Appendix 1

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 of the Institute of Applied
Mathematics and Computer Science
_____ A.V. Zamyatin
«16» 05 2022

Evaluation materials of the current control and intermediate certification in the discipline
(Evaluation tools by discipline)
Deep Learning – I

in the major of training
01.04.02 Applied mathematics and informatics

Orientation (profile) of training:
Big Data and Data Science

Tomsk – 2022

ET was implemented:
cand. tech. sciences,
Associate Professor of the Department
of Theoretical Foundations of Informatics — S.V. Aksenov

Reviewer:
cand. tech. sciences,
Associate Professor of the Department
of Theoretical Foundations of Informatics — O.V. Marukhina

Evaluation tools were approved at a meeting of the educational and methodological commission of the Institute of Applied Mathematics and Computer Science (EMC IAMCS).

Protocol dated 12.05.2022 № 4

Chairman of the EMC IAMCS,
Dr. tech. Sciences, Professor — S.P. Sushchenko

Evaluation tools (ET) are an element of the system for assessing the formation of competencies among students in general or at a certain stage of its formation.

The ET is developed in accordance with the work program (WP) of the discipline.

1. Competencies and training outcomes, obtained upon the discipline mastery

Competencies	**Competence indicator**	**Code and name of planned training outcomes that characterize the stages of competency formation**	**Criteria for evaluating training outcomes**			
			Excellent	Good	Satisfactory	Unsatisfactory

UK-1. Able to carry out a critical analysis of problem situations based on a systematic approach, develop an action strategy OPK-3. Able to develop mathematical models and analyze them when solving problems in the field of professional activity PC-6. Able to manage the receipt, storage, transmission, processing of large	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. IOPC-3.3. Develops and analyzes new mathematical models for solving applied problems of professional activity in the field of applied mathematics and informatics. 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	OR-1.1.1 The student will: - Know the procedures for identifying, forming and coordinating requirements for the results of analytical work using deep learning technologies OR-1.2.1 The student will: - Know the principles of planning and organizing analytical work using deep learning technologies OR-1.3.1 The student will be able to: - Prepare data for analytical work on big data research using deep learning methods OP-6.1.1 The student will be able to: - conduct analytical research and develop applications using deep learning technologies in accordance with customer requirements OP-6.1.2 The student will know the methods of data preparation for analytical work on the study of big data using deep learning methods OP-6.1.3 The student will own methods for conducting analytical research and developing applications using deep learning technologies in accordance with customer requirements	Thoroughly understands the basics of modern deep learning technologies for analyzing heterogeneous data (tables, images, signals, sequences) for building predictive models, preparing business intelligence, solving data presentation and processing problems. At a high level, he solves practical problems of designing applications that use deep learning using modern software tools (programming languages, libraries and frameworks), using machine and deep learning technologies to display interdependencies in data.	He is well versed in the basics of modern deep learning technologies for analyzing heterogeneous data (table, images, signals, sequences) for building predictive models, preparing business intelligence, solving data presentation and processing problems. Able to solve practical problems of designing applications using deep learning with the help of modern software tools (programming languages, libraries and frameworks), using machine and deep learning technologies to display	Poorly versed in the basics of modern deep learning technologies for analyzing heterogeneous data (tables, images, signals, sequences) for building predictive models, preparing business intelligence, solving data presentation and processing problems. With significant difficulties, he solves practical problems of designing applications that use deep learning using modern software tools (programming languages, libraries and frameworks), using machine and deep learning technologies to display interdependenci	Does not know the basics of modern deep learning technologies for analyzing heterogeneous data (tables, images, signals, sequences) for building predictive models, preparing business intelligence, solving data presentation and processing problems. Does not know how to use modern software tools (programming languages, libraries and frameworks) to solve practical problems, apply machine and deep learning technologies to display interdependencies in data.

2. Stages of competency formation and types of evaluation tools

№	Stages of competency formation (discipline sections)	Code and name of training outcomes	Type of evaluation tool (tests, assignments, cases, questions, etc.)
1	Section 1. Fundamentals and architectures of deep neural networks. Architectures of deep neural networks. Performing a laboratory work No. 1 (Implementation of deep fully connected neural network models)	OP-1.1.1, OP-1.2.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.1
2	Section 2: Develop applications that use deep learning. Design and deploy applications that use deep neural networks. Performing a laboratory work #2 (Development and deployment of applications with deep neural network models)	OP-1.1.1, OP-1.2.1, OP-1.3.1, OP-6.1.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.2
3	Section 3. Practical aspects of training deep neural networks. Techniques to Improve the Design and Training of Deep Neural Networks Perform Lab #3 (Tools to Improve the Design and Training of Deep Neural Networks)	OP-6.1.1, OP-6.2.1, OP-6.3.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.3
4	Section 4. Practical aspects of sequence modeling. Performing laboratory work No. 4 (Studying texts with deep neural network models)	OP-1.1.1, OP-1.2.1, OP-1.3.1, OP-6.1.1,OP-6.2.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.4
5	Section 5. Practical aspects of using deep neural networks in computer vision. Convolutional neural networks and autoencoders for solving computer vision problems. Completion of laboratory work No. 5 (Segmentation and classification of objects in images using deep neural networks)	OP-1.2.1, OP-1.3.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.5
6	Section 6. Practical aspects of using deep neural networks in natural language understanding problems. Using deep neural networks in natural language understanding problems. Complete Lab #6 (Implementing Chatbot Systems with Deep Learning Tools)	OP-1.2.1, OP-7.1.1, OP-6.2.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.6
7	Section 7. Deep generative models. Generative adversarial neural networks. Complete Lab #7 (Generating Images and Time Sequences)	OP-1.2.1, OP-6.1.1, OP-6.2.1, OP-6.3.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.7
8	Section 8. Deep Learning Research. Adaptation of neural network models. Reinforcement training. Performing laboratory work No. 8 (Research of the procedure for retraining neural networks)	OP-1.2.1, OP-6.1.1, OP-6.2.1, OP-6.3.1	Survey in the classroom, preparation for laboratory classes, public defense of laboratory work No.8
9	Intermediate certification (according to the results of laboratory work (min 70%) and	OP-1.3.1	Public presentation and protection of the results of an individual project.

	the presentation of an individual project - 2-3 min / person)	OP-6.1.1	

3. Typical control tasks or other materials necessary for the assessment of educational training outcomes

3.1. Typical tasks for conducting ongoing monitoring of progress in the discipline:

Laboratory work №1. "Implementation of Deep Fully Connected Neural Network Models"

The purpose of the work is to write a program in Python and R that builds and trains feed-forward neural networks that solve classification and regression problems (samples received from the teacher), it is required to select a non-redundant network architecture that works with an acceptable error level and visualize the process of model training. The results of the work should be included in the report.

Task description.

Write a Python program that trains a neural network classifier and regressor using the scikit-learn and keras libraries (optionally PyTorch). Take the options below as samples.

Select the features used in training and, if necessary, preprocess them. Divide the sample into training and test. In this work, it is necessary to study the operation of architectures and learning algorithms with different values of the structure and learning parameters (hyperparameters) of networks and choose the best values of the latter.

Write a short report on the work, including the program with comments, the quality values of the models. Choose the best model.

For your version of the regressor, you need to look at the last digit of your grade book (or student ID) number and make the following adjustments:

- if the last digit is 0 or 5: dataset – Forest fires (https://archive.ics.uci.edu/ml/datasets/Forest+Fires), predicted value – fire area (Area);
- if the last digit is 1 or 6: dataset - Wine Quality (https://archive.ics.uci.edu/ml/datasets/Wine+Quality) predicted value - quality (Quality), for a dataset with red wine, winequality-red .csv
- if the last digit is 2 or 7: dataset – Wine Quality (https://archive.ics.uci.edu/ml/datasets/Wine+Quality) predicted value – quality (Quality), for dataset with white wine, winequality-white .csv
- if the last digit is 3 or 8: dataset – Bicycle rental (https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset), predicted value – number of bike rentals per day (Area), day dataset .csv
- if the last digit is 4 or 9: dataset - Bicycle rental (https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset), predicted value - number of bike rentals per hour (Area), hour dataset .csv

Sample for the Covertype Data Set classifier (https://archive.ics.uci.edu/ml/datasets/Covertype).

To do this, you need to look at the last digit of your grade book number (or student ID) and make the following adjustments:

The class label is Cover_Type. Since it is necessary to create binary classifiers and there are 7 possible classes, it is first necessary to change the value of the Cover_Type label.

To do this, you need to look at the last digit of the number of your record book (or student card) and make the following adjustments: if the last digit is 0 or 5: replace label 0 with class A, labels 1, 2, 3, 4 replace with class B;

if the last digit is 1 or 6: replace label 1 with class A, labels 0, 2, 3, 4 replace with class B;

if the last digit is 2 or 7: replace label 2 with class A, labels 0, 1, 3, 4 replace with class B;

if the last digit is 3 or 8: replace label 3 with class A, labels 0, 1, 2, 4 replace with class B;

if the last digit is 4 or 9: replace label 4 with class A, labels 0, 1, 2, 3 replace with class B.

tevy models»

Laboratory work №2 "Developing and Deploying Applications with Deep Neural Network Models"

The purpose of the work is to develop a web application in Python using the trained neural network obtained in the previous work, presenting the user with a service for entering the studied data and outputting the result of data classification. The results of the work should be included in the report.

Laboratory work №3 "Tools for increasing the efficiency of designing and training deep neural networks"

The purpose of the work is to explore in Python and R languages methods for tuning parameters and hyperparameters of neural networks using different optimizers, enumeration of architectures for solving problems of multiclass, binary classification, as well as regression for samples provided by the teacher. The results of the work should be included in the report.

Laboratory work No. 4 "Research of texts with deep neural network models"

The purpose of the work is to write a Python program that performs text data classification based on the Keras library using NLTK tools: tokenization and lemmatization, vector text transformation for a sample received from a teacher. The results of the work should be included in the report.

Laboratory work No. 5 "Segmentation and classification of objects in images using deep neural networks"

The purpose of the work is to write a Python program that uses (optionally) either the PyTorch or Keras libraries that perform two tasks: 1) image classification, 2) localization of significant areas in the image, using convolutional neural networks. Samples of images received from the teacher. In the work, you should use the data augmentation procedure and test with the network obtained by transfer learning. The results of the work should be included in the report.

Task description

Write a Python program that trains a binary image classifier based on convolutional neural networks.

To do this, you need to prepare two folders with images belonging to two classes (images can be downloaded from the Internet). Perform data augmentation, for example, by rotating or scaling. Divide the sample into training and test.

Build a set of convolutional neural networks that differ in the number of layers, the alternation of layers, the presence of thinning, and train them on the prepared set of images.

Choose a network that allows you to classify images with sufficient quality, and does not have redundancy.

Write a short report on the work, including the program with comments, model tuning graphs and model metrics values (accuracy, completeness).

Laboratory work №6 "Implementing Chatbot Systems Using Deep Learning Tools"

The purpose of the work is to write a chatbot program in Python that performs the task of human interaction with the bot, using a neural network classifier of statements on one of the topics proposed by the teacher. The results of the work should be included in the report.

Task description

Write a Python program that interacts with the user, answering his questions about some discipline taught at the university.

1. Prepare a set of possible topics for questions on the discipline (at least 9 topics: what is studied, what tools are used, how is the test / exam, etc.)
2. For each topic, ask at least three questions and at least three possible answers.
3. Vectorize questions using any of the Text-To-Vec approaches.
4. Train the question classifier.
5. Build a system of interaction between a chatbot and a person. When asking a question that is close in meaning again, the chatbot will have to indicate that it has already answered this question and answer with a synonymous answer that was not used before.

Write a short conclusion about the most interesting results and errors that occurred during the work.

Laboratory work №7 Image and Time Sequence Generation

The purpose of the work is to write programs in Python that build and train a model for generating time sequences and images (samples of images and time sequences used for tuning were received from the teacher), it is required to select the GAN architectures used to create the result and evaluate its quality, visualize process of training models and lead quality metrics. The results of the work should be included in the report.

Laboratory work No. 8 "Research of the procedure for retraining neural networks"

The purpose of the work is to write a Python program that builds and trains a classifier model (a sample of images was received from a teacher), as well as to conduct experiments on retraining a neural network by adding new training data during the operation of the model. Show on which data sets the model retains the knowledge gained earlier, and on which it begins to lose memory of dependencies. The results of the work should be included in the report.

Topics of individual projects:

To strengthen the studied material, it is planned to carry out an individual project within hours of independent work. The project can be completed both individually and in a mini-group (2-3 people), provided that the amount of work will also be increased. At the end of the semester, a mini-presentation on the results of the work is presented for each project.

The theme of the individual project is related to the theme of the master's degree student. The purpose of the work is the use of deep learning methods in their scientific work.

Topics of surveys in the classroom:

Linked to the material of previous lectures, as well as the personal experience of students. Students can offer options for solving the problem set by the teacher, as well as solution tools.

Sample questions:

1. Which of the following neural network models is best suited for predicting time sequences?

a) Single-Layer Perceptron	b) CNN
c) LSTM	d) Multi-layer Perceptron

2. What is the name of several examples from the training set that are used to simultaneously calculate the gradient and weights of the network?

3. Why do models based on convolutional neural networks show the best performance in classifying objects in images compared to other models?

a)They are highly optimized for handling vectors with numeric rather than categorical features.	b)They have a wide range of feature space transformation tools that can be varied by the developer in the model.
c)They take into account the correlation of adjacent components of the vector	d)They use a significantly larger number of adjustable parameters compared to other models

3.2. Typical tasks for conducting intermediate certification in the discipline.

The student makes a presentation and also demonstrates the program code. Questions on the results can be asked by all students of the group, not only by the teacher.

4. Methodological materials that determine the procedures for evaluating training outcomes

4.1. Methodological materials for assessing the current control of progress in the discipline.

To assess the current performance, a rating system of assessment is used:

Table - Scoring for control elements

Elements of learning activity	Maximum score since the beginning of the semester	Assessed competence
Preparing for labs and defending a lab report	15*4=60	UK-1,OPK-3, PC-6
Protection of individual projects	40	UK-1,OPK-3, PC-6

4.2. Methodological materials for conducting intermediate certification in the discipline.

The mark for the exam is set on the basis of completed laboratory work and the presentation and defense of an individual project. A rating system is used to assess the current progress of students.

Table - Scoring for control elements

Elements of learning activity	Maximum score since the beginning of the semester	Assessed competence
Preparing for labs and	15*4=60	UK-1,OPK-3, PC-6

defending a lab report		
Protection of individual projects	40	UK-1,OPK-3, PC-6
Exam		

Recalculation of scores into intermediate performance assessments

Points on the checkpoint date	Mark
≥ 90% from the maximum points	5 (credited)
from 70% till 89% from the maximum points	4 (credited)
from 60% till 69% from the maximum points	3 (credited)
< 60% from the maximum points	2 (uncredited)