

# Feature Store

онлайн-«магазин» данных

A

The most phygital  
bank since 1990



Центр ML-технологий  
Центр продвинутой аналитики, Альфа-Банк

# Feature Store



The most phygital  
bank since 1990

Feature Store – это онлайн-«магазин» данных, ассортимент которого постоянно расширяется

## Ключевые особенности Feature Store:

- ★ Фичи (признаки) и данные хранятся централизованно – «в одном окне»
- ★ Данные для обучения разных моделей подбираются разными командами согласно единой логике
- ★ Все данные выверены и имеют высокое качество за счет встроенных инструментов мониторинга
- ★ Информация о фичах, старте расчетов, статусе работы сохраняется автоматически
- ★ Бесшовная доставка признаков (фичей) в промышленную среду для функционирования боевых моделей
- ★ Дополнительно: сбор большого количества метаданных, анализ «популярности» переменных, ведение мониторинга, статистики применения и прочее

Начнем  
сначала.  
Для чего  
нужны фичи  
и что это  
такое?

## Жизненный цикл данных

сырые данные

очищенные

нормализованные

▶ **преобразованные для потребления моделью - фича**

# 1

## Фичи для обучения модели

Историческая глубина

Большой объем данных

Не всегда высокое качество данных

~ 20 000 фичей

# 2

## Фичи для применения модели

Актуальные данные (или с относительно небольшой глубиной)

Обычно небольшой объем данных

Высокие требования к качеству данных

~ 1000 фичей

# Для чего нужен Feature Store?



Обеспечение качественных структурированных данных для построения и применения моделей



Переиспользование фичей и борьба с дублированием



Упрощение и ускорение процесса вывода моделей



Единая библиотека источников данных (ETL) и простая заливка фичей в горячее хранилище



Встроенные механизмы автоматического описания и обработки данных, и, как следствие, генерация фичей



Прозрачности связки Источники данных-Витрина-Модель



Мониторинг качества данных и предупреждение о падении расчетов



Оптимизация времени и ресурсов команды DS-специалистов

# Кому нужен Feature Store



## Дата-инженеры

Выполнение расчета прод фичей, создание широкой выборки фичей для моделей



## Дата-аналитики со стороны бизнеса

Работа с данными для аналитики



## DS

Работа с внутренними данными в банке, поиск фичей для переиспользования, описание шаблонов моделей



## ML-инженеры

Упрощение вывода моделей

## Необходимые компетенции для создания фичей

- Знание Python
- Понимание SQL и принципов работы с данными на Hadoop
- Понимание работы Spark
- Базовые знания инфраструктурных компонент (Git, Jenkins)

# Feature Store: основные компоненты

01

## FS ETL – единый процесс создания фичей

- Реализует бизнес-логику создания фичей
- Ведет мониторинг расчетов и качества данных

02

## FS REGISTRY – реестр метаданных о фичах

- Визуализация метаданных
- Отчеты по метрикам фичей

03

## FS DATAMART – единый интерфейс работы с фичами

- Получение данных в коде универсальным способом
- Мониторинг использования данных

# Детали. Бесшовность переноса фичей

## Обеспечение бесшовного внедрения модели

- DS прописывает данные для модели с помощью FS Datamart
- FS Datamart определяет контур, в котором запускается, и передает модели нужные данные

## Трансформатор фичей для онлайн-моделей

- Расчет на основе единой кодовой базы
- Быстрая обработка данных in-memory

01

Разработчик моделей не пишет код работы с данными

02

Разработчик не меняет код между тестовыми стендами

03

Разработчик не переписывает код для онлайн-моделей

04

Фичи бесшовно переносятся с контура разработки на контур применения



**Ускорение  
внедрения моделей  
в 3 раза**

# Детали. Проверка качества данных

01

Расчет метрик  
запуска



02

Установка параметров  
проверки

Мониторинг витрины **ALFA ВЕРСИЯ**

Параметры Отчеты

emorenko@alfabank.ru

Метрики фичей

НАЗВАНИЕ МЕТРИКИ	THRESHOLD, %
max	20
min	90
average	20
none	20

03

Построение отчетов

Мониторинг витрины **ALFA ВЕРСИЯ**

Параметры Отчеты

ДАТА ОТЧЕТА	ДАТА ПЕРВОГО РАСЧЕТА	ДАТА ВТОРОГО РАСЧЕТА	АЛЕРТ	ФИЧИ	CSV
2023-02-21 00:00	2022-10-03 00:00	2022-09-03 00:00	!		<a href="#">Скачать</a>



# Этапы развития и внедрения Feature Store в Альфа-Банке

0

## Небольшие DS-команды

Набор сохранённых и обновляемых витрин, из которых берутся данные для обучения моделей

### Эффект:

Повышение скорости разработки, отсутствие необходимости пересоздания фичей при создании новой модели

1

## Развивающиеся подразделения DS, средние компании

Регулярные проверки качества данных, реестр метаданных по фичам (исключение дублирования фичей)

### Эффект:

Повышение скорости и качества разработки моделей, снижение модельного риска

2

## Развитые подразделения DS, средние компании

CI / CD процесс создания ETL, преднастроенные инструменты манипуляции с фичами, версионирование фичей, удобный UI

### Эффект:

Промышленный и отказоустойчивый процесс работы с фичами при разработке моделей

3

## Промышленный DS в крупных компаниях

Бесшовный процесс миграции фичей со среды разработки в среду применения моделей, в т.ч. инструменты миграции фичей из offline storage в online процесс

### Эффект:

Промышленный и отказоустойчивый процесс разработки и внедрения фичей, сокращение TTM внедрения моделей

# Результаты. Feature Store

Feature Store – собственная разработка Департамента продвинутой аналитики Альфа-Банка

Используя небольшую команду энтузиастов и стандартную банковскую платформу Hadoop, всего **за год с нуля** нам удалось создать действующее решение, которое уже активно используется в моделях для бизнес-линий розничного бизнеса и малого и микро бизнеса

- ✓ Разработанный процесс расчета признаков (фичей) с помощью онлайн-трансформатора – **уникальное на рынке решение**
- ✓ Процесс сильно упрощает вывод моделей: сокращает время на написание кода **в 10 раз**, снижает ошибки и риски, связанные с некачественными данными
- ✓ И главное - все эти факторы значительно **снижают стоимость** разработки моделей