ABSTRACT

The thesis is presented in 78 pages. It contains 3 appendixes and bibliography of 27 references. Ten figures and 8 table are given in the thesis.

Topic relevance. In modern world, violating group anonymity problem as one of sources of uncovering purposely hidden information, obtaining which could harm proprietors of distributed data, is gaining huge importance. Rapid data volume growth, uncontrollable data spreading due to distributed data storage architecture and also a threat of data reconstructing from freely accessible ones demands more attentive position to violating group anonymity problem. Individual and group anonymity tasks favor creating educational data sets opportunities without danger of violating respondent's private rights.

Thesis connection to scientific programs, plans, and topics. The thesis was prepared according to the scientific research plan of the Applied Mathematics Department of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute."

Research goal and objectives. The goal of this thesis is to determine possibilities of violating group anonymity data by using machine learning techniques, explain constructed model structures and determining dependencies in decision making of chosen methods.

To accomplish this goal, the following objectives were reached:

– defining the state of development of the issues of the chosen scientific problem in the national and foreign literature;

– processing of statistical data of population census for determining key attributes needed for violating group anonymity;

– constructing and analysis of classification models;

– conducting accuracy analysis of constructed models;

– determining links between decisions made by different machine learning techniques.

Object of research is methods for violating group anonymity and also machine learning techniques, which are used for data classifying (decision trees, especially random forest, neural networks).

Subject of research is developing of mathematical model and software for assessing the possibility of violating group anonymity by machine learning techniques; explaining classifying model’s structures, which are built by different machine learning methods, their interdependencies.

Methods of research. To solve the task, the following methods were used: methods of machine learning, which are used for data classifying, decision trees, especially random forest, and neural networks, especially multi-layer perceptron.

Scientific contribution consists of the following:

– examined the possibility of violating group anonymity data by neural networks and random forests;

– examined the accuracy of constructed models, determined optimal characteristics of methods that are solving stated task;

– compared structure of constructed models and determined correspondences between classifying decisions made by random forest ensemble and space linear transformation made by multi-layer perceptron.

Practical value of obtained results. Obtained results could be applied for violating group anonymity data. Deep explanation of constructed model structures allows to mark a line between decisions made by random forest ensemble and neural network, therefore one could explain black-box structure – neural network – using white-box structure – random forest.

Approbation of the thesis results. Basic ideas and results of the research were presented at VI-th international practical-science internet-conference «Тенденції та вектор розвитку науки в сучасному світі» (2018).

Publications. Thesis results are published in 2 scientific works:

– in 2 papers in scientific journals included in the List of Professional Scientific Journals of Ukraine (technical sciences).

Keywords: group anonymity, random forest, neural network.