

Assignment 1

TASKS:

1. Load the dataset from Kaggle or any open source website. Datasets related to Kazakhstan are encouraged. The project must be completed individually. The selected dataset must be unique and cannot be reused by other Master's students within the same lecture cohort.
2. Train the dataset using the following machine learning algorithms: Ridge, Lasso, Elastic Net, k-Nearest Neighbors (KNN) Regression, Extra Trees Regression, Adaptive Boosting (Classic Boosting), Gradient Boosting Regression, High-Performance Boosting (XGBoost & LightGBM), and Specialized Boosting (CatBoost & HistGradientBoosting). Use k-fold validation. $k = 10$.
3. Evaluate the performance of machine learning models.
4. Fill in the learning results in Table 1.
5. Analyze the results.
6. Present as a report or an article.

Table 1.

Algorithm	Number of features	Number of targets	k-fold validation	Accuracy_score/Precision /Recall/F1 score (for classification tasks)	RMSE/R ² (for regression tasks)
Ridge					
Lasso					
Elastic Net					
KNN Regression					
Extra Trees Regression					
Adaptive Boosting (Classic Boosting)					
Gradient Boosting Regression					
High-Performance Boosting (XGBoost & LightGBM)					
CatBoost					
HistGradientBoosting					

Evaluation criteria:

90+ points: Use of a dataset relevant to Kazakhstan.

The article includes sections such as Introduction, Literature Review, Materials and Methods, Results and Discussion. References should be 10-15.

The Introduction section must clearly formulate the research problem and justify its relevance through a systematic Literature review of publications from the last five years.

The article, excluding the title, abstract, keywords (6 items), conclusion and bibliography, must contain 3500 words.

The article's originality should be greater than 97 percent; no artificial intelligence was used.

80 points: A report containing a link to the dataset, source code, and table 1.

60 points: Not all methods were applied, or the dataset was poorly chosen.