В качестве первого задания, я создал соревнование на Kaggle: https://www.kaggle.com/c/itmo2018ctml

Для участие в нём требуется пройти по этой ссылке: https://www.kaggle.com/t/b0c6b06a005c4ce888f42e9b2e15cd64

Это соревнование для обоих курсов. В качестве разминки и повторения. Вам требуется изучить какие-нибудь библиотеки по машинному обучению на ваш выбор. Например scikit-learn для python или WEKA для java. Постарайтесь как можно меньше реализовывать существующих алгоритмов (это вам предстоит ещё делать в последующих лабах. Цель задания научиться применять алгоритмы, принцип работы отдельных алгоритмов вы изучите далее по курсу.

В целом решение должно выглядеть так: Выберете несколько алгоритмов классификации и выбора признаков. Желательно каждый алгоритм учитывать несколько раз с различными гипер-параметрами. Разделите тренировочные данные на несколько частей для кросс-валидации. Используя кросс-валидацию найдите наилучшую комбинацию из классификатора и алгоритма выбора признаков. Составьте предсказание для тестового множества.