**Machine Learning Homework**

For this homework I experimented with these models: Logistic Regression, Decision Tree, Random Forest, K Nearest Neighbors, Support Vector Machines and Neural Networks.

I split the data 60/40 on training/testing. For y data was categorical. For preprocessing I used LabelsEncoder to convert it to numerical data. X data was preprocessed using MinMaxScaler.

I started with the simplest one – Logistic Regression. After applying Grid Search to it, I received a very decent result for the model score=0.8880. The same results I received for Support Vector Machine Model also after applying Grid search. K Nearest Neighbor Model showed the lowest model score of 0.8308. Decision Tree Model shower the second worse result for the model score of 0.8501.

I expected that Neural Network Model will show the best results, but its model score was slightly worse than for Random Forest Model; 0.8934 and 0.8984 accordingly. For Neural Network Model for y data preprocessing I used OneHotEncoding on top of LabelsEncoder to convert the values to Boolean format. Additionally, for Neural Network Model I experimented with adding one more layer to the model, but it did not work out. The model score for training data was close to 0.96, but for testing data it was lower than with one-layer model. I decided, that model probably was over trained.

**Conclusion**: The best model is Random Forest Model, with model score=0.8984.