

# Tinkoff Data Science Challenge

---

СТАНИСЛАВ СЕМЁНОВ

2016

# Задача 1. Выбор кредита Tinkoff.ru

---

Tinkoff.ru работает с сетью магазинов электроники, в которой присутствуют и другие банки.

Таблица содержит данные о кредитах на покупку электроники, которые были одобрены Tinkoff.ru.

Необходимо предсказать, выберет ли покупатель кредит от Tinkoff.ru.

# Задача 1. Выбор кредита Tinkoff.ru

	client_id	gender	age	marital_status	job_position	credit_sum	credit_month	tariff_id	score_shk	education	living_region
0	1	M	48	MAR	UMN	59998,00	10	1.60	0,770249	GRD	КРАСНОДАРСКИЙ КРАЙ
1	2	F	28	MAR	UMN	10889,00	6	1.10	0,248514	GRD	МОСКВА
2	3	M	32	MAR	SPC	10728,00	12	1.10	0,459589	SCH	ОБЛ САРАТОВСКАЯ
3	4	F	27	DIV	SPC	12009,09	12	1.10	0,362536	GRD	ОБЛ ВОЛГОГРАДСКАЯ
4	5	M	45	MAR	SPC	16908,89	10	1.10	0,421385	SCH	ЧЕЛЯБИНСКАЯ ОБЛАСТЬ
5	6	F	37	MAR	SPC	26788,00	10	1.10	0,484203	SCH	СТАВРОПОЛЬСКИЙ КРАЙ
6	7	M	34	MAR	SPC	10317,00	6	1.50	0,276012	SCH	ОБЛ НИЖЕГОРОДСКАЯ
7	8	F	26	UNM	SPC	47878,00	10	1.10	0,512525	GRD	МОСКОВСКАЯ ОБЛ
8	9	M	23	UNM	SPC	37577,16	12	1.60	0,642267	SCH	ХАНТЫ-МАНСИЙСКИЙ АВТОНОМНЫЙ ОКРУГ - ЮГРА
9	10	F	32	UNM	SPC	26268,00	10	1.10	0,465026	GRD	КРАЙ СТАВРОПОЛЬСКИЙ
10	11	M	26	MAR	SPC	16793,00	14	1.00	0,445430	SCH	САНКТ-ПЕТЕРБУРГ
11	12	F	26	UNM	SPC	18936,00	10	1.32	0,455774	GRD	РЕСП. БАШКОРТОСТАН
12	13	M	37	MAR	SPC	42289,00	10	1.60	0,691609	SCH	ОБЛ АРХАНГЕЛЬСКАЯ

# Задача 1. Выбор кредита Tinkoff.ru

---

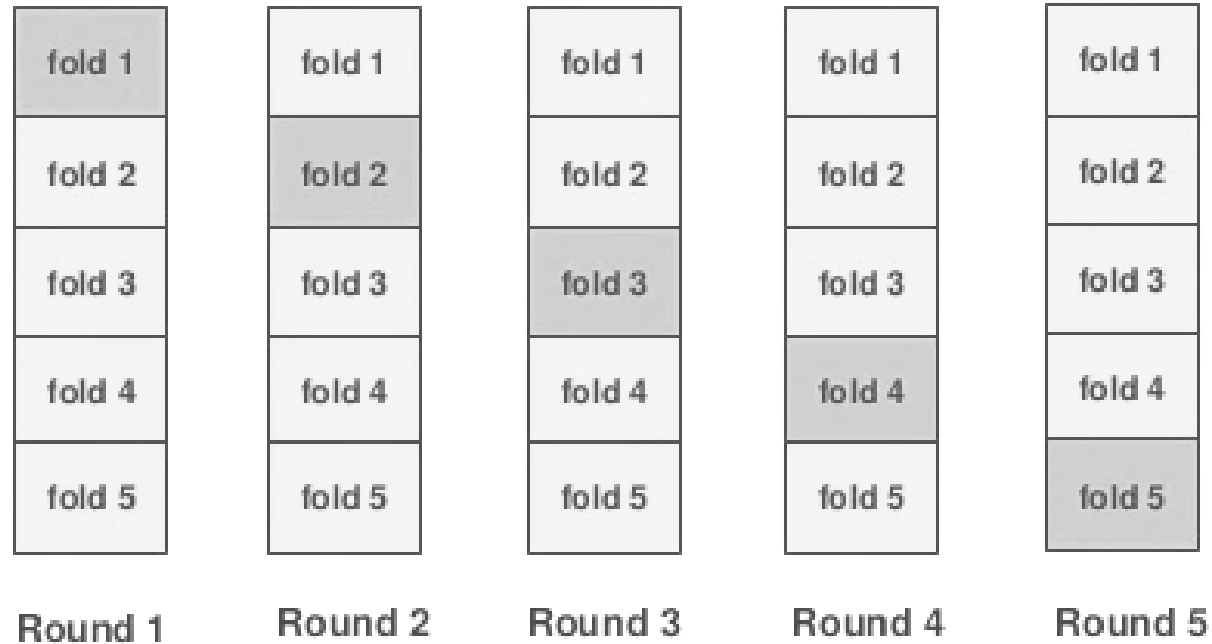
$$\text{Smoothed Likelihood} = \frac{\text{mean}(\text{target}) * \text{nrows} + \text{globalmean} * \text{alpha}}{\text{nrows} + \text{alpha}}$$

*globalmean* – среднее значение целевой переменной по всей выборке, *alpha* – коэффициент регуляризации

Город	Число строк	Likelihood	Smoothed Likelihood
Москва	100	0.74	0.72
Санкт-Петербург	50	0.66	0.64
Новосибирск	10	0.8	0.67
Ивановка	1	0.	0.57
Уфа	5	1.	0.68
Владивосток	0	-	0.6

# Задача 1. Выбор кредита Tinkoff.ru

---



## Задача 2. Восстановление позиции мерчанта

---

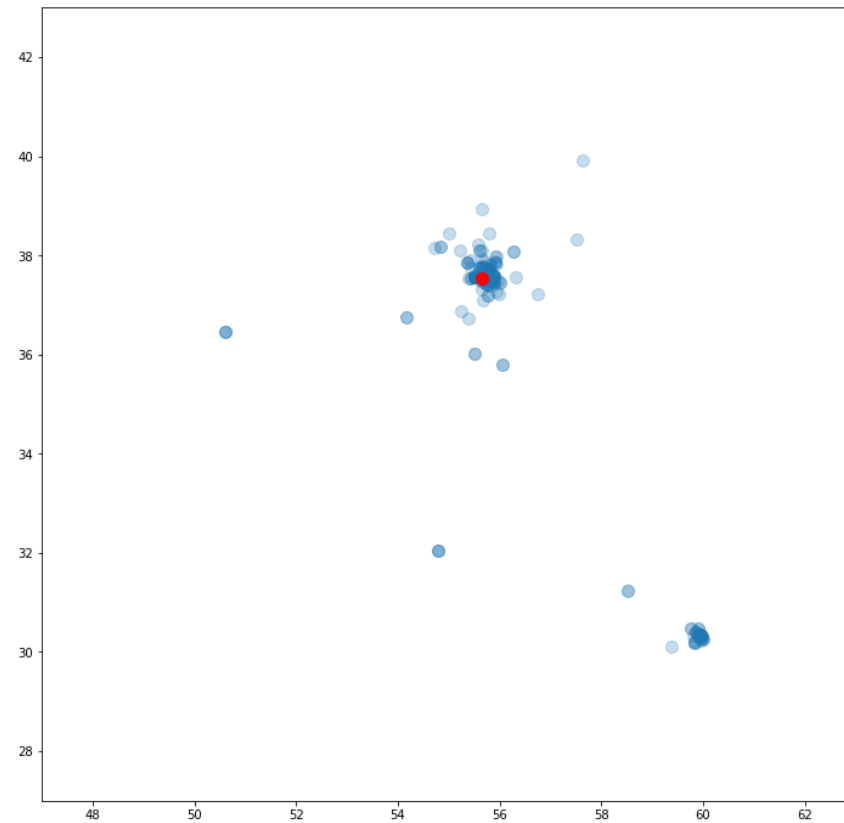
Мерчант — это место, принимающее платежи с использованием банковской пластиковой.

Для получения координат неизвестных мерчантов представители Tinkoff.ru в небольшой промежуток времени от покупки отправляют данные о своей геопозиции с мобильного телефона - производят чекин.

Необходимо по нескольким чекинам восстановить геопозицию мерчанта.

# Задача 2. Восстановление позиции мерчанта

---



## Задача 2. Восстановление позиции мерчанта

---

Строим обучающую выборку:

Объект – один чекин. Целевая переменная - если координата чекина входит в окрестность этого мерчанта.

Признаки: как много в окрестности находится других чекинов, расстояние до медианы чекинов, как много в окрестности находится других мерчантов, как он распределен по времени и т.д.

Обучаем XGBoost, затем для каждого мерчанта строим предсказания всех его чекинов и выбираем координаты того, где максимальная вероятность.



# Задача 2. Восстановление позиции мерчанта

---

Строим обучающую выборку:

Объект – это пара мерчант-мерчант. Целевая переменная - полное совпадение координат. Большая обучающая выборка - около 20 миллионов.

Признаки: число общих повторяющихся координат чекинов, времен, как далеко область одних чекинов отстоит от других и т.д.

Обучаем XGBoost, затем на тесте для каждого мерчанта строим пары со всеми мерчантами из трейна и если было какое-то сильно большое предсказание, то ставим в качестве ответа именно координату того мерчанта.

# Спасибо за внимание!

---

