

Выпускная квалификационная
работа на тему:

«Оценка дефолтности заемщиков кредитных организаций»

Выполнил студент 444 группы
Эль-Айясс Дани Валид

Научный руководитель:
Тихомиров Николай Петрович

Цель и задачи

Цель: построение скоринговой модели с помощью современных методов машинного обучения и формирование оптимального решения о выдаче кредита.

Задачи:

1. Изучить различные алгоритмы для построения скоринговой модели, их особенности и недостатки;
2. Рассмотреть основные подходы к оценке качества моделей;
3. Провести сопоставительный анализ всех методов и их результатов на основе метрик качества и изучить основные факторы, влияющие на кредитоспособность заемщика;
4. Обосновать выбор лучшего метода в рамках решения задачи кредитного скоринга и на его основе получить оценки вероятности дефолта для новых клиентов.

Информация о заемщиках

Характеристики заемщика

Номер клиента	Признак 1	...	Признак m
1	x_{11}	...	x_{1m}
...
n	x_{n1}	...	x_{nm}



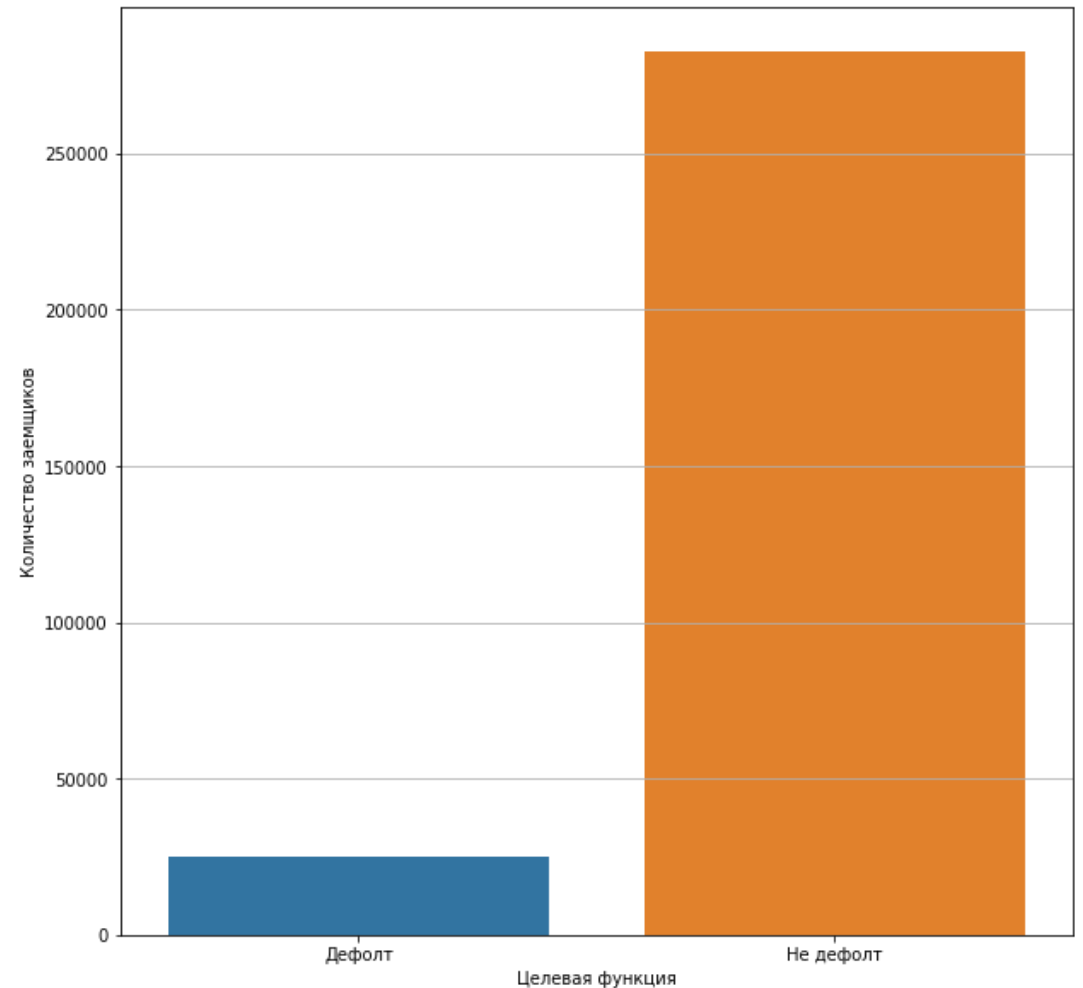
Дефолт

Признак m
y_1
...
y_m

Соотношение дефолтных и не дефолтных заемщиков

Количество всех заемщиков: 307511

Доля дефолтных заемщиков: 8%



Алгоритмы машинного обучения



Линейные модели:

- Логистическая регрессия
- Метод опорных векторов

Модели на деревьях решений:

- Случайный лес
- Градиентный бустинг

Другие модели:

- Метод k взвешенных соседей
- Наивный байесовский классификатор

Оценка моделей и сопоставительный анализ



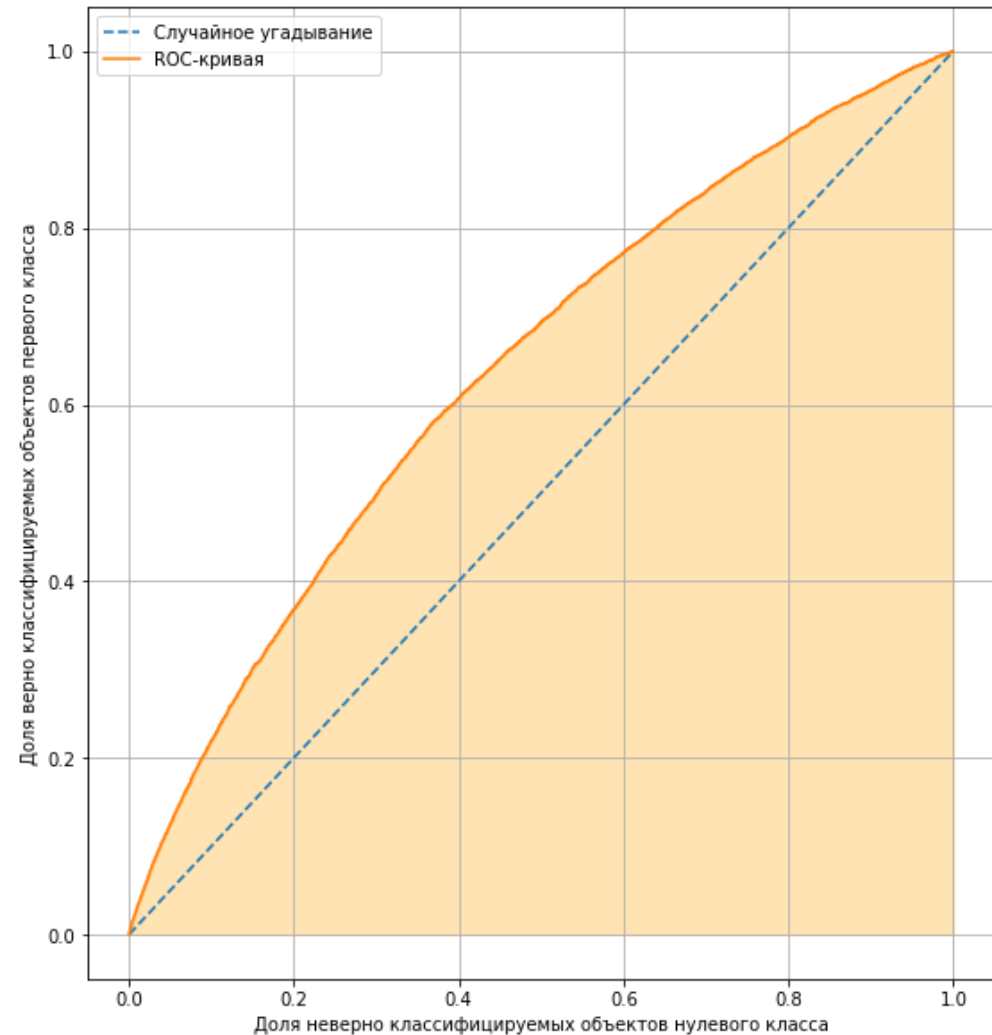
Алгоритм / Метрика	Площадь под ROC кривой	F -мера
Логистическая регрессия	0.61	0.19
Метод опорных векторов	0.61	0.19
Метод k ближайших взвешенных соседей	0.60	0.18
Наивный байесовский классификатор	0.59	0.18
Случайный лес	0.61	0.18
Градиентный бустинг	0.64	0.21

Интерпретация коэффициентов модели логистической регрессии

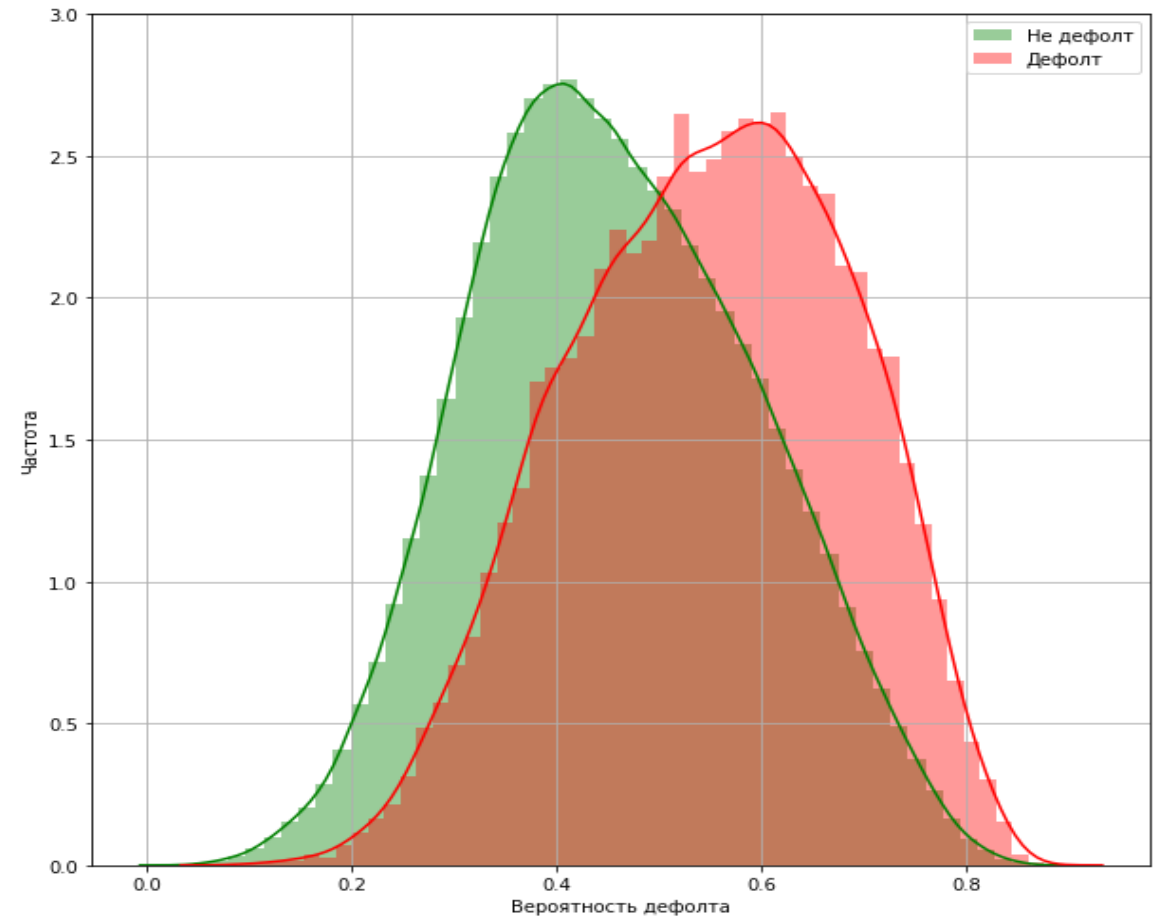
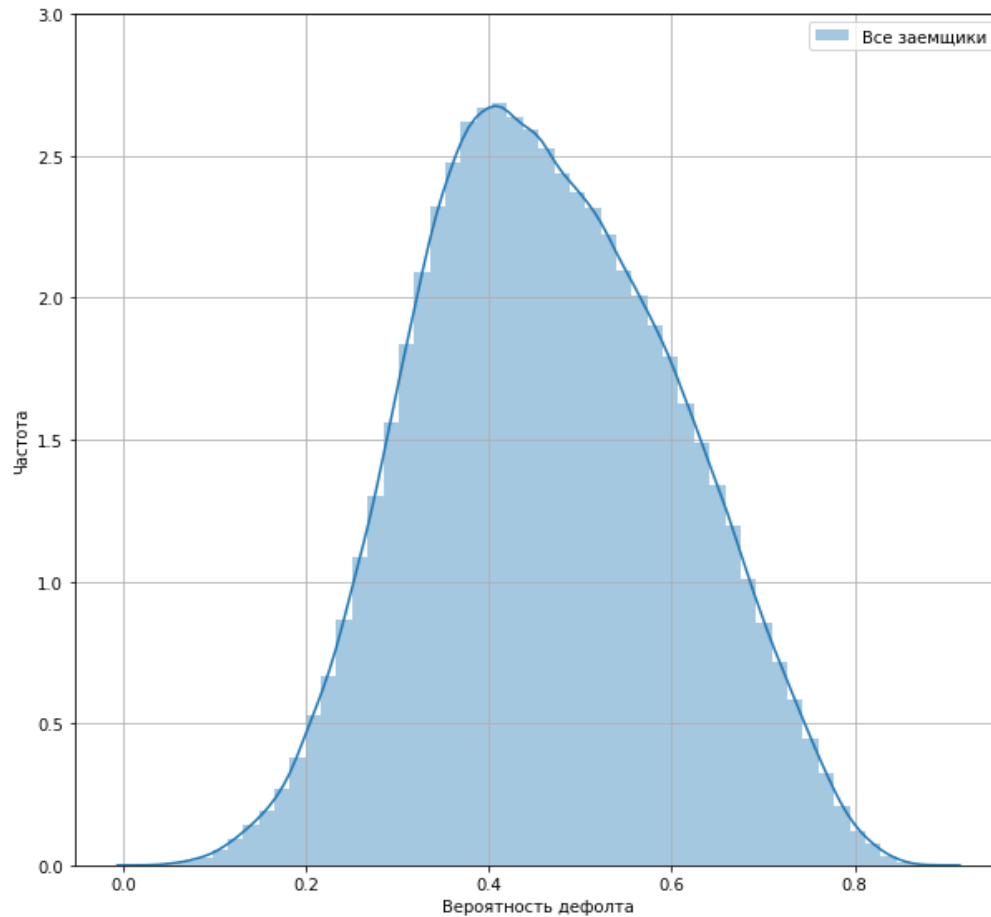
Фактор	Значение коэффициента
За сколько дней до подачи заявления клиент начал текущую работу	-2.71
Возраст клиента	-0.92
За сколько дней до подачи заявления клиент изменил документ, удостоверяющий личность, с которым он подал заявку на кредит	-0.55

ROC кривая алгоритма градиентного бустинга

Площадь под ROC кривой
равна 0.64



Распределение вероятностей дефолта резულიрующей модели



Наиболее важные характеристики заемщиков

- пол;
- наличие собственного автомобиля;
- возраст;
- количество проработанных дней на текущей работе до подачи заявления на выдачу кредита;
- количество дней, прошедших между изменением документа удостоверяющего личность и подачей заявления на выдачу кредита.

Спасибо за внимание!