

Rolling Drones - Innoagency Startup Recommendation

Задача проекта

Текущий проект - это персонализированная рекомендательная система по сервисам Москвы для стартапов

Quick Start

```
git clone https://github.com/NmadeleiDev/innagency-startup-recommendation.git
cd innagency-startup-recommendation

cp .env.example .env

docker-compose build
docker-compose up -d
```

или с использованием make

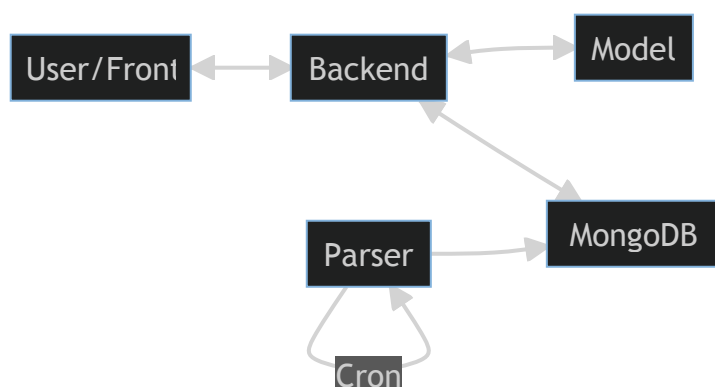
```
git clone https://github.com/NmadeleiDev/innagency-startup-recommendation.git
cd innagency-startup-recommendation

cp .env.example .env

make up
make up-c # с очисткой БД
```

Структура проекта

Папки - части общего проекта. Каждая часть опционально имеет внутри Dockerfile. Все части соединены docker-compose файлом в корне проекта.



ML

Папка ML содержит в себе Jupyter Notebooks для подготовки данных и генерации рекомендаций.

Всего в проекте используется 2 модели:

- Классификатор для получения наиболее подходящего типа сервиса для компании (Частный фонд, гос. фонд, акселератор, частный инвестор, корпоративный инвестор)
- Классификатор для получения наиболее подходящего сервиса для компании

Описание notebooks и файлов:

- dataprep.py - подготовка данных для обоих классификаторов
- investor_classifier.ipynb - создание и обучение модели рекомендации сервиса для компании
- type_classifier.ipynb - создание и обучение модели рекомендации типа сервиса для компании

Как готовили данные:

- Категориальные метрики: Onehot кодирование (<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelBinarizer.html>, <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>)
- Числовые метрики: Скейлинг (<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.OneHotEncoder.html>)

Проверка гипотез

Мы тестировали следующие гипотезы для классификации компании к сервисам:

- Метод линейной регрессии
- Метод опорных векторов
- Решающие деревья

Указанные методы сработали не очень хорошо. Мы решили, что стандартные методы плохо подходят для такой задачи и разработали свою кастомную модель.

Что сработало хорошо?

Мы создали кастомную модель. С помощью нее мы ищем подходящие сервисы для компании.

Под капотом она сделана следующим образом:

Системе предоставляют данные новой компании, а модель генерирует для нее "идеальный" сервис или фонд. После этого система ищет в БД существующий фонд/сервис, который наиболее близок по параметрам к "идеальному".

Работа нашей модели

Строго говоря, мы строим отображение из пространства компаний в пространство сервисов, и в последнем для каждой компании ищем ближайший сервис, используя L2 норму.

В TensorFlow использовали Sequential модель с 1 внутренним слоем в 64 нейрона (https://www.tensorflow.org/api_docs/python/tf/keras/Sequential)

Далее, при помощи пакета `numpy.linalg` и функции `norm` искали сервисы, близки к образу компании на пространство сервисов.

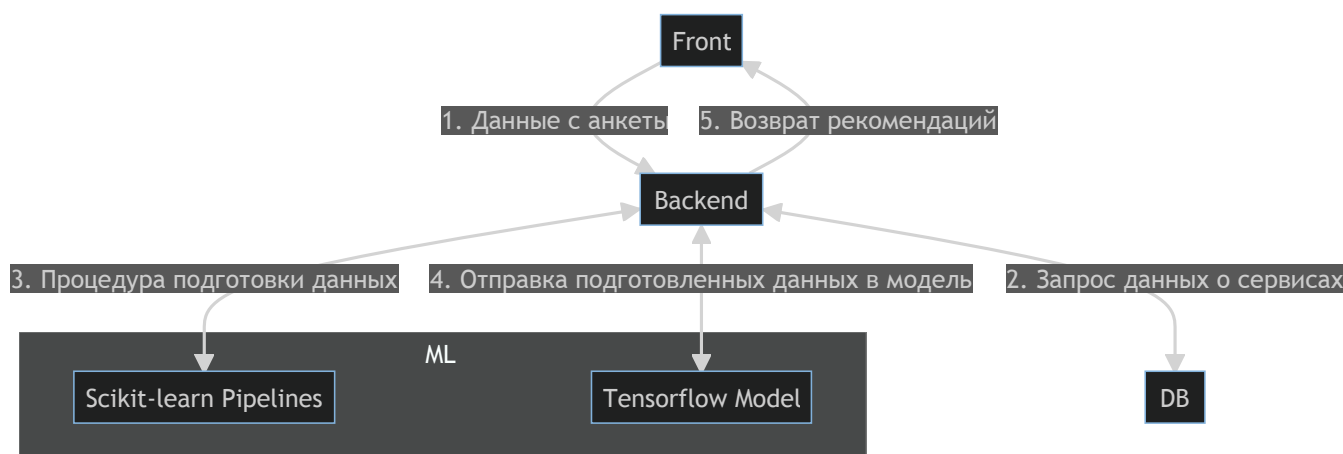
Backend

Папка `backend` содержит код (python{FastAPI}) серверной части веб-приложения, API для приема запросов от клиентской части веб-приложения (frontend) и логику работы с обученной моделью.

Бекенд взаимодействует с БД и производит следующие операции:

- Получает, вставляет или удаляет (CRUD) сущности (стартапы, компании, акселераторы и сервисы)
- Полученные данные подготавливает/трансформирует, используя процедуры, описанные в папке ML
- Трансформированные и подготовленные данные отправляет в обученную модель из папки ML и возвращает рекомендации клиенту.

Ниже представлена диаграмма взаимодействий компонент стека.



Front

Папка `front` содержит код (NodeJS{React}) клиентской части веб-приложения.

Parser

Папка `parser` содержит код для парсинга данных компаний по ИНН для пополнения датасета и дообучения модели.

DB

Так как данные разрежены (по многим атрибутам много пробелов), мы решили отказаться от реляционных БД в пользу NoSQL подобных. Поэтому в качестве базы данных взяли MongoDB.

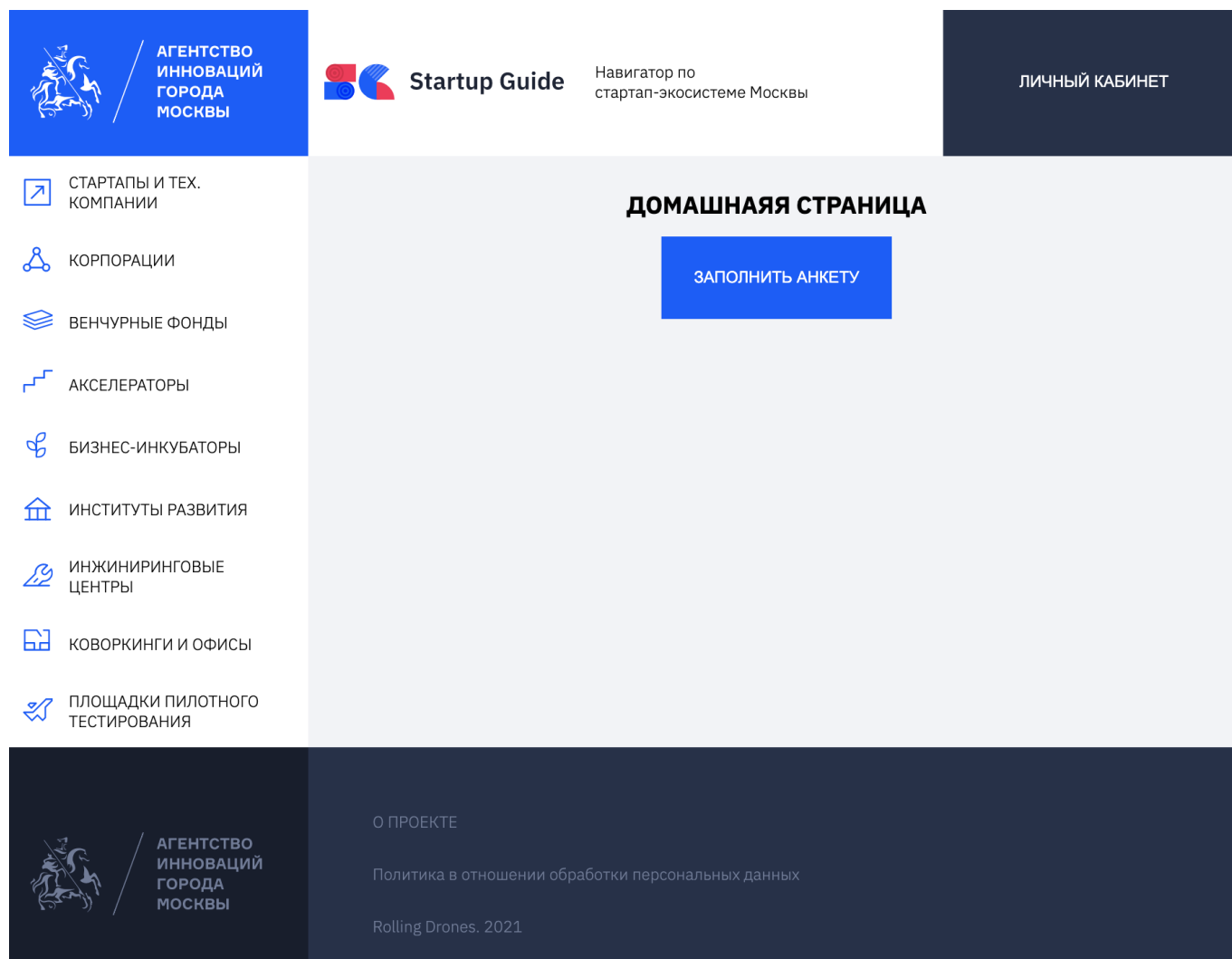
Документация по интерфейсу

Документация по интерфейсу в формате OpenAPI (Swagger) доступна по адресу <https://startup-guide.ml/backend/docs>

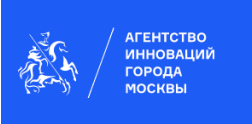
По адресу <https://startup-guide.ml> также доступна демо-версия проекта.

Пользовательская история / сценарий использования

Главный экран



Экран заполнения анкеты о стартапе



АГЕНТСТВО
ИННОВАЦИЙ
ГОРОДА
МОСКВЫ

Startup Guide

Навигатор по
стартап-экосистеме Москвы

ЛИЧНЫЙ КАБИНЕТ

СТАРТАПЫ И ТЕХ.
КОМПАНИИ

КОРПОРАЦИИ

ВЕНЧУРНЫЕ ФОНДЫ

АКСЕЛЕРАТОРЫ

БИЗНЕС-ИНКУБАТОРЫ

ИНСТИТУТЫ РАЗВИТИЯ

ИНЖИНИРИНГОВЫЕ
ЦЕНТРЫ

КОВОРКИНГИ И ОФИСЫ

ПЛОЩАДКИ ПИЛОТНОГО
ТЕСТИРОВАНИЯ

АНКЕТА

Основная информация

ИНН

3456789

Название компании

Компания

Дата основания компании

10/07/2021

Детали

Основной ОКВЭД

Основной ОКВЭД (номер)

Стадия развития компании

Расширение

Категория МСП

ЮЛ малое

Экспортер

нет

Участник инновационного кластера города Москвы

да

Участник Сколково

да

Инновационная компания

нет

Стартап

да

Теги

Технологический фокус компании

edtech
fintech
телекоммуникации
healthtech
космостим

Рынки

CleanTech
Consumer Goods & Services
Cybersecurity
E-commerce
EdTech

Технологии

AR/VR
Big Data
Аддитивные технологии
Беспилотники
Биометрия

Бизнес-модель

B2C
B2B
B2B2C
B2G
P2P

Данные о доходах

Чистая прибыль в год, \$

0

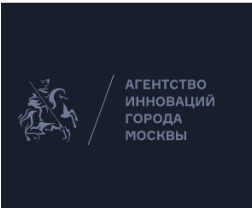
Налог на прибыль в год, %

0

Валовая прибыль в год, \$

0

ПОДОБРАТЬ ИНВЕСТОРА



АГЕНТСТВО
ИННОВАЦИЙ
ГОРОДА
МОСКВЫ


О ПРОЕКТЕ

Политика в отношении обработки персональных данных

Rolling Drones. 2021

Экран выданных рекомендаций сервисов Москвы

5 / 7



АГЕНТСТВО
ИННОВАЦИЙ
ГОРОДА
МОСКВЫ

СТАРТАПЫ И ТЕХ.
КОМПАНИИ

КОРПОРАЦИИ

ВЕНЧУРНЫЕ ФОНДЫ

АКСЕЛЕРАТОРЫ


БИЗНЕС-ИНКУБАТОРЫ

ИНСТИТУТЫ РАЗВИТИЯ

ИНЖИНИРИНГОВЫЕ
ЦЕНТРЫ

КОВОРКИНГИ И ОФИСЫ

ПЛОЩАДКИ ПИЛОТНОГО
ТЕСТИРОВАНИЯ



Startup Guide

Навигатор по
стартап-экосистеме Москвы

ЛИЧНЫЙ КАБИНЕТ

ИДЕАЛЬНЫЕ ИНВЕСТОРЫ

ПУБЛИЧНОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО "СБЕРБАНК РОССИИ"

Accelerator

ПУБЛИЧНОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО "СБЕРБАНК РОССИИ"

Accelerator

ПУБЛИЧНОЕ АКЦИОНЕРНОЕ ОБЩЕСТВО "БАНК УРАЛСИБ"

Accelerator

ПОКАЗАТЬ ЕЩЕ

SMG CAPITAL

VentureFund


RBF VENTURES

VentureFund

WINTER CAPITAL

VentureFund

ПОКАЗАТЬ ЕЩЕ




АГЕНТСТВО
ИННОВАЦИЙ
ГОРОДА
МОСКВЫ

О ПРОЕКТЕ

Политика в отношении обработки персональных данных

Rolling Drones. 2021

Экран дополнительной информации о сервисе



АГЕНТСТВО
ИННОВАЦИЙ
ГОРОДА
МОСКВЫ

СТАРТАПЫ И ТЕХ.
КОМПАНИИ

КОРПОРАЦИИ

ВЕНЧУРНЫЕ ФОНДЫ

АКСЕЛЕРАТОРЫ


БИЗНЕС-ИНКУБАТОРЫ

ИНСТИТУТЫ РАЗВИТИЯ

ИНЖИНИРИНГОВЫЕ
ЦЕНТРЫ

КОВОРКИНГИ И ОФИСЫ

ПЛОЩАДКИ ПИЛОТНОГО
ТЕСТИРОВАНИЯ



Startup Guide

Навигатор по
стартап-экосистеме Москвы

ЛИЧНЫЙ КАБИНЕТ

НАПРАВЛЕНИЕ
Accelerator

РЫНКИ


retailtech
e-commerce
fintech

СЕРВИСЫ

ТЕХНИЧЕСКИЙ ФОКУС

рост клиентской базы, лидогенерация
партнерские программы, увеличение ltv (life time value) клиентов
транзакционные сервисы
решения на стыке отраслей: банкинга и ритейла, банкинга и e-commerce

ПОДАТЬ ЗАЯВКУ



АГЕНТСТВО
ИННОВАЦИЙ
ГОРОДА
МОСКВЫ

О ПРОЕКТЕ

Политика в отношении обработки персональных данных

Rolling Drones. 2021

7 / 7