

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION
Federal State Autonomous Educational Institution
higher education
NORTH CAUCASIAN FEDERAL UNIVERSITY

I APPROVE

Dean of the Faculty of Mathematics
and computer science named after
Professor N.I. Chervyakov,
Candidate of Physical and Mathematical
Sciences, Associate Professor
Grobova T.A.

ASSESSMENT TOOLS FUND FOR DISCIPLINE
Machine Learning Programming Design Based on Python / Разработка программного
обеспечения для машинного обучения на основе Python

Direction of training	02.04.01 Mathematics and Computer Science
Directionality (profile)	Training Program for Electronic Information
Year of commencement of studies	2026
Form of study	full-time
Implemented in a semester	1

Introduction

1. Purpose

The assessment tools fund for the discipline "Machine Learning Programming Design Based on Python / Разработка программного обеспечения для машинного обучения на основе Python" is intended for an objective assessment of the level of development of competencies .

2. FOS is application To program discipline «Machine Learning Programming Design Based on Python / Разработка программного обеспечения для машинного обучения на основе Python".

3. Developer R. A. Voronkin, associate professor of the Department of Digital, Robotic Systems and Electronics at the Institute of Advanced Engineering.

4. An examination of the organophosphorus compounds was carried out.

Members of the expert group:

Chairman:

Poddubnaya N.A. – Chairman of the Teaching and Methodological Committee of the Faculty of Mathematics and Computer Science named after Professor N.I. Chervyakov

Members of the commission:

Gladkov A.V. – member of the teaching and methodological committee of the Department of Computational Mathematics and Cybernetics;

Andrukhiv L.V. – member of the teaching and methodological committee of the Department of Mathematical Modeling;

Kopytkova L.B. – member of the teaching and methodological committee of the department of mathematical analysis, algebra and geometry.

Representative of the employer organization: Lapin V.G. – Head of the Automated Control System Department of the Stavropol Regional Clinical Consultative and Diagnostic Center.

Expert opinion: the assessment tool fund is recommended for assessing the level of development of competencies during ongoing monitoring of academic performance and midterm assessment of students in the discipline **"Machine Learning Programming Design Based on Python / Разработка программного обеспечения для машинного обучения на основе Python"** .

April 15 , 2026

5. The validity period of the FOS is determined by the period of implementation of the educational program.

1. Description of the criteria for assessing competence at various stages of its development, description of assessment scales

Competence(s), indicator(s)	Levels of development of competencies,			
	The minimum level has not been reached (Unsatisfactory) 2 points	Minimum level (satisfactorily) 3 points	Intermediate level (Fine) 4 points	High level (excellent) 5 points
<i>OPK-1 Capable of finding, formulating and solving current and significant problems in applied and computer mathematics</i>				
Learning outcomes for the course (module): <i>Indicator:</i> ID-1 OPK-1 Finds and formulates significant problems of applied and computer mathematics	Inability to identify and formulate significant problems in applied and computational mathematics. Failure to understand the relationship between mathematical methods and practical problems in data analysis and machine learning. Problem statements are incorrect or absent.	Able to identify specific problems in applied and computer mathematics using a model or with the help of a teacher, but makes errors in their formulation and justification. Experiencing difficulty determining the practical significance of the problem under consideration.	Confidently identifies and formulates significant problems in applied and computer mathematics that arise when solving data analysis and machine learning problems. Able to justify the relevance of a problem and identify key areas for its solution.	Fully mastered the skills of identifying and formulating significant problems in applied and computer mathematics. Independently identifies relevant mathematical problems in machine learning and artificial intelligence, substantiates their practical and scientific significance, and proposes possible approaches to their solution.
Learning outcomes for the course (module): <i>Indicator:</i> ID-2 OPK-1 knows and uses methods for solving significant problems of applied and computer mathematics	Lacks knowledge of the basic methods of applied and computer mathematics used to solve data analysis and machine learning problems. Inability to select or apply a mathematical method to solve a given problem.	Knows individual methods of applied and computer mathematics and is able to apply them to solve typical problems. Allows errors in choosing a method and interpreting the results obtained.	Confidently uses applied and computational mathematics methods to solve data analysis and machine learning problems. Correctly selects mathematical methods based on the specifics of the problem and is able to justify their choices.	Fully proficient in modern applied and computer mathematics methods necessary for solving significant problems in artificial intelligence and machine learning. Independently selects the most effective solution methods, applies them effectively in practice, and critically evaluates the results.
<i>OPK-3 Capable of independently creating application software based on modern information technologies and</i>				

<i>network resources, including domestically produced ones</i>				
Learning outcomes for the course (module): <i>Indicator:</i> ID-1 _{OPK-3} Independently creates application software products	Incapable of independently developing software applications for data analysis and machine learning tasks. Makes significant errors in program structure, algorithm selection, and software implementation.	Able to develop simple software applications based on a template or using existing solutions. May make errors in the implementation of individual program components and struggles with integrating machine learning algorithms and data processing.	Confidently develops applied software products to solve common data analysis and machine learning problems. Proficient in the use of programming languages, libraries, and software tools, ensuring program performance and compliance with specified requirements.	Fully versed in the independent development of applied software products. Designs and implements effective software solutions for machine learning tasks, intelligently selects algorithms and development tools, and ensures the reliability, scalability, and quality of software products.
Learning outcomes for the course (module): <i>Indicator:</i> ID-2 _{OPK-3} Knows and can find modern information technologies, including domestically produced ones, necessary for the creation of applied software products .	Lacks familiarity with modern information technologies used to develop application software products. Incapable of independently selecting software tools, libraries, and development tools to solve a given problem .	Knowledgeable of certain modern information technologies and software tools, but struggles with their selection and justification. Can identify necessary technologies based on examples, but makes errors in assessing their applicability to a specific task.	Confidently navigates modern information technologies, including domestic software products and development tools. Capable of independently identifying, evaluating, and selecting technologies, tools, and libraries necessary for creating applied software products in the fields of data analysis and machine learning.	He is fully versed in modern approaches to searching, analyzing, and selecting information technologies, including domestic software solutions. He intelligently selects the most effective tools, platforms, and software for developing application software products, taking into account the specifics of the task, as well as performance, security, scalability, and software maintenance requirements.

The assessment of the level of development of competence in the discipline is carried out on the basis of the "Regulations on the current monitoring of academic performance and midterm assessment of students in higher education programs - bachelor's programs, specialist programs, master's programs - at the federal state autonomous educational institution of higher education "North Caucasus Federal University" in the current version.

ASSESSMENT TOOLS FOR CHECKING THE LEVEL OF DEVELOPMENT OF COMPETENCIES

Task number	Correct answer	Contents of the question	Competence
		Full-time education, Semester 1 (4)	
1	Artificial intelligence is a field that studies methods for creating intelligent systems.	Define the concept of "artificial intelligence"	OPK -1
2	Machine learning is a data analysis method in which a model is trained based on examples.	What is machine learning?	OPK -1
3	Classification	Which of the following types of problems is related to supervised learning? A) Clustering B) Dimensionality reduction C) Classification D) Data compression	OPK -1
4	With training, without training, with partial training	Name the main approaches to training in machine learning	OPK -1
5	1. Data collection; 2. Pre-processing; 3. Training; 4. Evaluation; 5. Implementation	Organize the stages of the machine learning project lifecycle: data preprocessing, model evaluation, data collection, model training, deployment	OPK -3
6	A training set is a data set containing input features and corresponding target values.	What is a training set in machine learning problems?	OPK -3
7	D) All of the above	What problems can be solved using AI? A) Disease diagnosis B) Facial recognition C) Text classification D) All of the above	OPK -1
8	A model is an algorithm trained on data and capable of making predictions.	What is a model in machine learning?	OPK -3
9	B) Regression	Which task involves predicting a numerical value? A) Classification B) Clustering C) Regression D) Association	OPK -1
10	Features are measurable characteristics of an object fed to the input of the model	What are features in machine learning?	OPK -3
11	To eliminate outliers, normalize and improve the quality of the model	Why is data preprocessing performed?	OPK -3

12	B) Training a model on labeled data	What does the term "supervised learning" mean? A) Supervised learning B) Training a model on unlabeled data C) Training a model on labeled data D) Post-deployment maintenance	OPK -1
13	Supervised learning, unsupervised learning, reinforcement learning	List the main types of training in machine learning	OPK -3
14	B) Unsupervised learning	What type of learning is used in clustering? A) Supervised learning B) Reinforcement learning C) Unsupervised learning D) Deep learning	OPK -3
15	Classification, regression, clustering, dimensionality reduction, association	Name at least four common machine learning tasks	OPK -3
16	B) Regression	What problem is solved by predicting a continuous value? A) Classification B) Clustering C) Regression D) Association	OPK -3
17	B) Detection of groups of similar objects	What does a clustering algorithm do? A) Predicts a number B) Determines a class C) Detects groups of similar objects D) Converts data to text	OPK -3
18	Supervised learning requires labeled data; unsupervised learning does not.	What is the fundamental difference between supervised and unsupervised learning?	OPK -3
19	D) All of the above	Which of the following tasks are related to supervised learning? A) Classification B) Regression C) Price prediction D) All of the above	OPK -3
20	A) The algorithm selects actions by receiving rewards	What happens in reinforcement learning? A) The algorithm selects actions by receiving rewards B) The model is trained on labeled data C) The model is trained without target values D) Only logic is used	OPK -3
21	A regression model predicts real values, while classification models predict categories.	What is the difference between a regression problem and a classification problem?	OPK -3
22	Association is the task of finding frequently occurring patterns in data.	Explain what an association problem is in machine learning	OPK -3

23	D) All of the above	Which of the following are machine learning tasks? A) Predicting equipment failures B) Analyzing customer behavior C) Image segmentation D) All of the above	OPK -3
24	Supervised learning	What type of learning is used in category and number prediction tasks?	OPK -3
25	Missing values, outliers, type mismatches, duplicates	Name the main problems encountered in raw data	OPK -3
26	Standardization, normalization, coding of categorical features	List the common data preprocessing methods	OPK -3
27	B) Filling with average value	Which method can be used to fill missing values in a numeric attribute? A) Deleting all rows B) Filling with random values C) Filling with the mean value D) Filling with the text "unknown"	OPK -3
28	A feature is a measurable characteristic of an object fed to the input of the model.	What is a feature in the context of machine learning?	OPK -1
29	D) Categorical and numerical	What types of features are usually distinguished in machine learning? A) Binary and text B) Numeric and real C) Logical and symbolic D) Categorical and numeric	OPK -3
30	Missing values can reduce the accuracy of the model and skew the training set.	Why is it important to handle missing values in a dataset?	OPK -3
31	One-hot encoding	Which categorical feature coding method is used to create binary columns for each unique value?	OPK -3
32	Changing the shape and structure of features to improve the efficiency of the model	What does the term "feature engineering" mean?	OPK -1
33	B) Selection of features, generation of new features, transformation of existing ones	What is included in the feature engineering process? A) Deleting the model B) Saving features to a file C) Selecting features, generating new features, transforming existing ones D) Inserting random features	OPK -1
34	D) Min-Max scaling	Which normalization method transforms values into the range 0 to 1? A) Z-transform B) Binary encoding C)	OPK -3

		Outlier removal D) Min-Max scaling	
35	Removing duplicates is necessary to prevent distortion of the data distribution.	Why is it necessary to remove duplicate rows in a dataset?	OPK -3
36	The target feature is the variable that the model is trying to predict.	What is a target feature in a supervised learning problem?	OPK -1
37	A relationship between one dependent variable and one or more independent variables	What does linear regression model?	OPK -3
38	$y = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_n x_n$	Provide the general equation for linear regression	OPK -3
39	B) Continuous variable	What type of variable does a linear regression model predict? A) Categorical B) Discrete C) Continuous D) Logical	OPK -3
40	Estimation of model coefficients based on a training sample	What does the process of training a linear regression model mean?	OPK -3
41	A) Least squares method	Which method is most often used to estimate the parameters of linear regression? A) Ordinary least squares B) Monte Carlo C) K-means D) Principal component analysis	OPK -3
42	Root mean square error	Name one of the metrics used to evaluate the quality of a linear regression model	OPK -3
43	D) The closer to 1, the better the model explains the variance of the target variable.	How is the coefficient of determination R^2 interpreted? A) It shows the probability of error B) It is always zero C) It depends only on the number of features D) The closer to 1, the better the model explains the variance of the target variable	OPK -3
44	Regularization reduces overfitting by penalizing large coefficients.	What is the role of regularization in linear regression?	OPK -3
45	Yes, in the presence of multicollinearity, outliers, or non-linearity	Are there any limitations to the use of linear regression? Provide examples.	OPK -3
46	Linear regression assumes a linear relationship between the features and the	What is the main assumption of the linear regression model?	OPK -3

	target variable.		
47	A) Features must be scaled	What should be done before training a linear regression to improve its accuracy? A) Features should be scaled B) Remove all categorical features C) Transform the target variable into a categorical variable D) Exclude missing values without checking	OPK-3
48	Comparison of predicted values with actual ones	How are the results of a linear regression model evaluated?	OPK-3
49	An optimization method in which the model parameters are updated after each random sample	What is stochastic gradient descent (SGD)?	OPK-3
50	C) One random example	How many objects are used in one step in stochastic gradient descent? A) The entire training set B) A mini-batch C) One random example D) Two examples with minimal error	OPK-3
51	To find the minimum of the loss function and update the model weights	Why is gradient descent used in machine learning?	OPK-3
52	The gradient indicates the direction of greatest increase in the function, so the movement is in the opposite direction	Why does gradient descent move in the opposite direction to the gradient when training a model?	OPK-3
53	B) Learning speed	What is the name of the parameter that determines the step length when updating weights? A) Regularization B) Learning rate C) Tree depth D) Feature size	OPK-3
54	Too large a learning rate can lead to divergence, too small a learning rate can lead to slow convergence.	How does learning rate affect the optimization process?	OPK-3
55	In SGD, parameters are updated more frequently, which can speed up training but make it noisier.	What is the difference between stochastic gradient descent and regular gradient descent?	OPK-3
56	Yes, for example: data shuffling, reducing the learning rate, using gradient averaging	Are there any measures to reduce SGD instability? Please provide examples.	OPK-3

57	A) A function that determines the error of the model	What is a loss function in the context of SGD? A) A function that determines the error of the model B) An accuracy score C) A classification result D) A number of incorrect answers	OPK-3
58	Mini-batch gradient descent is a compromise between SGD and full gradient descent.	What is mini-batch gradient descent and what makes it special?	OPK-3
59	C) Loss function, its gradient and learning rate	What is required for one step of stochastic gradient descent? A) Learning rate only B) Loss function only C) Gradient only D) Loss function, its gradient, and learning rate	OPK-3
60	Convergence means that the values of the model parameters stabilize and the error reaches a minimum.	What does it mean for gradient descent algorithm to converge?	OPK-3
61	Overfitting is a situation where a model performs well on the training set but performs poorly on new data.	What is overfitting in machine learning?	OPK-3
62	By adding a penalty for model complexity to the loss function	How does regularization help combat overfitting?	OPK-3
63	A) L1 and L2	Which of the following methods are types of regularization? A) L1 and L2 B) PCA and SVD C) BatchNorm and Dropout D) SGD and Adam	OPK-3
64	L1 regularization sets some weights to zero and facilitates feature selection, L2 distributes the penalty evenly	What is the difference between L1 and L2 regularization?	OPK-3
65	Regularization reduces the absolute values of the model's weights, thereby reducing its complexity.	What effect does regularization have on model parameters?	OPK-3
66	C) A method of evaluating a model in which the data is divided into several subsets	What is cross-validation? A) A model selection method B) An optimization algorithm C) A feature selection method D) A model evaluation method in which the data is divided	OPK-3

		into several subsets	
67	Model stability independent of the specific sample split	What is assessed by cross-validation?	OPK-3
68	Improves the reliability of model evaluation and helps identify overfitting	Why is it important to use cross-validation when training a model?	OPK-3
69	D) Split the data into K parts and sequentially train on K-1 parts with testing on the remaining	How does K-fold cross-validation work? A) Retraining on the same sample B) Using external data C) Selecting hyperparameters D) Splitting the data into K parts and sequentially training on K-1 parts and testing on the remaining	OPK-3
70	C) An overly complex model memorizes data precisely rather than generalizing patterns.	Why are complex models more likely to overfit? A) Due to lack of data B) Due to regularization C) Due to feature scaling D) An overly complex model memorizes data precisely rather than generalizing patterns	OPK-3
71	Yes, for example, using fewer features, simplifying the model, regularization, cross-validation	Is it possible to prevent overfitting? What are some ways?	OPK-3
72	B) On a test sample	On which dataset should the model be tested for overfitting? A) On the training set B) On the test set C) On the entire set D) On the set without the target feature	OPK-3
73	A model for solving classification problems that predicts the probability of class membership	What is logistic regression?	OPK-3
74	A) Sigmoid	What function is used in logistic regression to obtain the probability? A) Sigmoid B) Gaussian C) Polynomial D) Logarithmic	OPK-3
75	Predictions in logistic regression are the probabilities of an object belonging to the positive class	What is the output of logistic regression before thresholding?	OPK-3
76	B) Classification	For what task is logistic regression used? A) Regression B) Classification C) Clustering D) Ranking	OPK-3

77	D) A decision is made about the class based on a threshold (e.g. 0.5)	How does logistic regression transform probability into a predicted class? A) By the maximum weight of a feature B) By the number of features C) By the variance of the data D) A decision on the class is made based on a threshold (e.g. 0.5)	OPK-3
78	B) Prediction of categorical value	What is the main goal of logistic regression? A) Minimizing regression errors B) Finding outliers C) Predicting categorical values D) Finding cluster centers	OPK-3
79	D) Cross-entropy	What loss function does logistic regression use? A) Mean squared error B) L1 regularization C) Sum of weights D) Cross entropy	OPK-3
80	It compresses a linear combination of features into a range from 0 to 1	What is the role of sigmoid function in logistic regression?	OPK-3
81	Yes, for example, L2 regularization helps to avoid overfitting	Is regularization used in logistic regression? Give an example.	OPK-3
82	The weights of the model determine the contribution of each feature to the probability of assignment to a class	How are logistic regression coefficients interpreted?	OPK-3
83	B) Improving the interpretability and robustness of the model	What can regularization be used for in logistic regression? A) To speed up gradient descent B) To improve the interpretability and robustness of the model C) To improve the accuracy of the training set D) To increase the number of features	OPK-3
84	Logistic regression works well when classes are linearly separable.	In what cases does logistic regression give good results?	OPK-3
85	Accuracy	Which classification metric reflects the proportion of correct answers among all predictions?	OPK-3
86	Precision — accuracy; Recall — completeness	What do precision and recall metrics mean?	OPK-3
87	F1-measure is the harmonic mean between precision and recall	How is F1 score calculated?	OPK-3

88	D) 1.0	What is the value of F1-score for a perfectly accurate and complete classification? A) 0.0 B) 0.5 C) 0.8 D) 1.0	OPK-3
89	A) A matrix that shows the number of correct and incorrect predictions by class	What is a confusion matrix? A) A matrix that indicates the number of correct and incorrect predictions by class B) A method of data normalization C) A feature selection method D) The result of logistic regression	OPK-3
90	True Positive, True Negative, False Positive, False Negative	What four values are used in the confusion matrix?	OPK-3
91	Completeness is important when it is critical not to miss any positive cases.	In what tasks is recall more important than precision?	OPK-3
92	D) ROC-AUC	Which metric measures the area under the classifier error curve at different thresholds? A) MAE B) R^2 C) Accuracy D) ROC-AUC	OPK-3
93	ROC curve is being built By True Positive Rate and False Positive Rate values	How is the ROC curve constructed?	OPK-3
94	B) Recall	Which metric can be increased by lowering the classification threshold? A) Accuracy B) Recall C) Precision D) R^2	OPK-3
95	Takes into account both false positives and missed cases	Why is F1 score considered a balanced metric?	OPK-3
96	D) Precision and Recall	What two metrics does the F1-score combine? A) Accuracy and R^2 B) AUC and ROC C) TP and FPG D) Precision and Recall	OPK-3
97	A classification method in which an object is assigned to a class of the majority of its k nearest neighbors	What is the essence of the k-nearest neighbors (k-NN) method?	OPK-3
98	A) Euclidean distance	What distance measure is most commonly used in the k-NN method? A) Euclidean distance B) Correlation C) Product of features D) Cosine of an angle	OPK-3
99	The value of the parameter k determines the number of nearest objects participating	How does the parameter k affect the performance of the k-NN algorithm?	OPK-3

	in the voting		
100	Classes are defined based on the position of objects in the feature space	How is classification performed in the k-NN algorithm?	OPK-3
101	Support vector machine is an algorithm that finds the optimal boundary between classes.	What is Support Vector Machine (SVM)?	OPK-3
102	B) Hyperplane	What object is used in SVM to separate classes? A) Probability curve B) Hyperplane C) Cluster center D) Convex hull	OPK-3
103	Support vectors are objects that are closest to the separating hyperplane and determine its position.	What are support vector machines in SVM?	OPK-3
104	D) Linear, polynomial, radial	What kernels can be used in SVM? A) Only linear B) Only Gaussian C) Only logarithmic D) Linear, polynomial, radial	OPK-3
105	Parameter C controls the balance between maximizing clearance and minimizing error	What is the effect of regularization parameter C in SVM method?	OPK-3
106	Yes, for small k the model can overfit, for large k it can underfit	Can the k-NN method suffer from overfitting? Explain.	OPK-3
107	D) SVM is robust to outliers but sensitive to feature scaling	Which of the following is true for the SVM method? A) It is not sensitive to feature scales B) It is sensitive to the choice of distance metric C) It does not require regularization D) It is robust to outliers, but sensitive to feature scaling	OPK-3
108	Preprocessing includes feature normalization, which is especially important for SVM and k-NN	Why is it important to scale features before using SVM or k-NN?	OPK-3
109	An algorithm that constructs a structure from sequential conditions that separate data by features	What is a decision tree in machine learning?	OPK-3

110	A) Root node	From which element does the construction of a decision tree begin? A) Root node B) Leaf C) Random node D) Middle of the sample	OPK-3
111	Subdivision to improve the homogeneity of the target variable	What is the basic principle of constructing a decision tree?	OPK-3
112	Gini index, entropy, dispersion	Name the criteria used to evaluate the quality of splitting in decision trees	OPK-3
113	B) A node that does not contain any descendants	What is a leaf in a decision tree? A) The first node of the tree B) A node containing many objects C) A node containing no children D) Any intermediate node	OPK-3
114	Tree depth, minimum number of objects in a node, maximum number of features	What parameters can limit tree complexity and prevent overfitting?	OPK-3
115	D) Deep tree can overfit	Which of the following is true for decision trees? A) They require scaling B) They do not work with categorical features C) They do not overfit D) A deep tree can overfit	OPK-3
116	High interpretability, visualization, work with different types of features	Name the advantages of decision trees	OPK-3
117	B) Comparison of error distribution and visualization of tree structure	What methods are used to analyze a decision tree model? A) ROC curve B) Weight matrix C) Comparison of error distribution and visualization of tree structure D) Gradient descent	OPK-3
118	Yes, the tree can be displayed as a diagram with conditions and results	Is it possible to visualize a decision tree?	OPK-3
119	A) When using complex or overfitted trees	When can a decision tree produce poor results on new data? A) When using complex or overfitted trees B) When there are no categorical features C) When there is a linear relationship D) Only when using a random forest	OPK-3
120	The decision tree is the basis of many ensemble methods, such as random forest and gradient boosting.	What role do decision trees play in ensemble models?	OPK-3
121	Algorithms that combine multiple models	What are ensemble methods in machine learning?	OPK-3

	to improve prediction quality		
122	A random forest is an ensemble of decision trees trained on random subsamples of data and features.	What is the random forest algorithm?	OPK-3
123	A) Averaging (for regression) or voting (for classification)	What principle is used to combine predictions in a random forest? A) Averaging or voting B) Choosing the best tree C) Multiplying outputs D) Sorting by weights	OPK-3
124	Increasing model robustness, reducing overfitting, improving accuracy	What are the advantages of using ensembles over single models?	OPK-3
125	Gradient boosting builds trees sequentially, with each new tree learning from the mistakes of the previous one.	What is the basic idea behind gradient boosting?	OPK-3
126	B) Random forest and gradient boosting	Which of the following models are ensemble models? A) Logistic regression and SVM B) Random forest and gradient boosting C) Decision tree and PCAG A) k-NN and DBSCAN	OPK-3
127	Random forest trains trees in parallel, gradient boosting trains them sequentially	How does random forest differ from gradient boosting in terms of training process?	OPK-3
128	Overfitting can occur with too many trees, too much depth, and too high a learning rate.	Can gradient boosting overfit? Please provide reasons.	OPK-3
129	Parameters: n_estimators, learning_rate, max_depth And etc.	What are the main parameters that influence the quality of a gradient boosting model?	OPK-3
130	D) To assess the importance of features	What can feature_importances_ be used for in ensemble models? A) To adjust regularization B) To scale data C) To calculate metrics D) To evaluate feature importance	OPK-3
131	Yes, ensembles can significantly improve model quality and noise resilience.	Can an ensemble of models outperform a single model? Explain.	OPK-3
132	B) Averaging the predictions of several models	What does "bagging" mean in the context of ensembles? A) Training models sequentially B) Averaging	OPK-3

		predictions from multiple models C) Using only root nodes D) Manual feature selection	
133	Clustering is the task of dividing objects into groups based on similarity without using class labels.	What is clustering in machine learning?	OPK-3
134	A) KMeans	Which algorithm is most commonly used for clustering? A) KMeans B) Logistic regression C) Gradient boosting D) SVM	OPK-3
135	Cluster centers are the average coordinate values of all objects in a cluster.	What are cluster centers in KMeans method?	OPK-3
136	Redistributing objects to nearby cluster centers and updating the centers	Describe the iterative process of the KMeans algorithm.	OPK-3
137	B) Number of clusters	What parameter should be pre-set in the KMeans algorithm? A) Maximum depth B) Number of clusters C) Learning speed D) Number of features	OPK-3
138	Yes, for example, agglomerative hierarchical clustering	Are there clustering methods that do not require specifying the number of clusters? Give an example.	OPK-3
139	D) Dendrogram	What is the name of the tree that displays the process of combining objects during hierarchical clustering? A) Concordant matrix B) Diagram C) Cartogram D) Dendrogram	OPK-3
140	Yes, the scale of features affects the distances between points, and therefore the clustering result.	Does feature scaling affect clustering results? Explain.	OPK-3
141	C) The similarity of objects is determined by the distance between them	How is the similarity of objects determined during clustering? A) By cluster number B) By loss function C) By probability density D) By distance between them	OPK-3
142	Visualization helps to understand the structure of clusters and their separateness	Why visualize the clustering result in 2D space?	OPK-3
143	The quality of clustering can be assessed using silhouette, intra-cluster distances and	How can one evaluate the quality of clustering without class labels?	OPK-3

	other metrics.		
144	A) Presence of intersecting or elongated clusters	In what cases might the KMeans algorithm work inefficiently? A) Presence of intersecting or elongated clusters B) Real features C) The same number of objects in the classes D) A small number of features	OPK-3
145	The process of transforming data into a lower-dimensional space while preserving information content	What is dimensionality reduction?	OPK-3
146	A) Principal Component Analysis	How is deciphered abbreviation PCA? A) Principal Component Analysis B) Partial Classification Algorithm C) Primary Clustering Approach D) Projection of Categorical Attributes	OPK-3
147	Transformation of original features into new, orthogonal components with maximum variance	What does PCA do with features?	OPK-3
148	B) Preserve as much data variance as possible	What is the main goal of the PCA method? A) To simplify the model B) To preserve as much data variance as possible C) To reduce precision D) To eliminate categorical features	OPK-3
149	D) Number of principal components	What can be used as a parameter in dimensionality reduction using PCA? A) Learning speed B) Number of trees C) Kernel type D) Number of principal components	OPK-3
150	Yes, for example, t-SNE or UMAP are nonlinear methods for visualizing high-dimensional data	Are there nonlinear methods for dimensionality reduction? Give an example.	OPK-3
151	t-SNE preserves local data structure well and visualizes clusters	What is the advantage of t-SNE method for data visualization?	OPK-3
152	To identify data structure, clusters, outliers, and to evaluate the training of models	Why use data visualization after dimensionality reduction?	OPK-3
153	A) High dimensionality can hinder model training and visual analysis	Why is it important to reduce dimensionality before visualization? A) High dimensionality can hinder model	OPK-3

		training and visual analysis B) To increase training time C) To increase variance D) To replace all features with categorical ones	
154	Scatter plot, color coding, class labels	How can one visualize the result of dimensionality reduction in 2D?	OPK-3
155	Yes, dimensionality reduction can make data easier to interpret and speed up training.	Can dimensionality reduction improve model performance? Explain.	OPK-3
156	A) Percentage of explained variance	What is used to evaluate the informativeness of the principal components in PCA? A) Percentage of explained variance B) Correlation coefficient C) F1-measure D) Root mean square error	OPK-3
157	Hyperparameters are parameters that are set before training the model and are not changed during the training process.	What are hyperparameters in machine learning?	OPK-3
158	A) learning_rate, max_depth, n_estimators	Which from following are hyperparameters models gradient boosting ? A) learning_rate, max_depth, n_estimators B) precision, recall, accuracy C) weight_1, weight_2 D) fit, transform	OPK-3
159	Finding a combination of parameters that will give the best results for the model on the validation data	What is the purpose of hyperparameter selection?	OPK-3
160	C) GridSearchCV	Which method uses exhaustive search of all possible hyperparameter combinations? A) TrainTestSplit B) PCA C) RandomForest D) GridSearchCV	OPK-3
161	GridSearchCV and RandomizedSearchCV	Name two methods for automatic hyperparameter selection implemented in scikit-learn	OPK-3
162	GridSearchCV iterates over all combinations, while RandomizedSearchCV iterates over random subsets .	What is the difference between GridSearchCV and RandomizedSearchCV?	OPK-3

163	Cross-validation is used to evaluate the quality of a model under different hyperparameter values.	How are hyperparameter selection and cross-validation related?	OPK-3
164	D) Model quality on validation data	On what basis is the optimal set of hyperparameters selected? A) Minimum number of features B) Speed of training C) Level of overfitting D) Quality of the model on validation data	OPK-3
165	Overfitting is possible if the parameters are selected based on the test sample.	Can hyperparameter tuning lead to overfitting? Explain.	OPK-3
166	A) Selected before the model training begins	When are hyperparameters tuned? A) Before model training B) After training C) During forecasting D) During data normalization	OPK-3
167	Hyperparameter tuning helps improve the generalization ability of the model	How does proper hyperparameter tuning affect model quality?	OPK-3
168	Yes, especially when using limited computing resources and large parameter spaces	Does it make sense to use RandomizedSearchCV instead of GridSearchCV? Explain.	OPK-3

2. Description of the rating scale

Students' performance in each discipline is assessed during ongoing monitoring and midterm assessment.

3. Competency assessment criteria*

3.1. Assessment of OPK-1 competence

The OPK-1 competency is aimed at developing the ability to identify, formulate, and solve relevant problems in applied and computer mathematics that arise when developing machine learning software solutions in Python. This competency is developed through studying the mathematical foundations of machine learning, data analysis, model construction and evaluation, and the practical implementation of machine learning algorithms using modern software libraries and tools. Mastering this competency requires an understanding of the relationship between the mathematical formulation of a problem, the choice of a solution method, and the software implementation of the corresponding algorithm.

The criteria for achieving OPK-1 include:

- knowledge of the main problems of applied and computer mathematics arising in data analysis and machine learning;
- ability to formulate mathematical problems based on applied data processing problems;
- ability to select and justify mathematical methods for solving machine learning problems;
- understanding the principles of operation of mathematical models used in regression, classification, clustering and data dimensionality reduction;
- skills in applying mathematical methods in software implementation of machine learning algorithms in the Python language;
- ability to analyze the results of computational experiments and evaluate the correctness of the solutions obtained.

Competence is assessed through solving practical problems, completing laboratory work on programming machine learning algorithms, testing, analyzing cases, interpreting the results of computational experiments, and completing individual and group projects related to the development and research of machine learning models.

3.2. Assessment of OPK-3 competence

The OPK-3 competency is aimed at developing the ability to independently develop applied software tools for solving data analysis and machine learning problems using modern information technologies, software platforms, libraries, and network resources. This competency is developed through mastering the Python language and specialized libraries for data analysis and machine learning, as well as through practical and project-based assignments related to developing software solutions for data processing, model training, and results analysis.

The criteria for achieving OPK-3 include:

- knowledge of modern information technologies, software platforms and libraries used for developing machine learning systems;
- ability to find and reasonably select software, network resources and technological solutions for the implementation of applied tasks;
- ability to independently develop application software in Python to solve data analysis and machine learning problems;
- understanding the principles of constructing software solutions using the NumPy, pandas, matplotlib, scikit-learn libraries and other modern tools;
- skills in integrating software components, data processing and automation of computing processes;
- ability to evaluate the quality of the developed software product and ensure its compliance with the stated requirements.

Competence is assessed through laboratory work on Python programming, solving practical problems in data analysis and machine learning, developing software modules and applications, testing the created software tools, completing individual and group projects, as well as defending development results and demonstrating the functionality of the created software solutions.

3.3. Assessment of the level of development

The level of achievement of competencies is assessed on a four-point scale:

- **2 (low)** - competence is not developed, actions are fragmented or incorrect;
- **3 (sufficient)** - competence is formed at a basic level, there are inaccuracies;
- **4 (good)** - confident application of knowledge and skills, minor errors were made;
- **5 (high)** – competence is fully developed, decisions are justified, results are sustainable.

The OPC-1 and OPC-3 competencies are developed in close conjunction, ensuring that students are prepared to solve real-world problems in the field of data analysis and intelligent technologies.