Theoretical questions

1.Supervised learning problem statement. Task examples.
2. Write down the linear regression formula. Which quality metrics can be used for regression?
3. What is a gradient? Which property of gradient is used for function minimization?
4. What is regularization? How is it applied in the case of linear models?
5. What is validation data? Describe the cross-validation process.
6. Write down the formula for the linear classification model. Which loss functions can be used for its training?
7. Binary classification metrics: accuracy, F-score, precision, recall, ROC-AUC.
8. Describe the k-nearest-neighbors algorithm for classification. Which hyperparameters influence the prediction?
9. What is a decision tree? Describe how the decision tree makes predictions for classification/regression.

10. What is overfitting and underfitting? How to reduce overfitting?
11. What is the main idea of composition methods (bagging, random forest)? Describe the bagging process (with decision trees).
12. Write down the formula for final prediction in Random Forest and Gradient Boosting (how to combine basic algorithms predictions).
13. What is a neural network? Name several layers and activation functions that are used in neural networks.
14. Describe the back-propagation algorithm.
15. What is a convolution? How is the convolutional layer applied to an image?
16. What is a recurrent neural network? How does it work with a sequence of inputs?
17. Which text preprocessing methods do you know? What is a word co-occurrence matrix?
18. K-Means clustering algorithm. How does it work?