Ministry of Science and Higher Education of the Russian Federation
NATIONAL RESEARCH
TOMSK STATE UNIVERSITY (NR TSU)

Institute of Applied Mathematics and Computer Science

APPROVE:
Director
_____A. V. Zamyatin

Work program of the discipline

Applied Machine learning - I

in the major of training

01.04.02 Applied mathematics and informatics

Orientation (profile) of training:

Big Data and Data Science

Form of study
full-time

Qualification
Master

Year of admission
2023

Code of discipline in the curriculum: B1.O.07

AGREED:
Head of EP
_____ A.V. Zamyatin

Chairman of the EMC
_____ _S.P. Sushchenko

Tomsk – 2023

1. Purpose and planned results of mastering the discipline

The purpose of mastering the discipline is the formation of the following competencies:
- UK-1 - the ability to carry out a critical analysis of problem situations based on a systematic approach, to develop an action strategy;
- PC-6 - the ability to manage the receipt, storage, transmission, processing of big data.
The results of mastering the discipline are the following indicators of the achievement of competencies:
IUK-1.1 Identifies a problem situation, on the basis of a systematic approach, carries out its multifactorial analysis and diagnostics.
IUK-1.2 Carries out the search, selection and systematization of information to determine alternative options for strategic solutions in a problem situation.
IUK-1.3 Suggests and justifies the strategy of action, taking into account the limitations, risks and possible consequences.
IPK-6.1 Monitors and evaluates the performance of big data processing.
IPK-6.2 Uses methods and tools for receiving, storing, transmitting, processing big data.
IPK-6.3 Develops proposals to improve the performance of big data processing.

2. Tasks of mastering the discipline

The purpose of the discipline is to teach students to carry out work on the study of big data using artificial intelligence and big data technologies and develop intelligent systems using the tools of Python, R libraries, public cloud services, evaluate the effectiveness of their work and implement them in applications;

Discipline tasks:

1. To teach students to identify, form and agree on the requirements for the results of analytical work using artificial intelligence technologies and big data;

2. To teach students the principles of planning and organizing analytical work using artificial intelligence technologies and big data;

3. To teach students how to prepare data for analytical work on the study of big data using artificial intelligence methods;

4. Teach students to conduct analytical research and develop applications using artificial intelligence technologies and big data in accordance with customer requirements.

3. The place of discipline in the structure of the educational program

Discipline belongs to the mandatory part of the educational program.

4. Semester of mastering and form of intermediate certification in the discipline

First semester, exam.

5. Entrance requirements for mastering the discipline

For the successful mastering of the discipline, training outcomes are required in the following disciplines: "Intelligent systems", "Neural networks".

6. Implementation language

English.

7. Scope of discipline

The total labor intensity of the discipline is 5 credits, 180 hours, of which:
- lectures: 20 hours
- laboratory: 40 hours
  including practical training: 0 h.

The volume of independent work of the student is determined by the curriculum.

8. The content of the discipline, structured by topics

Topic 1. Exploratory data analysis

Preliminary data analysis. Visualization. Conducting exploratory data analysis.

Topic 2. Classification and regression models

Training of classifiers. Training of regressors. Evaluation of the quality of education. Acquaintance with software tools for designing machine learning systems. Building simple classification and regression models.

Topic 3. Ensembles of models

Selection of model hyperparameters. ensembles of models. Building ensembles of machine learning models for solving classification problems.

Topic 4. Working with feature space

Dimension reduction of feature space. Construction of signs. Data clustering. Evaluation of the information content of features and reduction of the dimension of the feature space for the analysis of objects from databases.

Topic 5. Fundamentals of neural network computing

Neural network architectures for heterogeneous data analysis problems. Application of neural networks for solving problems of classification and regression. Designing neural network models for classification and regression.

Topic 6. Image processing with convolutional neural networks

Convolutional neural networks. Visualization of the parameters and activity of the neural network. Designing neural network models for image classification.

Topic 7. Autoencoders

Transfer learning. Fundamentals of working with autoencoders. Improving the efficiency of deep neural networks. Designing neural network models of image segmentation.

Topic 8. Analysis of signals and time series

Vector representation of text data. Recurrent neural networks. Designing neural network models for time series analysis.

Topic 9. Using generative-competing models

Fundamentals of generative-competing models. Implementation of a generative neural network model.

Topic 10. Practical aspects of using reinforcement learning

Reinforcement training. Building artificial intelligence systems using reinforcement learning.

9. Ongoing evaluation

The ongoing evaluation is carried out by monitoring attendance, conducting tests, tests on lecture material, performing laboratory work, and is recorded in the form of a checkpoint at least once a semester.

10. The procedure for conducting and criteria for evaluating the intermediate certification

The theoretical material on the discipline is given in the form of lectures using standard multimedia demonstration tools in .pdf format. In practical classes, students solve problems of building artificial intelligence systems using the tools of the Python and R libraries. Current control on practical work is carried out in the form of a discussion of the algorithm and the results of its work.

Upon delivery of each practical work, the achievement of the competence of the IPK-3.1, assigned to the discipline, is checked.

The final grade is set as the arithmetic mean based on the results of control and practical work, rounded to the nearest integer.

11. Educational and methodological support

a) Electronic training course on the discipline at the electronic university "Moodle" - https://moodle.tsu.ru/course/view.php?id=221290

b) Assessment materials of the ongoing evaluation and intermediate certification in the discipline.

12. List of educational literature and Internet resources

a) main literature:

— Joel Grace. Data Science: Data science from scratch. 2nd edition. St. Petersburg: BHV-Petersburg, 2021.

- Sebastian Raska, Vahid Mirjalili. Python and machine learning. M.: Dialectics, 2020.

– Ameet V. Joshi. Machine Learning and Artificial Intelligence. Springer Nature Switzerland AG, 2020.

— Denis Rothman. Artificial Intelligence by Example. second edition. Packt Publishing, 2020.

– Stuart Russel, Peter Norvig. artificial intelligence. A Modern Approach. 4th edition. Hoboken: Pearson, 2021.

– Stuart Russel, Peter Norvig. artificial intelligence. A Modern Approach. 4th edition. Hoboken: Pearson, 2021.

— Andrew Glassner. Deep learning without math. Volume 1. Basics. Moscow: DMK Press, 2020.

— Andrew Glassner. Deep learning without math. Volume 2. Basics. Moscow: DMK Press, 2020.

– Ian Goodfellow, Joshua Bengio, Aaron Courville. Deep learning. Second color edition, revised. M.: DMK Press, 2018.

– Roman Shirkin. artificial intelligence. The Complete Beginners' Guide to Artificial Intelligence. ISBN: 9798609154415. Amazon KDP Printing and Publishing, 2020.

— François Chollet. Deep learning in Python. St. Petersburg: Peter, 2018.

13. List of information technologies

a) licensed and freely distributed software:

– Anaconda (Python, R), Tensorflow deep learning libraries, Keras, Pytorch, OpenAI services.

- publicly available cloud technologies (Google Colab.).

b) information reference systems:

– Electronic catalog of the TSU Scientific Library – http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system

– TSU electronic library (repository) – http://vital.lib.tsu.ru/vital/access/manager/Index

14. Logistics

Halls for lectures.

Classrooms for seminars, individual and group work, ongoing evaluation and intermediate certification.

Classrooms for independent work, equipped with computer technology and access to the Internet, to the electronic information and educational environment and to information reference systems.

15. Authors information

Aksyonov Sergey Vladimirovich, Ph.D. tech. Sciences, Associate Professor, Department of Theoretical Foundations of Informatics.