МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное

образовательное учреждение высшего образования

«Национальный исследовательский

Нижегородский государственный университет им. Н.И. Лобачевского»

Институт аспирантуры и докторантуры

УТВЕРЖДЕНО
решением президиума
Ученого совета ННГУ
от 14.12.2021 г.
протокол № 4

Рабочая программа дисциплины/ Work program of the course

Машинное обучение
Machine learning

Уровень высшего образования
Level of higher education
бакалавриат
bachelor's degree program
(бакалавриат / магистратура / специалитет)

Направление подготовки / специальность
Training direction / speciality
02.03.02 Фундаментальная информатика и информационные технологии
02.03.02 Fundamental Computer Science and Information Technology
(указывается код и наименование направления подготовки / специальности)

Направленность образовательной программы
Orientation of educational program
Общий профиль
General profile
(указывается профиль / магистерская программа / специализация)

Форма обучения
form of study
очная
full-time
(очная / очно-заочная / заочная)

Нижний Новгород, 2022 год
Nizhni Novgorod, 2022

1. The place of discipline in the structure of ОПОП

Дисциплина Б1.О.08 «Машинное обучение» относится к части ООП по направлению подготовки 02.03.02 «Фундаментальная информатика и информационные технологии», формируемой участниками образовательных отношений. Дисциплина читается студентам 4 курса в 8 семестре, 3 зачетных единицы, 108 часов, зачет.

The discipline "Machine learning" is one of the general education disciplines of the variable part of Block 1 of educational disciplines (Б1.В), is a compulsory discipline of professional disciplines and is studied in the 2nd year of study, in the 4th semester. Discipline code Б1.В.01.02.

The goals of mastering “Machine Learning” are to develop general professional (ОПК), universal (УК) and professional (ПК) competencies in graduate students in accordance with the requirements of the Federal State Educational Standards in the relevant field of study, taking into account the peculiarities of the scientific school of UNN and the needs of the labor market of the Nizhny Novgorod region.

The specific goals of mastering the discipline are expressed in the system of competencies, to the formation of which the implementation of ОПОП is intended.

№ варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
1	Блок 1. Дисциплины (модули) Часть, формируемая участниками образовательных отношений	Дисциплина Б1.В.08 «Машинное обучение» относится к части ООП направления подготовки 02.03.02 «Фундаментальная информатика и информационные технологии», формируемой участниками образовательных отношений.

2. The planned learning outcomes in the discipline, correlated with the planned results of mastering the educational program (competencies of graduates).

Таблица 1

Формируемые компетенции (код, содержание компетенции) / **Formed competencies** (code, content of competence)	**Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции / Planned learning outcomes for the discipline (module), in accordance with the indicator of achievement of competency**		**Наименование оценочного средства / Name of the evaluation tool**
	Индикатор достижения компетенции (код, содержание индикатора) / **Competency achievement indicator** (code, indicator content)	**Результаты обучения по дисциплине / Learning outcomes by the discipline**	
ПК-5 *Способен использовать современные инструментальные и вычислительные средства информационных*	***ПК-5.2. Знает*** *основные принципы автоматизации и компьютеризации процессов сбора и обработки физической информации /* ***Student knows*** *basic*	*KNOW: methods of critical analysis and assessment of modern scientific achievements, as well as methods of generating new ideas in solving research and practical problems, including in interdisciplinary fields; material of fundamental sections of mathematical*	*собеседование / interview*

технологий / Ability to use modern instrumental and computing tools of information technology	*principles of automation and computerization for collecting and processing of physical information*	*modeling of complex technical, organizational and social systems*	
	ПК-5.4. Умеет *обрабатывать полученные в ходе эксперимента данные с использованием современных информационных технологий; проводить численные расчеты физических величин при обработке экспериментальных результатов /* ***Student is able*** *to process data obtained in experiment with the use of modern information technology; make calculations of physical quantities when processing experimental results*	*BE ABLE to: analyze alternative options for solving research and practical problems and evaluate the potential gains / losses of the implementation of these options;* *present the results of research activities at a high level and respecting copyright;* *present the results of research activities at a high level and respecting copyright;* *use classical approaches to solving decision-making problems in various areas of human activity*	*тест / test* *задачи / tasks*

3. The structure and content of the discipline

3.1. The complexity of the discipline

	Full-time education
Total intensity	3 ЗЕТ
Curriculum hours	108
including	
classroom lessons (contact work): - lecture-type classes - seminar-type classes - laboratory type classes - current control (КСР)	21 10 10 1
Independent work	87
Intermediate certification - credit	0

3.2. Discipline content

The name and summary of sections and topics of the discipline	**Total hours**	Including	
		Contact work (work in cooperation with the teacher), hours. Of them	**Student's independent work, hours**

		lecture-type classes	seminar-type classes	laboratory type cla	Total	
Machine learning problems. Features. Supervised and unsupervised learning. Decision function (decision rule). The generalizing ability of the decision function (the problem of the quality of learning). Classification, regression, clustering. Examples of practical tasks.	6	1	0		1	5
Probabilistic formulation of the machine learning problem. The principle of minimizing empirical risk. Bayesian decision theory. The principle of maximum posterior probability. Regression function. Bayesian classifier.method.	6	1	0		1	5
Experimental methods for assessing the quality of education. Separation of data into training and test samples. Cross-validation.	6	1	0		1	5
Least squares method. Maximum likelihood method. Linear regression model. System of normal equations.	8	1	1		2	6
The problem of overfitting when solving the regression problem. Methods to combat overfitting: reduction in the number of parameters, regularization (ridge regression), lasso.	8	1	1		2	6
Nearest neighbors method. Risk estimation theorem.	8	1	1		2	6
Naive Bayesian classifier.	7	0	1		1	6
Linear Discriminant Analysis. Quadratic discriminant analysis.	7	0	1		1	6
Logistic regression. Neural networks. Stochastic Gradient Descent. Backpropagation. Regularization. Dropout. The concept of deep neural networks.	8	1	1		2	6
Support Vector Machine. Kernel trick.	7	0	1		1	6
Decision trees. CART method (classification and regression trees) for solving classification problems and regression reconstruction. Methods for handling missing values.	7	1	0		1	6
Ensembles of decision rules (classifiers). Boosting. AdaBoost algorithm. Boosting and additive models. Gradient boosting.Gradient Boosting Trees. Bagging. Random trests.	7	1	0		1	6
Unsupervised learning. Clustering. K-means, k-,edoids, Expectation	7	0	1		1	6

Maximization. DBSCAN						
Hierarchical clustering. Agglomerative and separating methods.	7	0	1		1	6
Foundations of the Vapnik – Chervonenkis theory. Bernstein's lemma. A theorem on the uniform convergence of empirical risk to the expected risk in the case of a finite class of decision rules. Substantiation of the principle of minimizing empirical risk. Vapnik – Chervonenkis dimension. Sauer's lemma. A theorem on the uniform convergence of empirical risk to expected risk in the case of finite Vapnik – Chervonenkis dimension. The principle of structural risk minimization.	8	1	1		2	6
Current control (KCP)	1			1		
Intermediate certification - credit						
Total	**180**	**10**	**10**	**1**	**20**	**87**

The current monitoring of progress is realized in the form of surveys in lecture-type lessons. Interim certification takes place in the traditional form (exam).

4. Educational and methodological support of independent work of students

Independent work of a student in the study of the discipline "Machine Learning" includes the implementation of tasks under the supervision of a teacher, preparation for intermediate certification in a traditional form.

Control questions and tasks for conducting current control and intermediate certification based on the results of mastering the discipline are given in clause 5.2.

Fund of assessment tools for intermediate certification by discipline (module),
including:

5.1. Description of the scales for assessing learning outcomes in the discipline "Machine Learning"

The level ofdevelopment of competencies (competency achievement indicator)	**scale formation evaluation competencies**						
	bad	**unsatisfactory**	**satisfactory**	**good**	**very good**	**excellent**	**excellent**
	Fail		credited				

Knowledge	Lack of knowledge of theoretical material. Inability to assess the completeness of knowledge due to the student's refusal to answer	The level of knowledge is below the minimum requirements. There were gross mistakes.	The minimum level of knowledge. It made many mistakes structurally unstable.	The level of knowledge in the amount corresponding to the training program. Allowed several non-robust error	level of knowledge in the volume corresponding to the training program. It allowed a few minor errors	level of in the knowledgethe volume corresponding to the training program, with no errors.	The level of knowledge in excess of theprogram training.
Skills	Lack of minimum skills. Inability to assess the availability of skills due to the failure of the student to answer	When solving standard tasks not demonstrated basic skills. There were gross mistakes.	Demonstrated basic skills. Solved typical tasks with minor errors. All tasks have been completed, but not in full.	demonstrated All basic skills have been. All major problems with minor errors have been solved. Completed all tasks, in full, but some with flaws.	demonstrated All basic skills have been. All major tasks have been solved. Completed all tasks, in full , but some with flaws.	demonstrated All basic skills have been, all basichave been solved taskswith some minor flaws, all tasks have beenin completedfull.	demonstrated All basic skills have been, all basichave been solved tasks. All tasks were completed, in full , without flaws.
Skills	Lack of knowledge of the material. Inability to assess the availability of skills due to the student's refusal to answer	when notsolving standard problems Basic skills weredemonstrated. There were gross mistakes.	There is a minimal set of skills for solving standard problems with some drawbacks.	Demonstrated basic skills in solving standard problems with some	Demonstrated basic skills in solving standard shortcomings.problems without errors and shortcomings.	Demonstrated skills in solving non-standard problems without errors and shortcomings.	Demonstrated a creative approach to solving non-standard problems.

Evaluation scale for intermediate certification

Assessment		**The level of training is**
credited	Excellent	All competences (parts of competencies), the formation of which theis directed discipline, are formed at a level not lower than "excellent"
	Excellent	All competences (parts of competencies), the formation of which theis directed disciplineat, are formed at a level not lower than "Excellent", while at least one competence is formed at the level of "excellent"
	Very good	All competencies (parts of competencies), the formation of which theis aimed disciplineat, are formed at a level not lower than "very good", while at least one competence is formed at the level "Very good"
	Good	All competencies (parts of competencies), the formation of which theis aimed disciplineat, are formed at a level not lower than "good", while at least one competence is formed at the level of "good"
	Satisfactory	All competencies (parts of competencies), the formation of which directed discipline, formed at the level of at least "satisfactory",at the same time whileleast one competency is formed at the level of "satisfactory"
Fail	Poor	At least one competency is formed at the level of "unsatisfactory", none of competencies is not formed at the level of "bad"
	Poor	At least one competency is formed at the level of "bad"

1.

Typical test items or other materials needed to assessoutcomes learning

5.2.1 Control questions for assessing the formation of competence

Вопросы	***Код формируемой компетенции***

Machine learning problems. Features. Supervised and unsupervised learning. Decision function (decision rule). The generalizing ability of the decision function (the problem of the quality of learning). Classification, regression, clustering.	ПК-5
Probabilistic formulation of the machine learning problem. The principle of minimizing empirical risk. Bayesian decision theory. The principle of maximum posterior probability. Regression function. Bayesian classifier.method.	ПК-5
Experimental methods for assessing the quality of education. Separation of data into training and test samples. Cross-validation.	ПК-5
Least squares method. Maximum likelihood method. Linear regression model. System of normal equations.	ПК-5
The problem of overfitting when solving the regression problem. Methods to combat overfitting: reduction in the number of parameters, regularization (ridge regression), lasso.	ПК-5
Nearest neighbors method. Risk estimation theorem.	ПК-5
Naive Bayesian classifier.	ПК-5
Linear Discriminant Analysis. Quadratic discriminant analysis.	ПК-5
Logistic regression. Neural networks. Stochastic Gradient Descent. Backpropagation. Regularization. Dropout. The concept of deep neural networks.	ПК-5
Support Vector Machine. Kernel trick.	ПК-5
Decision trees. CART method (classification and regression trees) for solving classification problems and regression reconstruction. Methods for handling missing values.	ПК-5
Ensembles of decision rules (classifiers). Boosting. AdaBoost algorithm. Boosting and additive models. Gradient boosting.Gradient Boosting Trees. Bagging. Random forests.	ПК-5
Unsupervised learning. Clustering. K-means, k-,edoids, Expectation Maximization. DBSCAN	ПК-5
Hierarchical clustering. Agglomerative and separating methods.	ПК-5
Foundations of the Vapnik – Chervonenkis theory. Bernstein's lemma. A theorem on the uniform convergence of empirical risk to the expected risk in the case of a finite class of decision rules. Substantiation of the principle of minimizing empirical risk. Vapnik – Chervonenkis dimension. Sauer's lemma. A theorem on the uniform convergence of empirical risk to expected risk in the case of finite Vapnik – Chervonenkis dimension. The principle of structural risk minimization.	ПК-5

5.2.2. Typical tasks for assessing the formation of the PC-4 competence

1. A training sample is given (see above). Using themethod, ridge regressionconstruct a polynomial model of the form $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ if the regularization parameter $\lambda = 2$.

2. Prove that in the case of a quadraticfunction loss, the minimum average risk is delivered by the conditional average. What is thein this case average risk?
3. Prove that if the loss function is equal to the modulus of the difference, then theminimum to the average risk conditional median delivers the. What is the average risk in this case?
4. Let the answer be given in the form of an analytical function **x XOR ((y XOR z) OR w)**, where w, x, y and z are TRUE or FALSE. Build a decision tree that predicts aresponse zero error.

5.2.3. Typical tasks for assessing the formation of the PC-3 competence

1. Download dataset Spam ((http://www-stat.stanford.edu/~tibs/ElemStatLearn/). Divide the data on the training and the test sample (according to the marks infile). spam.traintest Compare the quality of learning using support vector and *K* nearest neighbors.parameters of the models Select theat your discretion.
2. Load the Spam dataset ((http://www-stat.stanford.edu/~tibs/ElemStatLearn/). Divide the data into training and test samples (according to the labels in thefile spam.traintest) Compare the training quality using decision trees and method *K* nearest neighbors.parameters of the models Select theat your discretion.
3. Load the Spam dataset ((http://www-stat.stanford.edu/~tibs/ElemStatLearn/). Divide the data into training and test samples (according to the labels in thefile spam.traintest) Compare the training quality using decision trees and Support Vector Machine Select the parameters of the models at your discretion.
4. A training sample is given

*x*1	0	1	1	0	0	1	1	2	6
x2	3	3	1	0	1	1	2	3	1
y	0	0	0	0	1	1	1	1	1

Using the method of linear discriminant analysis for each class, construct a discriminant function and write the equation of the separating surface ...

5. A training sample is given (see table above). Using the method of quadratic discriminant analysis, construct discriminant functions.
6. A training sample is given (see table above). UsingNaïve Bayesian classifier to estimate the probability P$(Y = 1 \mid x1 = 1, x1, =2)$
7. A training sample is given:

x	-1	0	0	1	2
y	1	-2	1	7	8

Using the least squares method, construct a polynomial model of the form $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$.

Insert an example ticket

5.2.4. Example of an examination card

National Research Nizhny Novgorod State University

named after N.I. Lobachevsky

Institute / Faculty of Information Technologies of Mathematics and Mechanics

Department of Mathematical Support and Supercomputer Technologies

Discipline Computational Mathematics

Insert an example of a ticket

EXAMINATION TICKET № 1

1. Interpolation. Lagrange's formula.

2. Methods for finding the eigenvalues of a matrix and theorems about them.

1. Interpolation. The Formula Of Lagrange.

2. Methods for finding eigenvalues of a matrix and theorems about them.

Head department _____

Examiner _____

6. Educational-methodical and informational support of the discipline

a) Basic literature:

James G. et al. An introduction to statistical learning. New York: Springer, 2013.

Géron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. – O'Reilly Media, 2019.

b) Further reading

Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: Data Mining, Inference, and Prediction. 2nd Edition. Springer, 2009 URL: ***http://www-stat.stanford.edu/~tibs/ElemStatLearn/***

c) software and Internet resources

1. ScikitLearn Machine Learning Library https://scikit-learn.org/

7. Material and technical support of the discipline

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой (лекционного и семинарского типа), оснащенные оборудованием и техническими средствами обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

The premises are classrooms for conducting training sessions, provided by the program (lecture type), equipped with equipment and technical teaching aids.

Premises for independent work of students are equipped with computers with the ability to connect to the Internet and are provided with access to the electronic information and educational environment of UNN.

The program is compiled in accordance with the requirements of the Federal State Educational Standard of Higher Education in the direction of training 09.06.01 "Informatics and computer technology".

Программа составлена в соответствии с требованиями ФГОС ВО /ОС ННГУ _____.

Author of the program: _____Doctor of Physics and Mathematics, prof. N. Yu. Zolotykh

Reviewer: _____

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики от 01.12.2021 года, протокол № 2.