

REVIEW

By: Assoc. Prof. Atanas Georgiev Atanasov, PhD;

University of National and World Economy;

Academic position: Assoc. Prof.;

Professional field: 3.8. Economics;

Scientific specialty "Statistics and demography".

About: dissertation thesis for awarding of educational and scientific doctoral degree of professional field 3.8. Economics, scientific specialty: "Data Science", to Department of Statistics and Econometrics, Sofia University "St. Kliment Ohridski".

Author of dissertation thesis: **Gloria Ventsislavova Hristova**

Topic of the dissertation thesis: "An automated system for analysis of online communication with customers based on machine learning and natural language processing - structure, development and business applications".

Reason for presentation of the review: participation in the scientific jury on defence of the dissertation thesis in compliance with the Order RD 38-326/04.07.2022 of the Rector of Sofia University "St. Kliment Ohridski".

1. Information for the PhD student.

Gloria Hristova is a full-time PhD student at the Department of Statistics and Econometrics, dismissed with the right of defence.

2. General characteristics of the presented dissertation thesis.

The presented dissertation thesis comprises 245 standard pages basic text structured into an introduction, three chapters and a conclusion. Appendices are a list with the used abbreviations, glossary of terms, bibliography and annexes.

The introduction considered the relevance of the research topic. Definitions of the object, subject and main aim of the dissertation work were given. A number of research tasks and hypotheses were formulated. The structure of the dissertation was shortly presented.

In the first chapter, a literature review was made including text analysis, online chat communication analysis. Studies were considered in the field of the analysis of text data in Bulgarian, modelling of topics in text data, as well as the opinion and sentiment analysis expressed in text data.

The second chapter was dedicated to the methodology as under consideration was the algorithm for initial reading and structuring of the online chat communication /Module I/, incl. data description, modelling of topics in the online chat communication /Module II/, sentiment analysis in the online chat communication /Module III/ and a summary and visualization of the results /Module IV/.

The third chapter included the results of the empirical survey done. It was presented the sample of chats done, familiarization with the data, analysis of topics, summary and

discussions of the results of the modelling of topics in the online chat communication, client sentiment analysis, summary and discussion of the results of the client sentiment modelling and it ended with a summary and visualization of results.

The dissertation ended with a conclusion for summary of the main inferences of the study done.

The bibliography comprises 220 sources of Bulgarian and foreign authors and in the appendices are presented some basic distributions and interim results.

3. Evaluation of the results received.

In the third chapter the system for analysis of the online client communication was presented, which was tested within an actual real business case while using 38 thousand chats. The data used are topical and are for the period from January 22, 2019 till April 1, 2021. The online chat communication between customers and operators in the call centre of a large financial institution in Bulgaria was analysed as the main results and conclusions were presented and documented in the dissertation.

4. Evaluation of the scientific and applied scientific contributions.

The presented dissertation thesis is a scientific work which has new scientific and applied results which could be summarized as the following scientific and applied scientific contributions:

1. Establishing a wide literature source presenting a current picture for the development in the field of natural language processing in Bulgaria, as well as for the opportunities for analysis of text data in Bulgarian.

Only establishing of such a wide-scaled literature source although a comprehensive work in volume, it does not represent a scientific contribution, but analyzing the opportunities for analysis of text data in Bulgarian is a contribution without a doubt.

2. A methodology for interpretation of the results received has been established which is characterized by an applied significance and an opportunity for extrapolation of the steps in it. The methodology could be applied not only in the bank industry, but also in all other industries where such data are generated as a result of the business processes of the company, which is a scientific contribution beyond doubt.

3. Establishing automatization of the process of knowledge scraping of the client online chat communication done in the call centre. The illustrated quantitative analysis methods could be combined with set approaches for quality analysis so that the business could take better decisions characterized by higher rationality degree, which is applied contribution.

4. As additional contribution could be considered also the opportunity that the established automatized system could adapt itself in case of input of new data and the information could be updated in real time. It is essential that be noted that although the system is adaptable, a necessity exists of carrying out a periodic monitoring and update of the prognostic models versus the changing characteristics of the data during time.

5. Establishing of methodology for analysis of client online chat communication especially in Bulgarian. Specific algorithms and procedures for processing and analysis could be replicated by other researchers in this field and used aimed at automatizing and facilitating of the analysis of such type of data (e.g. establishing of libraries, modules, etc.)

6. A methodology for predicting of client satisfaction by online chat communication has been established, which is based only on text characteristics and grammar information scraped from the text.

7. Methodology for analyzing the topics of client interest has been established. Important techniques for processing and modelling have been pointed out, which are necessary for knowledge scraping of the type of data searched. Valuable conclusions have been done about the various levels of representation of the online chat communication in the call centre – a topic, which as far as the author knows, has not been touched by other researchers in this field.

5. Evaluation of the publications on the dissertation.

The PhD student has indicated five independent publications in English, which is significantly more than the required minimum. The publications correspond to the topic of the dissertation work.

6. Evaluation of the abstract.

The presented abstract reflects fully and accurately, both structurally and in terms of content, the dissertation work and the scientific-applied contributions therein.

7. Critical notes and recommendations.

I do not have essential critical notes or recommendations.

8. Conclusion.

The dissertation thesis with topic of: “An automated system for analysis of online communication with customers based on machine learning and natural language processing - structure, development and business applications” is a comprehensive scientific and applied scientific study. This work has been elaborated in compliance with the legal requirements for awarding of educational and scientific doctoral degree. In conclusion, I declare my statement “for” awarding of educational and scientific doctoral degree of professional field 3.8. Economics, scientific specialty: “Data Science”, to Gloria Hristova.

Date: August 22, 2022

Sofia

Signature:

Assoc. Prof. Atanas Atanasov, PhD