Copy and sign the following statement: "The results presented here are my own work. As instructed, I am not receiving any help from anybody nor using any literature, including course notes."

Write clearly and justify your answers. Incorrect statements in essay-type questions will incur negative points.

Time limit: 90 min.

(1) [5 points] *(essay-type question)* We have the following learning algorithms:

- **DT**: A single decision tree.
- **kNN**: k-nearest neighbors with Euclidean distance.
- **BOOST**: Boosted decision stumps (single-level trees).
- **FFNN**: Feed-forward neural network trained using backpropagation.

If you think some key detail about a learner is missing, you may assume which parameter was used or how it was implemented. However, the assumption must be sensible - something that could reasonably used as a default in some implementation (do not assume behavior that makes the learner clearly behave poorly or is not even feasible to implement). And, whatever you assume, applies to all the below questions.

Order the above learners from best to worst with respect to the following dimensions:

a. The time complexity of learning (consider both the number of observations and the number of input variables).

b. Time complexity of making a prediction (consider both the number of observations and the number of input variables).

c. Interpretability/ease of understanding the model and its predictions.

d. Difficulty of implementation if you had to implement it from scratch without third-party libraries.

Justify your answers.

(2) [5 points] *(essay-type question)* One technique for explaining a prediction model is to approximate the model with a decision tree (that is, find the decision tree that best mimics the model's predictions) and then explain it as if the model were the decision tree. List 2 other general explanation techniques that can be applied to prediction models. Discuss the advantages and disadvantages of these three techniques.

(3) [5 points] *(essay-type question)* Joe F. Random, a machine learning practitioner, is working on a classification problem. The constraints are as follows: He has data with 100 observations, 50 predictors (independent variables), and a binary class (dependent variable). He has to deliver a model that will be used in production without any possibility of modification. He can only choose between learner A or B, both of which have some continuous hyperparameters, and he can only use them in a black-box manner (train with some hyperparameters and use the trained model for predicting, repeating this an arbitrary number of times for different hyperparameters). Additionally, if he can't with some certainty conclude that learner B is better than learner A, learner A must be delivered.

Joe decided on the following training and evaluation process: He selected (uniformly at random) 20 different sets of hyperparameters for A and B. The estimated the performance of these 40 combinations of learner and hyperparameters using leave-one-out cross validation. Learner B with some set of hyperparameters had the best loss so he retrained learner B with those parameters on all 100 observations and delivered that model to production.

a. Does Joe's model evaluation process have any potential sources of bias (positive/negative?) or variance? That is, identify all sources of error when estimating true risk (how the delivered model is expected to perform in production) with Joe's performance estimate.

b. Are there any other potential flaws in Joe's learning process (how the models were trained and hyperparameters chosen)? Suggest a better learning and evaluation process that is still within the constraints of the problem.

(4) [5 points] Some machine learning-related mathematics:

a. Derive the Bayes estimator (in the continuous case) for quartic loss ($\ell(y, \hat{y}) = (y - \hat{y})^4$, where y is the true and $\hat{y}$ the predicted value).

b. Write the likelihood of a Poisson GLM with the logarithm link function.

c. Derive the posterior distribution for the Poisson likelihood with exponential prior.

Poisson PMF: $P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$. Exponential PDF: $p(x) = \lambda e^{-\lambda x}$.

The End.