

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 27.05.2026 14:42:39

Уникальный программный ключ:

ca953a0120d891083f939673078ef1a989dae18a

Федеральное государственное автономное образовательное учреждение высшего образования

«Российский университет дружбы народов имени Патриса Лумумбы»

Инженерная академия

(наименование основного учебного подразделения (ОУП) – разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

МАШИННОЕ ОБУЧЕНИЕ И АНАЛИЗ БОЛЬШИХ ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

27.04.04 УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ, МАШИННОЕ ОБУЧЕНИЕ И КОСМИЧЕСКИЕ НАУКИ

(наименование (профиль/специализация) ОП ВО)

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Машинное обучение и анализ больших данных» входит в программу магистратуры «Искусственный интеллект, машинное обучение и космические науки» по направлению 27.04.04 «Управление в технических системах» и изучается в 1 семестре 1 курса. Дисциплину реализует Кафедра механики и процессов управления. Дисциплина состоит из 5 разделов и 20 тем и направлена на изучение основ современной техногенной цивилизации и глобальных тенденциях смены научной картины мира, типах научной рациональности, системах ценностей, на которые ориентируются ученые, анализ основных мировоззренческих и методологических проблем, возникающих в науке на современном этапе ее развития; разбор основных методов решения типовых задач и знакомство с областью их применения в профессиональной деятельности.

Целью освоения дисциплины является формирование фундаментальных знаний и навыков применения методов решения задач, необходимых для профессиональной деятельности, повышение общего уровня грамотности студентов по дисциплине история и методология науки, сформировать представления о тенденциях исторического развития науки, а также современное представление об организации научно-исследовательской деятельности в выбранной области.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Машинное обучение и анализ больших данных» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ОПК-5	Способен проводить патентные исследования, определять формы и методы правовой охраны и защиты прав на результаты интеллектуальной деятельности, распоряжаться правами на них для решения задач в развитии науки, техники и технологии	ОПК-5.1 Знает методы и подходы к проведению патентных исследований, формы и методы правовой охраны и защиты прав на результаты интеллектуальной деятельности;; ОПК-5.2 Умеет распоряжаться правами на результаты интеллектуальной деятельности для решения задач в области развития науки, техники и технологий;; ОПК-5.3 Владеет методами и подходами к проведению патентных исследований, знает методы правовой охраны и защиты прав на результаты интеллектуальной деятельности.;
ОПК-6	Способен осуществлять сбор и проводить анализ научно-технической информации, обобщать отечественный и зарубежный опыт в области средств автоматизации и управления	ОПК-6.1 Знает основные методы сбора и проведения анализа научно-технической информации;; ОПК-6.2 Умеет анализировать и обобщать отечественный и зарубежный опыт в области средств автоматизации и управления;; ОПК-6.3 Владеет методами сбора и проведения анализа научно-технической информации, а также может обобщать отечественный и зарубежный опыт в профессиональной отрасли.;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Machine Learning and Big Data Mining» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Machine Learning and Big Data Mining».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ОПК-6	Способен осуществлять сбор и проводить анализ научно-технической информации, обобщать отечественный и зарубежный опыт в области средств автоматизации и управления		Research work / Научно-исследовательская работа; Undergraduate Training; Advanced Methods of Earth Remote Sensing;
ОПК-5	Способен проводить патентные исследования, определять формы и методы правовой охраны и защиты прав на результаты интеллектуальной деятельности, распоряжаться правами на них для решения задач в развитии науки, техники и технологии		Research work / Научно-исследовательская работа; Undergraduate Training; Dynamics and Control of Space Systems;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Машинное обучение и анализ больших данных» составляет «5» зачетных единиц.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			1
Контактная работа, ак.ч	34		34
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	17		17
Практические/семинарские занятия (СЗ)	0		0
Самостоятельная работа обучающихся, ак.ч.	110		110
Контроль (экзамен/зачет с оценкой), ак.ч.	36		36
Общая трудоемкость дисциплины ак.ч.	ак.ч.	180	180
	зач.ед.	5	5

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы*

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Introduction to Machine Learning and Data Processing. Software Tools for Data Mining and Machine Learning.	1.1	Introduction to Machine Learning and Data Processing. Formulation of Main Problem Classes in Machine Learning.	Machine learning as a scientific direction studying methods for constructing algorithms capable of learning from empirical data without strictly following predefined rules. Main components: features as independent variables, target variable as dependent variable, training sample. Classification of machine learning problems: supervised learning using labeled data, unsupervised learning in the absence of a target variable, reinforcement learning for behavior formation based on interaction with the environment.	ЛК, ЛР
		1.2	Regression and Classification; Clustering, Dimensionality Reduction.	Regression task as prediction of a continuous numerical value. Classification task as assigning an object to one of discrete categories. Clustering task as grouping objects based on their similarity in the absence of class labels. Dimensionality reduction task as transforming the feature space while preserving essential information and reducing the number of variables.	ЛК, ЛР
		1.3	Text Processing; Image Processing.	Methods for processing textual data: representing textual information numerically through the bag-of-words model and TF-IDF method. Main tasks of text analysis: document classification, sentiment analysis, entity extraction. Methods for image processing: representing visual information as pixel matrices. Main tasks of computer vision: object recognition, image segmentation, edge detection.	ЛК, ЛР
Раздел 2	Regression Analysis and Data Compression.	2.1	Regression Task. Minimizing the Squared Deviation. Regression Function: Conditional Mathematical Expectation.	Regression task as predicting the numerical value of a target variable based on input features. Minimizing the squared deviation as a standard approach to assessing regression quality. Regression function as the conditional mathematical expectation of the target variable given feature values. Relationship between the quadratic loss function and conditional mathematical expectation.	ЛК, ЛР
		2.2	Linear Regression and k-Nearest Neighbors Method. Overfitting and Underfitting.	Linear regression as a model assuming a linear relationship between features and the target variable, finding coefficients by minimizing mean squared error. k-nearest neighbors method as a non-parametric regression and classification method predicting based on the nearest objects in the training sample. Overfitting as excessive adjustment of the model to training data with loss of generalization ability. Underfitting as excessive simplification of a model unable to capture patterns in the data.	ЛК, ЛР
		2.3	Decomposition of Error into Noise, Bias, and Variance.	Decomposition of total prediction error into three components: noise as irreducible error due to data randomness, bias as systematic error due to incorrect model assumptions about the data, variance as error due to model sensitivity to small fluctuations in the training sample. Bias-variance tradeoff: simple models have high bias and low variance, complex models have the opposite.	ЛК, ЛР
Раздел 3	Outlier and Anomaly Detection. Data Cleaning and Regularization Technologies.	3.1	Outlier and Anomaly Detection. What Are Outliers, Types of Outliers.	Outliers as data objects that differ significantly from the main body of observations. Types of outliers: point outliers with an anomalous value of one feature, contextual outliers anomalous under specific conditions, collective outliers as groups of objects deviating from the overall distribution. Causes of outliers: measurement errors, rare events, equipment failures.	ЛК, ЛР
		3.2	Outlier Detection Methods. Anomaly Search.	Statistical outlier detection methods based on the three-sigma rule and interquartile range. Distance-based methods: Mahalanobis distance, nearest neighbor method. Density-based methods: local outlier factor. Machine learning for anomaly search: one-class classification, isolation forest, autoencoders.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
		3.3	Sample Censoring. Outlier Screening, Outlier Removal.	Sample censoring as the process of limiting extreme values to reduce the influence of outliers. Outlier screening as complete removal of anomalous records from the dataset. Criteria for outlier removal: exceeding threshold values, inconsistency with logical constraints, impossibility of verification. Risks when removing outliers: loss of useful information when mistakenly discarding rare but significant events.	ЛК, ЛР
		3.4	Data Cleaning and Regularization Technologies. Main Types of Regularization.	Data cleaning as a set of measures to identify and correct errors, omissions, and inconsistencies in data. Regularization as a technology for combating overfitting by adding a penalty term to the loss function. Main types of regularization: L1-regularization lasso leading to sparse solutions, L2-regularization ridge uniformly reducing weights, Elastic Net as a combination of L1 and L2.	ЛК, ЛР
		3.5	Dimensionality Reduction Method. Feature Selection Methods.	Dimensionality reduction as the process of reducing the number of features while preserving essential information. Feature selection methods: filter methods evaluating the importance of each feature independently, wrapper methods searching feature subsets based on model quality, embedded methods with feature selection within the learning process. Principal component analysis as a linear transformation of original features into new uncorrelated components.	ЛК, ЛР
Раздел 4	Clustering and Classification Technologies. Neural Networks. Genetic Algorithms	4.1	Clustering and Classification Technologies. K-means. EM Algorithm.	Clustering as an unsupervised task of grouping objects based on their similarity. K-means method with iterative assignment of objects to nearest cluster centers and recalculation of centers. Choosing the number of clusters using the elbow method and silhouette coefficient. EM algorithm for probabilistic clustering assuming a mixture of Gaussian distributions. Alternating E-step for estimating object membership and M-step for recalculating distribution parameters.	ЛК, ЛР
		4.2	Other Clustering Methods. Classification Tasks. Bayesian Classifier.	Agglomerative hierarchical clustering with sequential merging of the nearest objects into clusters. DBSCAN as a density-based clustering method identifying regions of high density. Classification as the task of assigning an object to one of predefined classes. Bayesian classifier based on Bayes' theorem and the assumption of feature independence naive Bayes classifier.	ЛК, ЛР
		4.3	Linear Methods for Classification. Logistic Regression, Maximum Likelihood.	Linear classification methods with class separation by a linear boundary in feature space. Logistic regression as a binary classification method predicting class membership probability through a sigmoid function. Maximum likelihood as a learning criterion for logistic regression instead of minimizing the squared error. Interpretation of logistic regression coefficients as the influence of features on the log odds ratio.	ЛК, ЛР
		4.4	Neural Networks: General Architecture. Multilayer Networks. Backpropagation.	General architecture of a neural network: input layer of features, one or more hidden layers with non-linear activation functions, output layer with dimensionality matching the task. Multilayer networks as universal approximators of any continuous function. Forward pass for computing network outputs given input data. Backpropagation as an algorithm for computing gradients of the loss function with respect to network weights through sequential application of the chain rule from output layer to input layer.	ЛК, ЛР
		4.5	Stochastic Gradient Descent. Genetic Algorithms.	Stochastic gradient descent as an optimization method for training neural networks with weight updates based on a random mini-batch of examples rather than the entire dataset. Advantages of stochastic gradient descent: faster convergence, ability to work with large datasets, escaping local minima. Genetic algorithms as evolutionary optimization methods simulating natural selection processes. Basic operations of a genetic algorithm: selection of the best solutions, crossover for exchanging chromosome parts, mutation for random	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				changes.	
Раздел 5	Feature Detection; Data Normalization. Fuzzy Sets. Bayesian Networks	5.1	Feature Extraction.	Feature extraction as the process of creating new informative features from raw data. Difference between feature extraction and feature selection: extraction creates new features, selection chooses a subset of existing ones. Feature extraction methods for images: corner and edge detectors, histograms of oriented gradients. Feature extraction methods for texts: n-grams, topic modeling, word vector representations.	ЛК, ЛР
		5.2	Feature Transformations. Data Normalization. Data Normalization Methods.	Feature transformations as bringing data to a form convenient for model training. Data normalization as scaling features to a specific range or distribution. Necessity of normalization for methods sensitive to feature scale: gradient descent, principal component analysis, distance-based methods.	ЛК, ЛР
		5.3	Min-Max Normalization. Z-score Normalization. Decimal Scaling.	Min-max normalization with linear transformation of features into a given range, usually from zero to one. Preservation of the original distribution shape with min-max normalization. Z-score normalization as centering and scaling to unit variance: subtracting the mean and dividing by the standard deviation. Transforming features to a standard normal distribution. Decimal scaling as dividing all feature values by a power of ten to shift the decimal point. Simplicity and reversibility of decimal scaling.	ЛК, ЛР
		5.4	Fuzzy Sets. Bayesian Networks. Bayesian Inference Tasks. Method for Constructing a Fuzzy Bayesian Network.	Fuzzy sets as an extension of classical set theory with element membership specified by a continuous function from zero to one. Membership functions for describing linguistic variables: high, medium, low. Bayesian networks as probabilistic graphical models representing dependencies between variables through directed acyclic graphs. Bayesian inference tasks: computing posterior probabilities of variables given observations. Method for constructing a fuzzy Bayesian network: combining fuzzy logic for working with imprecise data and Bayesian networks for probabilistic inference. Application in decision support systems with incomplete or fuzzy information.	ЛК, ЛР

* - заполняется только по ОЧНОЙ форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве ____ шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. James, G. et al. An introduction to statistical learning. – Springer, 2013. – 426 pp
2. Trevor Hastie, Robert Tibshirani, et al., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition, 2017
3. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования : учебное пособие / В. В. Вьюгин. — Москва : МЦНМО, 2014. — 304 с.

Дополнительная литература:

1. Bruce, P. C., & Bruce, A. (2017). Practical Statistics for Data Scientists : 50 Essential Concepts (Vol. First edition). Sebastopol, CA: O'Reilly Media
2. Molnar, C. (2018). iml: An R package for Interpretable Machine Learning
3. Explainable and interpretable models in computer vision and machine learning. (2018)
4. Комбинаторика и теория вероятностей, учебное пособие, 99 с., Райгородский, А. М., 2013

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров
 - Электронно-библиотечная система РУДН – ЭБС РУДН <http://lib.rudn.ru/MegaPro/Web>
 - ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
 - ЭБС Юрайт <http://www.biblio-online.ru>
 - ЭБС «Консультант студента» www.studentlibrary.ru
 - ЭБС «Троицкий мост»
2. Базы данных и поисковые системы
 - электронный фонд правовой и нормативно-технической документации <http://docs.cntd.ru/>
 - поисковая система Яндекс <https://www.yandex.ru/>
 - поисковая система Google <https://www.google.ru/>
 - реферативная база данных SCOPUS <http://www.elsevier.com/scopus/>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля*:

1. Курс лекций по дисциплине «Machine Learning and Big Data Mining».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИКИ

Доцент

Должность

РУКОВОДИТЕЛЬ БУП

Заведующий кафедрой

Должность

РУКОВОДИТЕЛЬ ОП ВО

Профессор

Должность

Салтыкова О.А.

Фамилия И.О

Разумный Ю.Н.

Фамилия И.О

Разумный Ю.Н.

Фамилия И.О