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

Academy of Engineering

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

COURSE SYLLABUS

MACHINE LEARNING AND BIG DATA ANALYSIS

(name of discipline/module)

Recommended by the Didactic Council for the Education Program:

27.04.04 CONTROL IN TECHNICAL SYSTEMS

(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:

DATA SCIENCE AND SPACE ENGINEERING

(name (profile/specialization) EP HE)

2024

1. GOAL OF DISCIPLINE MASTERING

The discipline "Machine Learning and Big Data Mining" is included in the master's program "Data Science and Space Engineering" in the direction 27.04.04 "Control in Technical Systems" 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)

Cipher	Competence	Indicators of Competency Achievement (within this discipline)
GPC-5	Able to conduct patent research, determine forms and methods of legal protection and protection of rights to the results of intellectual activity, manage rights to them to solve problems in the development of science, technology and technology	GPC-5.1 Knows methods and approaches to conducting patent research, forms and methods of legal protection and protection of rights to the results of intellectual activity;; GPC-5.2 Able to manage rights to the results of intellectual activity to solve problems in the field of development of science, technology and engineering;; GPC-5.3 Knows methods and approaches to conducting patent research, knows methods of legal protection and protection of rights to the results of intellectual activity.;
GPC-6	Capable of collecting and analyzing scientific and technical information, summarizing domestic and foreign experience in the field of automation and control equipment	GPC-6.1 Knows the basic methods of collecting and analyzing scientific and technical information;; GPC-6.2 Able to analyze and generalize domestic and foreign experience in the field of automation and control equipment;; GPC-6.3 Knows methods of collecting and analyzing scientific and technical information, and can also generalize domestic and foreign experience in the professional industry.;

3. PLACE OF DISCIPLINE IN THE STRUCTURE OF HE EP

Discipline "Machine learning and big data analysis" refers to the mandatory part 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

Cipher	Name of competency	Previous disciplines/modules, practices*	Subsequent disciplines/modules, practices*
GPC-5	Able to conduct patent research, determine forms and methods of legal protection and protection of rights to the results of intellectual activity, manage rights to them to solve problems in the development of science, technology and technology		Research work / Scientific research work; Undergraduate practice / Pre-graduate practice; Dynamics and Control of Space Systems;
GPC-6	Capable of collecting and analyzing scientific and technical information, summarizing domestic and foreign experience in the field of automation and control equipment		Advanced Methods of Earth Remote Sensing; Research work / Scientific research work; Undergraduate practice / Pre-graduate practice;

* - 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 "5" 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
Contact work, ac.ch.	*34*		*34*
Lectures (LC)	17		17
Laboratory work (LR)	17		17
Practical/seminar sessions (SZ)	0		0
Independent work of students, ac.ch.	*110*		*110*
Control (exam/test with assessment), academic degree.	*36*		*36*
Total labor intensity of the discipline	**ac.ch.**	*180*	*180*
	credit units	*5*	*5*

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 University - EBS RUDN Universityhttp://lib.rudn.ru/MegaPro/Web
        - EBS "University Library Online"http://www.biblioclub.ru
        - EBS Lawhttp://www.biblio-online.ru
        - EBS "Student Consultant"www.studentlibrary.ru
        - EBS "Trinity Bridge"
    2. Databases and search engines
        - electronic fund of legal and regulatory technical documentationhttp://docs.cntd.ru/
        - Yandex search enginehttps://www.yandex.ru/
        - search systemGoogle https://www.google.ru/
        - abstract databaseSCOPUS 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:

		Saltykova Olga
Assistant professor		Alexandrovna
Position	*Signature*	*Last name I.O.*

HEAD OF DEPARTMENT:

Head of the department		Razumny Yuri Nikolaevich
Position	*Signature*	*Last name I.O.*

HEAD OF EP HE:

Professor		Razumny Yuri Nikolaevich
Position	*Signature*	*Last name I.O.*