

# Лабораторная работа по модулю “Построение моделей”

В данной работе предлагается решить задачу мультиклассовой классификации для датасета флагов различных стран (<https://archive.ics.uci.edu/ml/datasets/Flags>)

1. Загрузить данные из папки Data Folder
2. Провести описательный анализ предложенных данных
3. Предложить стратегию по работе с категориальными переменными
4. Провести углубленный анализ данных (корреляции переменных, визуализация взаимосвязей)
5. Следует ли увеличивать/уменьшать размерность признакового множества (какой подход выдает лучшие результаты)
6. Следует ли балансировать классы путем oversampling/undersampling? (улучшает ли подход результаты)
7. Сделать кросс-валидацию данных с использованием подхода K-fold ( $n\_folds=3$ )
8. Решить задачу мультиклассовой классификации и предсказать религию страны (religion) по деталям ее флага
9. Рассчитать Feature Importances для переменных модели
10. Проверить качество классификации с использованием следующих метрик: Accuracy, F1-Score, Precision, Recall
11. Построить Confusion Matrix для ваших результатов
12. Загрузить ipython notebook с результатами работы на github репозиторий

\*Пункты, показавшиеся вам сложными, либо которые вы не знаете можно пропускать, но с пометкой в ноутбуке, что показалось сложным и непонятным в данном пункте для вас.