



ML Boot Camp IV

Задача “с секретом”



Вячеслав Введенский



Вступление

mlbootcamp.ru

"Оценка производительности"

"Выход из он-лайн игры"

"Бинарные данные"

Задача "с секретом"

Python

или R

sklearn, matplotlib, jupyter notebook

Постановка задачи

Задача “с секретом”

223 признака

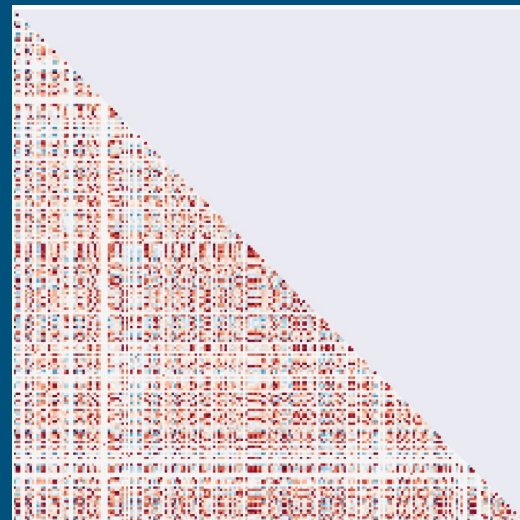
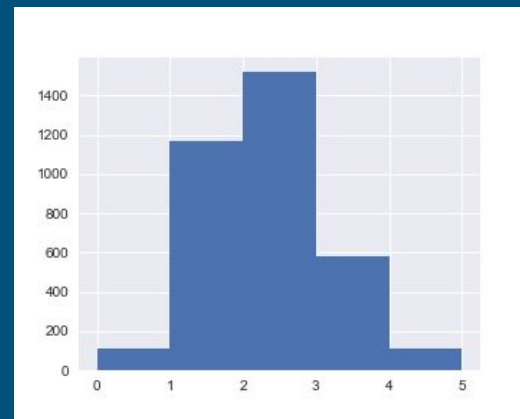
5 классов

train: 3489

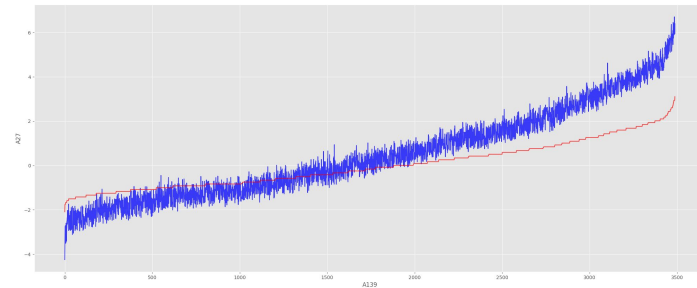
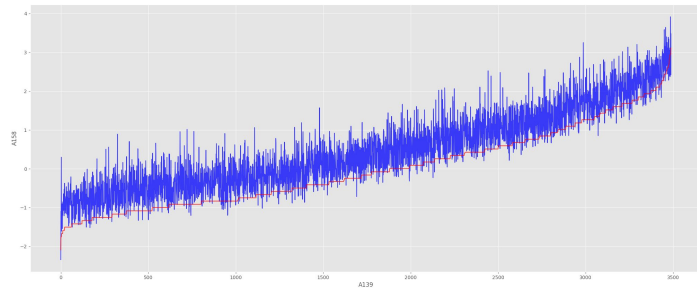
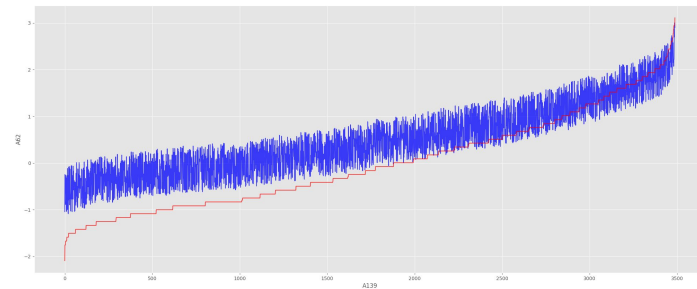
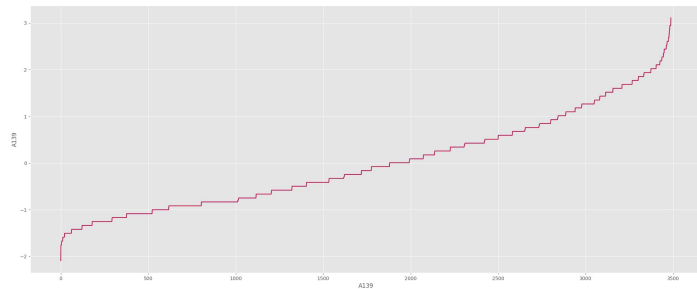
test: 2327 (60:40)

metrics: accuracy !

5 попыток в сутки



Отбор признаков



Отбор признаков

Метод пристального
всматривания

A11

A12

A77

A80

A97

A98

A116

A132

Линейные зависимости признаков

A139

A157

A183

A201

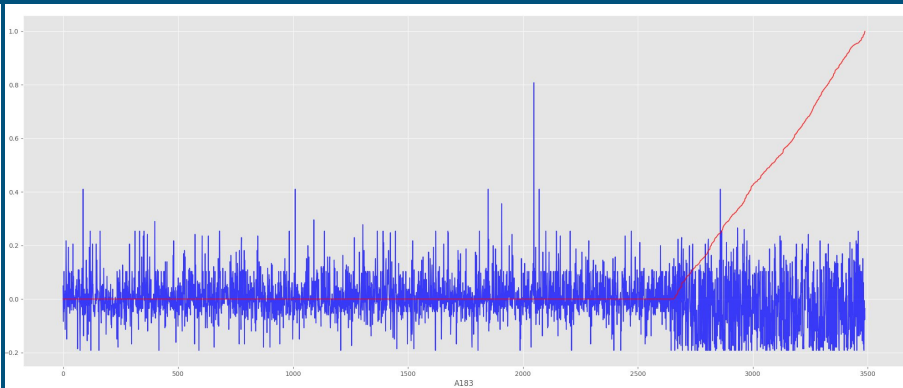
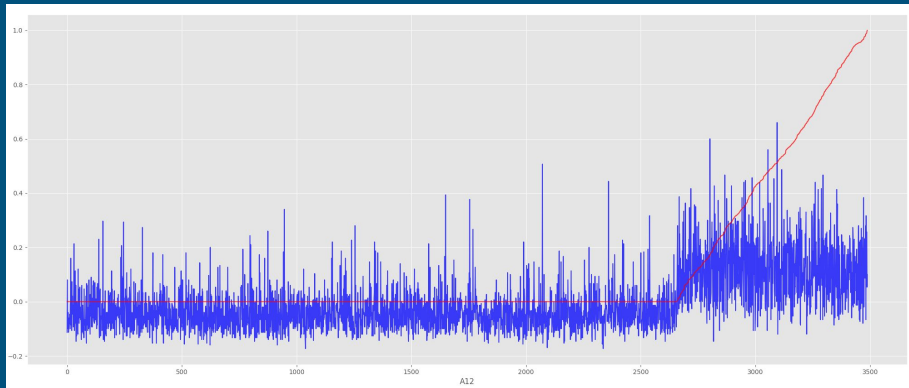
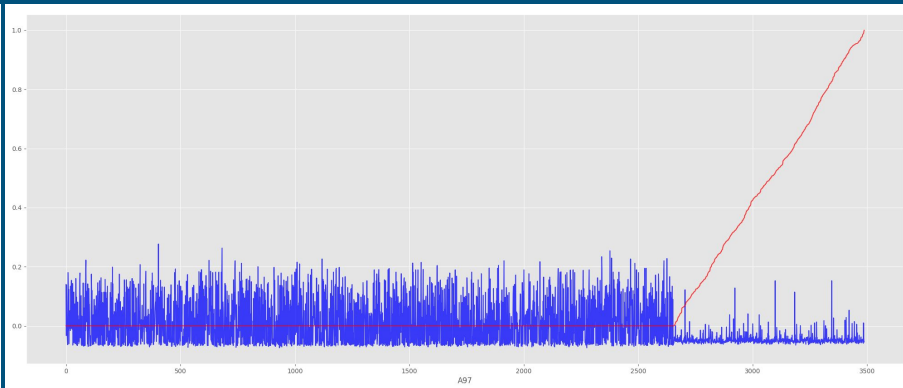
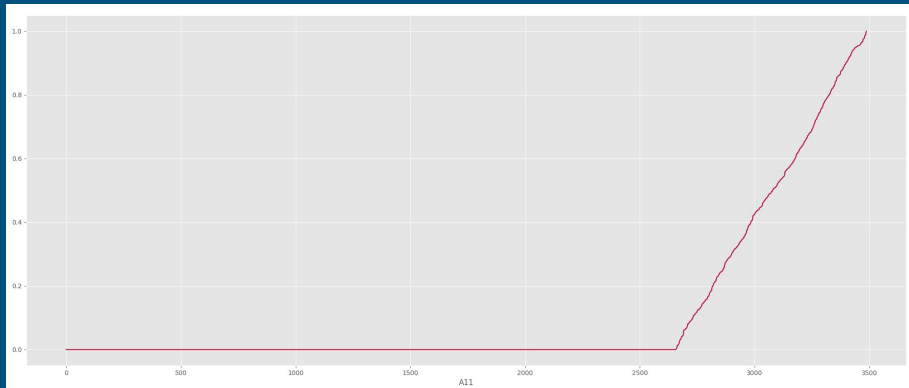
“Жадный” алгоритм

Sequential feature selection

Ассоциативные правила

На базе алгоритмов

A11



Работа с данными

Выбросы

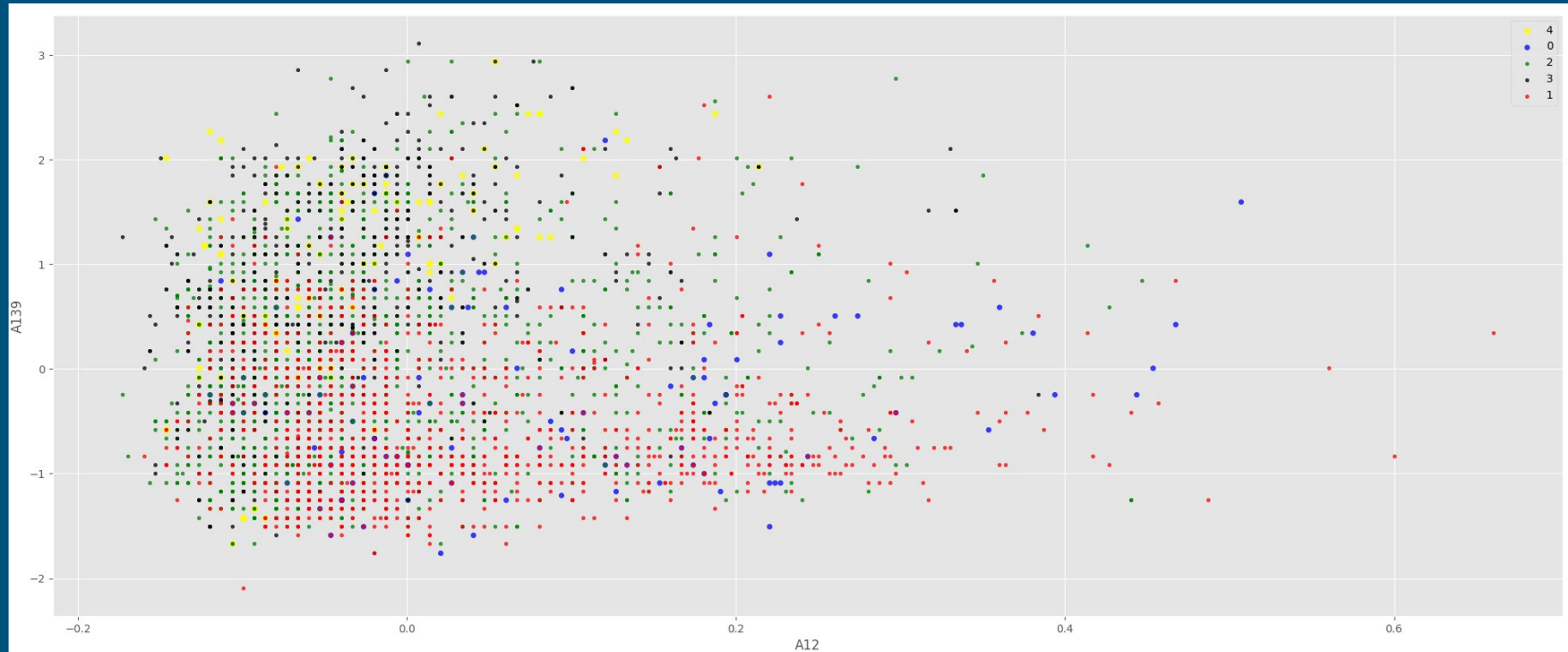
0, 4 - класс

A11 и 0, 4 класс

“Упорядочивание”

Поиск “грааля” и заблуждения

Общая картина



Генерация признаков

A12	A97	A132-PI*A157	A97-PI*A157	A201-PI*A97
A183-PI*A132	A183-PI*A116	A77-PI*A183	A12-PI*A183	
A139-PI*A12	A157-PI*A11	A157-PI*A77		
A77*A157+A116*A201				

Решение

Леса

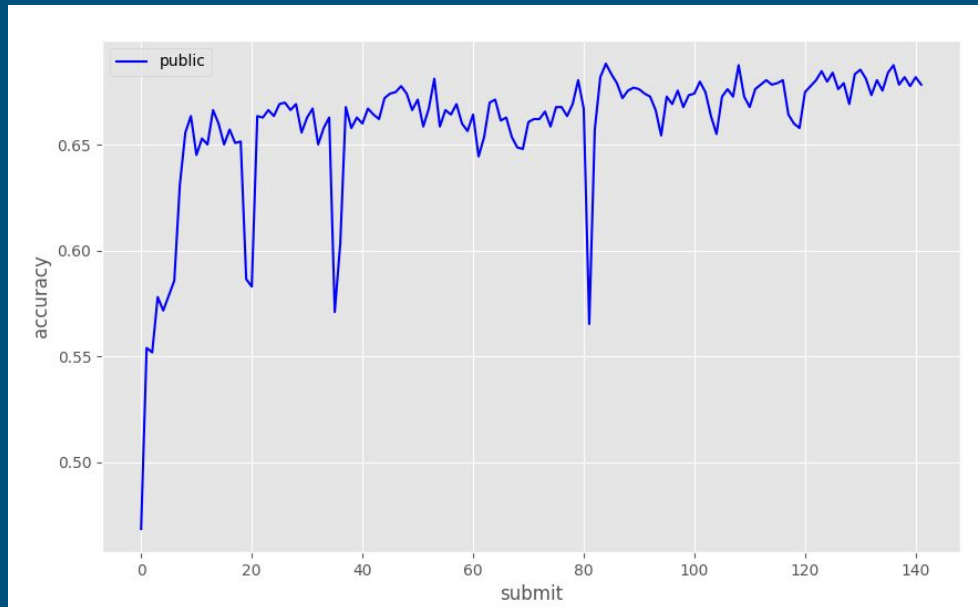
Соседи

XGBoost

Регрессии :(












Голосование

Голосование на разных
признаках














Результаты

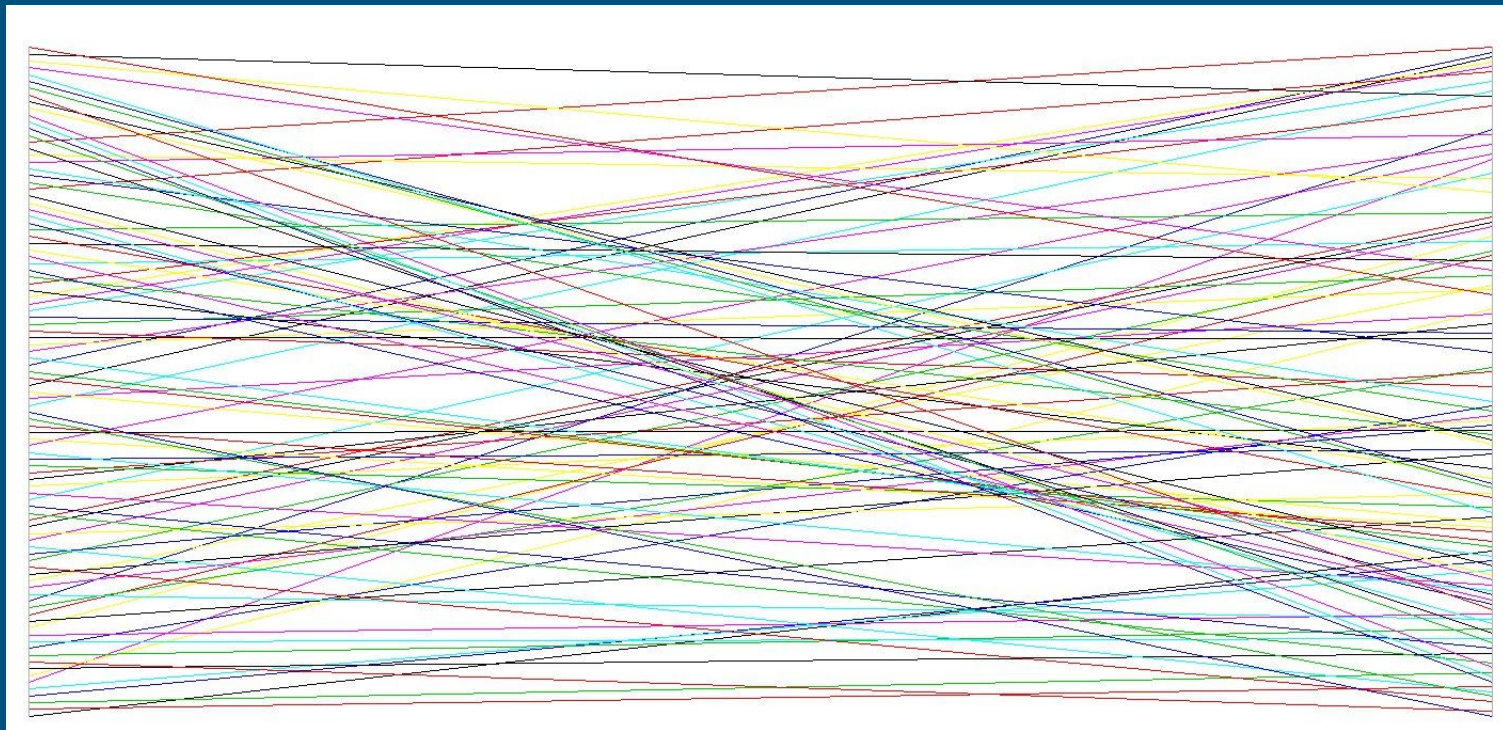
public

1		Валерий Бабушкин	0,6968198
2		Артём Дубинич	0,6925795
3		Александр Киселев	0,6918728
4		Вячеслав Введенский	0,6883392
5		Иван Тамгин	0,6840989
6		Олег Бессонов	0,6833922
7		Дмитрий Антипов	0,6833922
8		Денис Шевченко	0,6833922
9		Святослав Ковалёв	0,6833922
10		Мирас Амир	0,6833922
11		Алексей Тихонов	0,6826855

private

1		Александр Иванов	0,6622807
2		Святослав Ковалёв	0,6611842
3		Иван Черданцев	0,6600877
4		Dasha Chirkina	0,6600877
5		Артём Мазеев	0,6600877
6		Konstantin Nikolaev	0,6600877
7		Артём Дубинич	0,6600877
8		Владимир Иванов	0,6589912
9		Алексей Козловцев	0,6589912
10		Александр Киселев	0,6589912
11		Дмитрий Богачев	0,6589912

Результаты



Расшифровка задачи

archive.ics.uci.edu/ml/datasets/Wine+Quality

1 - fixed acidity (фиксированная кислотность)

2 - volatile acidity (летучая кислотность)

3 - citric acid (лимонная кислота)

4 - residual sugar (остаточный сахар)

5 - chlorides (хлориды)

6 - free sulfur dioxide (свободный диоксид серы)

7 - total sulfur dioxide (общий диоксид серы)

8 - density (плотность)

9 - pH (кислотность)

10 - sulphates (сульфаты)

11 - alcohol (спирт)

Output variable :

12 - quality (score between 0 and 10)

Спасибо за внимание !
