

## I. IDENTIFICATION DATA

<b>Thesis name:</b>	<b>Application of machine learning methods to solve the problem of user identification on various digital devices</b>
<b>Author's name:</b>	<b>Elena Ivanova</b>
<b>Type of thesis :</b>	master
<b>Faculty/Institute:</b>	Faculty of Electrical Engineering (FEE)
<b>Department:</b>	Department of Computer Science
<b>Thesis reviewer:</b>	Rustem Faridovitch Salimov
<b>Reviewer's department:</b>	Kazan Federal University, Institute of Computer Mathematics and Information Technologies, Department of Mathematical Statistics

## II. EVALUATION OF INDIVIDUAL CRITERIA

<b>Assignment</b>	<b>challenging</b>
<i>Evaluation of thesis difficulty of assignment.</i>	
<p>Nowadays everyone uses many digital devices for different purposes. User's identity becomes fragmented which is not good for advertising. This thesis is focusing on different machine learning techniques that could be applied to link computers and mobile devices that belong to one person. The task was hosted on Kaggle. Data includes both user's behavior on mobile devices/computers and used IP addresses.</p> <p>The main goal of the thesis was to construct an algorithm that for any device predicts a list of cookies of the same user. Algorithm uses machine learning techniques to obtain greater accuracy of predictions.</p>	

<b>Satisfaction of assignment</b>	<b>fulfilled</b>
<i>Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.</i>	
<p>Author describes the problem and the goal, but the introduction is rather short. Then she carried out an analysis of the literature devoted to the problem and based on them chose two approaches to its solution.</p> <p>Theoretical part gives all the explanations of methods and terms that were used in the thesis.</p> <p>Author provides detailed description of the construction of the final algorithm, as well as the methods and libraries that were applied. Each stage of development is accompanied by results and conclusions. These chapters are well written. However, author considered only gradient boosting methods as the main learning algorithm. Probably, other models could improve the results.</p> <p>All structure is logical and consistent.</p>	

<b>Method of conception</b>	<b>correct</b>
<i>Assess that student has chosen correct approach or solution methods.</i>	
Author chose a standard approach in machine learning to develop an algorithm	

<b>Technical level</b>	<b>A - excellent.</b>
<i>Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.</i>	
Author showed a good level of knowledge in various ML methods and applied them properly to solve the task	

<b>Formal and language level, scope of thesis</b>	<b>A - excellent.</b>
<i>Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.</i>	
Diploma thesis is well formed and good structured. Language level is above average.	

<b>Selection of sources, citation correctness</b>	<b>B - very good.</b>
<i>Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own</i>	

*results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.*

All resources were properly cited. Author used all relevant resources according to the theme; most of them are quite new.

### **Additional commentary and evaluation**

*Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.*

The value of the final error at the level of the 5th place in the contest on the Kaggle site indicates the good quality of the final algorithm.

### **III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION**

*Questions for defense:*

- *Why didn't you use other machine learning models for solving the task?*

I evaluate handed thesis with classification grade **A - excellent**.

Date: **4.6.2018**

Signature: