

# Прогнозирование покупок клиентами «Ростелеком»

# Бизнес-ценность



Более точное и успешное предложение  
тарифа клиентам



Возможность учитывать больше  
факторов и значения

# Аналитика



- дисбаланс классов ("0" - 97,5%)
- наибольшее количество предложений приходится на Москву и Санкт-Петербург (более 10к)
- лучший конверсия из предложения в покупку приходится на Мурманскую область, Ненецкий АО, Республику Коми (более 10%)
- зависимость покупки тарифа от периода и региона предложения

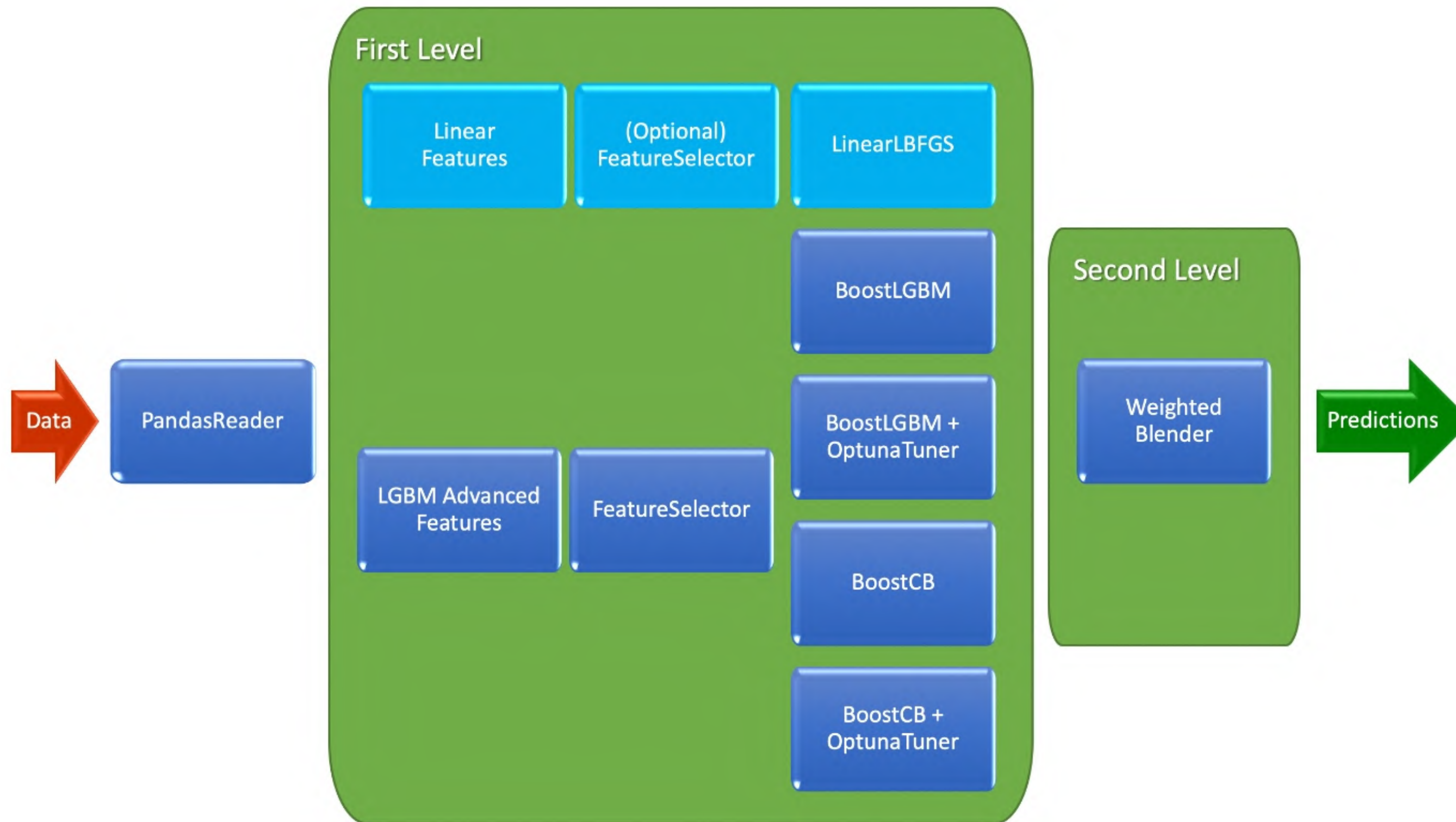


# Данные

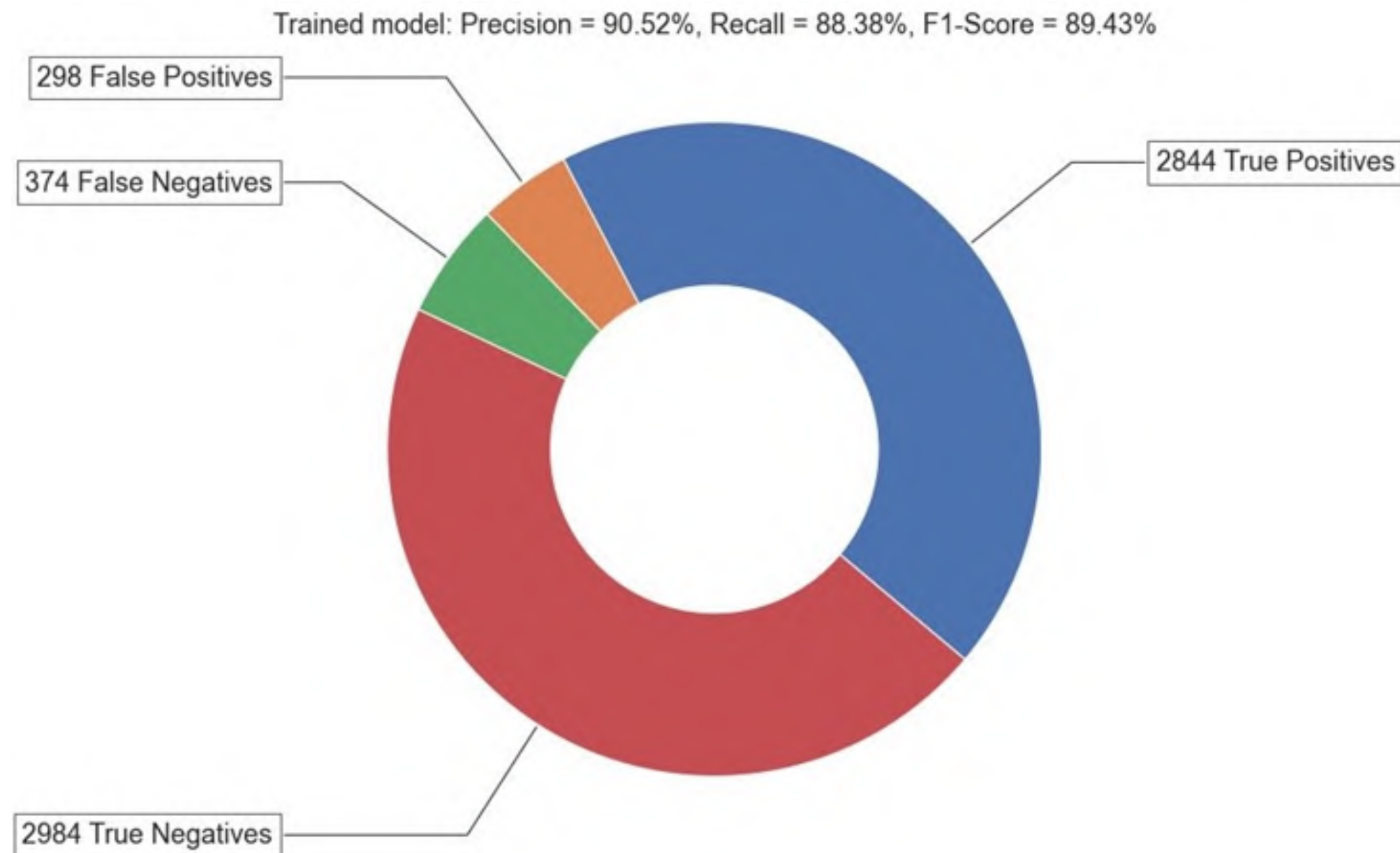
- **Росстат** (федеральные округа, население, средняя зарплата)
- **OSM** (удаленность от крупных городов, от регионального центра, наличие и количество телекоммуникационных станций)
- **Генерация новых фич** из имеющихся данных и более детального анализа (группировка и агрегация по периодам, регионам, городам и подсчет различных статистик)

# Что использовали

- LogReg, Random Forest, SVM
- XGBoost, CatBoost, LGBM
- TabNet
- Stritified K-Fold CV
- Out-of-time CV
- Ансамбли, блендинги



# Метрики



# Уникальность решения

- использование AutoML
- генерация новых фич
- Out-of-time CV
- Робастная и устойчивая модель (we hope (ಠ\_ಠ))



# Дальнейшие улучшения

- Более детальная аналитика
- Улучшение модели (модели по регионам)
- Обогащение данных

# Команда



Дмитрий Гайфуллин  
CV-Engineer



Иван Мишин  
ML-Engineer



Анна Ефимова  
CV-Engineer

28.08.2022

# Q&A

