**Предварительная оценка успешности учащихся на курсе**

**Автор:** Дейкина Софья Алексеевна

**Наставник:** Светлова Виктория Сергеевна

**Аннотация**

**Цель:** создать программу, которая будет успешно прогнозировать результативность выполнения курса.

**Тип проекта:** прикладной (практико-ориентированный)

**Продукт проекта:** программа, которая предсказывает успешность выполнения курса

**Введение**

**Цель:** создать программу, которая будет успешно прогнозировать результативность выполнения курса. В этом заинтересованы авторы курсов, которые хотят определить ту группу людей, которые, действительно, могут закончить курс успешно, и, например, предложить им особые условия или работать с ними дополнительно.

**Проблема:** проблема нахождения оптимального метода оценки успешности учащихся на курсе.

**Актуальность:** прогноз результативности выполнения курса может помочь автору уже на небольшом сроке увидеть тех людей, которые в будущем смогут закончить курс успешно.

**Задачи:**

* предобработка данных
* выбор оптимального алгоритма для обучения
* реализация программного кода и обучения на выборке
* анализ полученных результатов

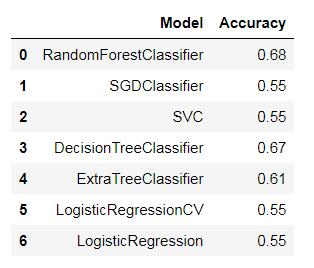
**Теоретическая часть**

**1 этап –** Предобработка данных

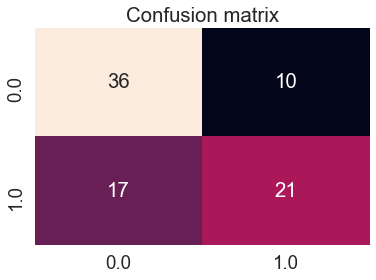
Для решения данного кейса были предоставлены разные данные по курсу. Задачей первого этапа было отобрать все нужные параметры и сформировать единый датасет.

**2 этап –** Выбор алгоритма

В ходе этого этапа был проведён сравнительный анализ существующих моделей. Результаты представлены в таблице:



Для решения этой задачи я выбрала алгоритм RandomForest.



**Практическая часть**

**3 этап –** Обучение и реализация

Для реализации проекта в качестве языка программирования был выбран – Python. Библиотекой для обучения модели – RandomForestClassifier. Библиотекой для работы с данными – Pandas, Numpy.

Алгоритм:

После обработки данных:

1. определить X и y
2. разбить на тестовую и тренировочную выборки
3. обучить модель
4. проанализировать результат

Ссылка на GitHub с исходным кодом:

<https://github.com/deisof/BV2021>

**Вывод**

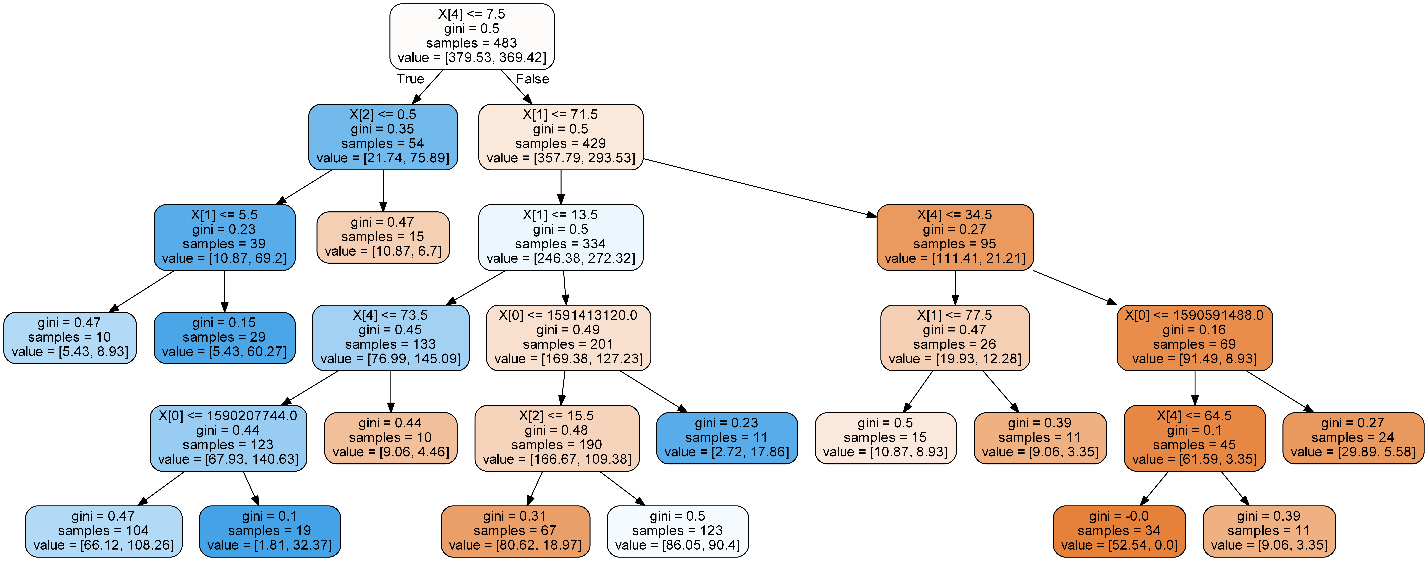
В данном задании нужно было выполнить работу по предсказанию успешности выполнения курса учеником.

В ходе работы была создана модель, которая предсказывает результативность выполнения курса учеником по результатам первой недели работы.

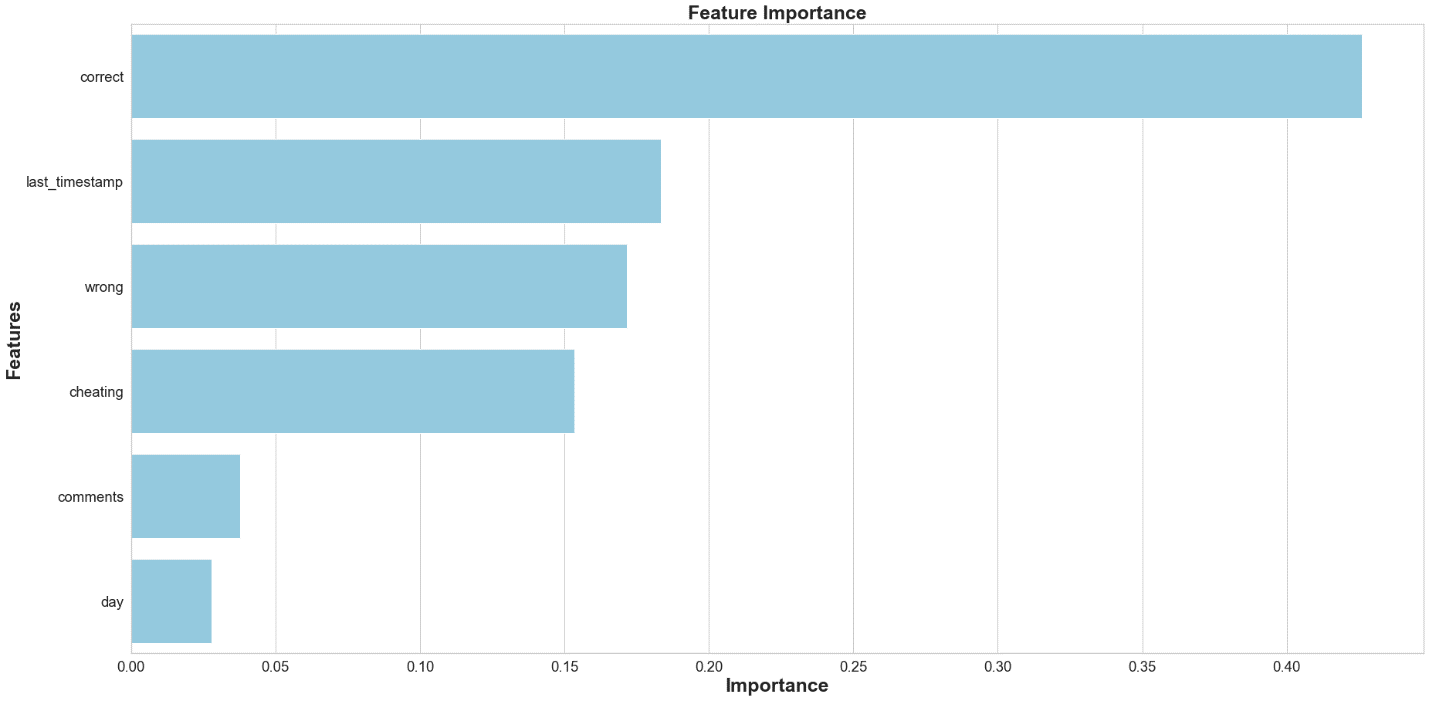
Итог:

Roc на тест 0.795

*Визуализация дерева решений*



*Важные параметры*



**Список литературы**

Курс Введение в Data Science и машинное обучение

<https://stepik.org/course/4852/syllabus>

Документация Randomforest

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>