



# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ПРОВОДИТ ТАМОЖЕННЫЙ КОНТРОЛЬ

Команда «НЕЙМАРК»  
г. Нижний Новгород





# ПРОБЛЕМАТИКА

01

**ВРЕМЕННЫЕ ЗАТРАТЫ**  
на определение кода товара

02

**СУБЪЕКТИВНЫЕ ОШИБКИ**  
должностных лиц  
таможенных органов

03

**СЛОЖНОСТЬ ПОНИМАНИЯ**  
классификационных кодов

04

**НЕКОРРЕКТНОСТЬ НАЧИСЛЕНИЯ**  
таможенных платежей

05

**ВОЗМОЖНЫЕ ИЗДЕРЖКИ**  
связанные с неправильным  
определением кода товара

06

**НЕКОРРЕКТНОСТЬ РАБОТЫ СИСТЕМЫ**  
при прохождении алгоритмов  
автоматического выпуска товаров





# АНАЛИЗ ДАННЫХ

## Анализ:

Товарная группа (2 цифры) - 96 групп

Товарная позиция (4 цифры) - 1222 позиции

## Проблема:

29 товарных позиций отсутствуют в наборе данных

199 позиций имеют < 50 примеров на позицию

## Решение:

Предсказывать не только товарную позицию (4 цифры), но и товарную группу (2 цифры);

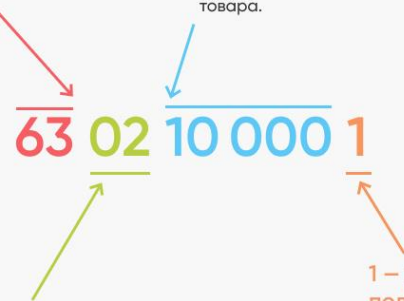
на основе товарной группы улучшать предсказание

63 — код группы

Показывает обобщённую категорию товаров.

10 000 — код субпозиции

Показывает технологию изготовления, функцию или физическую характеристику товара.



02 — код товарной позиции

Показывает конкретный вид товара.

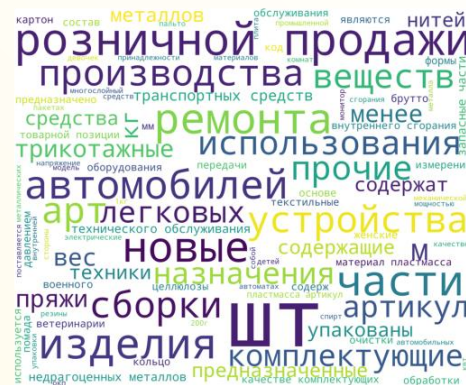
1 — код подсубпозиции

Показывает химические, визуальные особенности.



# ПРЕДОБРАБОТКА ДАННЫХ

- Приведение к нижнему регистру
- Удаление знаков препинания
  - Удаление предлогов
- Лемматизация, стемминг



Полученное облако слов до и после обработки



# ПОДХОДЫ К РЕШЕНИЮ

Товарная группа (2 цифры)

Метод	Accuracy	Presicion	Comments
SGD	0.90	0.91	На 80% датасета
FCN	0.85	0.86	На 30% датасета
<b>LSTM</b>	<b>0.94</b>	<b>0.94</b>	На 30% датасета
DeepPavlov RuBert	0.67	0.71	На 10% датасета

Товарная позиция (4 цифры)

Метод	Accuracy	Presicion	Comments
SGD	0.84	0.88	На 80% датасета
DeepPavlov RuBert	0.48	0.59	На 10% датасета
<b>Каскадное решение 2-4 цифры</b>	<b>0.90</b>	<b>0.89</b>	На 100 %

## КАСКАДНАЯ МОДЕЛЬ

- Верхняя сеть предсказывает код группы.  
На выбор: *SGD*, *LSTM*, *DeepPavlov* (качество выше 94%)
- По коду группы выбирается сеть, которая предсказывает номер товарной позиции внутри группы.

При обучении нам наиболее интересная метрика **Precision**

Мы стараемся обучить нашу систему хорошо отличать каждый наш класс от других.



# РЕШЕНИЕ

63 02 10 000 1

## Этап 1.

### Предсказание кода группы

С помощью описания товара SGD, LSTM, DeepPavlov RuBERT предсказывают 2 цифры

## Этап 2.

### Предсказание кода подгруппы

С помощью описания товара и предсказанного кода группы SGD предсказывает 4 цифры

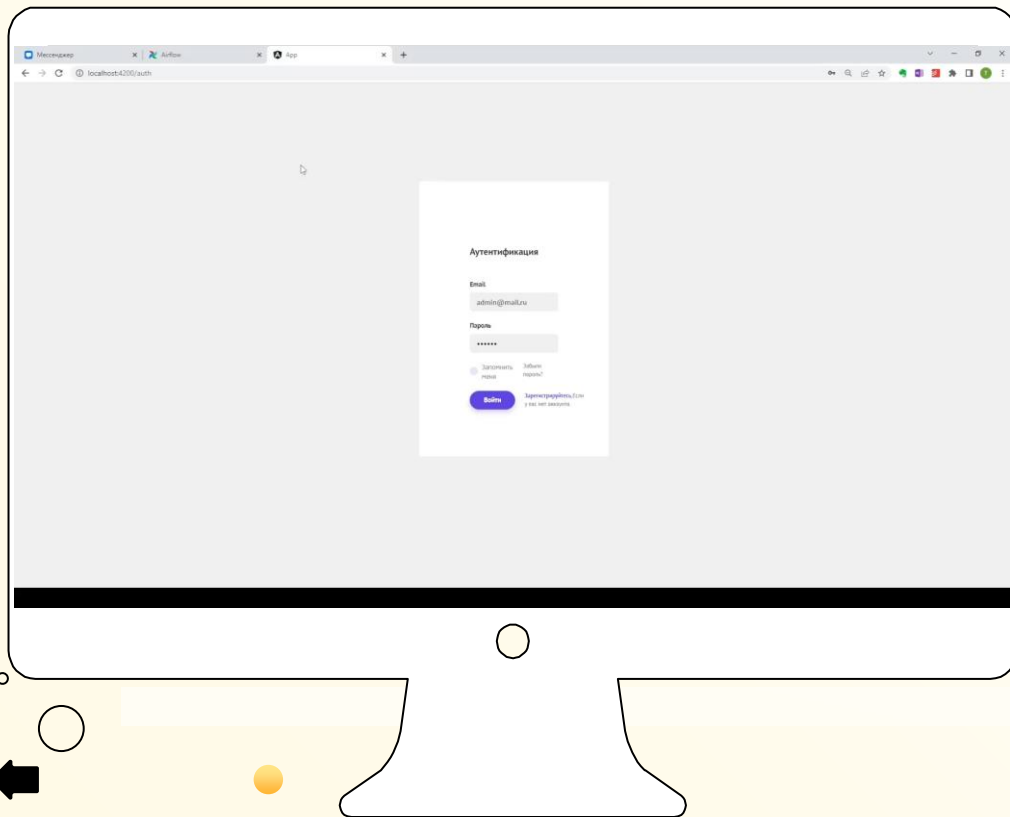
## Этап 3.

### Предсказание товарного кода

По описанию товара строится его описание в текстовых токенах (вектор токенов), затем вектор сравнивается с векторами всех описаний товарной субпозиции, выбирается ближайший



# НАШЕ РЕШЕНИЕ



- Алгоритмы по предсказанию товарной группы, алгоритмы по предсказанию кода товарной позиции, алгоритм по предсказанию полного кода ТН ВЭД (10 цифр)
- Модуль по дообучению модели на поступающих данных
- Вся система разворачивается из Docker автоматически
- Проработка пользовательских сценариев
- Реализация прототипа системы в виде web-сервиса
- Система адаптивна и масштабируема
- Точность нашего решения составляет >94%
- Отсутствие импортного ПО и библиотек

Решение:



<http://194.58.123.92:3000/>

Демо:



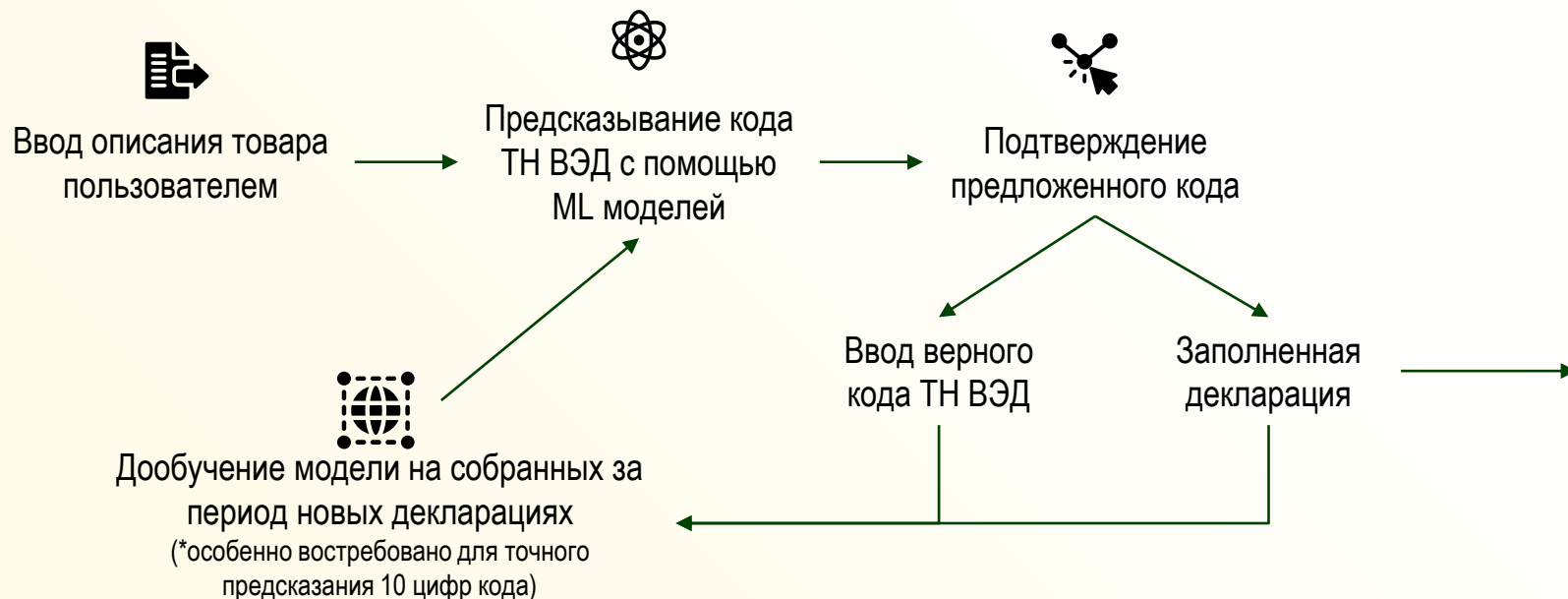
<https://clck.ru/psaKe>

Мониторинг, оркестрация, обработка данных

<http://194.58.123.92:8080/>



# АЛГОРИТМ РЕШЕНИЯ







## УНИКАЛЬНОСТЬ РЕШЕНИЯ



### Возможность добавления в каскад следующих уровней

- Каскадная система позволяет добавлять позиции рекуррентно, выделяя из датасета подмножества тренировочные данные по интересующей товарной позиции/субпозиции



### Возможность дообучения внутри товарных групп

- При поступлении новых данных можно дообучить только часть товарных позиций внутри товарных групп (первые кандидаты на дообучение - группы 50, 80, 88, 97, 93, 36)



### Apache Airflow для контроля и запуска дообучения модели



### Личный кабинет таможенного работника и участника ВЭД

## МАСШТАБИРУЕМОСТЬ



### Подсказка участнику ВЭД верного кода по описанию



### Расчет суммы налога



### Вычленение информации об объеме, весе и др.





# КОМАНДА «НЕЙМАРК»



**Багиров Миrabбас**

Fullstack-разработчик

bagirov\_mirabbas@mail.ru



**Карклин Тимофей**

DataScientist, backend-разработчик

timkar164@gmail.com



**Васильев Евгений**

ML-специалист

evgeny.vasiliev@itmm.unn.ru



**Бородинa Татьяна**

Frontend-разработчик

borodinat102@yandex.ru





# СПАСИБО ЗА ВНИМАНИЕ!

Капитан **Мираббас Багиров**  
+7(960) 185-99-55  
bagirov\_mirabbas@mail.ru

Команда «**НЕЙМАРК**»  
г. Нижний Новгород

