Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Механико-математический факультет

УТВЕРЖДАЮ:
Декан
_____Л. В.Гензе
«30» ____00____ 2022 г.

Рабочая программа дисциплины

Методы машинного обучения с использованием Python

по направлению подготовки

01.04.01 Математика

Направленность (профиль) подготовки :
Математический анализ и моделирование (Mathematical Analysis and Modelling)

Форма обучения
Очная

Квалификация
Магистр

Год приема
2022

Код дисциплины в учебном плане: Б1.В.3.ДВ.02.02

СОГЛАСОВАНО:
Руководитель ОП
_____А.В.Старченко

Председатель УМК
_____Е.А.Тарасов

Томск – 2022

Programm arranged					
Senior lecturer					
_____ Strebkova Ekaterina Aleksandrovna					
	(sign)				
Reviewer					
Professor, Adv. Doc. of Phys. and Math.					
_____ Starchenko Aleksandr Vasilyevich					
	(sign)				
The working program of the discipline "Department of Computational Mathematics and Computer Modeling" was developed in accordance with the SUOS of NR TSU: *Independently established educational standard of NR TSU in the field of training 01.04.01 Mathematical analysis and modeling (Approved by the Scientific Council of NR TSU, Protocol No. 03 dated 28.03.2019)*					

The working program was approved at the meeting of the MMF UMK

Protocol No. 1 of January 30, 2020

1. The purpose of mastering the discipline

The main purpose of the discipline is the formation of students' theoretical knowledge and practical skills on the basics of machine learning implemented with Python, mastering the tools, models and methods of machine learning, as well as acquiring skills in data research and the development of mathematical models, methods and algorithms for data analysis.

Objectives:

- form theoretical knowledge on the basics of machine learning for the construction of formal mathematical models and interpretation of modeling results;

- develop skills and abilities to use Python libraries to develop machine learning algorithms;

- development and using of mathematical and information tools based on machine learning methods in scientific and practical activities.

2. The place of discipline in the structure MEP

The discipline belongs to the variable disciplines of the professional cycle of Block 1 "Disciplines/modules".

Prerequisites of the discipline – programming, mathematical logic, discrete mathematics.

Post-requirements of the discipline: research work, execution and defense of the final qualifying work.

3. Competencies and learning outcomes formed as a result of mastering the discipline

Table 1

Competence	**Indicator ща Competence**	**Code and name of learning outcomes** (planned results of training in the discipline, characterizing the stages of competence formation)
PC 1: - The student is able to apply computer science to solve the problems of the main branches of mathematics: algebra, geometry, functional and complex analysis, probability theory and mathematical statistics, computational mathematics.	IPC 1.1: - The student understands how theory and practice in various branches of mathematics and computer science affect each other. IPC 1.2: - The student uses a wide range of possibilities of computer science and computational mathematics to solve problems of the main sections of mathematics	LO-1.1: The student is able to solve mathematical modeling problems using machine learning methods and implement its with Python. LO-1.2: The student has methods for different classes of problems solved using machine learning algorithms, selecting and combining them when analyzing and solving specific theoretical and applied problems. LO-1.3: The student is able to use Python libraries necessary for solving machine learning tasks.
PC 2: - The student is able to use modern computer technology to solve fundamental and applied problems of mathematics and mechanics.	IPC 2.1: - The student uses modern computing systems to solve professional tasks. IPC 2.2: - The student has the basics of constructing machine	LO-2.1: The student is able to formulate the problem statement in terms of mathematical modeling methods, choose the optimal solution method with justification of the choice made. LO-2.2: The student has the skills to build and verify

	learning methods for different classes of problems for solving problems of mathematics and mechanics	the quality of machine learning models in Python and interpret the results in terms of the applied field for solving problems of mathematics and mechanics in order to gain new knowledge and conclusions.

4. Structure and content of the discipline

4.1. The structure and labor intensity of the types of academic work in the discipline

The total labor intensity of the discipline is 4 credit units, 144 hours.

Table 2

Type of educational work	Labor intensity in academic hours	
Total labor intensity	1-st semetr	total
Contact work (CW):	35,85	35,85
Lectures(L):	16	16
Practical classes (P)	16	16
Group consultations	1,6	1,6
Intermediate certification	0	0
Independent work of the student:	92,4	92,4
Study of the main libraries for implementing machine learning methods in Python	24	24
Learning machine learning theory	32,65	32,65
Preparation of reports on laboratory work.	20	20
Preparation and passing of the credit	15,75	15,75
Type of intermediate certification	**Credit**	**2,25**

4.2. The content and labor intensity of the discipline sections

Table 3

Class code	The name of sections and topics and their content	Type of educational work, classes, control	Semester	hours in electronic form	Total (hr.)	Literature	Learning result(s) code(s)
	Sction 1. Lecture part						
1.1	Introduction to machine learning. The main stages of solving the machine learning problem. Examples of applied tasks. Types of training: supervised, unsupervised, with reinforcement. The main types of tasks. The main problems of machine learning: insufficient training sample size, data gaps, overfitting/underfitting.	L	1	0	2		LO -1.2, LO -2.1
1.2	Solving the regression problem. The least squares method. Error measurement in regression tasks. Linear regression. The solution to the problem of overfitting: L1- regularization (Lasso), L2- regularization (ridge regression), elastic net (L1+L2). Configuring hyperparameters of the algorithm using cross-validation.	L	1	0	2		LO -1.2, LO -2.1
1.3	Solving the classification problem. Linear classification model. Logistic regression as a binary classifier. Loss function (classification errors). Confusion matrix (classification error matrix). Classification quality metrics: accuracy (percentage of correct answers), precision (accuracy), recall (completeness), F1-measure. AUC-ROC is the area under the error curve. Metric classification is the nearest neighbor method (kNN).	L	1	0	2		LO -1.2, LO -2.1
1.4	Solving the clustering problem. Clustering tasks and approaches. Conditions for clustering tasks. Clustering algorithms: K-Means, EM algorithm, Agglomerative clustering, DBSCAN. Comparison of algorithms. Quality assessment: internal and external assess-	L	1	0	2		LO -1.2, LO -2.1

	ments.						
1.5	Tree Models: Decision Trees, Random Forest	L	1	0	2		LO -1.2, LO -2.1
1.6	Ensemble methods. Random forest as an example of bagging. AdaBoost . Gradient boosting in regression and classification problems. Stacking.	L	1	0	2		LO -1.2, LO -2.1
1.7	Neural networks and deep learning. Introduction to neural networks. The main components of the neural network (activation functions, dropout, etc.). Gradient descent. Reverse propagation of the error. Ready-made architectures.	L	1	0	4		LO -1.2, LO -2.1
	Section 2. Practical part						
2.1	The main libraries needed for machine learning in Python. NumPy library for optimized calculations on arrays of data. Operations on data in the Pandas library. Visualization using the Matplotlib library. Introduction to the Scikit-Learn machine learning library. Complition of individual task No. 1 according to section 2.1.	P	1	0	2		LO -1.3
2.2	Working with data. Data types, work with gaps, initial processing and visualization. Feature engineering, Feature Extraction, Feature transformations: non-numeric data coding, normalization and calibration. Search for outliers. Completion of individual task No. 1 according to section 2.2.	P	1	0	2		LO -1.3
2.3	Writing and debugging a program on an individual task No. 1. Preparing of report.	IW	1	0	4		
2.4	Solution of the regression problem by example. Completion of individual task No. 2 according to section 2.4.	P	1	0	2		LO -1.1, LO-1.2, LO -2.2
2.5	Solution of the classification problem by example. Completion of individual task No. 2 according to section 2.5.	P	1	0	2		LO -1.1, LO -1.2, LO -2.2
2.6	Writing and debugging a program on an individual task No. 2. Preparing of report.	IW	1	0	4		
2.7	Solution of the clustering problem by example. Completion of individual task No. 3 according to section 2.7.	P	1	0	2		LO -1.1, LO -1.2, LO -2.2
2.8	Writing and debugging a program on an individual	IW	1	0	2		

	task No. 3. Preparing of report.						
2.9	The use of tree structures for solving problems of classification and regression. Completion of individual task No. 4 according to section 2.9.	P	1	0	2		LO -1.1, LO -1.2, LO -2.2
2.10	Writing and debugging a program on an individual task No. 4. Preparing of report.	IW	1	0	2		
2.11	Using an Ensemble Approach to Solve Machine Learning Problems by Example. Completion of individual task No. 5 according to section 2.11.	P	1	0	2		LO -1.1, LO -1.2, LO -2.1, LO -2.2
2.12	The practice of building a neural network to solve the problem of machine learning by example. Completion of individual task No. 5 according to section 2.12.	P	1	0	2		LO -1.1, LO -1.2, LO -2.1, LO -2.2
2.13	Writing and debugging a program on an individual task No. 5. Preparing of report.	IW	1	0	8		
	Section 3. Consultation						
3.1	Consultations	CW	1	0	1,6		
	Section 4. Intermediate certification						
4.1	Admission of the credit in the discipline	CW	1	0	2,25		

5. Educational technologies, educational, methodological and informational support for the development of the discipline

During the implementation of the discipline, classical educational technologies are used - practical exercises; independent study of the recommended literature and the gradual implementation of an individual task; intermediate certification in the form of checking individual tasks.

Independent work includes: the theoretical development of the lecture course, the practical implementation of assignments and individual assignments, preparation for a test with an assessment. To perform independent work, access to the information resources of the course is provided:

- lecture materials;
- a list of questions for self-testing knowledge and preparing for the exam.
- a list of references, including textbooks and books on the subjects studied in the course.

All laboratory work and individual assignments are selected in such a way as to maximally stimulate the psychological attitude of mathematics students to form a connection between mathematical theory and its practical application. The report for each laboratory work includes a theoretical part, a completed practical task and an analysis of the results.

5.1. Literature and educational and methodological support

a) List of basic educational literature.

1. Lubanovic B. Simple Python. Modern programming style. 2nd ed. – St. Petersburg: St. Petersburg, 2021. – 592 p.
2. Muller A., Guido S. introduction to machine learning using Python. - St. Petersburg: Peter, 2016. – 392 p.
3. Gruc J. Data Science from scratch. 2nd ed., reprint and additional – St. Petersburg: BHV-Petersburg, 2021 – 416s.
4. Burkov A. Machine learning without unnecessary words. — St. Petersburg: Peter, 2020. — 192 p.
5. Nikolenko S., Kadurin A., Arkhangelskaya E. Deep learning. — St. Petersburg: Peter, 2018. — 480 p.

b) List of additional educational literature.

1. Vander P. J. Python for complex tasks: Data science and machine learning. St. Petersburg: St. Petersburg, 2020. – 576c.
2. Albon C. Machine learning with Python. Cookbook. St. Petersburg: BHV, 2020 – 384s.
3. Watt J. Machine learning refined: Foundations, algorithms, and applicatios. St. Petersburg: BHV-Petersburg, 2022 – 640s.
4. Trask A. Grokking deep learning. – St. Petersburg: Peter, 2020. – 352 p.

5.2. Databases and information and reference systems, including foreign ones

- https://e.lanbook.com/book/131723
- https://e.lanbook.com/reader/book/160142/#209

5.3. List of licensed and software

1) OS Windows 7 or Windows 10 https://www.microsoft.com/ru-ru/software-download/windows10
2) Anaconda (distribution of the Python) https://www.anaconda.com/products/distribution

5.4. Equipment and technical means of training

The classrooms of the MMF educational and computing laboratory are used for laboratory work and independent work. When performing individual tasks, independent and laboratory work, free and licensed software is used:

- Microsoft Office 2010 (reports preparation);
- Jupyter Notebook (it is supplied in the Anaconda distribution package for Python).

6. Methodological guidelines for students on the development of the discipline

For the successful development of the material, students need to attend classes, and during independent work use the main and additional literature, databases and information and reference systems that are presented in the list of references. The independent work of students consists in repeating the material from practical classes and self-study of additional questions, a deeper analysis of the topic with the help of literature.

7. Teaching staff implementing the discipline

Strebkova Ekaterina Aleksandrovna, senior lecturer.

8. Teaching language

Russian, English

Update sheet

Work program discipline «Methods of Machine Learning with Python»

Program Fundamentals of Research in Mathematics and Computer Science

Direction 01.04.01 Mathematical analysis and modeling

Chapter (subsection) in which changes are made	Reasons for change	Brief description of the changes	Date and number of the protocol of the meeting of the educational and methodological commission