



**OSA  
HYBRID  
PLATFORM**

# Прикладной искусственный интеллект в ритейле



**ВПЕРЕД С ЛЮБОВЬЮ!**

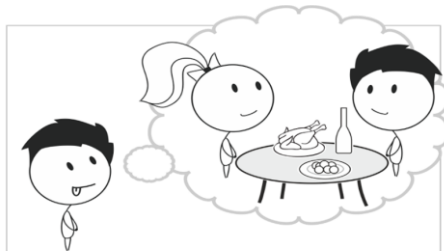
# С чего все начиналось в ритейле



# Мировой рынок ритейла теряет 500 млрд долларов в год



# Проблема: негативный покупательский опыт и упущенные продажи



Это Петя. Он хочет приготовить ужин и разделить романтический вечер с женой



#### Ритейлер

Инвестировал \$ в:

- Склад
- Магазин
- Персонал



#### Производитель

Инвестировал \$ в:

- Производство
- Логистику
- Отдел продаж

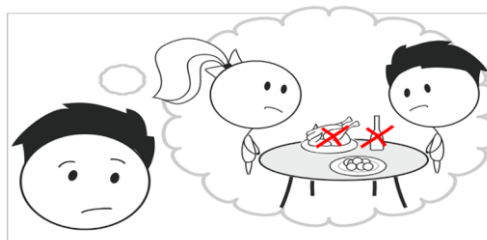
И ритейлер и производитель вложили много труда, чтобы удовлетворить потребности Пети



Петя написал список покупок и идет в любимый супермаркет



Но нужного товара в магазине НЕТ



?!?!?!



#### Проблемы для производителя и ритейлера

- Упущенные продажи
- Неэффективные инвестиции
- Огорченный покупатель, потеря лояльности

**ТЕРЯЮТ ВСЕ**



**Сервис на основе машинного обучения,  
обеспечивающий:**

**постоянную доступность товара на полке  
в оптимальном количестве**

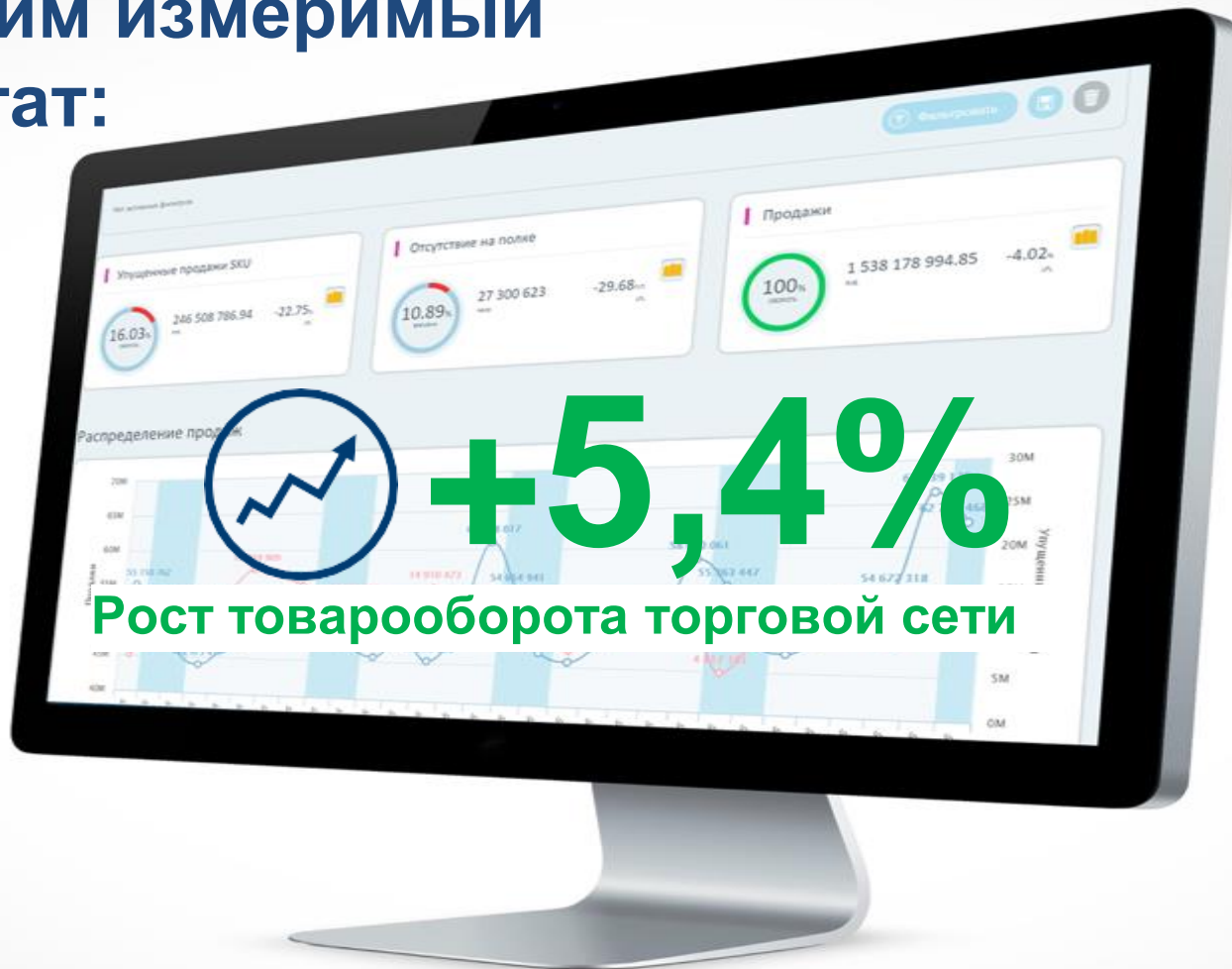


**в режиме  
реального времени**

# Приносим измеримый результат:



OSA  
HYBRID  
PLATFORM



# Лидеры выбирают наш сервис:



# Информационные активы в OSA Hybrid Platform





# Прогноз спроса



# Прогноз спроса

## Проблематика:

- Прогноз большого количества временных рядов (в одном магазине 7,000 – 35,000 товаров)
- Грязные данные (периоды отсутствия продаж, промо)

## Задача:

- Недельный прогноз
- Один период вперед
- Около 250 тысяч временных рядов

# Прогноз спроса

## Что мы тестировали:

- ARIMA
- Нейронная сеть типа “Многослойный персептрон”
- XGBoost
- Адаптивная селективная модель

## Метрика

Truncated MAPE –  
truncated mean absolute  
percentage error

**Что победило?**

# Прогноз спроса

$$\text{MAPE} = \frac{\sum_{i=1}^n \frac{|A_i - F_i|}{|A_i|}}{n}, \text{ где } A - \text{ факт, } F - \text{ форекаст}$$

$$\text{Truncated MAPE} = \frac{\sum_{i=1}^n \begin{cases} \frac{|A_i - F_i|}{|A_i|}, \frac{|F_i|}{|A_i|} < 2 \\ 1, \frac{|F_i|}{|A_i|} \geq 2 \end{cases}}{n}$$

# Прогноз спроса

## ARIMA (Autoregressive integrated moving average)

Модель  $ARIMA(p, d, q)$  для нестационарного временного ряда  $X_t$  имеет вид:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t$$

где  $\varepsilon_t$  - стационарный временной ряд;

$c, a_i, b_j$  - параметры модели.

$\Delta^d$  - оператор разности временного ряда порядка  $d$  (последовательное взятие  $d$  раз разностей первого порядка - сначала от временного ряда, затем от полученных разностей первого порядка, затем от второго порядка и т.д.)

Также данная модель интерпретируется как  $ARMA(p + d, q)$ - модель с  $d$  единичными корнями. При  $d = 0$  имеем обычные  $ARMA$ -модели.

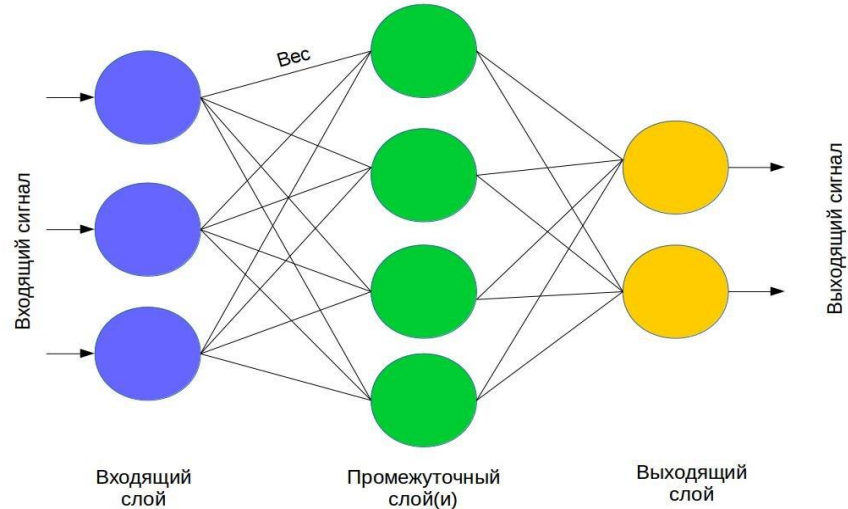
**Метрика** - не все стационарные ряды

Truncated MAPE = 0.4384

# Прогноз спроса

Нейронная сеть типа “Многослойный персептрон”

- два слоя по 40 нейронов
- около 25 фич



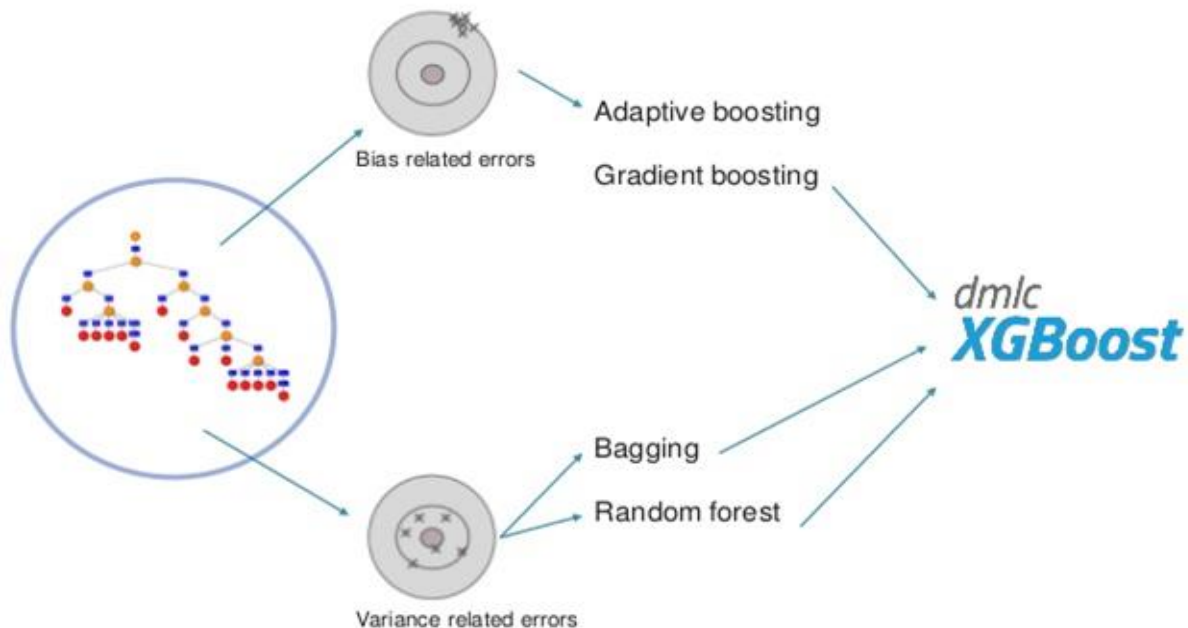
## Метрика

Truncated MAPE = 0.3427

# Прогноз спроса

## XGBoost

- настройка гиперпараметров
- около 25 фич



## Метрика

Truncated MAPE = 0.3342

# Прогноз спроса

## Адаптивная селективная модель

Пусть имеется  $k$  моделей прогнозирования,  
 $\hat{y}_{j,t+d}$  — прогноз  $j$ -й модели на момент  $t + d$ ,  
 $\varepsilon_{jt} = y_t - \hat{y}_{jt}$  — ошибка прогноза в момент  $t$ ,  
 $\tilde{\varepsilon}_{jt} := \gamma|\varepsilon_{jt}| + (1 - \gamma)\tilde{\varepsilon}_{jt}$  — экспоненциально сглаженная ошибка.

Лучшая модель в момент времени  $t$ :

$$j_t^* = \arg \min_{j=1, \dots, k} \tilde{\varepsilon}_{jt}.$$

Адаптивная селективная модель:

$$\hat{y}_{j,t+d} := \hat{y}_{j_t^*, t+d}$$

Требуется подбор  $\gamma$ , рекомендация:  $\gamma = 0.01 \dots 0.1$ .

## Метрика

Truncated MAPE = 0.3217





## Прогноз спроса

# Адаптивная селективная модель

**Победитель как по точности, так и по скорости!**

# Определение промо-акции



# Определение промо-акции

## Проблематика:

- Размеченные данные
- Качество данных

## Что было опробовано:

- Логистическая регрессия
- XGBoost
- Catboost

## Метрика

ROC AUC



Что победило?

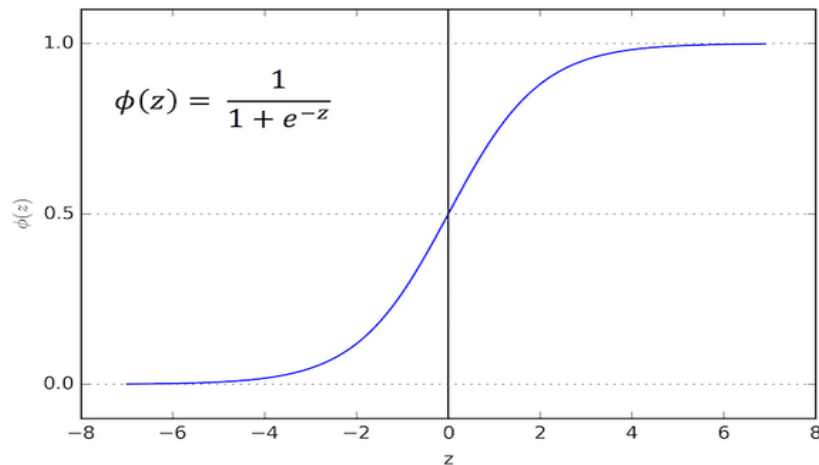
# Определение промо-акции

Логистическая регрессия

Старая добрая база

Метрика

ROC AUC = 0.6543



$z = \theta^T x$ , где

$\theta^T$  - коэффициенты регрессии,

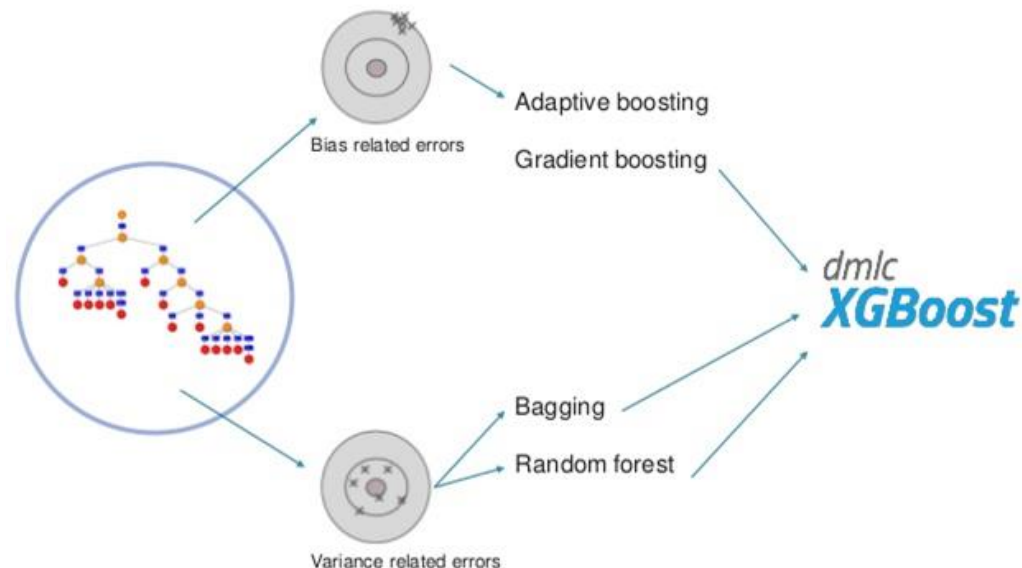
$x$  - параметры.

# Определение промо-акции

XGBoost

Звезда Kaggle  
соревнований

Метрика  
ROC AUC = 0.7717



# Определение промо-акции



Yandex  
CatBoost

## Основные фиши:

- поддержка категориальных данных
- хорошие результаты на дефолтных параметрах
- забывчивые деревья решений

## Метрика

ROC AUC = 0.7434

# Определение промо-акции

**XGBoost**

**Победитель!**

# Связанные товары





# Связанные товары

## Проблематика:

- Стоп лист товаров

## Задача:

- Определить товар, который влияют на спрос другого

Генераторы



+



Каннибалы



vs.



OSA  
HYBRID  
PLATFORM

# Связанные товары

## Что мы тестировали:

- Ассоциативные правила
- Нейронная сеть plu2vec

## Метрика

Экспертная оценка 100 пар связанных товаров на адекватность и суммарное количество

**Что победило?**

# Связанные товары

## Ассоциативные правила

## Стандарт индустрии

**Метрика**  
52%

Support

$$\text{supp}(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|}$$

Confidence

$$\text{conf}(X \Rightarrow Y) = \text{supp}(X \cup Y) / \text{supp}(X)$$

Lift

$$\text{lift}(X \Rightarrow Y) = \frac{\text{supp}(X \cup Y)}{\text{supp}(X) \times \text{supp}(Y)}$$

Conviction

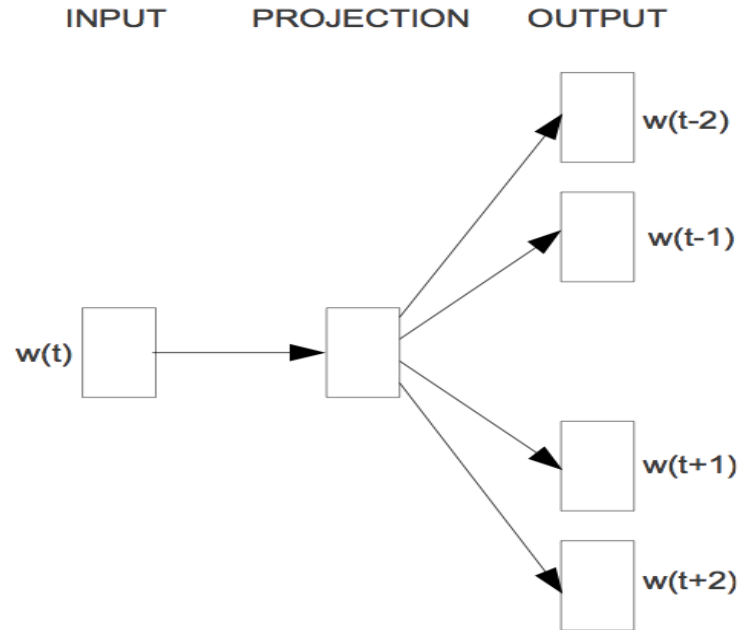
$$\text{conv}(X \Rightarrow Y) = \frac{1 - \text{supp}(Y)}{1 - \text{conf}(X \Rightarrow Y)}$$

# Связанные товары

plu2vec

- Пары товаров берутся из чеков
- Результат валидируется экспертно

**Метрика**  
85%



**Skip-gram**

# Связанные товары

**plu2ves**

**Победитель!**

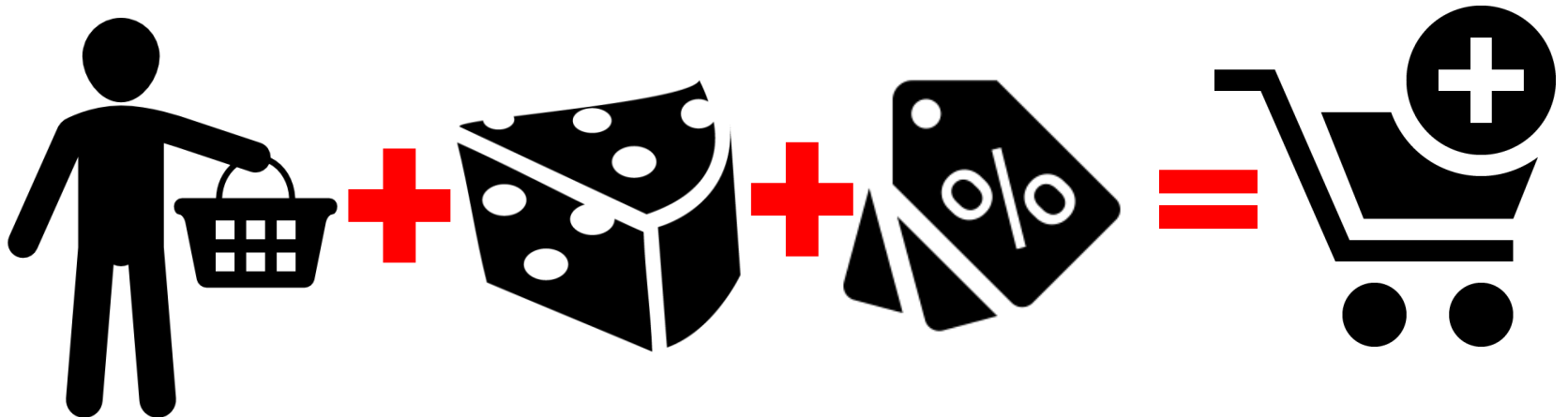
# Целевые маркетинговые кампании



# Рекомендательные системы для целевых маркетинговых кампаний

## Задача:

Найти идеальный треугольник человек-товар-предложение, который увеличивает суммарную покупку за период



# Рекомендательные системы для целевых маркетинговых кампаний

- Ассоциативные правила
- Коллаборативная фильтрация
- LifeStyle: кластеризация на основе атрибутов каталога мастер-данных
- История чеков по карте лояльности – выпадающие продукты
- Воронка
- Нейронная сеть

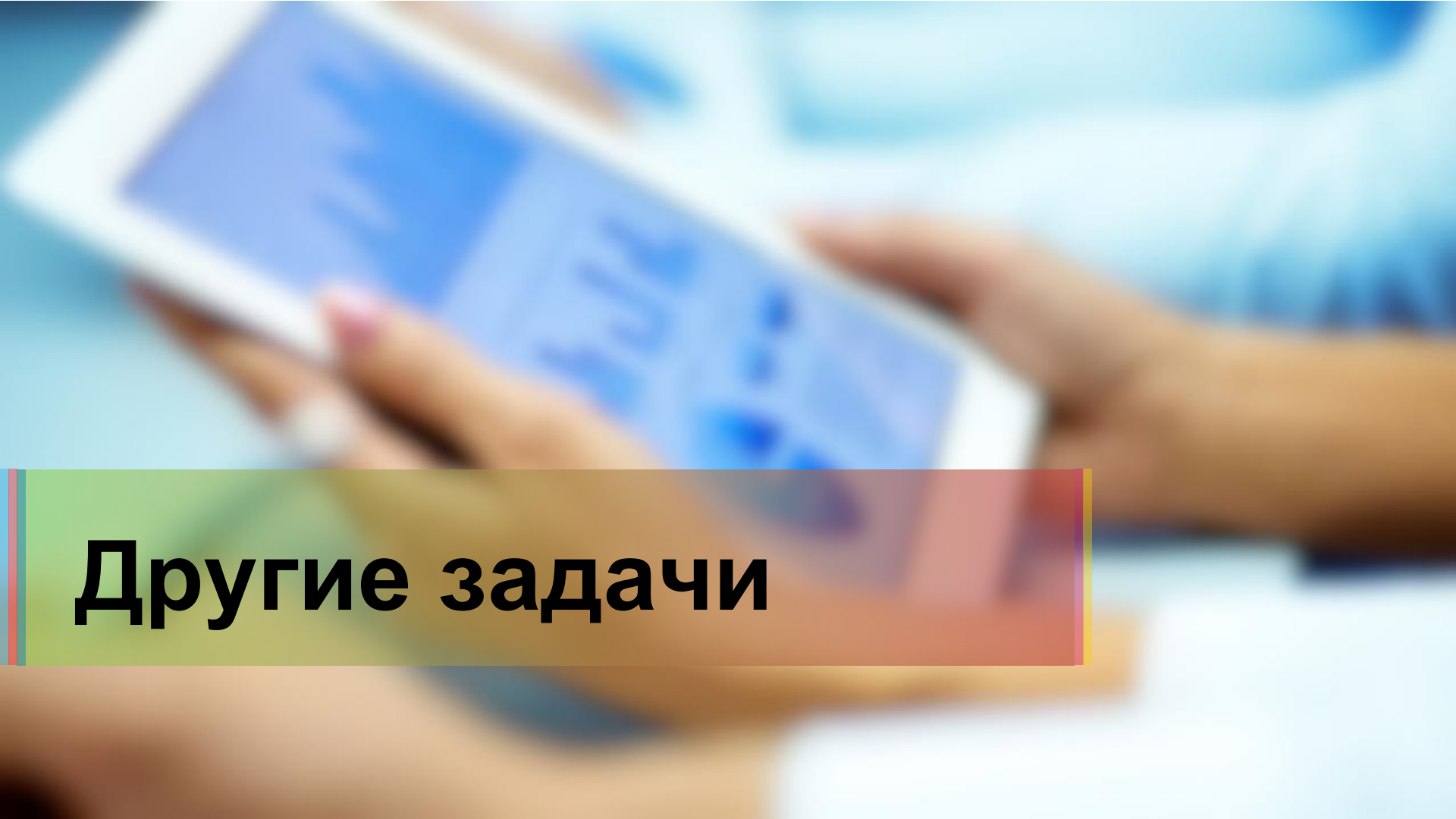


# Рекомендательные системы для целевых маркетинговых кампаний

LifeStyle: кластеризация на основе атрибутов каталога мастер-данных

- Сформировано 27 сегментов на основе атрибутов, присвоенных товарам
- Количество кластеров отобрано экспертно
- Алгоритм кластеризации k-means



A person's hands are holding a white tablet computer. The screen displays a blue interface with some text and icons, though it is out of focus. The background is a blurred light blue.

**Другие задачи**

# Ряд других задач, которые решаются с использованием машинного обучения

- **Управление ценообразованием**  
задача - максимизация продаж и прибыли



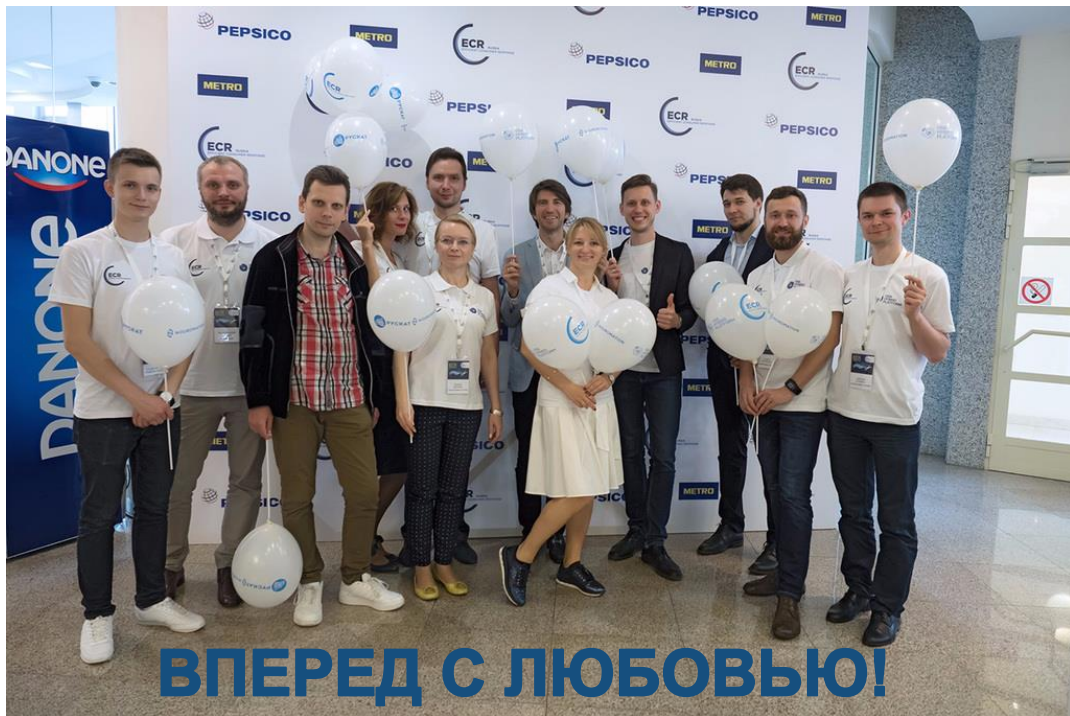
- **Управление запасами**  
задача - оптимизация запасов в магазинах и распределительных центрах

# Ряд других задач, которые решаются с использованием машинного обучения

- Управление цепочкой поставок
- задача - оптимизация цепочки поставок



# Создаем лучшую data science команду в мире в индустрии ритейл



Data sensing

# Благодарю за внимание!

**Тел.:** +38 063 36 55 953  
**Email:** o.potapenko@osahp.com  
**Skype:** lyapen88  
**Facebook:** fb.me/oleksii.potapenko  
**LinkedIn:** www.linkedin.com/in/oleksii-potapenko/

