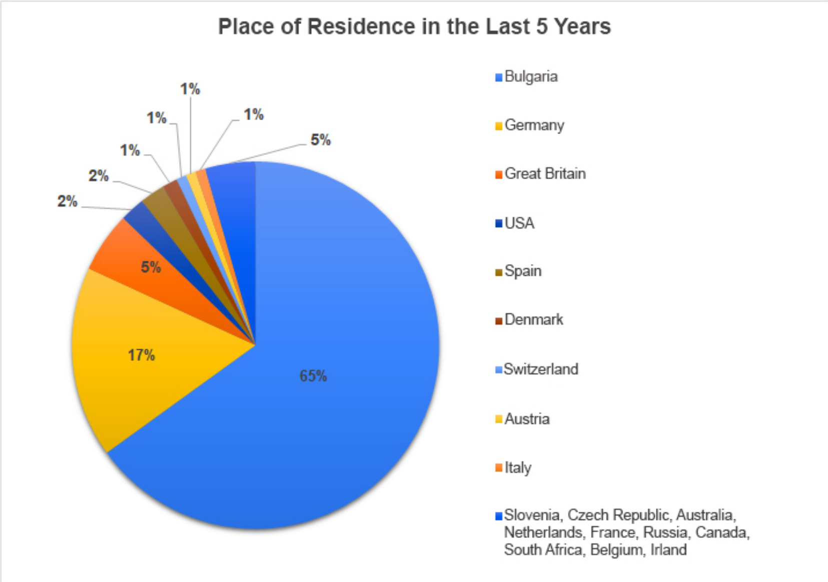


Figure 3.1: Place of residence in the last 5 years

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

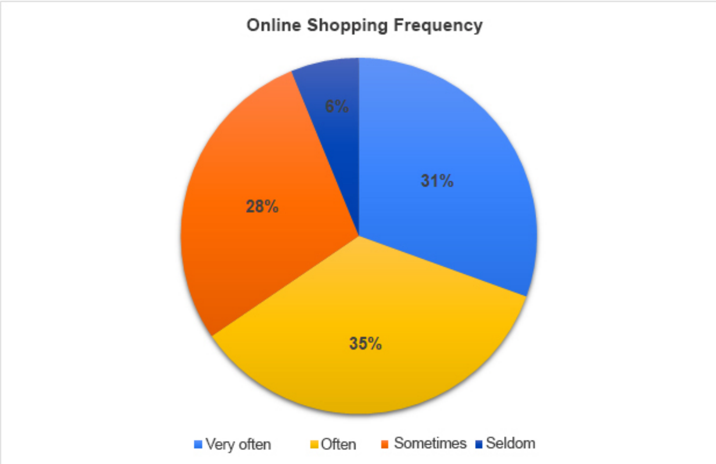


Figure 3.2: Online shopping frequency

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 PSPP[1] 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.

[1]GNU 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 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.**			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 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.**			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 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.**			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 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.**			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 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.**			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 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.**			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 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.**			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 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.**			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 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.**			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 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.**			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
Accuracy 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

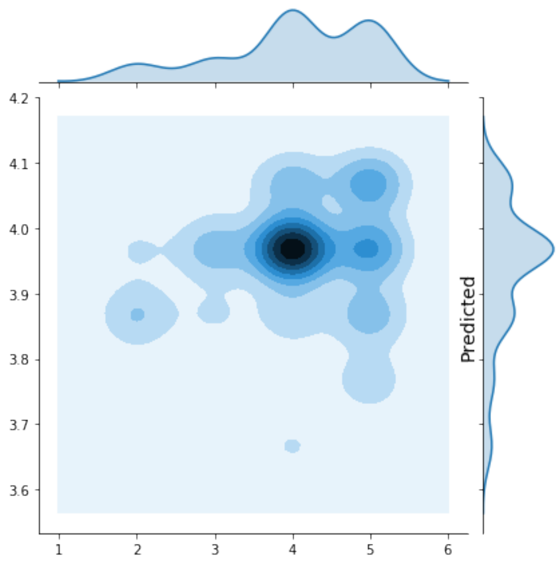


Figure 3.3: Linear Regression - prediction of checking for alternative (secure) payment methods and risk averseness, data distribution

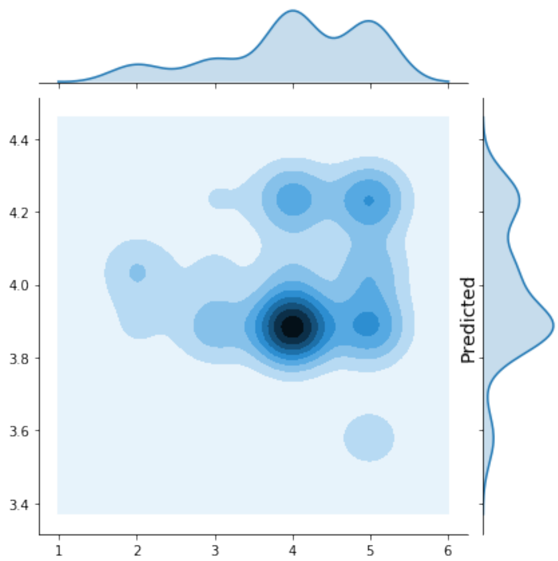


Figure 3.4: Decision Trees - prediction of checking for alternative (secure) payment methods and risk averseness, data distribution

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.

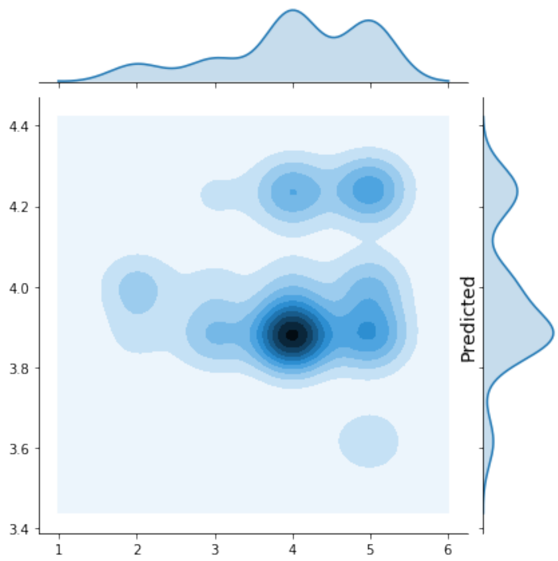
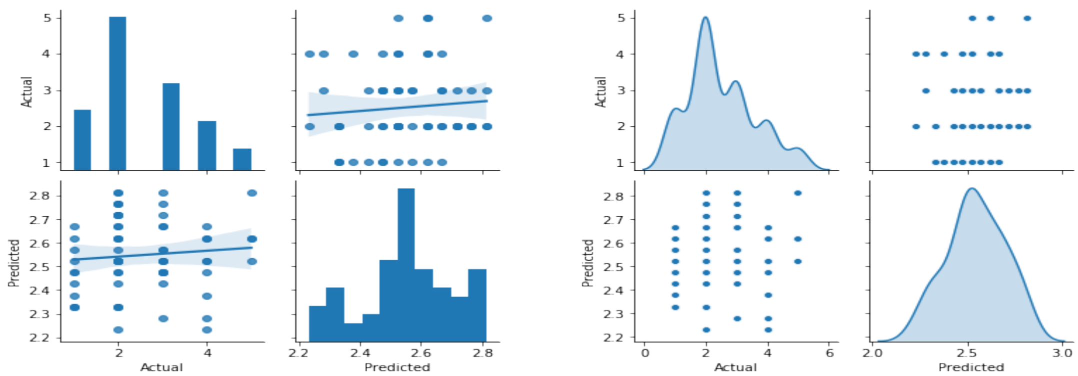


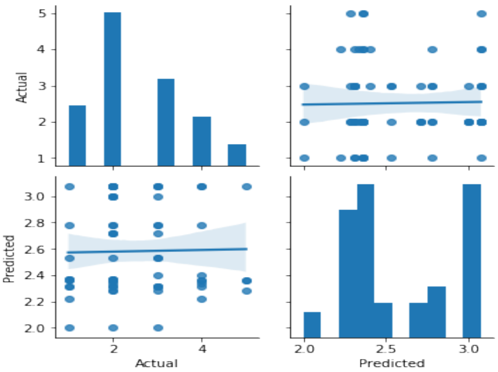
Figure 3.5: Random Forest - prediction of checking for alternative (secure) payment methods and risk averseness, data distribution



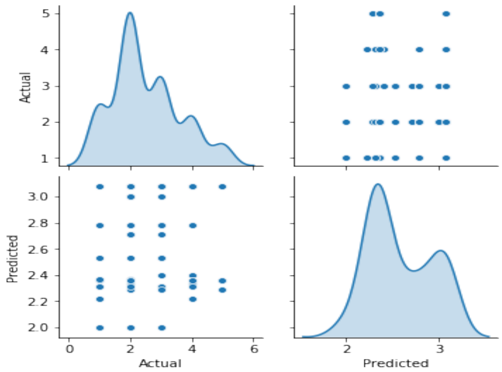
(a) Regression (b) Density

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

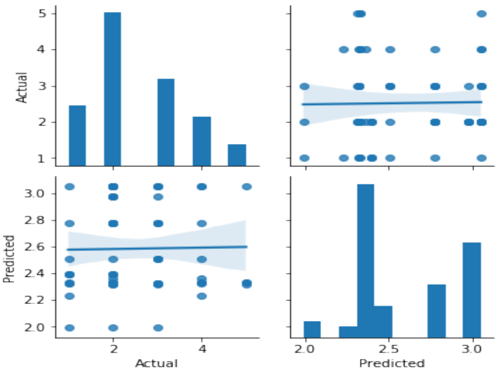


(a) Regression

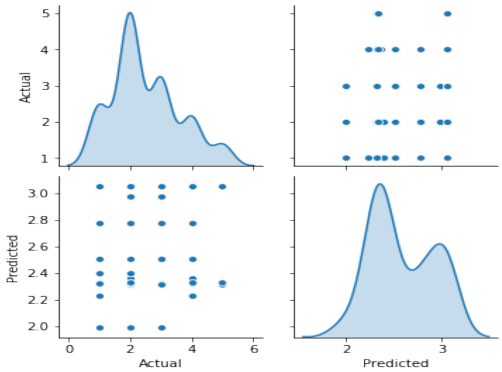


(b) Density

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



(a) Regression



(b) Density

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.

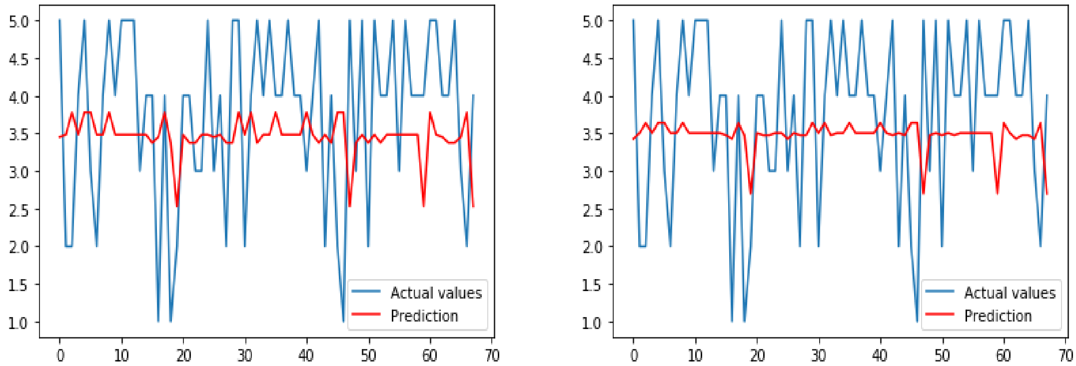
In the proposed optimization, *GridSearchCV* goes through all the combinations 10 times because the value of the cross-validation generator is set to 10 (cv=10). In this case, there is a total of 120 fits.

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].

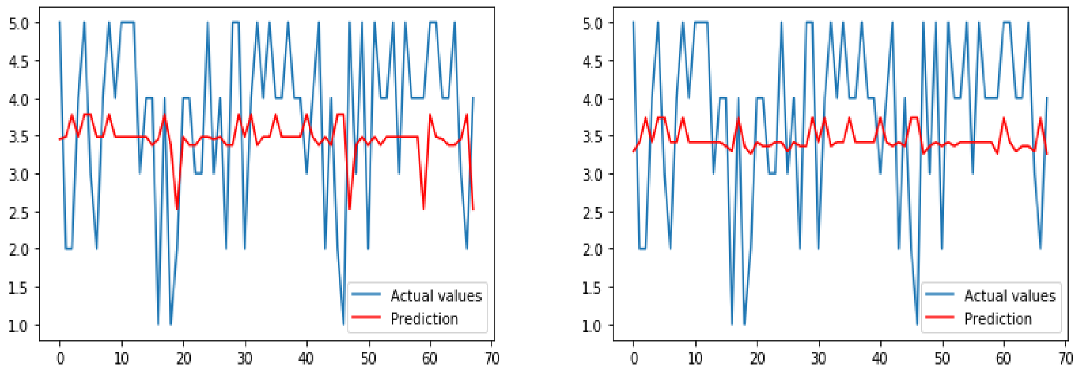
In the proposed optimization, TPOT has to evaluate 1 100 configurations as the population size is set to 100 and the number of iterations to the run pipeline optimization process is set to 10 (population_size + (generations x offspring_size)). By default, the number of offspring to produce in each genetic 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 "17. 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%.



(a) Random Forest (b) GridSearchCV

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



(a) Random Forest (b) TPOT

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, whereat 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 articel (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%). Neverless, 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 accordingly 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, decsion 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-Ekonomska Istraživanja*, Vol. 27(1), 2014, p. 267–279.

[12] Donthu, 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 $//< https : //www.sas.upenn.edu/ \ cavitch/pdf - library/Freud_S E_T hree_E ssays_c omplete.pdf >$

[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., McLaughlin, 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 Uniwersitetu 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* >