Deploy in production

Or, pipeline to save the world

SKB Kontur

- B2B
- 40+ products
- 7700+ employee
- 1000+ devs
- 1200+ tech support

Products

- Tax reporting
- Accounting, salary, personnel
- Electronic document management
- Trade
- Checking counterparties
- Electronic bidding and procurement
- Interaction with GIS

ML Tasks

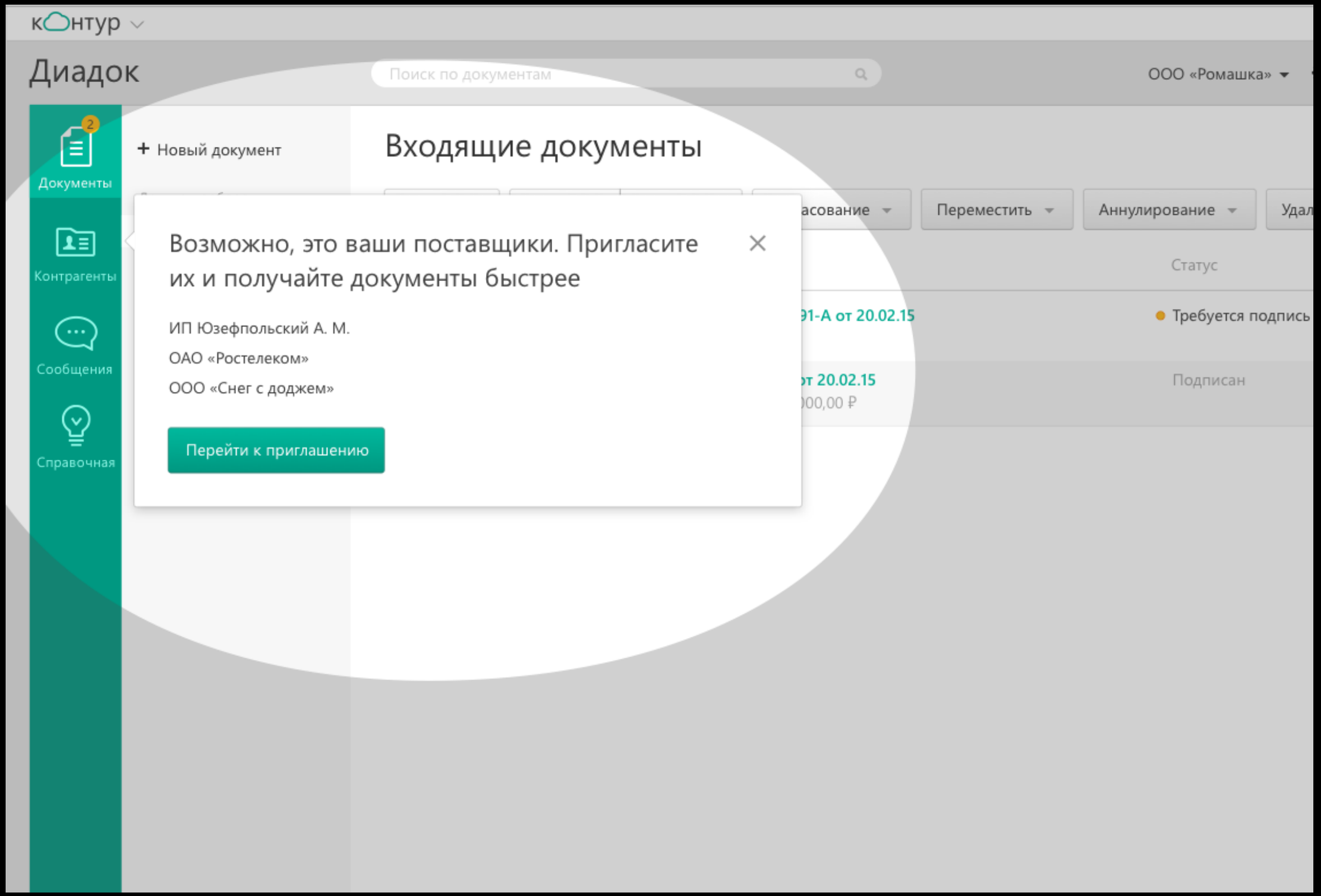
Добавление товара

Как заполнить карточку товара

Наименование: Макароны увелка 450г

Весовой товар

Группа: Бакалея > Макаронные изделия Изменить группу


контур
Диадок
Поиск по документам
ООО «Ромашка»
Документы
Контрагенты
Сообщения
Справочная
+ Новый документ
Входящие документы
асование
Переместить
Аннулирование
Удал
Статус
91-А от 20.02.15
Требуется подпись
от 20.02.15
000,00 ₽
Подписан
Возможно, это ваши поставщики. Пригласите их и получайте документы быстрее
ИП Юзефпольский А. М.
ОАО «Ростелеком»
ООО «Снег с доджем»
Перейти к приглашению

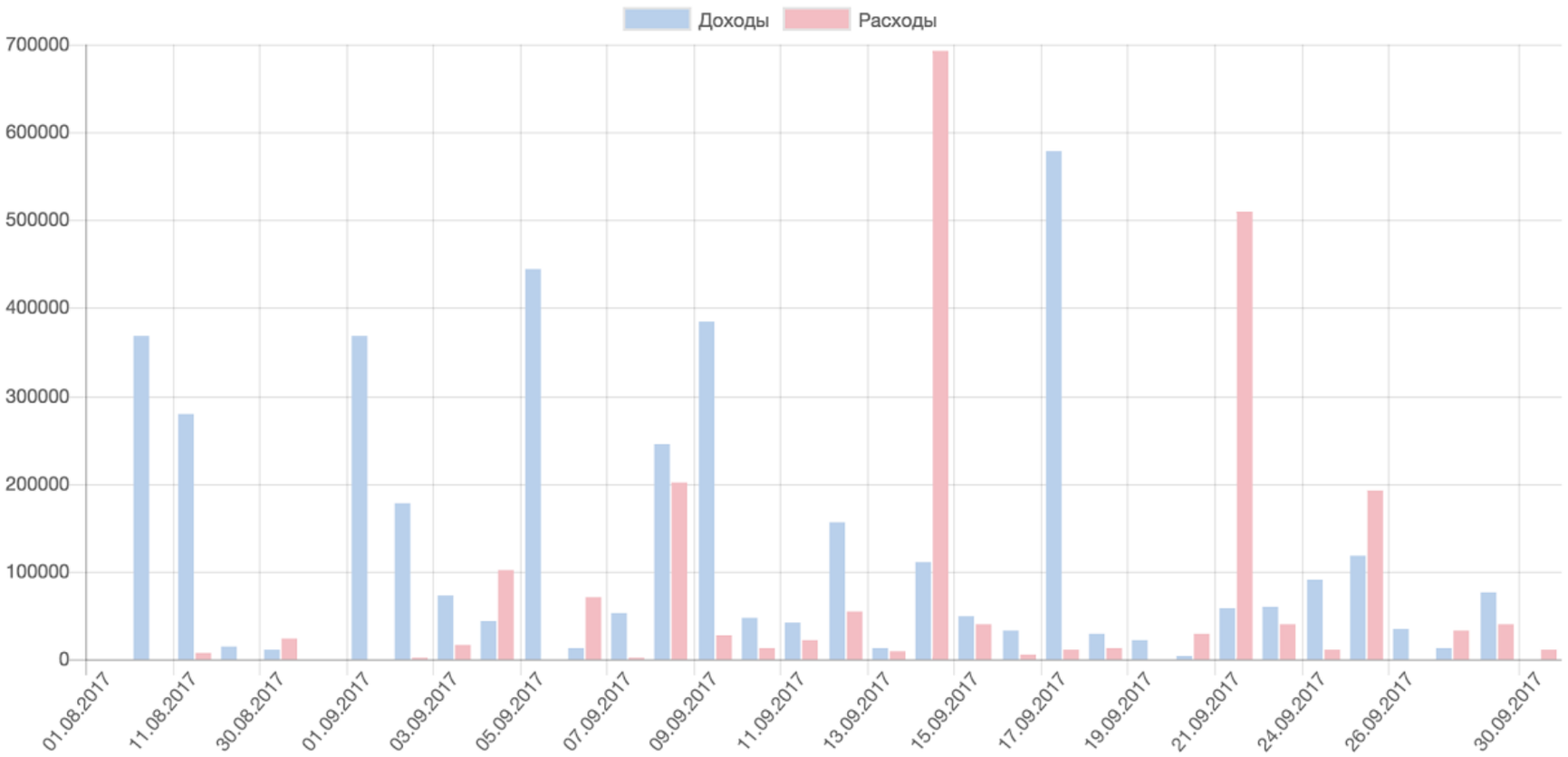
Аналитика

Если у вас есть пожелания по развитию сервиса — напишите нам об этом

Загрузить банковскую выписку

По всем счетам

Период 01.08.2017 — 01.10.2017

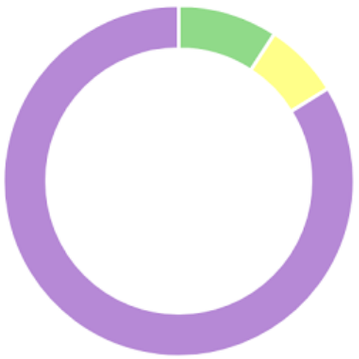


Остаток
1 952 200,71 ₽

Доходы
4 006 178,76 ₽

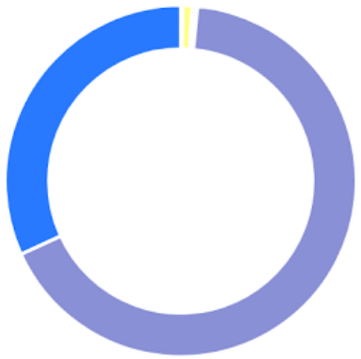
Расходы
2 178 187,08 ₽

Доходы



Прочие поступления денег	367 810 ₽ (9.18%)
Мат. помощь от учредителей	280 000 ₽ (6.99%)
Оплата от покупателей, клиентов	3 358 368,76 ₽ (83.83%)

Расходы



ПФР: взносы на обязательное медицинское страхование в ФФОМС	3 796,85 ₽ (0.17%)
ПФР: страховая часть пенсионных взносов	19 356,48 ₽ (0.89%)
УСН	7 425 ₽ (0.34%)
Оплата услуг банка	2 784 ₽ (0.13%)
Оплата поставщикам	1 451 431,92 ₽ (66.63%)
Выплаты сотрудникам	693 392,83 ₽ (31.83%)

АО "ПФ "СКБ Контур"

Более подробная информация будет доступна, если войти в систему. Получить демо-доступ Купить

Акционерное общество "Производственная фирма "СКБ Контур"
21.06.2017

Действующее предприятие

ИНН 6663003127
КПП 668601001
ОГРН 1026605606620
ОКПО 00242766

Дата образования: 26 марта 1992

620017
Свердловская обл
г Екатеринбург
пр-кт Космонавтов, 56 61 Осмотреть
19.12.2008

Налоговый орган: 6686
Дата постановки на учет: 1 января 2012

Телефон
[redacted]
ЕГРЮЛ, Росстат
[redacted]
Добавлен пользователем
Показать все (10)

Генеральный Директор
[redacted]
11.02.2010 ИНН [redacted]

Основной вид деятельности:

Арбитражные дела

Роль	За 12 месяцев	За всё время
Ответчик	[redacted] (4)	[redacted] (26)
Истец	[redacted] (6)	[redacted] (27)
Другое	[redacted] (9)	[redacted] (82)

Финансы на конец 2016 года

Баланс — [redacted] ↑ [redacted]%
Выручка — [redacted] ↑ [redacted]%
Чистая прибыль — [redacted]

Подробности 2005...2016

Госзакупки

Участник — [redacted] руб. (2382)
Контракт заключен — [redacted] руб. (2164)

Филиалы и представительства

[redacted]
03.10.2008

[redacted]

Связанные организации

АНО ДПО "УЦ Скб Контур"
АНО Клуб "Контур-Видео"
ЗАО СКБ "Контур" Челябинский филиал

Все связи (21)

Могут быть представлены как актуальные, так и исторические

Лицензии

Роскомнадзор - связь (4)
По данным ЕГРЮЛ (2)

Связанные товарные знаки

Упоминания в товарных знаках — 36

Логотип скрыт | Логотип скрыт | Л

Проверки

2017 год — 1
Выявлены нарушения — 1

Tech Stack

- Langs
  - Reality: .Net
  - Data Science: + Python, Java/Scala, C++
- Databases
  - Reality: MongoDb, Cassandra, MsSql
  - Data Science: + Hadoop/Spark
- Deploy
  - Reality: Windows, Octopus, TeamCity, Houston, Vostok
  - Data Science: + Linux, Docke

Principle Scheme

- Data Science as a service
  - Task
  - Hypothys
  - Research
  - MVP
  - Evaluation
  - Integration
  - Support and extension

Problem


jupyter

Research model

!=

Production model

Solution

Scikit-Learn Pipeline

Concept

- Every preprocessing step is sk-learn compatible transformer
- Preprocessing + decision model as Pipeline

Concept

- All params a accessible from outside (easy grid search)
- Model a serialisable

Research

Deploy


Model
Push Model
Service
Store Model
Update Model
DataBase

Practice

- Create class for your transformation
- Inherit sklearn's BaseEstimator and TransformerMixin
- Implement fit, transform
- Do not compute ANYTHING in __init__

Practice

```
class MyTransformer(BaseEstimator, TransformerMixin):
    def __init__(self, value):
        # no logic here
        self.value = value

    def fit(self, X: np.array, y=None):
        return self

    def transform(self, X: np.array):
        # logic here
        if self.value < 0:
            self.value = 0

        return X * value
```

Practice

- Use FeatureUnion for parallel transformation
- Use memory arg in Pipeline to cache transformation (for grid search)
- Use sklearn-pandas Mapper for DataFrame

Practice

```
model = Pipeline(
    steps=[
        ('fill_na', FillNa()),
        ('union', FeatureUnion([
            ('text_vect', CountVectoriser()),
            ('have_name', TextMatcher())
        ])),
        ('clf', LGBMClassifier)
    ]
)
```

Practice

- Create separate library and store all transformation in it
- Share this lib across projects

Practice

- Use joblib to serialise model
  joblib from sklearn not compatible with separate joblib, choose one
- Use database to store model
- Use docker to deploy your service
- Try Pipenv and Pipefile
- Try Kafka for data streams

You good