

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Ястребов Олег Александрович  
Должность: Ректор  
Дата подписания: 30.05.2024 13:48:58  
Уникальный программный ключ:  
ca953a0120d891083f939673078ef1a989dae18a

**Federal State Autonomous Educational Institution of Higher Education  
"Russian Peoples' Friendship University named after Patrice Lumumba"**

**Engineering Academy**

(name of the main educational unit (POU) - developer of the EP HE)

**DISCIPLINE WORK PROGRAM**

**MACHINE LEARNING AND BIG DATA ANALYSIS**

(name of discipline/module)

**Recommended by MSSN for the following areas of training/specialty:**

**01.04.02 APPLIED MATHEMATICS AND INFORMATION SCIENCE**

(code and name of the area of training/specialty)

**The discipline is mastered as part of the implementation of the main professional educational program of higher education (OP HE):**

**BALLISTIC DESIGN OF SPACE COMPLEXES AND SYSTEMS**

(name (profile/specialization) EP HE)

## 1. GOAL OF DISCIPLINE MASTERING

The discipline “Machine Learning and Big Data Mining” is included in the master’s program “Ballistic Design of Space Complexes and Systems” in the direction 01.04.02 “Applied Mathematics and Computer Science” and is studied in the 1st semester of the 1st year. The discipline is implemented by the Department of Mechanics and Control Processes. The discipline consists of 9 sections and 27 topics and is aimed at studying the fundamental principles of clustering and its basic techniques, classification and its basic techniques, frequent itemset mining and association rules, feature selection and dimensionality reduction, outlier detection, recommender systems and algorithms, ensemble clustering and classification, multimodal relational clustering, artificial neural methods and stochastic optimization, elements of statistical learning; analysis of the basic methods for solving typical problems and familiarization with the scope of their application in professional activities.

The purpose of mastering the discipline is to develop fundamental knowledge and skills in applying problem solving methods necessary for professional activities, to increase the overall level of literacy of students in the discipline Machine Learning and Big Data Mining, to familiarize them with a new rapidly evolving field and provide practical knowledge experience in analysis of real world data.

## 2. REQUIREMENTS FOR THE RESULTS OF MASTERING THE DISCIPLINE

Mastering the discipline “Machine learning and big data analysis” is aimed at developing the following competencies (parts of competencies) in students:

*Table 2.1. List of competencies formed in students when mastering the discipline (results of mastering the discipline)*

<b>Cipher</b>	<b>Competence</b>	<b>Indicators of Competency Achievement (within this discipline)</b>
UK-1	Able to critically analyze problem situations based on a systematic approach and develop an action strategy	UK-1.1 Analyzes the task, highlighting its basic components;; UK-1.2 Identifies and ranks the information required to solve the task;; UK-1.3 Searches for information to solve a given problem using various types of requests;; UK-1.4 Offers options for solving a problem, analyzes the possible consequences of their use;;
UK-7	Able to search for the necessary sources of information and data, perceive, analyze, remember and transmit information using digital means, as well as using algorithms when working with data received from various sources in order to effectively use the information received to solve problems; evaluate information, its reliability, build logical conclusions based on incoming information and data	UK-7.1 Searches for the necessary sources of information and data, perceives, analyzes, remembers and transmits information using digital means, as well as using algorithms when working with data received from various sources in order to effectively use the received information to solve problems;; UK-7.2 Evaluates information, its reliability, builds logical conclusions based on incoming information and data.;
PC-1	Able to formulate goals and objectives of scientific research in the field of applied mathematics and computer science, computer technology and modern programming technologies, select methods and means for solving problems	PC-1.1 Has fundamental knowledge acquired in the field of mathematical and (or) natural sciences, programming and information technology;; PC-1.2 Can find, formulate and solve standard problems in their own research activities in the field of applied mathematics and computer science, computer technology and modern programming technologies;;

<b>Cipher</b>	<b>Competence</b>	<b>Indicators of Competency Achievement (within this discipline)</b>
		PC-1.3 Has practical experience in research activities in the field of applied mathematics and computer science, computer technology and modern programming technologies.;
PC-2	Able to apply modern theoretical and experimental methods for developing mathematical models of studied objects and processes related to professional activities in the field of training and participate in their implementation in the form of software products	PC-2.1 Knows modern theoretical and experimental methods for developing mathematical models, innovative design tools and elements of architectural solutions of information systems;; PC-2.2 Can develop and implement algorithms for mathematical models based on languages and application packages for modeling;; PC-2.3 Has practical experience in developing options for implementing information systems using innovative tools.;

### 3. PLACE OF DISCIPLINE IN THE STRUCTURE OF HE EP

Discipline "Machine learning and big data analysis" refers to the part formed by the participants in educational relations of block 1 "Disciplines (modules)" of the educational program of higher education.

As part of the educational program of higher education, students also master other disciplines and/or practices that contribute to achieving the planned results of mastering the discipline "Machine Learning and Big Data Analysis".

*Table 3.1. List of components of EP HE that contribute to achieving the planned results of mastering the discipline*

<b>Cipher</b>	<b>Name of competency</b>	<b>Previous disciplines/modules, practices*</b>	<b>Subsequent disciplines/modules, practices*</b>
UK-7	Able to search for the necessary sources of information and data, perceive, analyze, remember and transmit information using digital means, as well as using algorithms when working with data received from various sources in order to effectively use the information received to solve problems; evaluate information, its reliability, build logical conclusions based on incoming information and data		Pre-Graduation Internship in Industry; Practical Training and Research in Dynamics and Control of Space Systems (online from RUDN Mission Control Center) / Research work; Technological Training; Advanced Methods of Remote Sensing and Geoinformation Systems;
UK-1	Able to critically analyze problem situations based on a systematic approach and develop an action strategy		Advanced Methods of Remote Sensing and Geoinformation Systems; Structures & Materials Modeling; System Design; Dynamics and Control of Space Systems; Project "Drone Systems Engineering. Part 1"; Practical Training in Receiving Remote Sensing Data from Satellites and its

Cipher	Name of competency	Previous disciplines/modules, practices*	Subsequent disciplines/modules, practices*
			Interpretation (online from RUDN Mission Control Center) / Research; Practical Training and Research in Dynamics and Control of Space Systems (online from RUDN Mission Control Center) / Research work; Technological Training; Pre-Graduation Internship in Industry;
PC-1	Able to formulate goals and objectives of scientific research in the field of applied mathematics and computer science, computer technology and modern programming technologies, select methods and means for solving problems		Pre-Graduation Internship in Industry; Practical Training in Receiving Remote Sensing Data from Satellites and its Interpretation (online from RUDN Mission Control Center) / Research; Practical Training and Research in Dynamics and Control of Space Systems (online from RUDN Mission Control Center) / Research work; Technological Training; Advanced Methods of Remote Sensing and Geoinformation Systems; System Design; Dynamics and Control of Space Systems;
PC-2	Able to apply modern theoretical and experimental methods for developing mathematical models of studied objects and processes related to professional activities in the field of training and participate in their implementation in the form of software products		Advanced Methods of Remote Sensing and Geoinformation Systems; System Design; Project "Drone Systems Engineering. Part 1"; Pre-Graduation Internship in Industry; Practical Training in Receiving Remote Sensing Data from Satellites and its Interpretation (online from RUDN Mission Control Center) / Research; Practical Training and Research in Dynamics and Control of Space Systems (online from RUDN Mission Control Center) / Research work; Technological Training;

\* - to be filled out in accordance with the competency matrix and SUP EP VO

\*\* - elective disciplines/practices

#### 4. SCOPE OF DISCIPLINE AND TYPES OF STUDY WORK

The total labor intensity of the “Machine Learning and Big Data Mining” discipline is “4” credit units.

*Table 4.1. Types of educational work by periods of mastering the educational program of higher education for full-time study.*

Type of educational work	TOTAL,ac.ch.		Semester(s)
			1
<i>Contact work, ac.ch.</i>	20		20
Lectures (LK)	10		10
Laboratory work (LR)	10		10
Practical/seminar sessions (SZ)	0		0
<i>Independent work of students, ac.ch.</i>	88		88
<i>Control (exam/test with assessment), academic degree.</i>	36		36
<b>Total labor intensity of the discipline</b>	<b>ac.ch.</b>	<b>144</b>	<b>144</b>
	<b>credit units</b>	<b>4</b>	<b>4</b>

## 5. CONTENT OF DISCIPLINE

Table 5.1. Contents of the discipline (module) by type of academic work

Section number	Name of the discipline section	Contents of the section (topic)		Type of educational work*
Section 1	Introduction to Machine Learning and Data Mining	1.1	Introduction to modern data analysis	LK, LR
		1.2	Machine learning. Data Mining and Knowledge Discovery in Data Bases	LK, LR
Section 2	Clustering and its basic techniques	2.1	The task of clusterization	LK, LR
		2.2	K-means and its modifications (k-medoids and fuzzy cmeans clustering)	LK, LR
		2.3	Density-based methods: DB-scan and Mean Shift	LK, LR
		2.4	Hierarchical clustering	LK, LR
		2.5	Criteria of quality	LK, LR
Section 3	Classification and its basic techniques	3.1	The task of classification	LK, LR
		3.2	1-Rules. K-Nearest Neighbors approach	LK, LR
		3.3	Naïve Bayes. Decision Trees. Logistic Regression	LK, LR
		3.4	Quality assessment: precision, recall, F - measure, loss-function, confusion-matrix, cross- validation and learning curves (ROC, lift etc.)	LK, LR
		3.5	Multi-class and multi-label classification	LK, LR
Section 4	Frequent Itemset Mining and Association Rules	4.1	Frequent itemsets. Apriori and FP-growth algorithms	LK, LR
		4.2	Association rules. Interestingness measures: support and confidence. Closed itemsets	LK, LR
		4.3	Connection with Lattice Theory and Formal Concept Analysis. Applications	LK, LR
Section 5	Feature Selection and Dimensionality Reduction. Outlier detection	5.1	Feature selection versus feature extraction and generation	LK, LR
		5.2	Singular Value Decomposition, Latent Semantic Analysis and Principal Component Analysis. Boolean Matrix Factorization	LK, LR
		5.3	Outlier and novelty detection techniques	LK, LR
Section 6	Recommender Systems and Algorithms	6.1	Collaborative filtering. User-based and item-based methods. Slope one	LK, LR
		6.2	Association rules based and bicluster-based techniques. Quality: MAE, precision and recall assessment	LK, LR
		6.3	SVD-based approaches: pureSVD, SVD++ and time-SVD. Factorization machines	LK, LR
Section 7	Ensemble Clustering and Classification	7.1	Ensemble methods of clusterization for k-means partitions' aggregation	LK, LR
		7.2	Ensemble methods of classification: Bagging, Boosting, and Random Forest	LK, LR
Section 8	Multimodal relational clustering	8.1	Biclustering. Spectral co-clustering. Triclustering	LK, LR
		8.2	Two-mode networks. Folksonomies and resource-sharing systems. Multimodal approaches	LK, LR
		8.3	Applications: Community detection in Socail Network Analysis and gene expression analysis	LK, LR
Section 9	Artificial Neural Methods and Stochastic Optimization. Elements of Statistical Learning	9.1	Artificial Neural Networks. Basic ideas of Deep Learning. (Stochastic) gradient descent. Statistical (Bayesian) view on Machine learning	LK, LR

\* - to be filled out only for full-time education: LC – lectures; LR – laboratory work; SZ – practical/seminar classes.

## 6. MATERIAL AND TECHNICAL SUPPORT OF DISCIPLINE

Table 6.1. Material and technical support of the discipline

Audience type	Auditorium equipment	Specialized educational/laboratory equipment, software and materials for mastering the discipline (if necessary)
Lecture	An auditorium for conducting lecture-type classes, equipped with a set of specialized furniture; board (screen) and technical means of multimedia presentations.	
Computer class	A computer class for conducting classes, group and individual consultations, ongoing monitoring and intermediate certification, equipped with personal computers ([Parameter] pcs.), a whiteboard (screen) and technical means for multimedia presentations.	
For independent work	An auditorium for independent work by students (can be used for seminars and consultations), equipped with a set of specialized furniture and computers with access to EIOS.	

\* - the audience for independent work of students is MANDATORY!

## 7. EDUCATIONAL, METHODOLOGICAL AND INFORMATIONAL SUPPORT OF DISCIPLINE

### *Main literature:*

1. Han, J., Kamber, M., Pei, J. Data Mining: Concepts and Techniques, Third Edition. – Morgan Kaufmann Publishers, 2011. – 740 pp.

### *Additional literature:*

1. Hall, M., Witten, Ian H., Frank, E. Data Mining: practical machine learning tools and techniques. – 2011. – 664 pp

### *Resources of the information and telecommunications network “Internet”:*

1. EBS of RUDN University and third-party EBS, to which university students have access based on concluded agreements

- Electronic library system of RUDN - EBS RUDN

<http://lib.rudn.ru/MegaPro/Web>

- EBS “University Library Online” <http://www.biblioclub.ru>

- EBS Yurayt <http://www.biblio-online.ru>

- EBS “Student Consultant” [www.studentlibrary.ru](http://www.studentlibrary.ru)

- EBS “Trinity Bridge”

2. Databases and search engines

- electronic fund of legal and regulatory technical documentation

<http://docs.cntd.ru/>

- Yandex search engine <https://www.yandex.ru/>

- Google search engine <https://www.google.ru/>

- SCOPUS abstract database <http://www.elsevierscience.ru/products/scopus/>

*Educational and methodological materials for students’ independent work when mastering a discipline/module\*:*

1. A course of lectures on the discipline “Machine learning and big data analysis.”

\* - all educational and methodological materials for students' independent work are posted in accordance with the current procedure on the discipline page in TUIS!

## **8. ASSESSMENT MATERIALS AND POINT-RATING SYSTEM FOR ASSESSING THE LEVEL OF COMPETENCIES FOR A DISCIPLINE**

Evaluation materials and point-rating system\* for assessing the level of development of competencies (parts of competencies) based on the results of mastering the discipline“Machine learning and big data analysis” are presented in the Appendix to this Work Program of the discipline.

\* - OM and BRS are formed on the basis of the requirements of the relevant local regulatory act of RUDN University.



**DEVELOPER:**

Assistant professor

*Position, PBU*

*Signature*

Saltykova Olga

Alexandrovna

*Last name I.O.*

**HEAD OF BUP:**

Head of the department

*Position PBU*

*Signature*

Razumny Yuri Nikolaevich

*Last name I.O.*

**HEAD OF OP VO:**

Professor

*Position, PBU*

*Signature*

Razumny Yuri Nikolaevich

*Last name I.O.*