

MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE RUSSIAN FEDERATION
Federal State Autonomous Educational Institution for Higher Education
«North-Caucasus Federal University»

APPROVED BY

Dean of FPE of NCFU

Candidate of Technical Sciences,

Associate Professor A.E. Verisokin

ASSESSMENT TOOLS FOR THE DISCIPLINE (MODULE)
« Machine Learning and Data Analytics for Power Engineering »

Field of study	<u>13.04.02 Electric Power Engineering and Electrical Engineering</u>
Specialization (Profile)	<u>Training Program of Electrical Engineering</u>
Year of commencement of studies	<u>2026</u>
Mode of study	<u>full-time</u>
Implemented in semester(s)	<u>2</u>

Foreword

1. Purpose: This assessment tools (AT) document is intended to assess the level of competency formation during ongoing progress control and intermediate assessment of students in the discipline " Machine Learning and Data Analytics for Power Engineering ".

2. The AT is an appendix to the syllabus of the discipline "Machine Learning and Data Analytics for Power Engineering".

3. Developer: A.A.Stepanova, Candidate of Technical Sciences, Associate Professor at the Department of Automated Electric Power Systems and Power Supply

4. The AT has undergone expert review.

Members of the expert group:

Chairperson: Yu. G. Kononov, Doctor of Technical Sciences, Professor, Head of the Department of Automated Electric Power Systems and Power Supply.

Committee members:

D.A. Kostyukov, Candidate of Technical Sciences, Associate Professor, Department of Automated Electric Power Systems and Power Supply.

A.V. Petrov, Candidate of Technical Sciences, Associate Professor, Department of Automated Electric Power Systems and Power Supply.

Representative of the employer organization:

S.A. Filippov, Candidate of Technical Sciences, Head of the Leninsky District Department of the Stavropol Interdistrict Branch of the PJSC «Stavropolenergosbyt».

Expert opinion: The AT is recommended for assessing the level of competency formation during ongoing progress control and intermediate assessment of students in the discipline "Machine Learning and Data Analytics for Power Engineering ".

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

1. Description of indicators and criteria for assessment at various stages of their formation, description of grading scales

Competency(ies), indicator(s)	Levels of competency formation			
	Minimum level not achieved (Unsatisfactory) – 2 points	Minimum level (Satisfactory) – 3 points	Medium level (Good) – 4 points	High level (Excellent) – 5 points
PC-3. Capable of managing the power system operating mode of an electrical power supply system using digital technologies				
Learning outcomes for the discipline (module): Indicator: ID-3 PC-3. Applies knowledge of digital technologies in the processes of generation, transmission, and distribution of electrical energy;	Fails to apply machine learning and Python analytics to power system processes; lacks basic knowledge of data processing tools in the energy sector; cannot solve the assigned task.	Applies basic machine learning tools to simple forecasting or diagnostic tasks in power systems; requires instructor guidance when dealing with complex energy data scenarios.	Confidently applies standard machine learning pipelines to manage power processes; correctly selects and uses predictive technologies for typical load forecasting and diagnostics with only minor inaccuracies.	Expertly applies advanced machine learning and MLOps technologies to optimize power systems; proposes innovative, well-justified solutions for complex forecasting, diagnostics, and industrial deployment..

ASSESSMENT TOOLS FOR CHECKING THE LEVEL OF COMPETENCY FORMATION

Task No.	Correct Answer	Question Content	Competency
1.	2	Which Python library is the primary tool for working with tabular data and performing time series aggregation operations in the energy sector? 1. Matplotlib 2. pandas 3. NumPy 4. SciPy	PC-3
2.	3	Which type of database is most suitable for storing high-frequency SCADA telemetry data from millions of power system sensors? 1. Relational database (PostgreSQL) 2. Graph database (Neo4j) 3. Time-series database (InfluxDB, TimescaleDB) 4. Document-oriented database (MongoDB)	PC-3
3.	3	Which Exploratory Data Analysis (EDA) method is most effective for identifying seasonal patterns in hourly electricity consumption data? 1. Distribution histogram 2. Scatter plot 3. Time series decomposition (seasonal decomposition) 4. Correlation matrix	PC-3

4.	2	Which of the following features is the most important for load forecasting in hot weather conditions? 1. Day of the week 2. Air temperature 3. Moon phase 4. Month number	PC-3
5.	2	Which technique should be applied to handle gaps in telemetry data if the gaps are short-term (a few minutes)? 1. Deleting rows with missing values 2. Linear interpolation 3. Replacing with zero values 4. Replacing with the annual average	PC-3
6.	2	Which model is most suitable for load forecasting considering daily and weekly seasonality? 1. Linear regression 2. SARIMA (Seasonal ARIMA) 3. Logistic regression 4. K-Means	PC-3
7.	2	Which machine learning algorithm belongs to gradient boosting methods and is widely used for forecasting solar power plant generation? 1. Random Forest 2. XGBoost 3. K-Nearest Neighbors (KNN) 4. Support Vector Machine (SVM)	PC-3
8.	3	Which deep learning architecture is most effective for long-term time series forecasting considering long-term dependencies? 1. Multilayer Perceptron (MLP) 2. Convolutional Neural Network (CNN) 3. LSTM (Long Short-Term Memory) 4. Autoencoder	PC-3

9.	3	Which validation method is correct for evaluating the quality of a time series forecasting model to avoid "future data leakage"? 1. K-fold cross-validation 2. Stratified cross-validation 3. Temporal (rolling window) cross-validation 4. Leave-one-out cross-validation	PC-3
10.	3	Which metric is most informative for evaluating load forecast accuracy if it is important to understand the relative error in percentage terms? 1. MAE (Mean Absolute Error) 2. MSE (Mean Squared Error) 3. MAPE (Mean Absolute Percentage Error) 4. R² (Coefficient of Determination)	PC-3
11.	2	Which unsupervised machine learning method is most suitable for automatic anomaly detection in transformer equipment operation? 1. Linear regression 2. Isolation Forest 3. Logistic regression 4. Gradient boosting	PC-3
12.	1	Which clustering algorithm is used to segment electricity consumers by consumption profiles? 1. K-Means 2. Random Forest 3. Gradient Boosting 4. Linear Regression	PC-3
13.	1	Which model interpretability method allows evaluating the contribution of each feature to the model's prediction for a specific forecast? 1. SHAP (SHapley Additive exPlanations) 2. PCA (Principal Component Analysis) 3. K-Means 4. Cross-validation	PC-3

14.	2	Which tool is designed for experiment tracking, model versioning, and ensuring reproducibility in MLOps? 1. Jupyter Notebook 2. MLflow 3. pandas 4. Matplotlib	PC-3
15.	2	Which task relates to power system optimization using machine learning? 1. Consumer clustering 2. Forecasting optimal load dispatch among generators 3. Data visualization 4. Data cleaning	PC-3
16.		What specific temporal and meteorological features would you engineer for an hourly power load forecasting model?	PC-3
17.		To identify hidden trends and cyclical patterns in hourly electricity consumption data during Exploratory Data Analysis (EDA), the method of time series _____ is applied.	PC-3
18.		The process of creating new informative features based on raw telemetry data (e.g., extracting moving averages or statistics for the previous day) is called _____.	PC-3
19.		For load forecasting taking into account explicit daily and weekly seasonality, the most suitable algorithm from the class of classical statistical models is _____.	PC-3
20.		The recurrent neural network architecture specifically designed to solve the vanishing gradient problem and effectively capture long-term dependencies in time series is _____.	PC-3
21.		To correctly evaluate the quality of a time series forecasting model and prevent "future data leakage", _____ cross-validation is used instead of standard K-fold.	PC-3
22.		The unsupervised machine learning algorithm that effectively isolates outliers and is most suitable for automatic anomaly detection in transformer equipment operation is _____.	PC-3

23.		The interpretability method based on cooperative game theory that allows evaluating the contribution of each feature to a specific prediction of a "black box" model is _____.	PC-3
24.		The predictive equipment diagnostics task aimed at forecasting the time until a critical failure occurs is called _____ estimation.	PC-3
25.		The open-source platform for managing the machine learning lifecycle (MLOps), used for experiment tracking, model versioning, and ensuring reproducibility, is _____.	PC-3
26.		When deploying ML models into real-time power system control systems, a critically important operational constraint is computational _____, which is the minimum time required for the model to generate a forecast.	PC-3
27.	1-B, 2-D, 3-A, 4-C	Match the Python library with its primary purpose in energy data analytics: 1. pandas 2. scikit-learn 3. TensorFlow 4. Matplotlib A. Training and building deep neural networks (LSTM, CNN) B. Processing and analyzing tabular data, working with time series C. Data visualization and plotting D. Implementation of classical machine learning algorithms and pipelines	PC-3
28.	1-B, 2-D, 3-A, 4-C	Match the forecasting method with its characteristic: 1. SARIMA 2. Prophet 3. XGBoost 4. LSTM A. Ensemble method based on decision trees with gradient boosting B. Statistical model accounting for seasonality and autoregression C. Recurrent neural network for long-term dependencies D. Facebook model for forecasting with multiple seasonality	PC-3
29.	1-C, 2-A, 3-D, 4-B	Match the neural network architecture with its application in power engineering: 1. CNN (Convolutional Neural Network) 2. LSTM 3. Transformer 4. MLP (Multilayer Perceptron) A. Time series forecasting with memory of long-term dependencies B. Classification and regression for tabular energy data C. Extracting local patterns from equipment vibration signals D. Modern architecture for long-horizon forecasting with attention mechanism	PC-3

30.	1-B, 2-A, 3-D, 4-C	Match the machine learning method with the equipment diagnostics task: 1. K-Means 2. Isolation Forest 3. PCA 4. RUL Regression A. Anomaly detection by isolating outliers in SCADA data B. Consumer segmentation by consumption profiles C. Forecasting time until critical equipment failure D. Dimensionality reduction of vibration data for state visualization	PC-3
31.	1-C, 2-D, 3-A, 4-B	Match the metric with its interpretation: 1. MAE 2. RMSE 3. MAPE 4. R^2 A. Mean Absolute Percentage Error, convenient for business interpretation B. Coefficient of Determination, shows proportion of explained variance C. Mean Absolute Error in units of the target variable D. Root Mean Squared Error, sensitive to large outliers	PC-3
32.	1-C, 2-B, 3-A, 4-D	Match the interpretability method with its description: 1. SHAP 2. LIME 3. Feature Importance 4. Partial Dependence A. Global feature importance based on impurity reduction in trees B. Local interpretation by approximating the model around the prediction C. Explanation based on cooperative game theory for each prediction D. Visualization of the effect of a single feature on model prediction	PC-3
33.	1-B, 2-A, 3-D, 4-C	Match the MLOps tool with its function: 1. MLflow 2. Docker 3. FastAPI 4. Airflow A. Model containerization for reproducible deployment B. Experiment tracking, parameters, and model artifacts C. Orchestration of machine learning pipelines D. Creating REST API for integrating models into industrial systems	PC-3

34.	1-B, 2-C, 3-D, 4-A	Match the data type with its example in a power system: 1. Time series 2. Geospatial data 3. SCADA telemetry 4. Graph data A. Electrical network topology and connections between nodes B. Hourly electricity consumption of a substation C. Coordinates of wind turbines and local solar maps D. Real-time voltage, current, and power from sensors	PC-3
35.	1-D, 2-C, 3-A, 4-B	Match the task with the type of machine learning: 1. 24-hour load forecast 2. Cyberattack detection at a substation 3. Consumer segmentation 4. Remaining useful life forecast A. Unsupervised learning (clustering) B. Supervised learning (regression) C. Supervised learning (classification) D. Supervised learning (time series forecasting)	PC-3

2. Description of the grading scale

Within the framework of the rating system, student performance in each discipline is assessed during ongoing progress control and intermediate assessment. The rating system for assessing student knowledge is based on the use of a set of control measures (checkpoints) for testing the material covered, optimally located throughout the entire time interval of studying the discipline. The principles of the rating system for assessing student knowledge are based on the provisions described in the "Regulations on the organization of the educational process based on the rating system for assessing student knowledge" at North Caucasus Federal University.

The rating assessment system is not provided for students studying in master's degree programs, or for students studying in bachelor's degree programs by correspondence or part-time (evening) forms of study.

3. Criteria for assessing competencies

The grade "Excellent" is awarded to a student who has deeply and firmly mastered the program material, presents it exhaustively, consistently, clearly, and logically coherently, is able to closely link theory with practice, handles tasks, questions, and other types of knowledge application with ease, and does not struggle to answer when assignments are modified.

The grade "Good" is awarded to a student who has a solid command of the material, presents it competently and substantively without making significant inaccuracies in answering questions, correctly applies theoretical principles to practical questions and tasks, and possesses the necessary skills and techniques for their execution.

The grade "Satisfactory" is awarded to a student who has knowledge of only the basic material but has not mastered its details, makes inaccuracies, insufficiently correct formulations, and violates logical sequence in presenting the program material.

The grade "Unsatisfactory" is awarded to a student who does not know a significant part of the program material and makes substantial errors.