

Обо мне

Меня зовут Антон Сметанин, я развиваюсь в области машинного обучения и аналитики данных. В данный момент учусь на 3 курсе ВШЭ. Участвую в хакатонах и соревнованиях Kaggle. В сентябре 21 года прошел в финал хакатона FinoDays. На 2 курсе писал курсовую работу по теме банковского скоринга и успешно защитил ее. В процессе обучения в институте был освоен SQL. Основным языком является Python, так же мне знакомы языки разметки HTML, CSS. В свободное время увлекаюсь просмотром фильмов и чтением книг.

Образование

Высшая Школа Экономики (НИУ ВШЭ) Бизнес-Информатика, GPA: 3.52/4.00	Москва Сентябрь 2019–Настоящее время
Высшая Школа Экономики ФКН ПМИ Машинное обучение 1	Москва Сентябрь 2021–Настоящее время
Яндекс Практикум Специалист по Data Science	Москва Лето 2020–Март 2021
Московский физико-технический институт (МФТИ) Курс по машинному обучению, основной поток (3 курс, 2020)	Москва Лето 2020–2021

Хакатоны

GSB Data Science Practice Финалист, 6 место	Москва Ноябрь 2021–Ноябрь 2021
FINODAYS Финалист	Москва Сентябрь 2021–Настоящее время

Проекты

Полный список моих проектов можно посмотреть на github.com/Stuksus

• Аналитика

– [Исследование объявлений о продаже квартир](#)

- * **Задача** — установить параметры на основании данных из сервиса Яндекс.Недвижимость. Это позволит построить автоматизированную систему: она отследит аномалии и мошенническую деятельность.
- * **Примененные библиотеки и технологии:** pandas, numpy, matplotlib

• Регрессия

– [Прогнозирование прибыли от нефтяных скважин](#)

- * **Задача** — спрогнозировать наиболее прибыльный для бурения скважины регион
- * **Примененные библиотеки и технологии:** pandas, numpy, matplotlib, tqdm, sklearn: (LinearRegression, Pipeline, StandardScaler)

- **Классификация**

- [Прогнозирование оттока клиентов банка](#)

- * **Задача** — спрогнозировать отток клиентов банка

- * **Примененные библиотеки и технологии:** pandas, numpy, matplotlib, tqdm, sklearn: (LogisticRegression, RandomForestClassifier, DecisionTreeClassifier, GridSearchCV)

- [Оценка экономического эффекта от внедрения моделей машинного обучения в банковском секторе](#)

- * **Задача** — Разработать и внедрить модель кредитного скоринга более лучшего качества. Оценить экономический эффект от внедрения новой модели

- * **Примененные библиотеки и технологии:** pandas, numpy, matplotlib, tqdm, CatBoost, WOE, sklearn: (LogisticRegression, PCA, RandomForestClassifier, SVC, Pipeline, StandardScaler, ROC AUC), imblearn: (SMOTE)

- [Решение в рамках хакатона GSB Data Science Practice](#)

- * **Задача** — Предсказать вероятность дефолта(PD) компаний

- * **Примененные библиотеки и технологии:** pandas, numpy, matplotlib, tqdm, WOE, sklearn: (LogisticRegression, Pipeline, StandardScaler, ROC AUC), imblearn: (SMOTE)

- **Обработка текста**

- [Классификация токсичных комментариев](#)

- * **Задача** — классифицировать токсичные комментарии

- * **Примененные библиотеки и технологии:** pandas, numpy, matplotlib, tqdm, sklearn: (LogisticRegression, RandomForestClassifier, SVC, TfidfVectorizer, CountVectorizer), nltk: (stopwords, WordNetLemmatizer)

- **Временные ряды**

- [Прогнозирование количества заказов такси](#)

- * **Задача** — спрогнозировать количество заказов на следующий час

- * **Примененные библиотеки и технологии:** pandas, numpy, matplotlib, CatBoost: (CatBoostRegressor), LightGBM, sklearn: (RandomForestRegressor, TimeSeriesSplit, GridSearchCV), statsmodels: (seasonal decompose)

Языки

- **Русский:** Носитель

- **Английский:** intermediate