IV межвузовская конференция

Бизнес-аналитика. Использование аналитической платформы Loginom (Deductor) в учебном процессе вуза

Опыт использования серверных компонентов аналитической платформы Deductor Enterprise

при решении учебных и практических задач
на кафедре «Информатика и программное обеспечение»
Брянского государственного технического университета

Подвесовский Александр Георгиевич
заведующий кафедрой, к.т.н., доцент
apodv@tu-bryansk.ru

Лагерев Дмитрий Григорьевич
к.т.н., доцент
lagerevdg@mail.ru

Бабурин Артем Николаевич
ассистент
ababurin@bk.ru

г. Москва, 27 июня 2017 г.

Содержание

- Введение
- Проекты, выполненные с использованием серверных компонентов платформы Deductor Enterpise
  - Выявление групп риска в рамках мониторинга наркоситуации в Брянской области
  - Программная поддержка полного цикла социологического исследования
  - Поиск единомышленников в социальной сети VK
- Итоги работы. Дальнейшие планы и пожелания

Содержание

- **Введение**
- Проекты, выполненные с использованием серверных компонентов платформы Deductor Enterprise
  - Выявление групп риска в рамках мониторинга наркоситуации в Брянской области
  - Программная поддержка полного цикла социологического исследования
  - Поиск единомышленников в социальной сети VK
- Итоги работы. Дальнейшие планы и пожелания

Кафедра «Информатика и программное обеспечение»: образовательный процесс

Бакалавриат

02.03.03 – Математическое обеспечение и администрирование информационных систем

09.03.01 – Информатика и вычислительная техника

- Программное обеспечение вычислительной техники и автоматизированных систем

09.03.04 – Программная инженерия

- Разработка программно-информационных систем

Магистратура

09.04.01 – Информатика и вычислительная техника

- ***Компьютерный анализ и интерпретация данных***
- Информационное и программное обеспечение вычислительных систем

09.04.04 – Программная инженерия

- Проектирование программно-информационных систем

Аспирантура

05.13.18 – Математическое моделирование, численные методы и комплексы программ

05.13.10 – Управление в социальных и экономических системах

Магистерская программа «Компьютерный анализ и интерпретация данных»

- Реализуется в рамках направления «Информатика и вычислительная техника» с 2009 г.
- К настоящему моменту выпущено 34 магистра
- Основной целью является углубленная (на базе соответствующих направлений бакалавриата) подготовка профессиональных разработчиков программного обеспечения и системных аналитиков со специализацией в следующих областях
  - **обработка и анализ больших объемов данных, методы машинного обучения**
  - **модели и методы поддержки принятия решений**
  - **интеллектуальные системы на основе мягких вычислений**
  - **обработка и анализ изображений, машинное зрение**
  - **цифровая обработка сигналов**

Ключевые дисциплины учебного плана

1. Базовые дисциплины направления

- Методы оптимизации
- Теория принятия решений
- Компьютерное моделирование
- Теория систем и системный анализ
- Технология проектирования, разработки и верификации программного обеспечения

2. Математический аппарат, методология и инструменты анализа данных

- Статистический анализ данных
- **Интеллектуальный анализ данных**
- Теория нейронных сетей
- Интеллектуальные системы

3. Специальные разделы и приложения методологии анализа данных

- **Хранилища данных**
- Системы с параллельной обработкой данных
- Обработка и анализ изображений
- Системы машинного зрения
- Цифровая обработка сигналов

Интеллектуальный анализ данных: общие сведения о курсе

- ***Цели курса***
  - изучение современных информационных технологий, предназначенных для интеллектуального анализа данных
  - формирование целостного представления об анализе и интерпретации данных, как о процессе поиска, так и о методологии применения скрытых в них закономерностей для достижения поставленных целей
- ***Структура курса***
  - изучается на 1-м курсе во 2-м семестре
  - лекции – 17 часов
  - лабораторные работы – 34 часа
  - курсовой проект, экзамен
- ***Теоретическая часть:***
  - структура и возможности аналитических систем
  - основные методы интеллектуального анализа и предобработки данных
  - процесс ETL
  - ансамбли моделей
- ***Практическая часть:***
  - использование **Deductor Studio Academic / Professional / Enterprise**

История использования платформы Deductor на кафедре

- С начала 2000-х – эпизодическое использование Deductor Studio Academic в рамках отдельных лабораторных работ по некоторым дисциплинам
- **2009 – начало систематического использования Deductor Studio Academic в курсе «Интеллектуальный анализ данных» в магистратуре**
- 2015 – начало регулярного взаимодействия с компанией BaseGroup в рамках партнерской программы
- 2016, март – приобретение Deductor Studio Professional
- **2016, июнь – участие с докладом в III межвузовской конференции**
  - https://basegroup.ru/system/files/events/lagerev.pdf
- **2016, август – получение лицензии на использование Deductor Studio Enterprise, Deductor Integration Server, Deductor Analytic Server**
- 2016, октябрь – прохождение доцентом Лагеревым Д.Г. тренинга «Разработка скоринговых моделей»
- 2016-2017 – обучение и сертификация преподавателей
  - доцент Лагерев Д.Г. – прохождение электронного курса, сдача экзамена
  - ассистент Бабурин А.Н. – прохождение электронного курса

Содержание

- Введение
- **Проекты, выполненные с использованием серверных компонентов платформы Deductor Enterprise**
  - Выявление групп риска в рамках мониторинга наркоситуации в Брянской области
  - Программная поддержка полного цикла социологического исследования
  - Поиск единомышленников в социальной сети VK
- Дальнейшие планы

Наша команда

Бондарева Инна

Козлов Евгений

Герасимчук Иван

Тупикина Екатерина

Журин Владислав

Самородов Илья

Содержание

- Введение
- **Проекты, выполненные с использованием серверных компонентов платформы Deductor Enterprise**
  - **Выявление групп риска в рамках мониторинга наркоситуации в Брянской области**
  - Программная поддержка полного цикла социологического исследования
  - Поиск единомышленников в социальной сети VK
- Итоги работы. Дальнейшие планы и пожелания

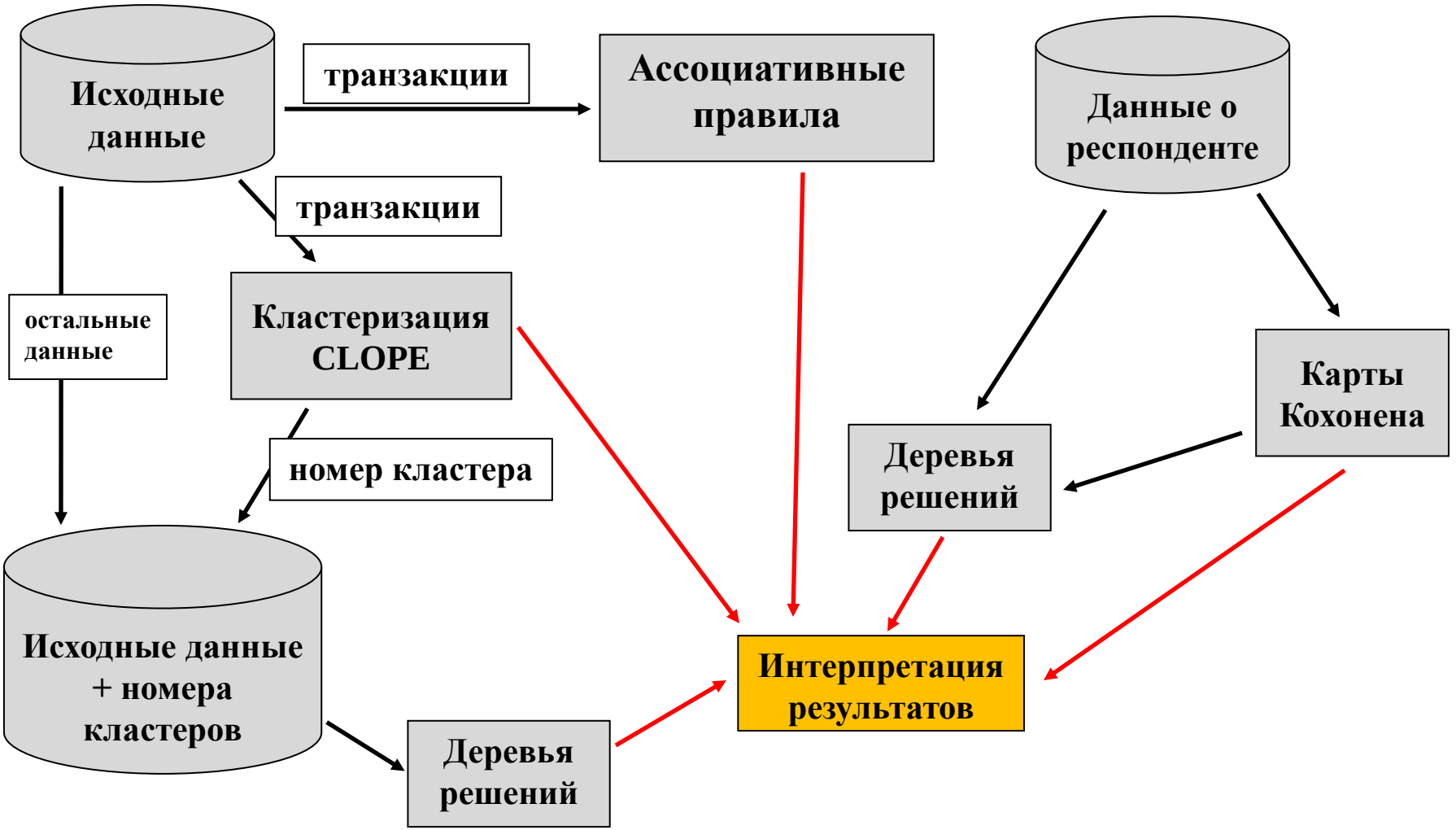
Проект «Мониторинг наркоситуации». Описание проекта

- Исходные положения
  - Анализ и оценка состояния наркоситуации в Брянской области
  - Проводится ежегодно по заказу УФСКН РФ по Брянской области
  - Исполнитель – кафедра социально-гуманитарных дисциплин Брянского филиала Российской академии народного хозяйства и государственной службы
    - Мы приняли участие благодаря имеющимся научным связям с данной кафедрой
  - Основной метод – социологический опрос населения, в том числе пациентов наркологических клиник
    - Единая форма анкеты (в рамках РФ)
    - Рекомендована статистическая обработка результатов анкетирования
  - Социологический опрос проводился в 2013-2017 гг.
    - 2013 г. – 1875 респондентов
    - 2014 г. – 2211 респондентов
    - 2015 г. – 1915 респондентов
    - 2016 г. – 1208 респондентов
    - 2017 г. – проект планируется продолжить, но данные пока недоступны
- Цель проекта – повышение эффективности обработки данных за счет применения методов интеллектуального анализа

Проект «Мониторинг наркоситуации». Структура данных

- Анкета содержит следующие типы вопросов
  - Данные о респонденте (пол, возраст, образование)
  - Жизненные ориентиры респондента (наиболее острые проблемы, жизненные ценности, проведение свободного времени)
  - Вопросы, касающиеся здоровья респондента (оценка здоровья, наличие вредных привычек)
  - Отношение респондента к наркотикам и наркомании
- Структура анкеты в 2013-2016 гг. незначительно менялась. Общее число вопросов – от 37 до 42, среди которых
  - с ответами в категориальной шкале – от 32 до 36 вопросов
  - с ответами в порядковой шкале – от 1 до 2 вопросов
  - с ответами в числовой шкале – от 3 до 4 вопросов

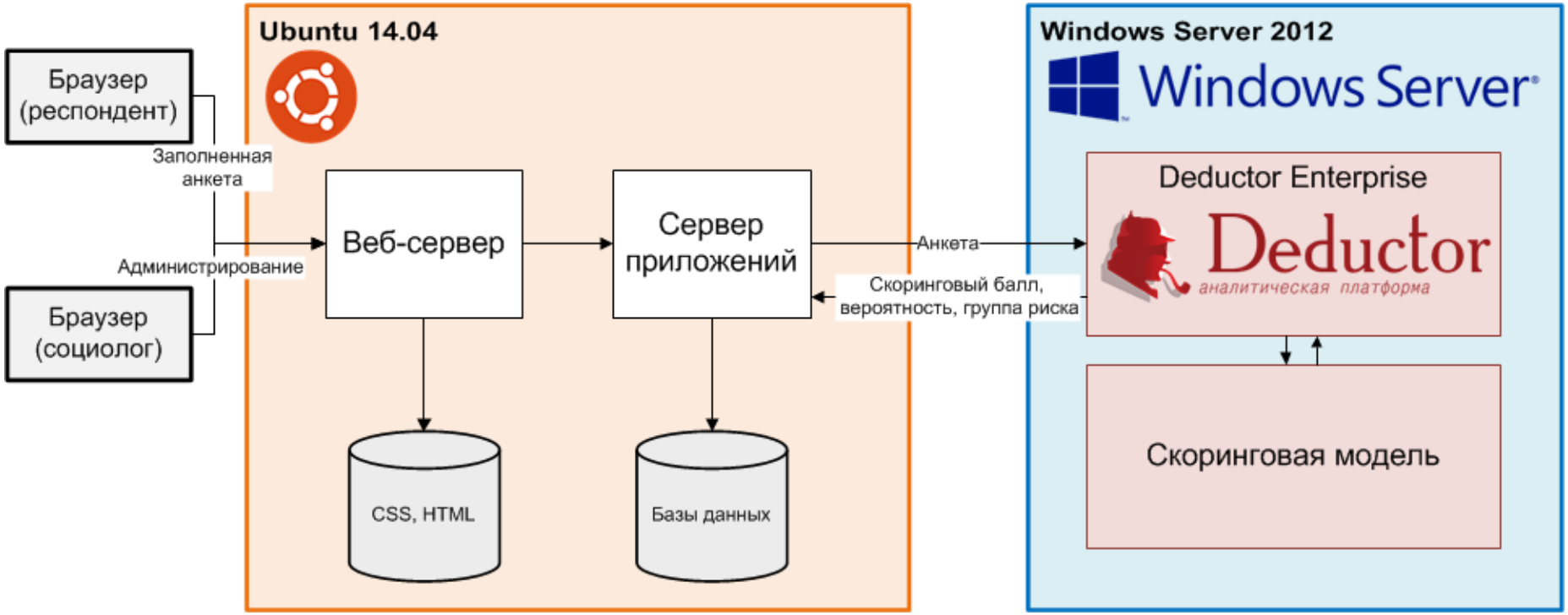
Проект «Мониторинг наркоситуации». Ансамбль моделей



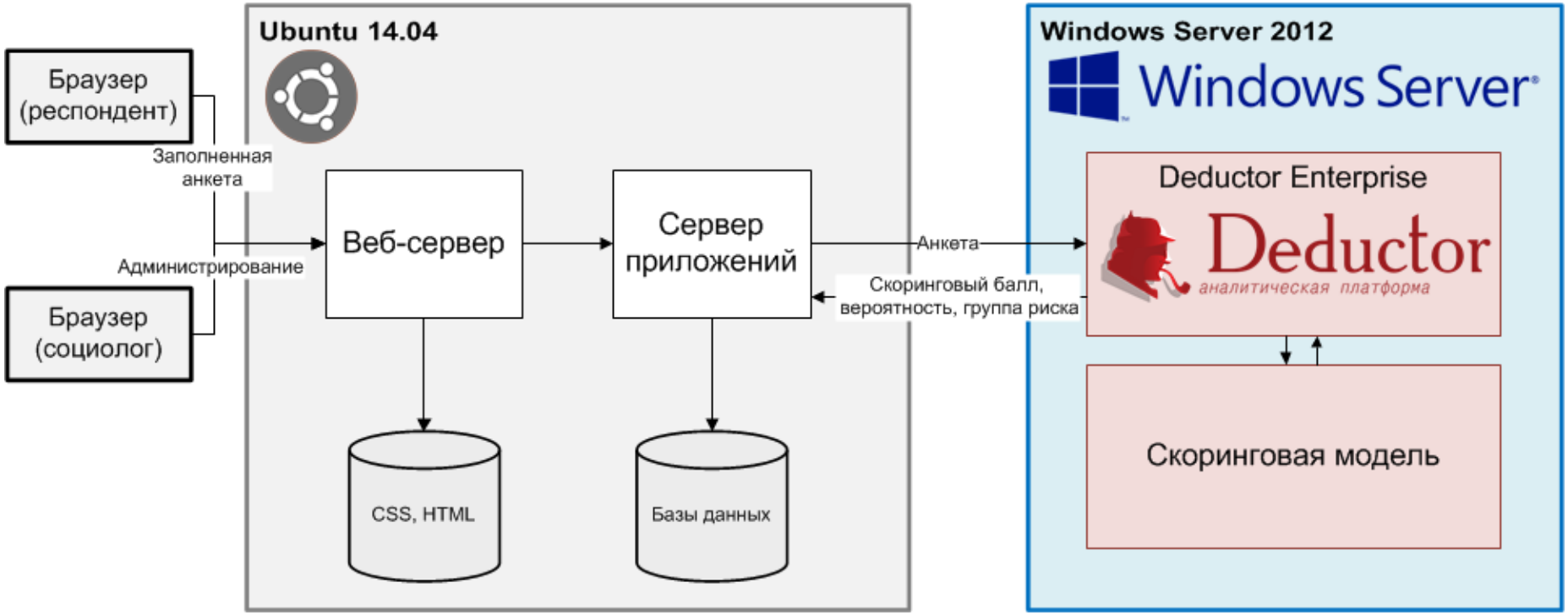
Постановка задачи

- Проблемы:
  - социально приемлемые ответы респондентов (особенно подростков)
  - ложные ответы на вопросы, касающиеся наркотиков и проблемы наркомании
  - …
- **Идея:** построить скоринговую модель, которая выделяла бы группы риска относительно наркозависимости на основе анкеты, не содержащей вопросов, напрямую связанных с наркотикам и проблемой наркомании
  - эти вопросы исключены из анкеты
  - добавлены фиктивные вопросы на тему компьютерных игр (для отвлечения внимания), которые не будут подавать в сценарий скоринговой модели

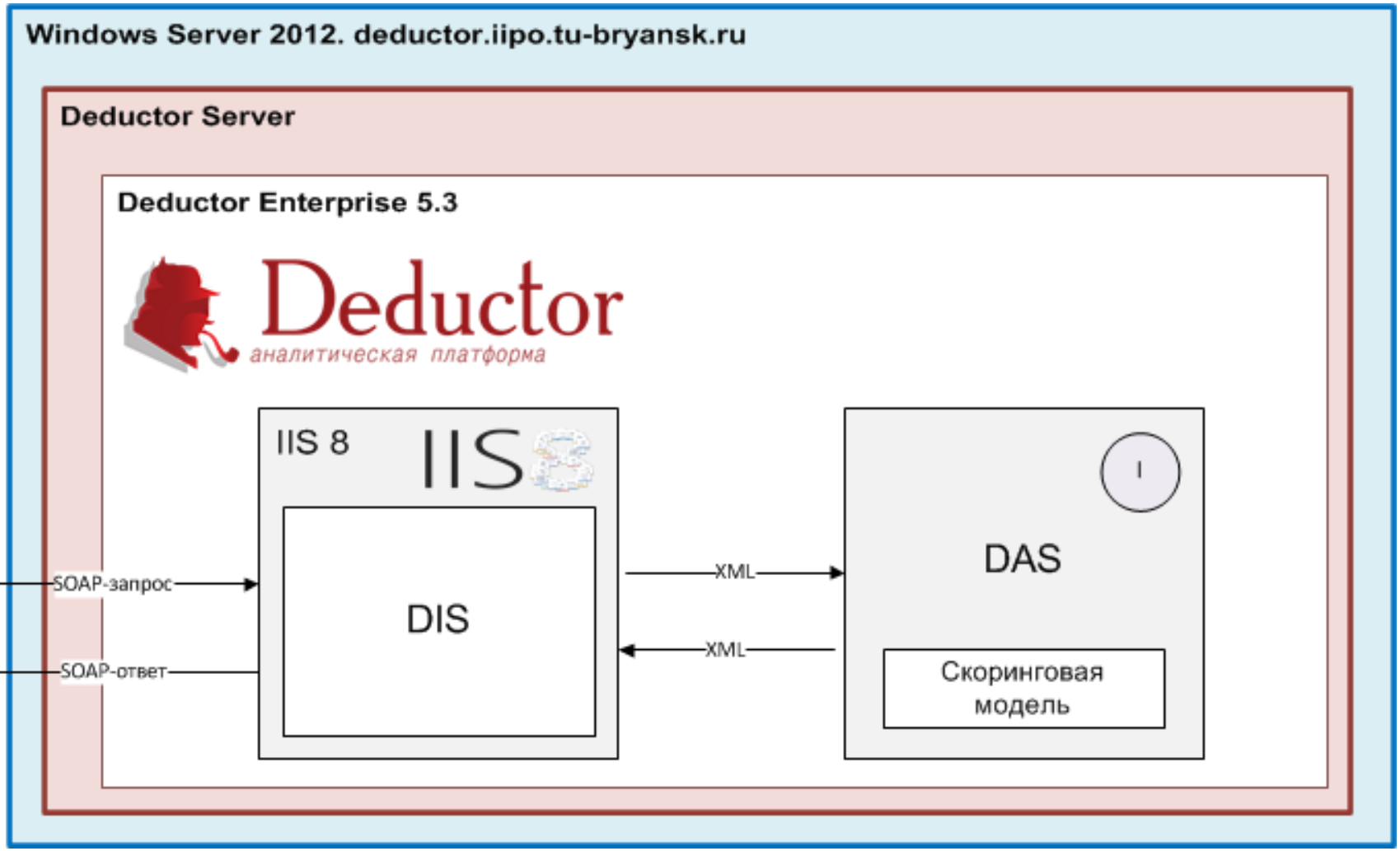
Архитектура приложения



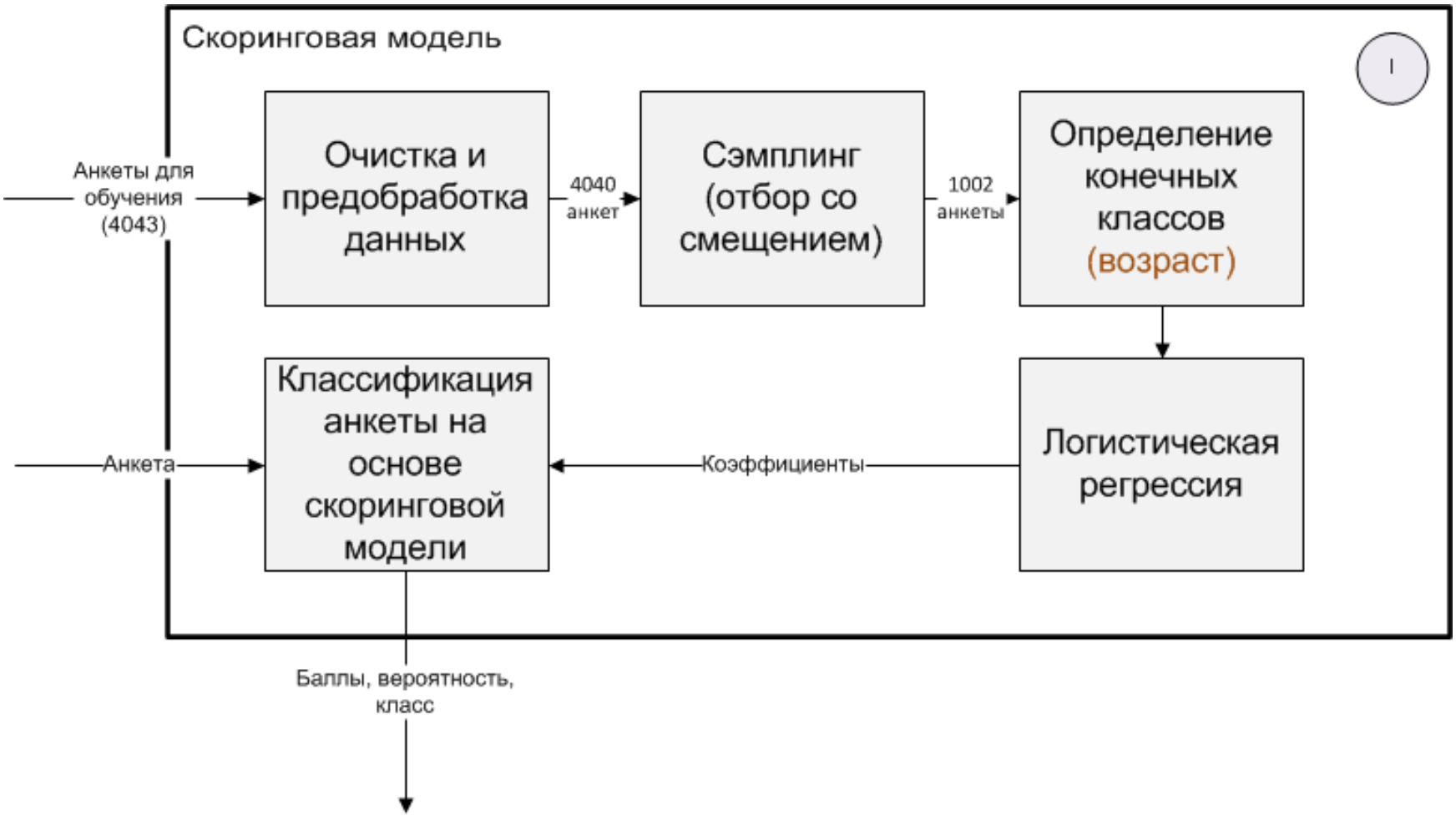
Архитектура приложения



Архитектура подсистемы взаимодействия с внешними клиентами



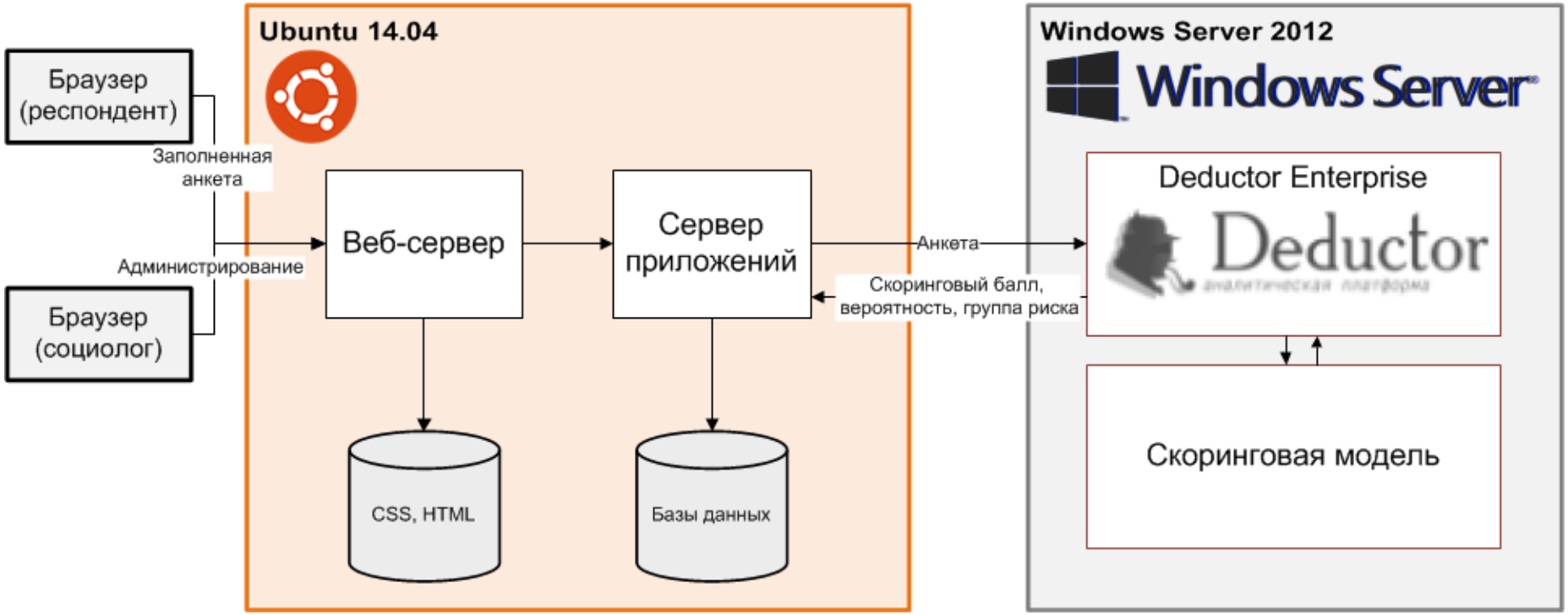
Структура скоринговой модели



Скоринговые баллы

Вариант ответа	Скоринговый балл
Материальное положение – высоко обеспеченная семья	53,69
Материальное положение – не обеспеченная самым необходимым	48,95
Оценка уровня здоровья (настроение, «жизненные силы») – плохое, скорее плохое	48,58
Социальное положение – руководитель отдела, подразделения	43,90
Пол – мужской	39,08
Наличие вредных привычек – имеются	30,68
….	….

Архитектура приложения



Архитектура визуальной и административной подсистем

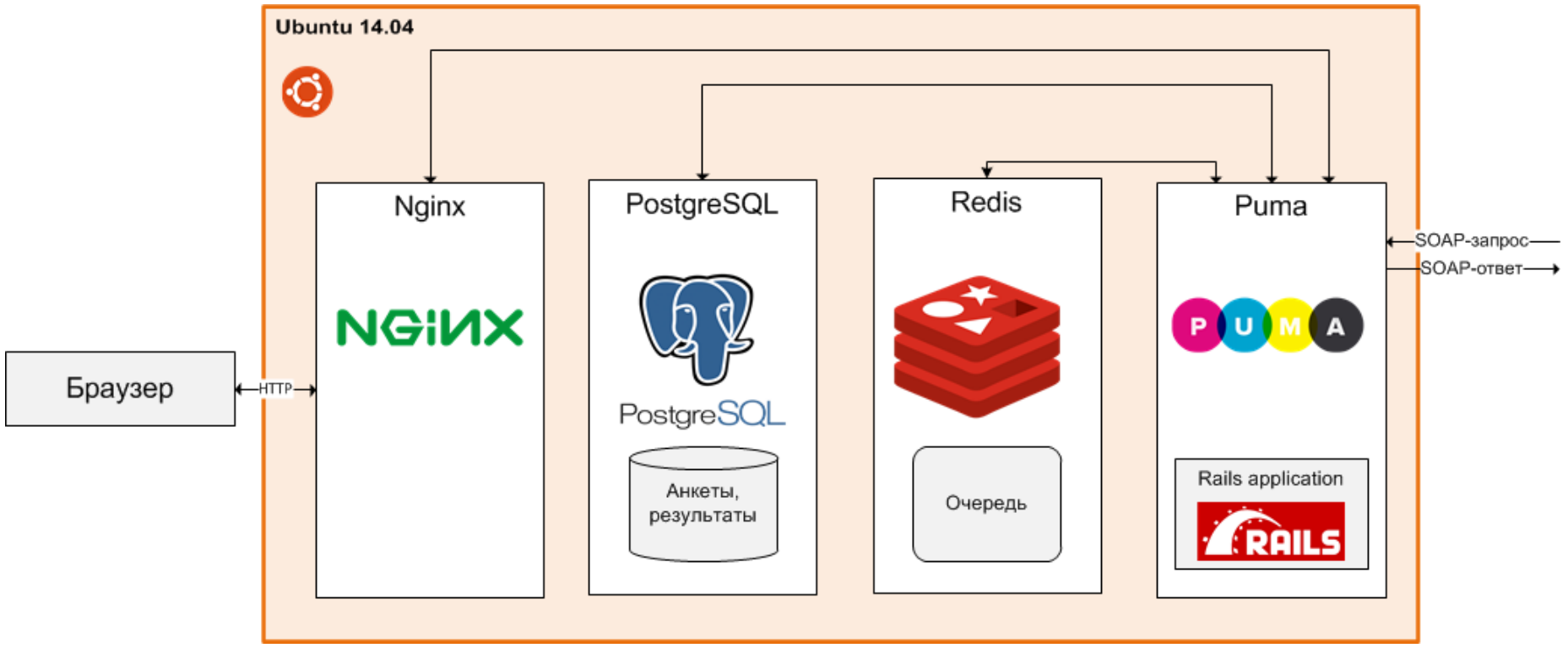
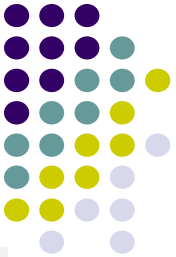


Схема данных XML

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema      xmlns:xs="http://www.w3.org/2001/XMLSchema"
                                        xmlns:tns="drug_add" targetNamespace="drug_add"
                                        elementFormDefault="qualified"
                                        attributeFormDefault="unqualified">
        <xs:complexType name="Input">
                <xs:sequence>
                                        <xs:element name="q1_1" type="xs:int" />
                                        <xs:element name="q1_2" type="xs:int" />
                                        <xs:element name="q1_3" type="xs:int" />
                                        ……
                                        <xs:element name="social" type="xs:int" />
                                        <xs:element name="material" type="xs:int" />
                                        <xs:element name="Target" type="xs:int" />
                </xs:sequence>
        </xs:complexType>
        <xs:complexType name="Output">
                <xs:sequence>
                                        <xs:element name="Target_OUT" type="xs:boolean"/>
                                        <xs:element name="Target_prob" type="xs:float"/>
                                        <xs:element name="Target_ball" type="xs:float"/>
                </xs:sequence>
        </xs:complexType>
        <xs:element name="input" type="tns:Input"/>
        <xs:element name="output" type="tns:Output"/>
</xs:schema>
```

Описание WSDL-сервиса

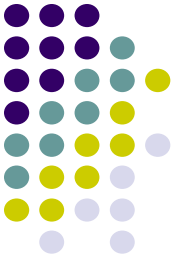


deductor.iipo.tu-bryansk.ru/DIS_drug/Service.svc?wsdl

This XML file does not appear to have any style information associated with it. The document tree is shown below.


```
▼<wsdl:definitions xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/" xmlns:wsx="http://schemas.xmlsoap.org/ws/2004/09/mex" xmlns:wsu="http://docs.oasis-open.org/wss/2004/01/oasis-
200401-wss-wssecurity-utility-1.0.xsd" xmlns:wsa10="http://www.w3.org/2005/08/addressing" xmlns:wsp="http://schemas.xmlsoap.org/ws/2004/09/policy"
xmlns:wsap="http://schemas.xmlsoap.org/ws/2004/08/addressing/policy" xmlns:msc="http://schemas.microsoft.com/ws/2005/12/wsdl/contract"
xmlns:soap12="http://schemas.xmlsoap.org/wsdl/soap12/" xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/08/addressing" xmlns:wsam="http://www.w3.org/2007/05/addressing/metadata"
xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:tns="http://www.basegroup.ru/DeductorIntegrationServer" xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
xmlns:wsaw="http://www.w3.org/2006/05/addressing/wsdl" xmlns:soapenc="http://schemas.xmlsoap.org/soap/encoding/" name="ServiceContract"
targetNamespace="http://www.basegroup.ru/DeductorIntegrationServer">
 ▼<wsdl:types>
  ▼<xsd:schema targetNamespace="http://www.basegroup.ru/DeductorIntegrationServer/Imports">
     <xsd:import schemaLocation="http://deductor.iipo.tu-bryansk.ru/DIS_drug/Service.svc?xsd=xsd0" namespace="drug_add"/>
     <xsd:import schemaLocation="http://deductor.iipo.tu-bryansk.ru/DIS_drug/Service.svc?xsd=xsd1" namespace="http://www.basegroup.ru/warehouses/XsdDbConnection4"/>
     <xsd:import schemaLocation="http://deductor.iipo.tu-bryansk.ru/DIS_drug/Service.svc?xsd=xsd2" namespace="http://www.basegroup.ru/DeductorIntegrationServer"/>
   </xsd:schema>
  </wsdl:types>
 ▼<wsdl:message name="drug_addMessage">
    <wsdl:part name="parameters" element="tns:drug_add"/>
  </wsdl:message>
 ▼<wsdl:message name="drug_addResponseMessage">
    <wsdl:part name="parameters" element="tns:drug_addResponse"/>
  </wsdl:message>
 ▼<wsdl:message name="ServiceClass_drug_add_SOAPExceptionFault_FaultMessage">
    <wsdl:part name="detail" element="tns:SOAPException"/>
  </wsdl:message>
 ▼<wsdl:portType name="ServiceClass">
  ▼<wsdl:operation name="drug_add">
     <wsdl:input wsaw:Action="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/drug_add" name="drug_addMessage" message="tns:drug_addMessage"/>
     <wsdl:output wsaw:Action="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/drug_addResponse" name="drug_addResponseMessage"
     message="tns:drug_addResponseMessage"/>
     <wsdl:fault wsaw:Action="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/drug_addSOAPExceptionFault" name="SOAPExceptionFault"
     message="tns:ServiceClass_drug_add_SOAPExceptionFault_FaultMessage"/>
    </wsdl:operation>
  </wsdl:portType>
 ▼<wsdl:binding name="BasicHttpBinding_ServiceClass" type="tns:ServiceClass">
    <soap:binding transport="http://schemas.xmlsoap.org/soap/http"/>
  ▼<wsdl:operation name="drug_add">
     <soap:operation soapAction="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/drug_add" style="document"/>
   ▼<wsdl:input name="drug_addMessage">
      <soap:body use="literal"/>
    </wsdl:input>
   ▼<wsdl:output name="drug_addResponseMessage">
      <soap:body use="literal"/>
    </wsdl:output>
   ▼<wsdl:fault name="SOAPExceptionFault">
      <soap:fault name="SOAPExceptionFault" use="literal"/>
    </wsdl:fault>
    </wsdl:operation>
  </wsdl:binding>
 ▼<wsdl:service name="ServiceContract">
  ▼<wsdl:port name="BasicHttpBinding_ServiceClass" binding="tns:BasicHttpBinding_ServiceClass">
     <soap:address location="http://deductor.iipo.tu-bryansk.ru/DIS_drug/Service.svc/ServiceClass"/>
   </wsdl:port>
  </wsdl:service>
</wsdl:definitions>
```

Тело HTTP-запроса (SOAP)



```
<soap:Envelope
   xmlns:soap="http://schemas.xmlsoap.org/soap/envelope/"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema">
   <soap:Body>
  <drug_add xmlns="http://www.basegroup.ru/DeductorIntegrationServer">
      <variables xmlns=""/>
      <data xmlns="">
          <input xmlns="http://www.basegroup.ru/warehouses/XsdDbConnection4">
              <q1_1 xmlns="drug_add">1</q1_1>
              <q1_2 xmlns="drug_add">1</q1_2>
              <q1_3 xmlns="drug_add">0</q1_3>
              <q1_4 xmlns="drug_add">1</q1_4>
              ..……..……..
              <sex xmlns="drug_add">2</sex>
              <age xmlns="drug_add">19</age>
              <education xmlns="drug_add">4</education>
              <social xmlns="drug_add">1</social>
              <material xmlns="drug_add">3</material>
              <Target xmlns="drug_add">1</Target>
          </input>
      </data>
  </drug_add>
</soap:Body>
</soap:Envelope>
```

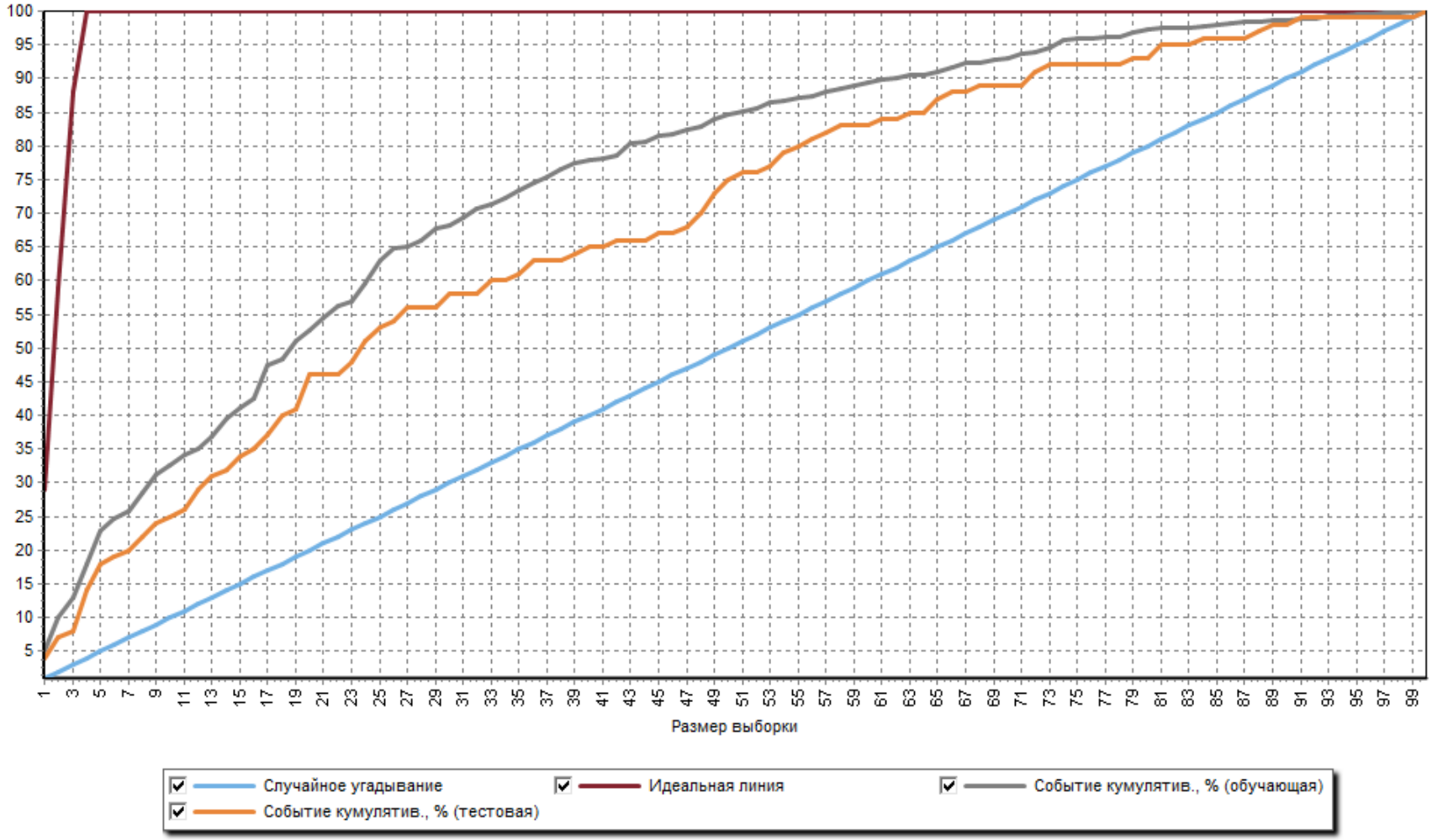
Тело HTTP-ответа (SOAP)

```
<s:Envelope xmlns:s="http://schemas.xmlsoap.org/soap/envelope/">
   <s:Body xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
      <drug_addResponse
xmlns="http://www.basegroup.ru/DeductorIntegrationServer">
         <output xmlns="http://www.basegroup.ru/warehouses/XsdDbConnection4"
xmlns:DRGD="drug_add">
            <DRGD:Target_OUT>false</DRGD:Target_OUT>
            <DRGD:Target_prob>0.072048804057169</DRGD:Target_prob>
            <DRGD:Target_ball>413.382858296961</DRGD:Target_ball>
         </output>
      </drug_addResponse>
   </s:Body>
</s:Envelope>
```

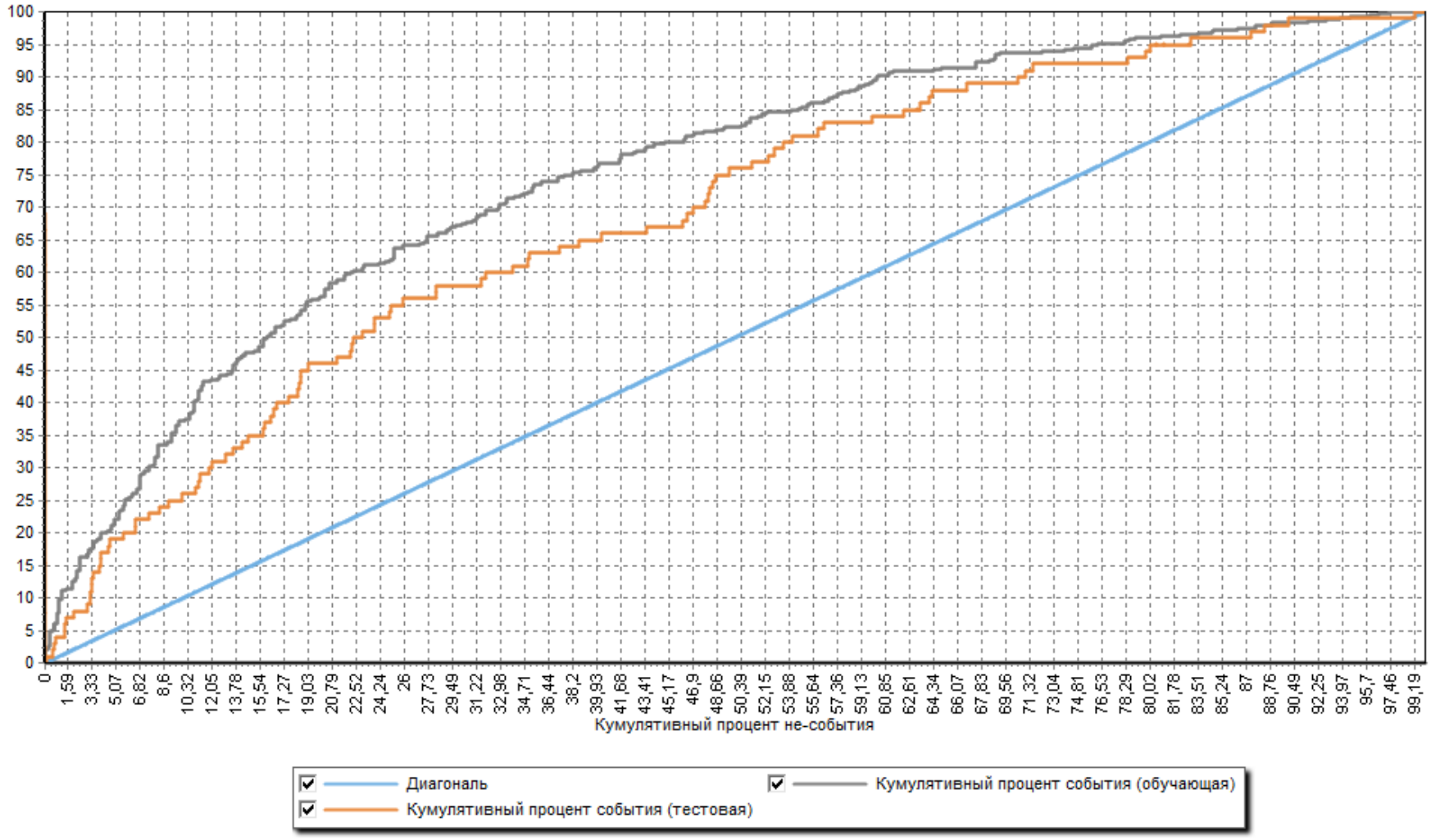
Оценка качества моделей

- Построено 4 скоринговых модели, которые отличались методом отбора переменных
- Для каждой модели были построены
  - ROC-кривые
  - CAP-кривые
  - диаграммы, отражающие шансы события/не-события
- Выбрана скоринговая модель с лучшими показателями
  - Метод отбора переменных – прямой отбор
  - Использована в сценарии Deductor Enterprise в качестве рабочей модели

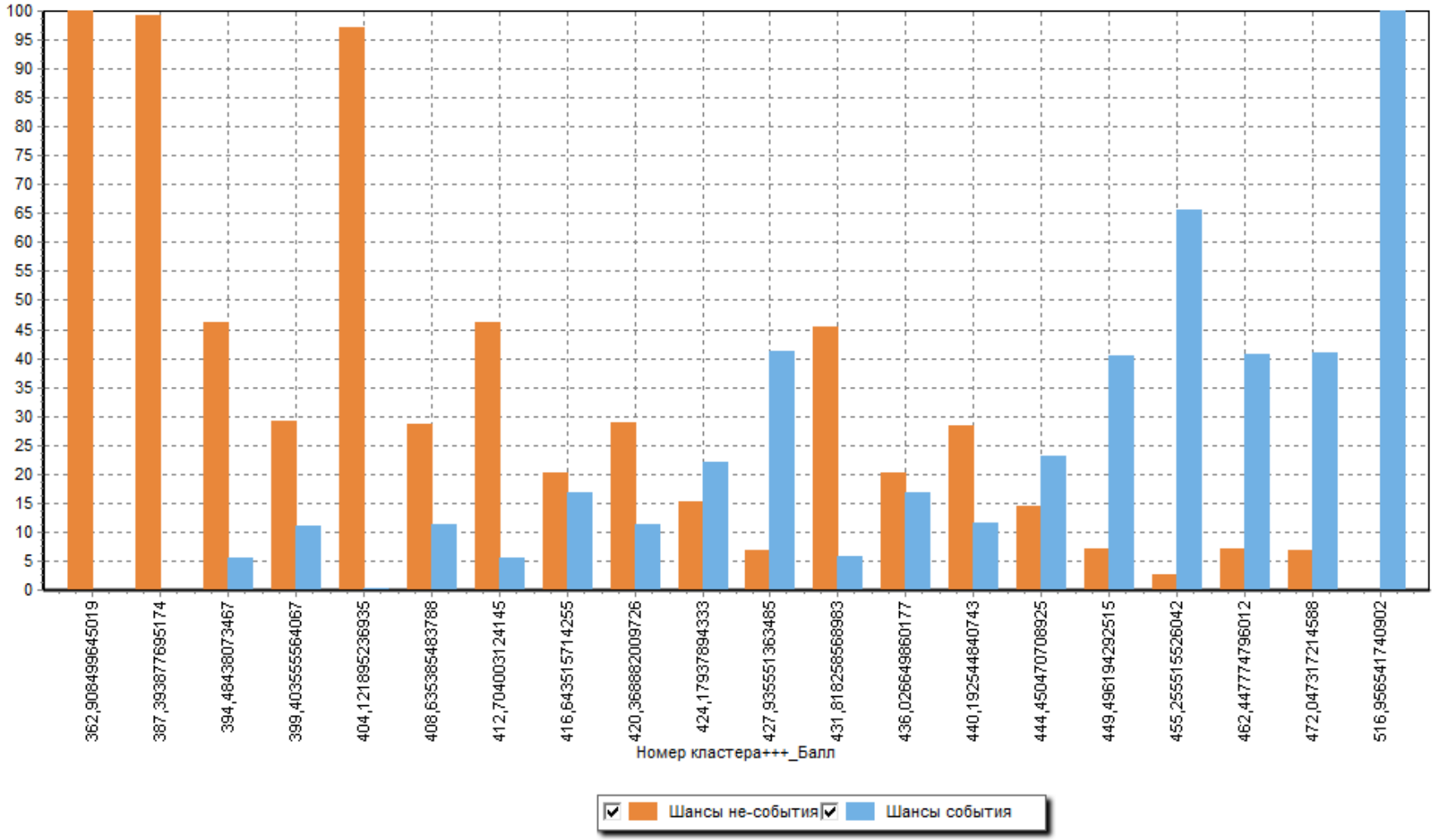
CAP-кривые


100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99
Размер выборки
Случайное угадывание
Идеальная линия
Событие кумулятив., % (обучающая)
Событие кумулятив., % (тестовая)

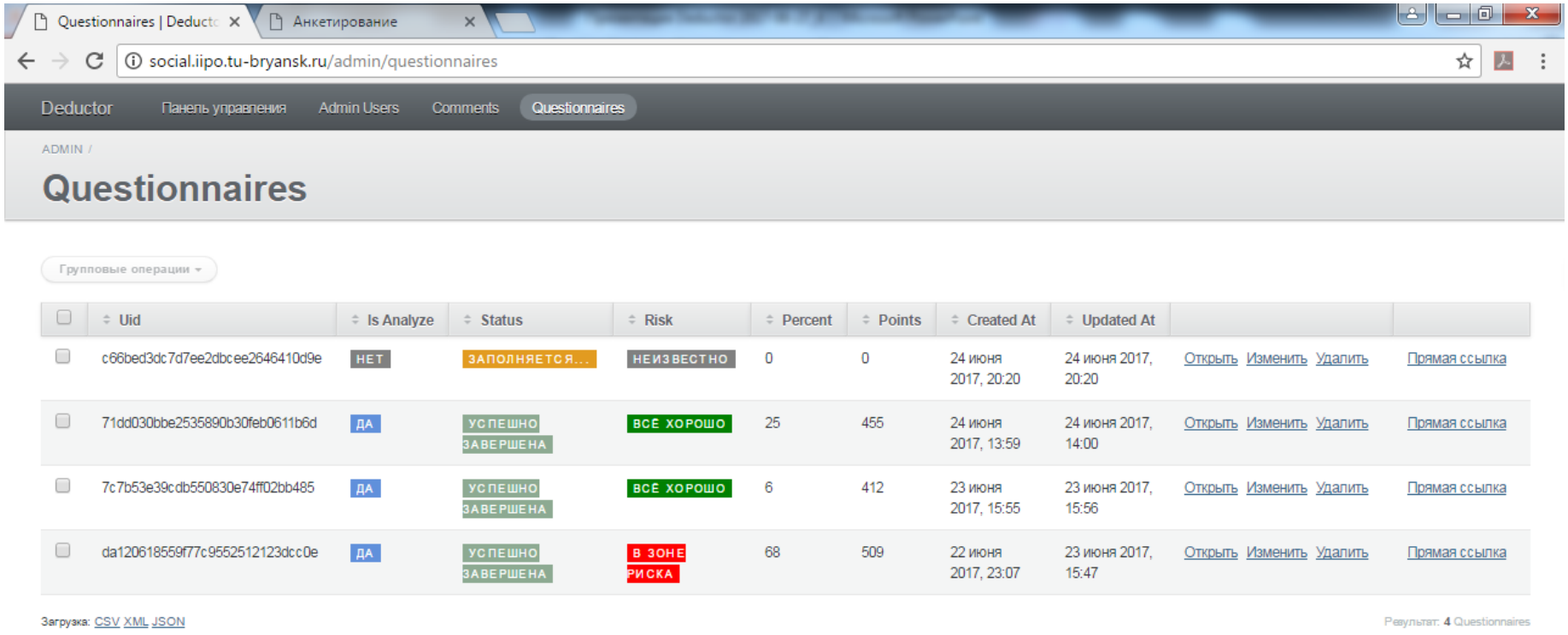
ROC-кривые


100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
1,59
3,33
5,07
6,82
8,6
10,32
12,05
13,78
15,54
17,27
19,03
20,79
22,52
24,24
26
27,73
29,49
31,22
32,98
34,71
36,44
38,2
39,93
41,68
43,41
45,17
46,9
48,66
50,39
52,15
53,88
55,64
57,36
59,13
60,85
62,61
64,34
66,07
67,83
69,56
71,32
73,04
74,81
76,53
78,29
80,02
81,78
83,51
85,24
87
88,76
90,49
92,25
93,97
95,7
97,46
99,19
Кумулятивный процент не-события
Диагональ
Кумулятивный процент события (обучающая)
Кумулятивный процент события (тестовая)

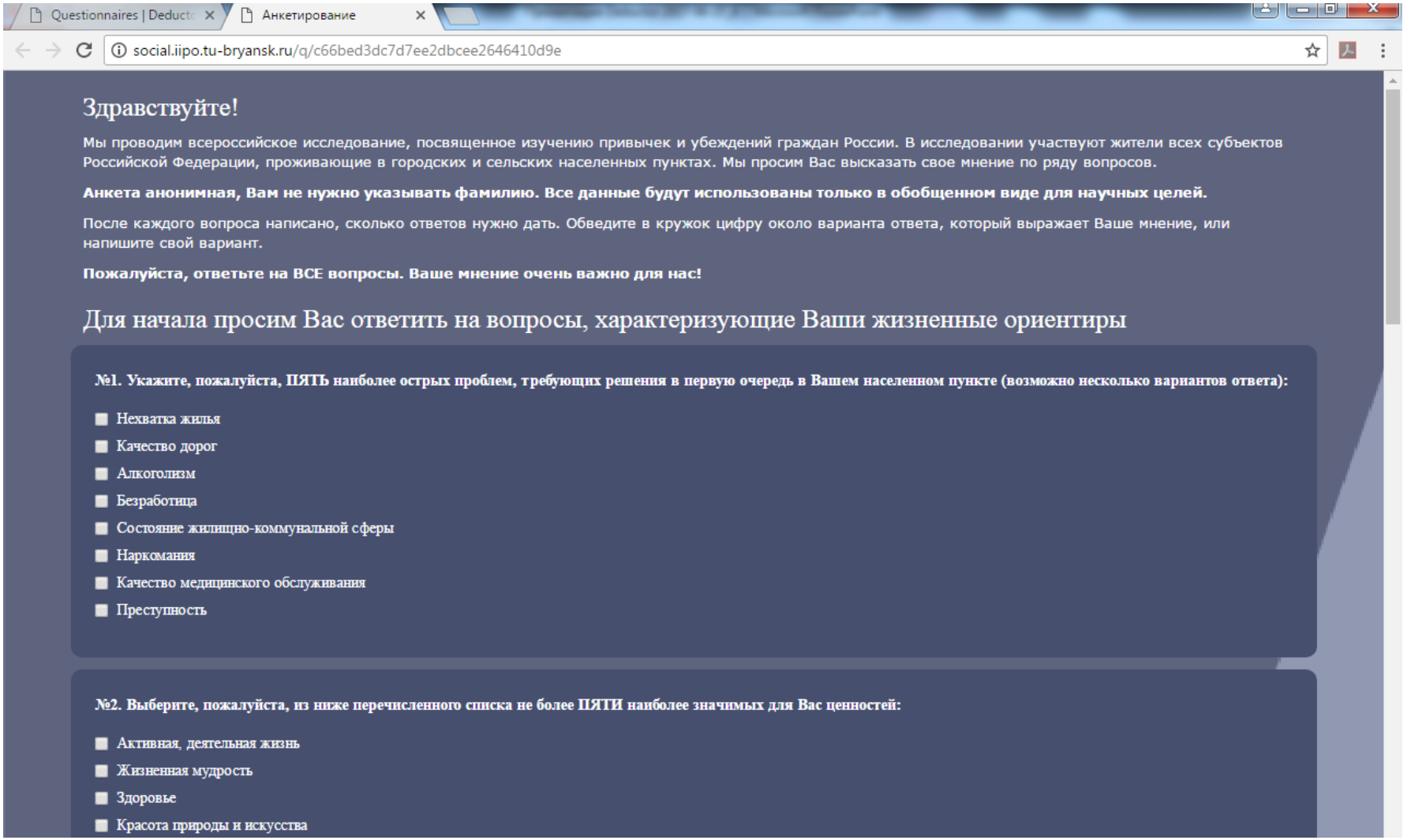
Шансы события/не-события



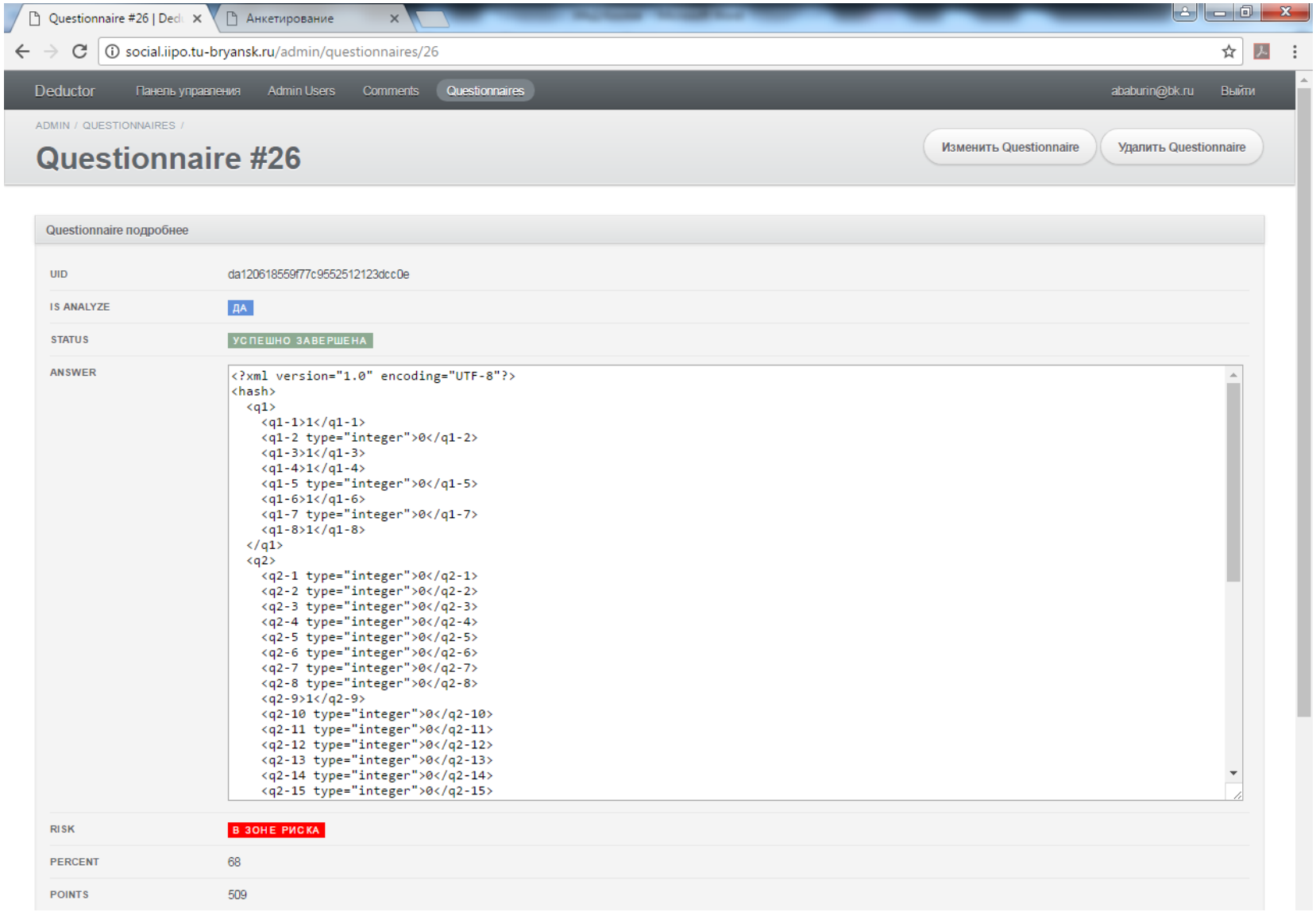
Панель администрирования



Фрагмент анкеты



Информация о заполненной анкете



Итоги и перспективы

- Построена скоринговая модель для выявления группы риска относительно наркозависимости
- Разработаны соответствующие веб-приложения, которые используют серверные компоненты аналитической платформы Deductor Enterprise:
  - веб-приложение для анкетирования респондентов
  - веб-приложение для социолога (просмотр результатов опроса и классификация респондентов)
- В ближайшее время планируется
  - представить проект заказчику
  - выполнить экспериментальную проверку на реальном социологическом опросе
  - подготовить пресс-релиз для сайта компании BaseGroup Labs
  - опубликовать полученные результаты

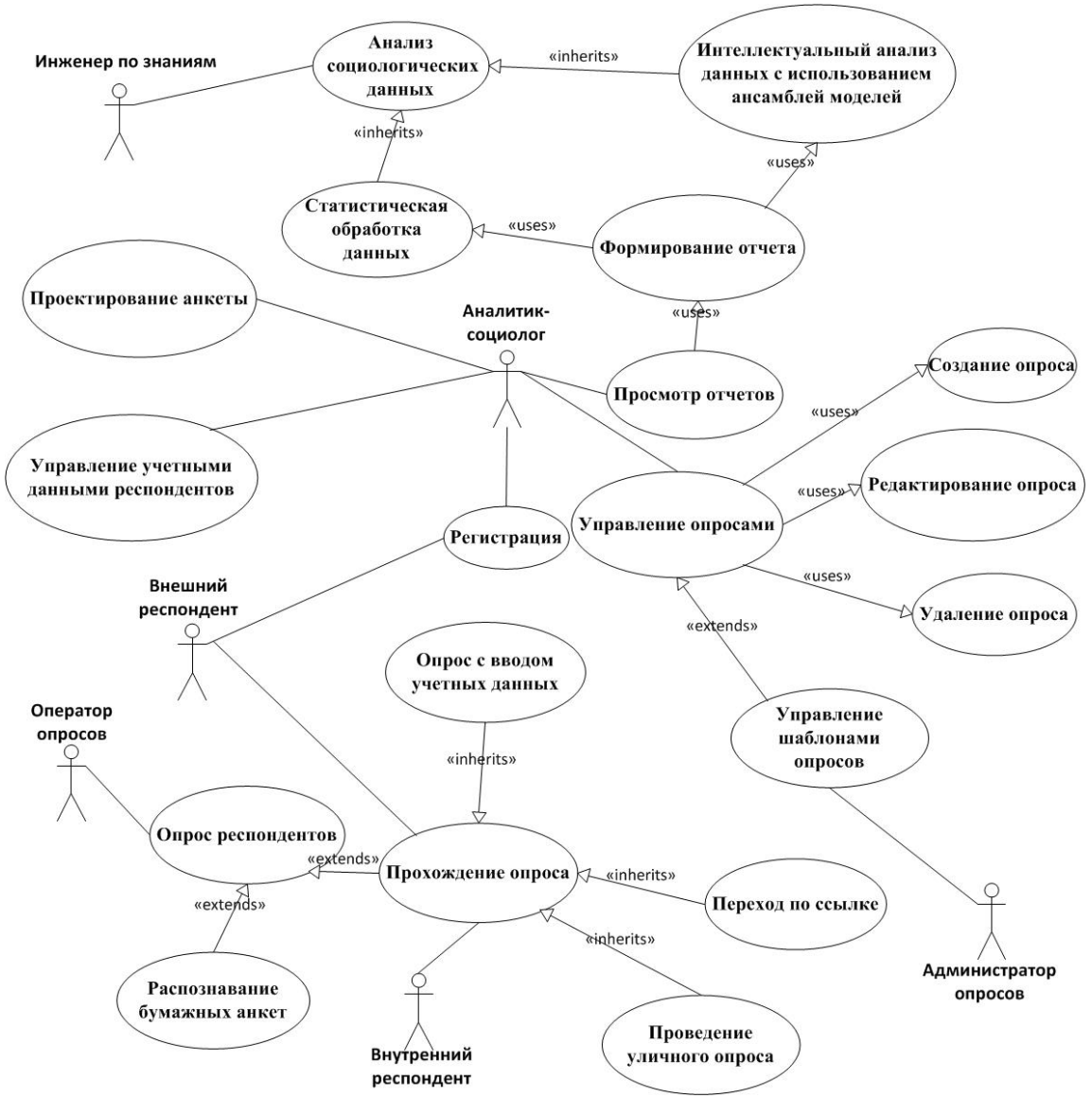
Содержание

- Введение
- **Проекты, выполненные с использованием серверных компонентов платформы Deductor Enterprise**
  - Выявление групп риска в рамках мониторинга наркоситуации в Брянской области
  - **Программная поддержка полного цикла социологического исследования**
  - Поиск единомышленников в социальной сети VK
- Итоги работы. Дальнейшие планы и пожелания

О проекте

- Выполняется в рамках кандидатской диссертации Бабурина А.Н.
- Рабочее название «Разработка методов и программных средств комплексной автоматизации социологических исследований с использованием ансамблей моделей интеллектуального анализа данных»
  - Специальность 05.13.10 «Управление в социальных и экономических системах»
- **Основная идея:** комплексная автоматизация полного цикла социологического исследования, начиная от составления анкеты и заканчивая выдачей результатов интеллектуального анализа заказчику.

Диаграмма вариантов использования

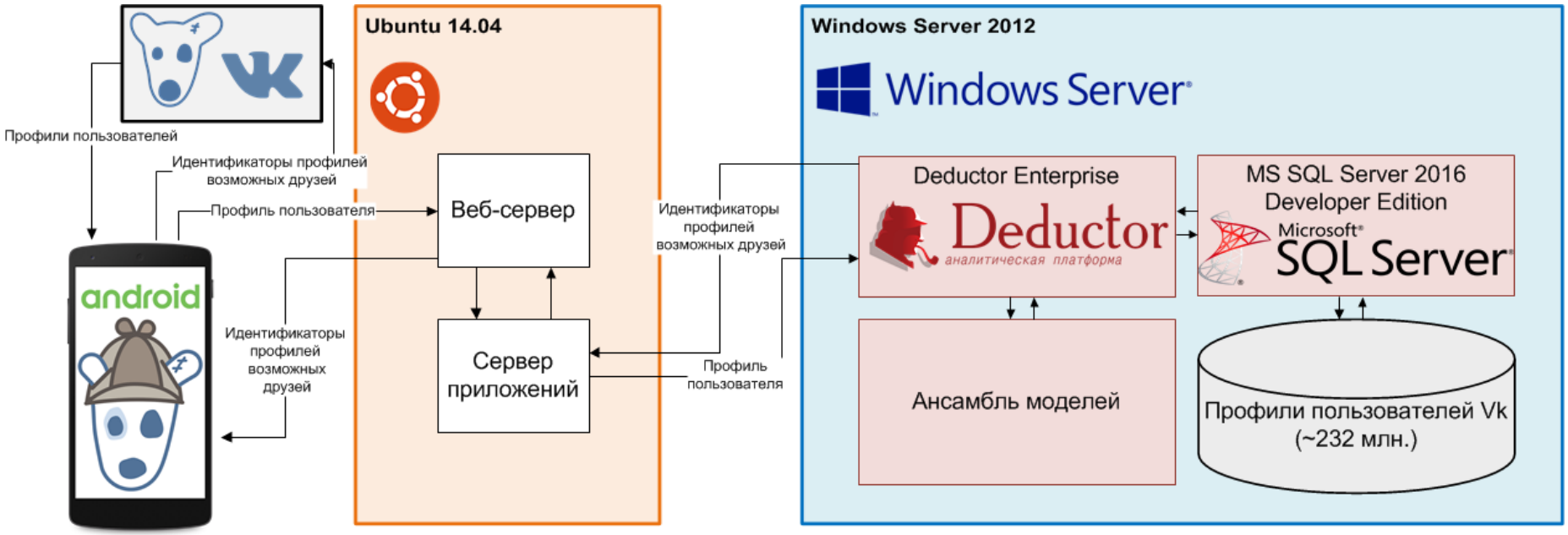


Архитектура программного комплекса

Содержание

- Введение
- **Проекты, выполненные с использованием серверных компонентов платформы Deductor Enterprise**
  - Выявление групп риска в рамках мониторинга наркоситуации в Брянской области
  - Программная поддержка полного цикла социологического исследования
  - **Поиск единомышленников в социальной сети VK**
- Итоги работы. Дальнейшие планы и пожелания

Архитектура программного комплекса

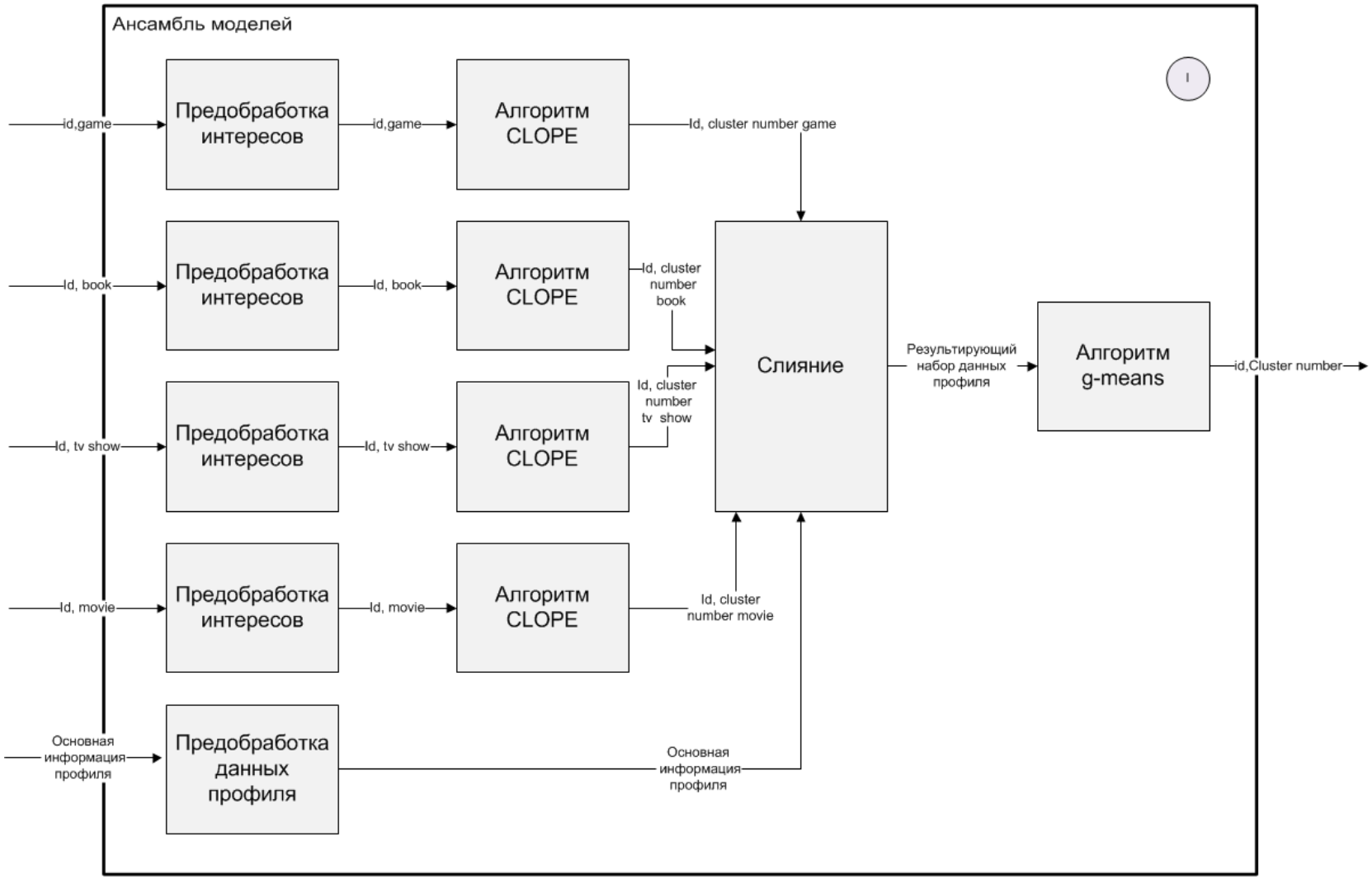


* *Выражаем благодарность компании **«АйТи Про»** (www.itprocomp.ru) и лично **Бондареву Борису Игоревичу** (главному архитектору) за помощь в получении данных из социальной сети VK*

Конфигурация сервера

- Процессор: Intel Xeon E5-2620 v2 2,1 ГГц
- ОЗУ: 16 Гб
- Объем жесткого диска: 1 Тб
- ОС: MS Windows Server 2016 R2
- СУБД: MS SQL Server 2016 Developer Edition
- Аналитическая платформа: Deductor Enterprise

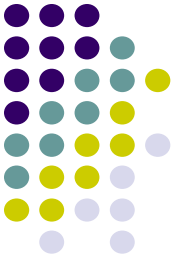
Ансамбль моделей



Предобработка интересов

- Все операции выполнялись средствами Deductor, кроме преобразования строк в формате JSON (в VK так хранятся данные о интересах пользователя) в табличный вид, пригодный для импорта в Deductor
- Операция осложнялась вольным написанием названий интересов: (WoW, World of Warcraft, worldofwarcraft, World of Warcraft – BurningCrusade, Wor croft, Wor craft, Wor of Kraft, wor workrawt, worald of warkraft, …)
- Очистка и предобработка данных:
  - Строки приводятся к нижнему регистру.
  - Замена незаполненных интересов словом «нет»
  - Удаляются все символы кроме букв, так же удаляются буквы «i» и «v», так как пользователи часто используют их в качестве цифр
  - Удаляются пустые строки
  - Замена по словарю наиболее частых ошибок. Например Wor craft → warcraft
  - Отбрасываются редкие варианты (менее 1000 вхождений)

Примеры результатов работы алгоритма CLOPE



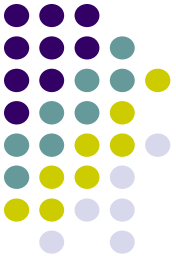
- Кластер 13:
  - Футбол
  - Твистер
  - Теннис
  - Бильярд
  - Волейбол
  - Шахматы
- Кластер 148:
  - Ролевые
  - Взрослые
  - Сексуальные
- Кластер 280:
  - Зомбиферма
  - Need for Speed Carbon
  - Паук
  - Warcraft
- Кластер 309:
  - Баскетбол
  - Гарри Поттер
  - RPG
  - Теннис
  - Приключения
  - Компьютерные игры
  - Метро
  - Футбол
  - Секс
  - Игры
  - Бильярд
  - Рыбалка
  - Sims
  - Спорт
  - Жизнь
  - Маджонг

Предобработка для G-Means

- Данная таблица состоит из следующих столбцов:
  - Идентификатор пользователя
  - Дата рождения
  - Пол
  - Город проживания
  - Страна проживания

- Алгоритм предобработки:
  - На основе даты рождения определяются:
    - возраст
    - знак зодиака
    - знак животного по восточному календарю
  - Выполняется слияние с таблицами, полученными по результатам работы алгоритмов CLOPE

Создание веб-сервиса для обработки набора записей



D:\!Analys\VK\scheme.xsd - Notepad++

File Edit Search View Encoding Language Settings Macro Run Plugins Window ?

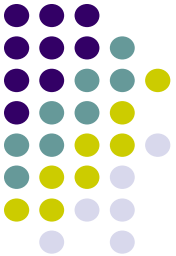
scheme.xsd

```
11              <xs:element name="sex" type="xs:int" />
12          </xs:sequence>
13      </xs:complexType>
14      <xs:complexType name="Output">
15          <xs:sequence>
16              <xs:element name="id" type="xs:int"/>
17              <xs:element name="id_animal" type="xs:int"/>
18              <xs:element name="id_horoscope" type="xs:int"/>
19              <xs:element name="cluster_number_game" type="xs:int"/>
20              <xs:element name="cluster_number_book" type="xs:int"/>
21              <xs:element name="cluster_number_movie" type="xs:int"/>
22              <xs:element name="cluster_number_tv" type="xs:int"/>
23              <xs:element name="cluster_number" type="xs:int"/>
24              <xs:element name="distance" type="xs:float"/>
25          </xs:sequence>
26      </xs:complexType>
27      <xs:element name="input" type="tns:Input"/>
28      <xs:element name="output" type="tns:Output"/>
29
30      <xs:complexType name="InputRowsType">
31          <xs:sequence>
32              <xs:element name="RowItem" type="tns:Input" maxOccurs="unbounded"/>
33          </xs:sequence>
34      </xs:complexType>
35      <xs:element name="InputRows" type="tns:InputRowsType"/>
36      <xs:complexType name="OutputRowsType">
37          <xs:sequence>
38              <xs:element name="RowItem" type="tns:Output" maxOccurs="unbounded"/>
39          </xs:sequence>
40      </xs:complexType>
41      <xs:element name="OutputRows" type="tns:OutputRowsType"/>
42
43  </xs:schema>
```

eXtensible Markup Language file | length : 1 755 lines : 43 | Ln : 42 Col : 1 Sel : 0 | 0 | Windows (CR LF) | UTF-8 | INS

ENG 20:30 24.06.2017

Описание WSDL-сервиса



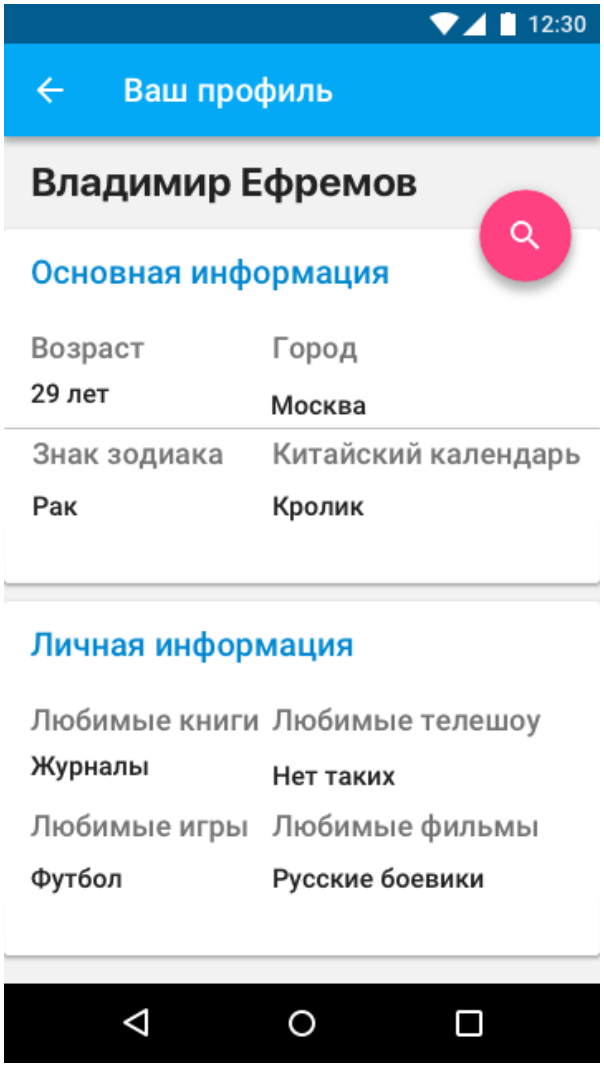
deductor.iipo.tu-bryansk ×

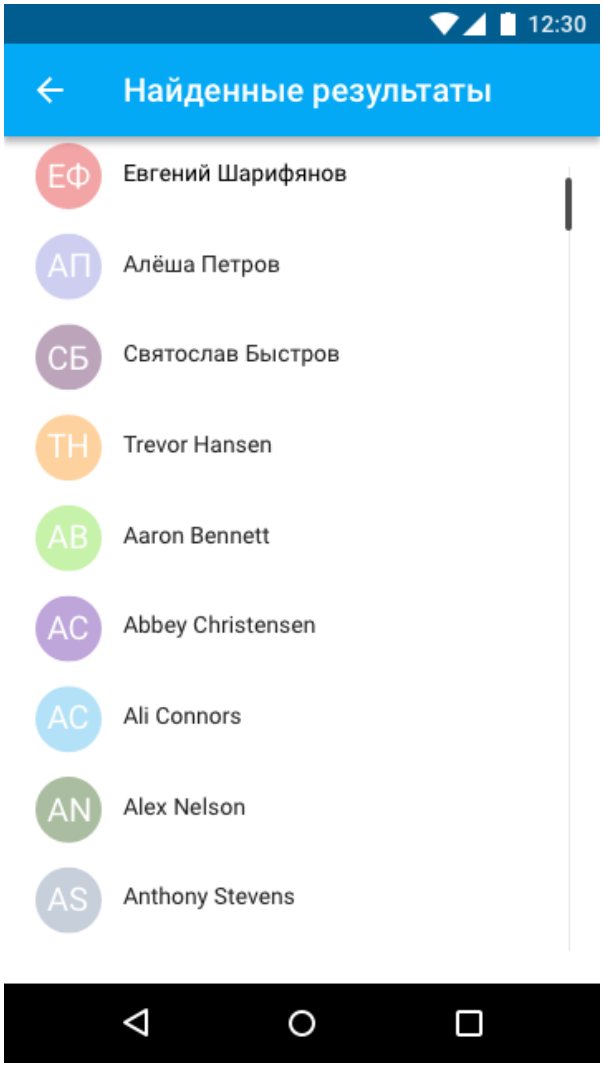
deductor.iipo.tu-bryansk.ru/vk/service.svc?wsdl

This XML file does not appear to have any style information associated with it. The document tree is shown below.

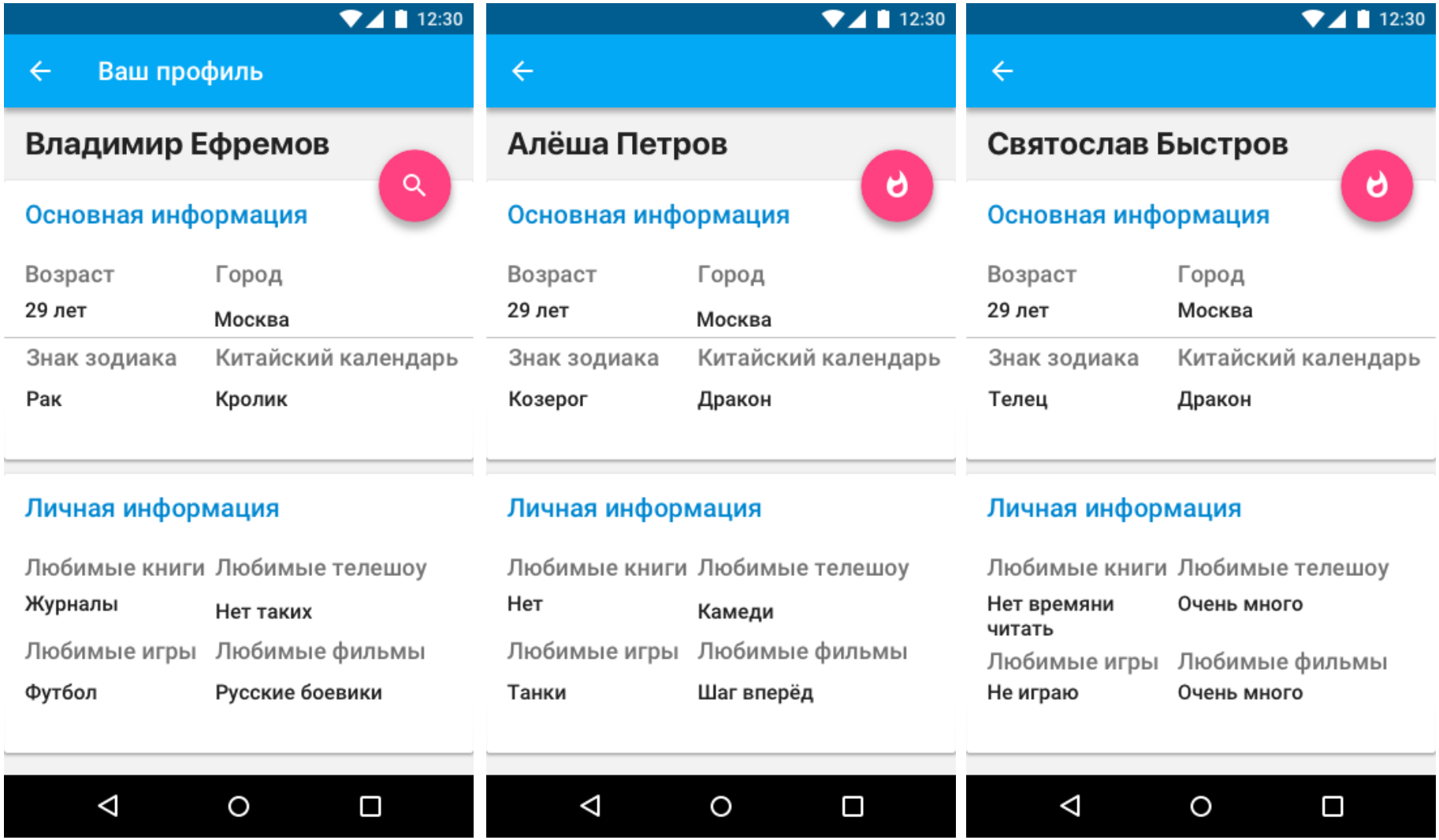

```
<wsdl:definitions xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/" xmlns:wsx="http://schemas.xmlsoap.org/ws/2004/09/mex" xmlns:wsu="http://docs.oasis-open.org/wss/2004/01/oasis-200401-wss-wssecurity-utility-1.0.xsd" xmlns:wsa10="http://www.w3.org/2005/08/addressing"
 xmlns:wsp="http://schemas.xmlsoap.org/ws/2004/09/policy" xmlns:wsap="http://schemas.xmlsoap.org/ws/2004/08/addressing/policy" xmlns:msc="http://schemas.microsoft.com/ws/2005/12/wsdl/contract" xmlns:soap12="http://schemas.xmlsoap.org/wsdl/soap12/"
 xmlns:wsa="http://schemas.xmlsoap.org/ws/2004/08/addressing" xmlns:wsam="http://www.w3.org/2007/05/addressing/metadata" xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:tns="http://www.basegroup.ru/DeductorIntegrationServer"
 xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/" xmlns:wsaw="http://www.w3.org/2006/05/addressing/wsdl" xmlns:soapenc="http://schemas.xmlsoap.org/soap/encoding/" name="ServiceContract" targetNamespace="http://www.basegroup.ru/DeductorIntegrationServer">
  <wsdl:types>
    <xsd:schema targetNamespace="http://www.basegroup.ru/DeductorIntegrationServer/Imports">
      <xsd:import schemaLocation="http://deductor.iipo.tu-bryansk.ru/vk/Service.svc?xsd=xsd0" namespace="DIS_vk"/>
      <xsd:import schemaLocation="http://deductor.iipo.tu-bryansk.ru/vk/Service.svc?xsd=xsd1" namespace="http://www.basegroup.ru/warehouses/vk"/>
      <xsd:import schemaLocation="http://deductor.iipo.tu-bryansk.ru/vk/Service.svc?xsd=xsd2" namespace="http://www.basegroup.ru/DeductorIntegrationServer"/>
    </xsd:schema>
  </wsdl:types>
  <wsdl:message name="DIS_vkMessage">
    <wsdl:part name="parameters" element="tns:DIS_vk"/>
  </wsdl:message>
  <wsdl:message name="DIS_vkResponseMessage">
    <wsdl:part name="parameters" element="tns:DIS_vkResponse"/>
  </wsdl:message>
  <wsdl:message name="ServiceClass_DIS_vk_SOAPExceptionFault_FaultMessage">
    <wsdl:part name="detail" element="tns:SOAPException"/>
  </wsdl:message>
  <wsdl:portType name="ServiceClass">
    <wsdl:operation name="DIS_vk">
      <wsdl:input wsaw:Action="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/DIS_vk" name="DIS_vkMessage" message="tns:DIS_vkMessage"/>
      <wsdl:output wsaw:Action="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/DIS_vkResponse" name="DIS_vkResponseMessage" message="tns:DIS_vkResponseMessage"/>
      <wsdl:fault wsaw:Action="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/DIS_vkSOAPExceptionFault" name="SOAPExceptionFault" message="tns:ServiceClass_DIS_vk_SOAPExceptionFault_FaultMessage"/>
    </wsdl:operation>
  </wsdl:portType>
  <wsdl:binding name="BasicHttpBinding_ServiceClass" type="tns:ServiceClass">
    <soap:binding transport="http://schemas.xmlsoap.org/soap/http"/>
    <wsdl:operation name="DIS_vk">
      <soap:operation soapAction="http://www.basegroup.ru/DeductorIntegrationServer/ServiceClass/DIS_vk" style="document"/>
      <wsdl:input name="DIS_vkMessage">
        <soap:body use="literal"/>
      </wsdl:input>
      <wsdl:output name="DIS_vkResponseMessage">
        <soap:body use="literal"/>
      </wsdl:output>
      <wsdl:fault name="SOAPExceptionFault">
        <soap:fault name="SOAPExceptionFault" use="literal"/>
      </wsdl:fault>
    </wsdl:operation>
  </wsdl:binding>
  <wsdl:service name="ServiceContract">
    <wsdl:port name="BasicHttpBinding_ServiceClass" binding="tns:BasicHttpBinding_ServiceClass">
      <soap:address location="http://deductor.iipo.tu-bryansk.ru/vk/Service.svc/ServiceClass"/>
    </wsdl:port>
  </wsdl:service>
</wsdl:definitions>
```

Пример результатов работы в мобильном приложении (1)





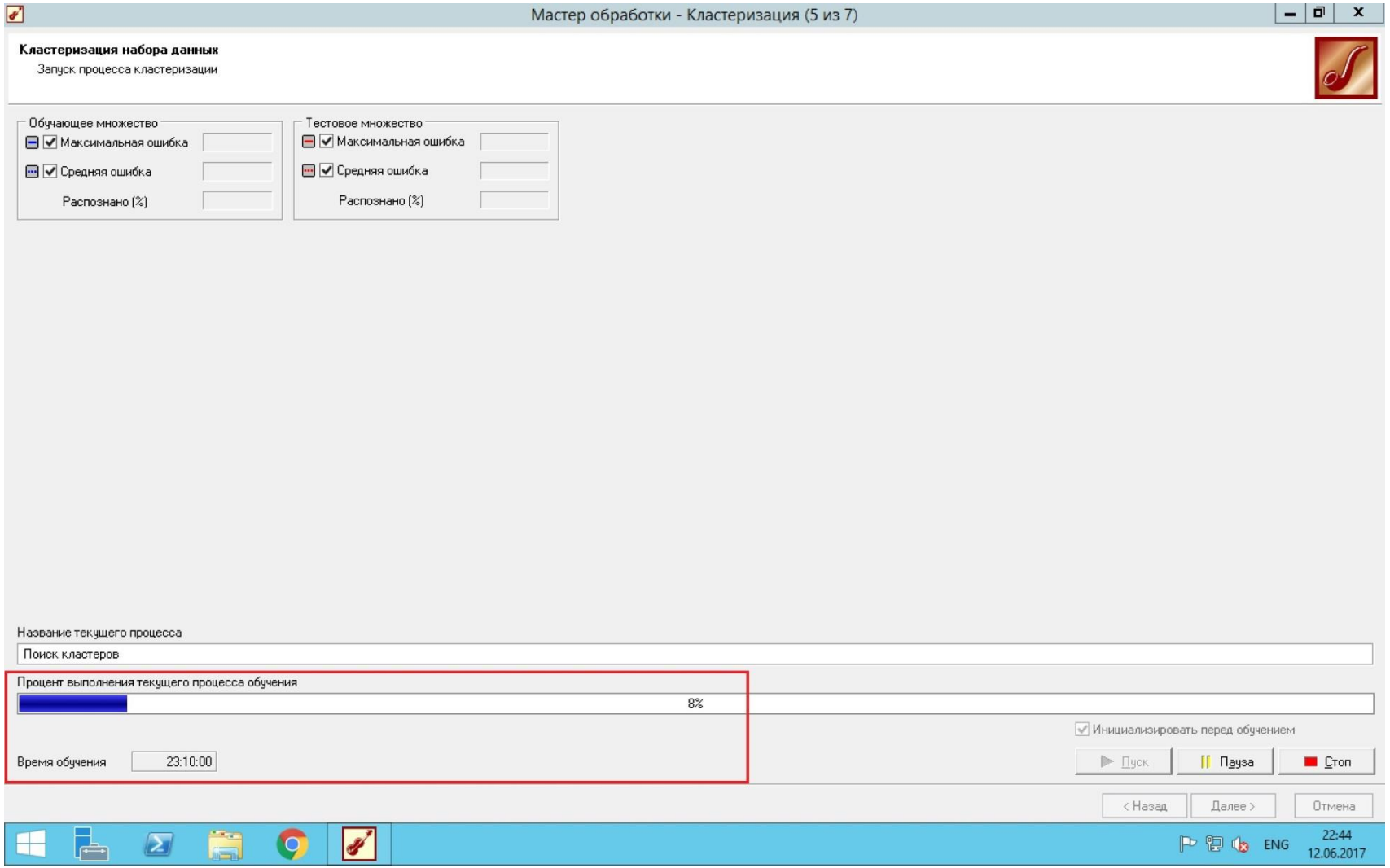
Пример результатов работы в мобильном приложении (2)



Перспективы развития проекта

- Поддержание в актуальном состоянии базы данных профилей пользователей социальной сети ВКонтакте
- Улучшение качества предобработки
  - Применение более совершенных словарей
  - Применение функции вычисления значения Дамерау-Левенштейна для определения похожести строк с целью идентификации опечаток
- Поиск оптимального алгоритма кластеризации профилей
  - Опыты с другими алгоритмами (EM, карты Кохонена, …)
  - Экспериментальная проверка с целью выявления наилучшего
- Ранжирование выходного набора данных для повышения релевантности результатов подбора людей со схожими интересами
  - По расстоянию между местами проживания в километрах
  - По длине цепочки друзей (количество промежуточных «звеньев»)

8% за 23 часа…

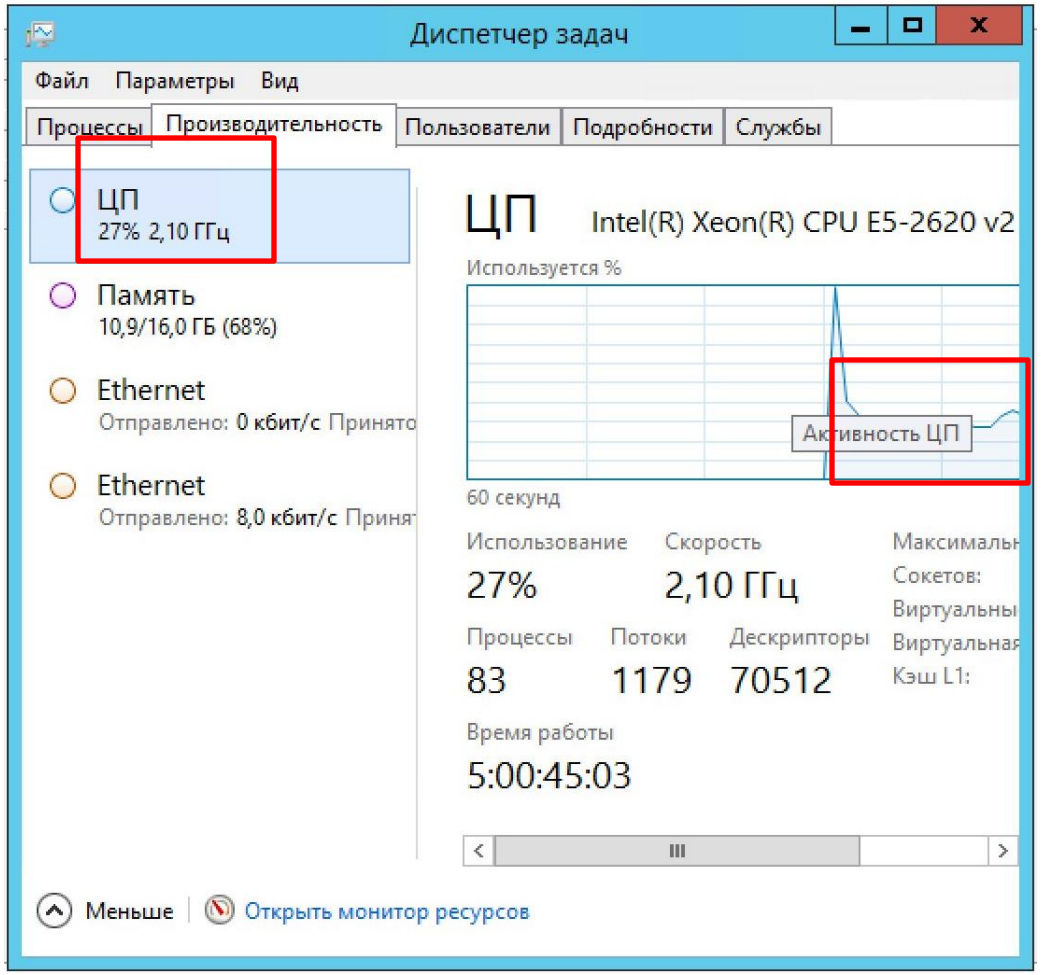

Мастер обработки - Кластеризация (5 из 7)
Кластеризация набора данных
Запуск процесса кластеризации
Обучающее множество
Максимальная ошибка
Средняя ошибка
Распознано (%)
Тестовое множество
Максимальная ошибка
Средняя ошибка
Распознано (%)
Название текущего процесса
Поиск кластеров
Процент выполнения текущего процесса обучения
8%
Время обучения
23:10:00
Инициализировать перед обучением
Пуск
Пауза
Стоп
< Назад
Далее >
Отмена
ENG
22:44
12.06.2017

Затраты времени

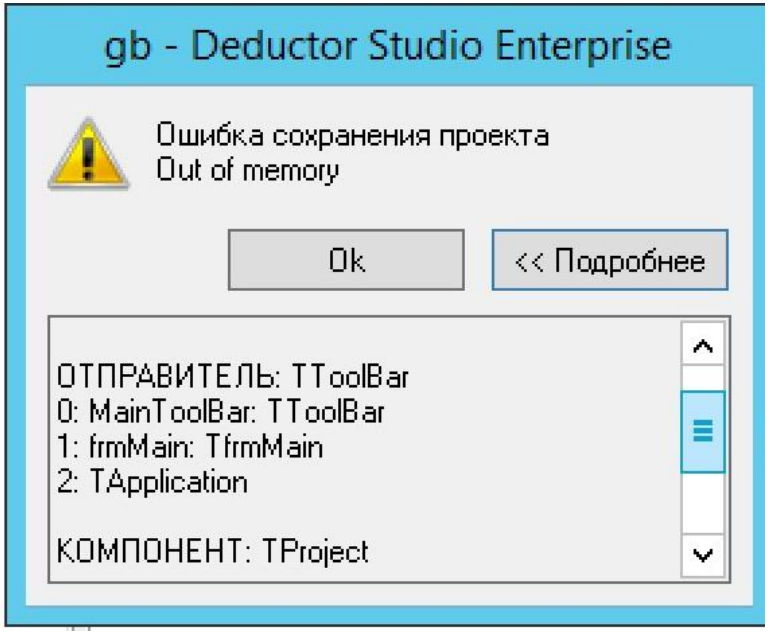
- Кластеризация транзакций методом CLOPE занимала **от 24 до 72 часов**
- Кластеризация g-means: 11 столбцов данных, настройки не менялись:
  - 5 000 000 строк, ограничение 1000 кластеров, длительность ~ **54 часа**.
  - 300 000 строк, итоговое количество кластеров 2213, длительность ~ **25 часов**.
- После выбора алгоритма имела место долгая предобработка
  - Занимала несколько часов
  - На это время Deductor «зависал» и не реагировал на действия пользователя
  - При этом не отражался никакой визуальный индикатор

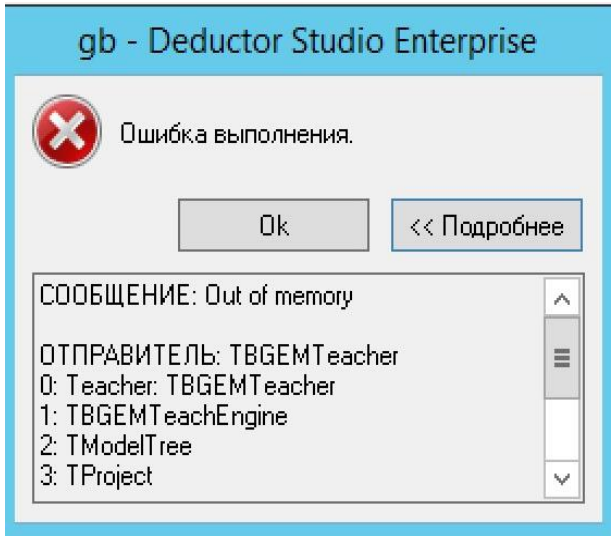
Выполнение кластеризации

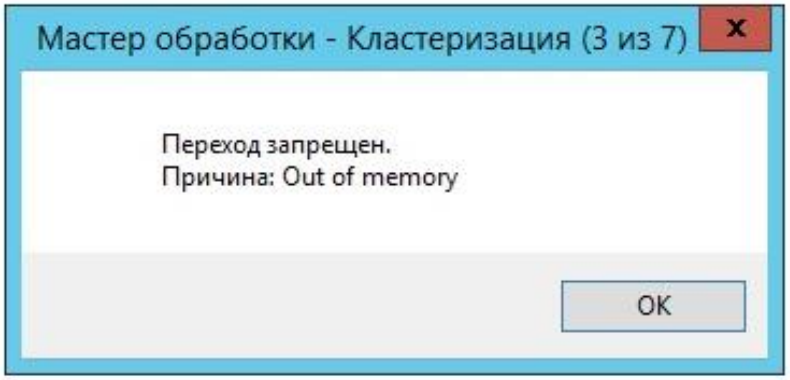
Deductor использует только **одно** ядро, остальные бездействуют

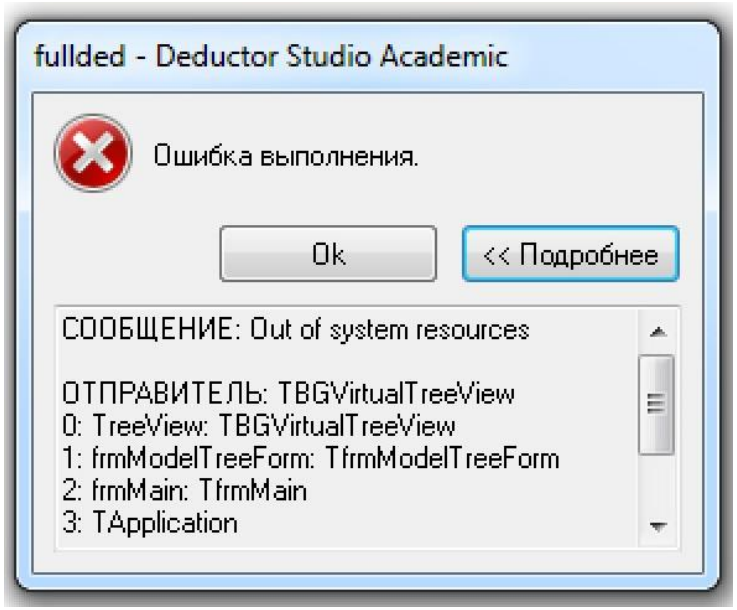


Out of memory…

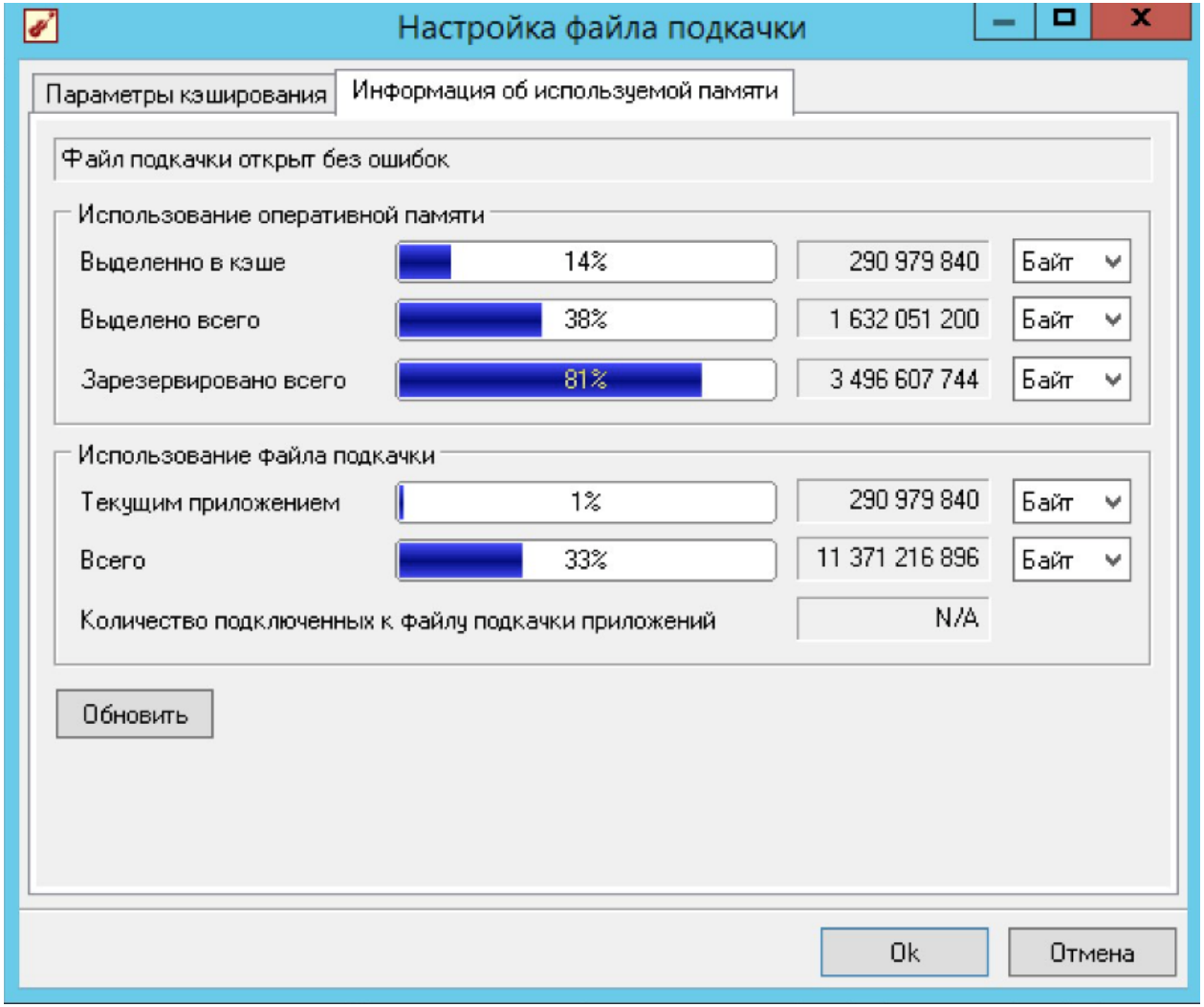




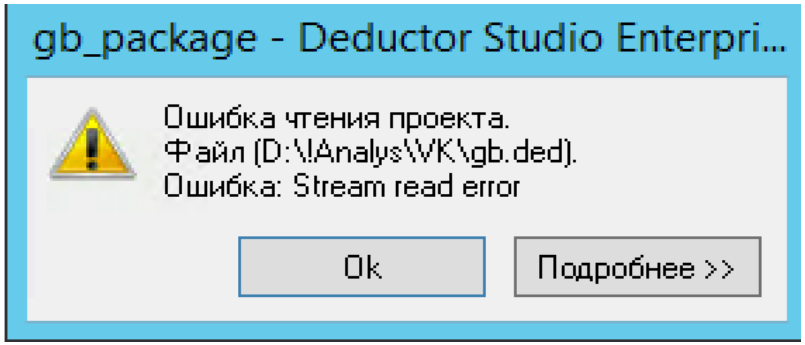


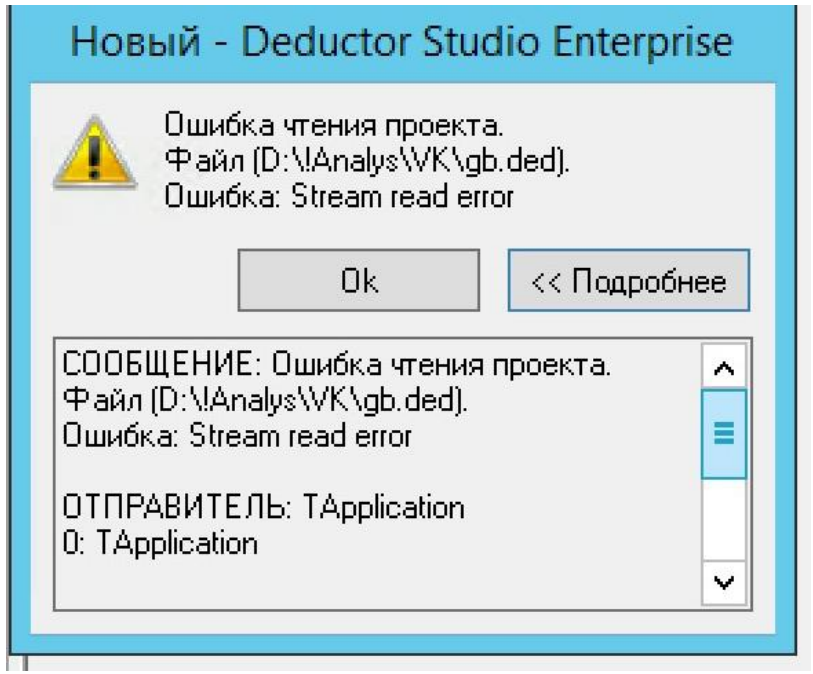


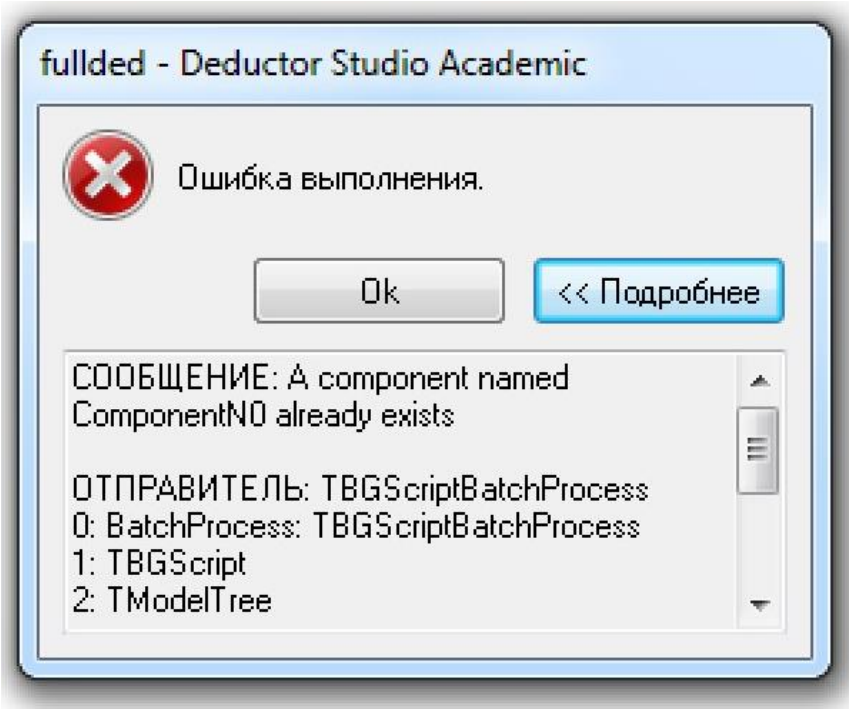
Настройки файла подкачки…



Прочие проблемы…







Проблемы DIS при работе совместно с кластеризацией

DIS_log_2017-06-21_14-20-43 {a9de46aa-877a-4fa4-9487-20e45583b40a} — Блокнот

Файл Правка Формат Вид Справка

```
Текст исключения: Не удалось установить подключение к Deductor Server
Класс исключения: System.Exception
Трасса
   в DeductorIntegrationServer.Core.ServiceClass.ConnectDeductorServer(String address, Int32 port, LogWriter
logWriter, Int32& activeApplicationCount)
   в DeductorIntegrationServer.Core.ServiceClass.ExecuteProject(XDocument inputData, LogWriter logWriter,
XDocument& outputData)
   в DeductorIntegrationServer.Core.ServiceClass.InternalExecute(XDocument inputData, LogWriter logWriter,
XDocument& outputData, Exception& exception)
   в DeductorIntegrationServer.Core.ServiceClass.Execute(InputData inputData)
---------------------------------------------
Текст исключения: Подключение не установлено, т.к. конечный компьютер отверг запрос на подключение
82.179.88.29:4386
Класс исключения: System.Net.Sockets.SocketException
Трасса
   в System.Net.Sockets.Socket.Connect(IPAddress[] addresses, Int32 port)
   в System.Net.Sockets.Socket.Connect(String host, Int32 port)
   в DeductorClient.DeductorConnector.Connect()
   в DeductorIntegrationServer.Core.ServiceClass.ConnectDeductorServer(String address, Int32 port, LogWriter
logWriter, Int32& activeApplicationCount)
```

Содержание

- Введение
- Проекты, выполненные с использованием серверных компонентов платформы Deductor Enterprise
  - Выявление групп риска в рамках мониторинга наркоситуации в Брянской области
  - Программная поддержка полного цикла социологического исследования
  - Поиск единомышленников в социальной сети VK
- **Итоги работы. Дальнейшие планы и пожелания**

Планы и пожелания прошлого года
(были представлены на конференции в июне 2016)

- **Обучение и сертификация**
  - Пройти обучение двум преподавателям дисциплины «Интеллектуальный анализ данных»
  - Пройти сертификацию этим преподавателям
  - Рассмотреть возможность сертификации студентов в дистанционном режиме
- **Расширение количества лицензий Deductor Studio Professional**
  - В настоящее время имеется одна лицензия, и для комфортного выполнения курсовых проектов этого недостаточно
- **Учет упоминавшихся ранее особенностей обучения магистрантов *со специализацией в области программирования***
  - Курсовые проекты и магистерские диссертации должны быть направлены не просто на анализ данных, а на *разработку приложений анализа данных*, что возможно только при наличии серверных компонентов Deductor
  - При получении такого опыта обязуемся рассказать об этом на следующей конференции

Итоги работы за прошедший год

- **Обучение и сертификация**
  - доцент Лагерев Д.Г. – 3 сертификата BaseGroup
  - ассистент Бабурин А.Н. – 1 сертификат BaseGroup
  - сертификация 5-и магистрантов
- **Расширение количества лицензий Deductor Studio Professional**
  - Получена лицензия на Deductor Studio Enterprise, Deductor Integration Server, Deductor Analytic Server
- **Учет упоминавшихся ранее особенностей обучения магистрантов *со специализацией в области программирования***
  - Представлены 3 проекта, ориентированные на *разработку приложений анализа данных*, с использованием серверных компонентов Deductor
  - Один из них можно отнести к категории Big Data

Дальнейшие планы и пожелания

- **Продолжение обучения и сертификации преподавателей**
- **Использование Deductor Enterprise и/или Loginom студентами и аспирантами**
  - для выполнения магистерских и кандидатских диссертаций
  - для выполнения курсовых проектов (продолжить практику использования)
- **Получение дополнительных лицензий для Deductor Enterprise**
  - В настоящее время имеется одна лицензия, и для эффективной работы с большими данными этого недостаточно
  - При наличии только одной лицензии невозможно выполнять лабораторные работы по разработке сервисов DIS
- **Переход на платформу Loginom**
  - Обучение работе с платформой (в рамках мастер-класса 28 июня)
  - Получение доступа к платформе
  - Развертывание платформы Loginom на кафедральном сервере (для работы с большими данными)

IV межвузовская конференция

Бизнес-аналитика. Использование аналитической платформы Loginom (Deductor) в учебном процессе вуза

СПАСИБО ЗА ВНИМАНИЕ !

Подвесовский Александр Георгиевич
заведующий кафедрой, к.т.н., доцент
apodv@tu-bryansk.ru

Лагерев Дмитрий Григорьевич
к.т.н., доцент
lagerevdg@mail.ru

Бабурин Артем Николаевич
ассистент
ababurin@bk.ru

г. Москва, 27 июня 2017 г.