

Прогнозирование целевых действий на сайте

(эксперименты в обучении аналитиков)

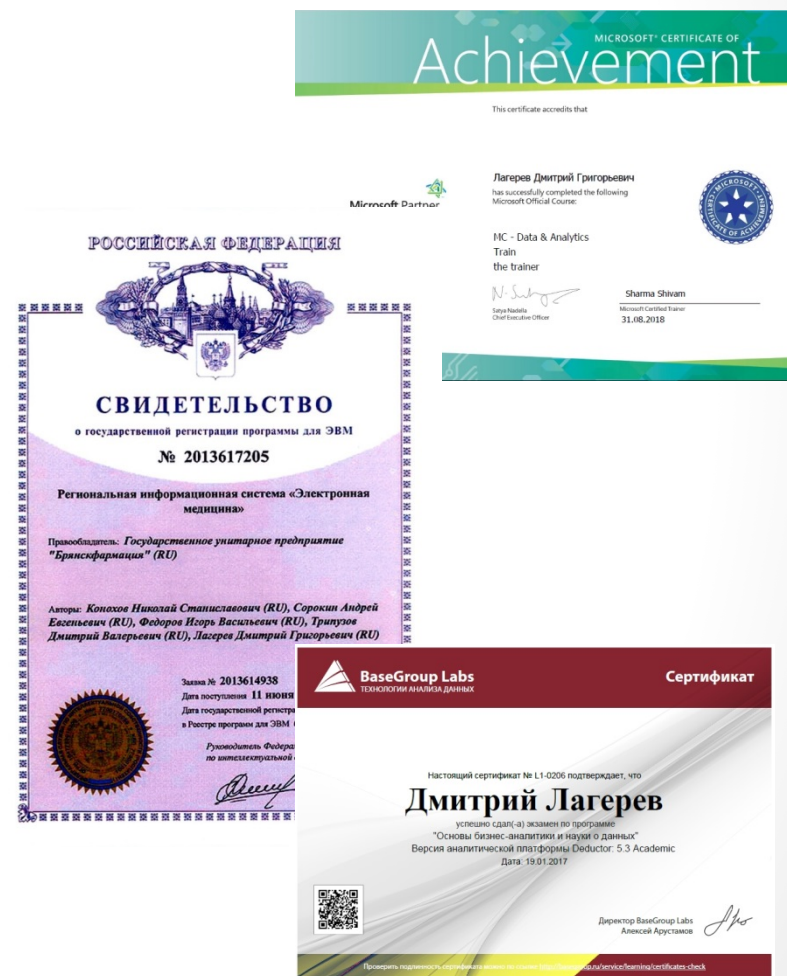
кандидат технических наук, доцент
Лагерев Дмитрий Григорьевич



О себе



- Кандидат технических наук, доцент
- Доцент кафедры «Информатика и программное обеспечение» ФГОУ ВО «Брянский государственный технический университет»
- Руководитель проектов в сфере управленческой аналитики ООО «АйТи Про»
- Сотрудничаю с компанией Loginot более **10 лет**



Что сдерживает внедрение продвинутой бизнес-аналитики?



Дефицит специалистов!

Последствия

Бизнес-аналитика доступна лишь крупным компаниям и требует больших бюджетов:



1. Отсутствует собственная экспертиза в сфере бизнес-аналитики
2. Бизнес-аналитика сводится к консолидации данных и дашбордам
3. Аналитики неадекватно оценивают возможности современных инструментов и с опаской относятся к реализации сложных проектов

Возможные решения

- Отказ от бизнес-аналитики
- Упрощение проектов, использование функционала MS Excel или средств визуализации Tableau, Power BI...
- Использование аналитических платформ
- Обучение и развитие собственных сотрудников
- **Подготовка квалифицированных специалистов**



Кого мы готовим?



Кафедра готовит профессиональных разработчиков ПО и системных аналитиков со специализацией в следующих областях:

- обработка и анализ больших объемов данных
- модели и методы поддержки принятия решений
- обработка и анализ изображений, машинное зрение
- ...

Полузнание хуже незнания

Алгоритмы, языки, технологии и инструменты важны

НО, если при решении практических задач ограничиться анализом **СИНТЕТИЧЕСКИХ** данных, вместо специалиста получается **ПОЛУЗНАЙКА**, который считает себя специалистом и хочет получать высокую зарплату.



Как мы готовим специалистов?

До курса «Интеллектуальный анализ данных»:

- Математические курсы
- Языки программирования
- Алгоритмы и структуры данных
- Хранилища данных

В курсе «Интеллектуальный анализ данных»:

- Алгоритмы анализа данных
- Аналитические пакеты и библиотеки
- Синтетические данные для лабораторных и практических работ
- **Только реальные данные для курсового и дипломного проекта**

Loginom Хакатон 2019

Повышение точности
расчета риска просрочки
(невозврата) займа на
основе обогащения поля
«Место работы» из анкет
клиентов МФО



LOGINOM
ХАКАТОН 2020



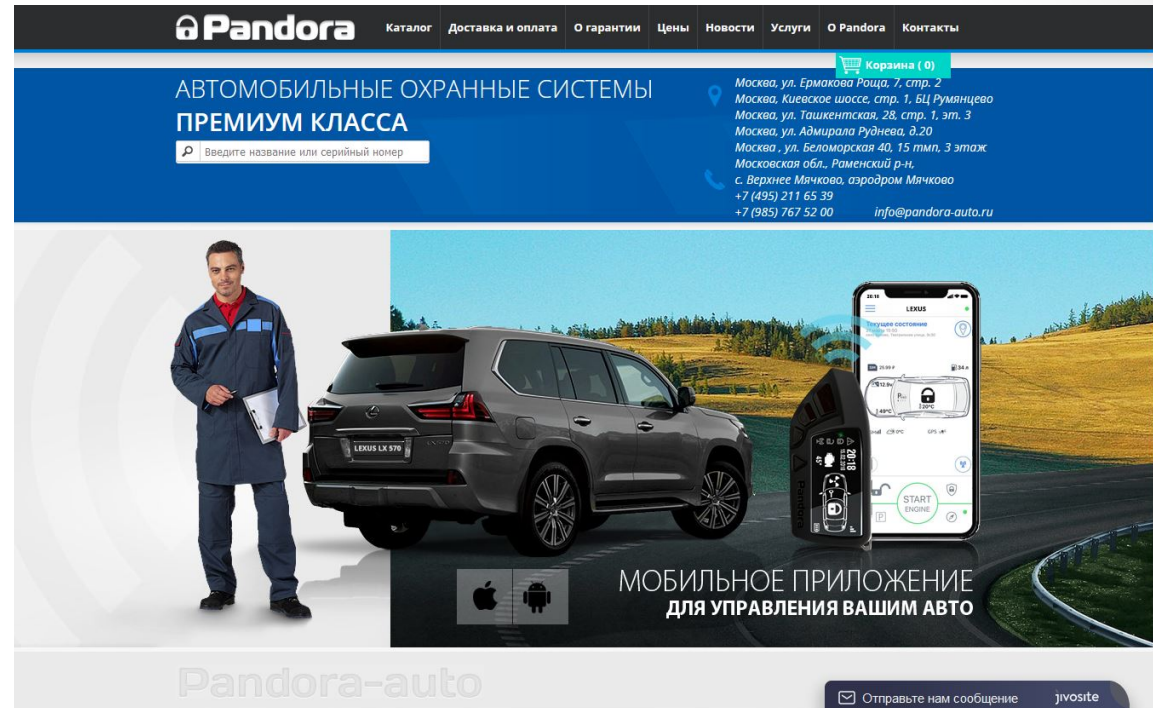
Описание проекта

Цель проекта

Прогнозирование вероятности достижения посетителем интернет-магазина «Сигнализации Pandora» маркетинговых целей в момент нахождения на сайте

Перспективы практического применения

Стимулирование перспективных клиентов к достижению маркетинговых целей



Наша команда



- Савостин Игорь
- Полякова Марина
- Кузьмин Сергей
- Курилов Аркадий



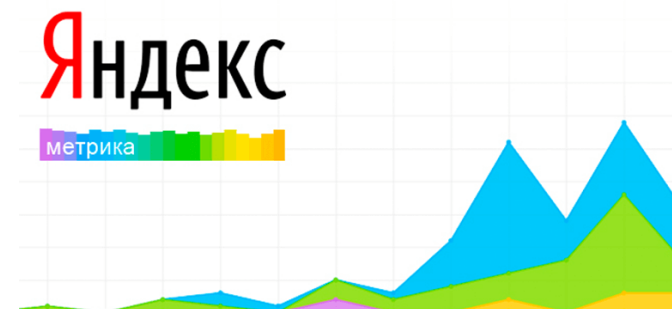
Магистранты направления
«Информатика вычислительная
техника», профиль
«Компьютерный анализ и
интерпретация данных»



Исходные данные

- Партнер по проекту компания Idea-Promotion Агентство комплексного интернет-маркетинга
- Сбор данных для анализа осуществлялся сервисом Яндекс.Директ
- Данные получены специалистами компании Idea-Promotion из сервиса Яндекс.Метрика

По умолчанию Яндекс.Метрика выдает предобработанные и агрегированные данные, дальнейший анализ которых малоперспективен



Исходные данные (фрагмент)

ym:s:visitID;ym:s:dateTime;ym:s:isNewUser;ym:s:startURL;ym:s:endURL;ym:s:pageViews;ym:s:visitDuration;ym:s:bounce;ym:s:regionCity;ym:s:goalsID;ym:s:goalsSerialNumber;ym:s:lastTrafficSource;ym:s:lastReferralSource;ym:s:lastSearchEngineRoot;ym:s:lastSocialNetwork;ym:s:deviceCategory;ym:s:browser1307260240131786591;2018-03-18 20:13:22;1;https://signalki.ru/catalog/ksenon-parktronik/182-parkmaster-8-dj-3233/;https://signalki.ru/catalog/ksenon-parktronik/182-parkmaster-8-dj-3233/;1;0;1;Lipetsk;[];[];organic;yandex.ru;yandex;;2;yandex_browser173338331697914587;2018-01-27 18:40:33;1;https://signalki.ru/catalog/predpuskovye-otopiteli/175-tajmer-easystart-t/;https://signalki.ru/catalog/predpuskovye-otopiteli/175-tajmer-easystart-t/;1;0;1;;[];[];organic;google.ru;google;;1;chrome18299069828925766411;2018-01-13 14:31:07;1;https://signalki.ru/https://signalki.ru/catalog/arkan/arkan-control/;8;318;0;Saint Petersburg;[16225015];[1];organic;yandex.ru;yandex;;1;chrome2672374073567228104;2018-05-18 02:44:58;1;https://signalki.ru/catalog/sputnikovye-sistemy/arkan/https://signalki.ru/catalog/sputnikovye-sistemy/arkan/80-arkan-control-/;8;506;0;Kemerovo;[16225015];[1];organic;yandex.ru;yandex;;1;chrome

Задачи команды

Понять потребности бизнеса

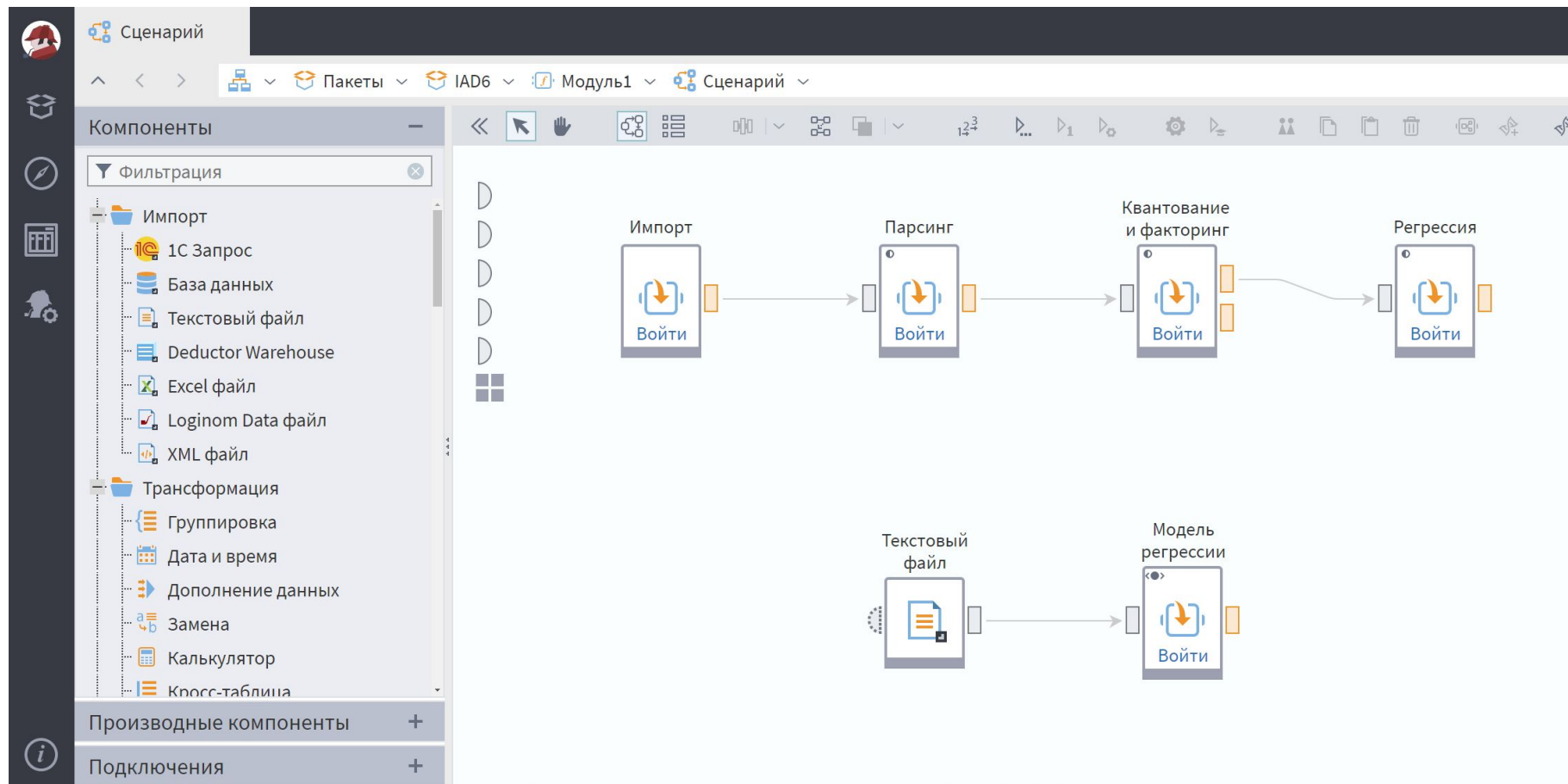
Получить данные

Выполнить предобработку данных

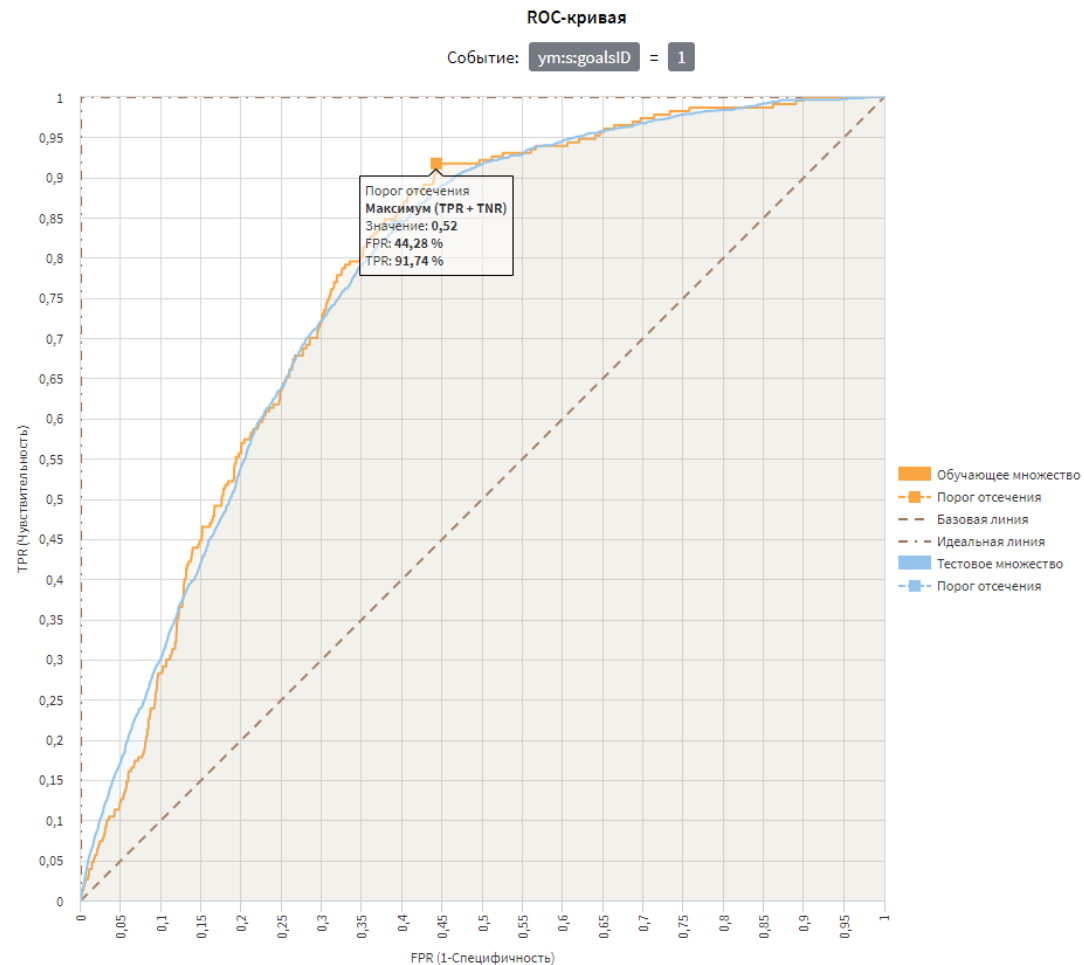
Построить ансамбль моделей

Оценить качество

Сделать выводы понятные для бизнеса



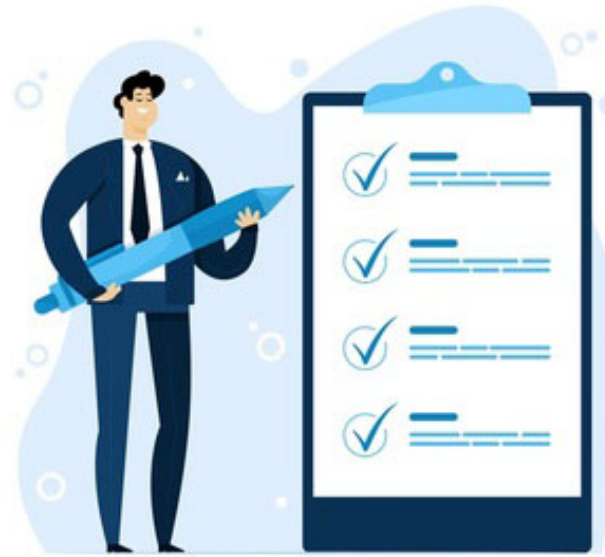
Используемые в проекте методы



- Анализа временных рядов
- Поиска ассоциативных правил
- Кластеризации
- Классификации
- Построения скоринговых моделей

Итоговая точность модели на тестовом множестве: **77%**

Выводы



- Выявлена зависимость количества посещений от **времени суток**
- Вероятность попадания в воронку продаж **выше**, если пользователь пришел из социальных сетей или тематического сообщества
- Вероятность попадания в воронку продаж больше, если пользователь зашел на сайт с **ПК**, чем с мобильного или другого устройства

Результаты проекта

Разработанный сервис размещен в сети Интернет:

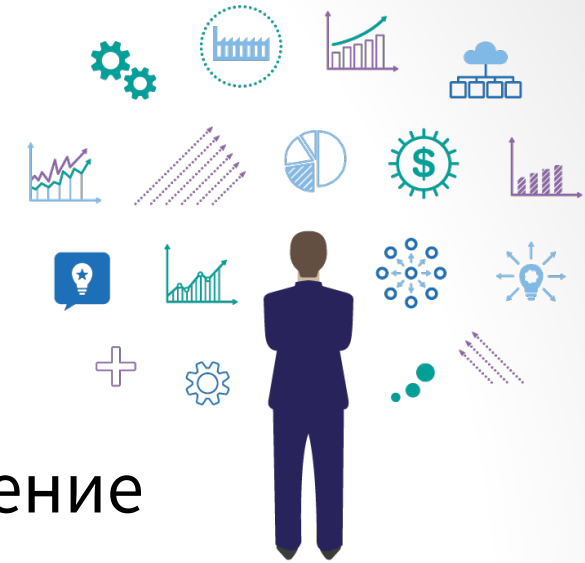
http://deductor.iipo.tu-bryansk.ru/lgi/Service.svc/Rest/IAD6/scoring_model

```
id: "2562c087-4cb6-48b7-884f-29df105f331b"
name: "deductor"
description: null
auth: null
events: null
variables: []
▶ order: [...]
  folders_order: []
  folders: []
▼ requests:
  ▼ 0:
    id: "c799ff05-2a13-457b-a434-e8f288424375"
    ▼ name: "http://deductor.iipo.tu-bryansk.ru/lgi/Service.svc/Rest/IAD6/scoring_model"
      ▼ url: "http://deductor.iipo.tu-bryansk.ru/lgi/Service.svc/Rest/IAD6/scoring_model"
        description: ""
        data: []
        dataMode: "raw"
        ▶ headerData: [...]
        method: "POST"
        pathVariableData: []
        queryParams: []
        ▶ auth: {...}
        events: null
        folder: null
        currentHelper: "basicAuth"
        ▶ helperAttributes: {...}
        collectionId: "2562c087-4cb6-48b7-884f-29df105f331b"
        ▼ rawModeData:
          "{\n\t\"DataSet\": {\n\t\t\"Rows\": [\n\t\t\t{\n\t\t\t\t\t\"source\":
          \n\t\t\t\t\t\"ym:s:dateTime;ym:s:isNewUser;ym:s:startURL;ym:s:endURL;ym:s:pageViews;ym:s:visitDuration;ym:s:bounce;ym:s:regionCity;ym:s:visitID;ym:s:goalsID;ym:s:goalsDateTime;ym:s:lastSearchEngineRoot;ym:s:browser;ym:s:de
          \\\r\n2001-07-13 15:01:13;0;http://\\pandora-auto.ru\\catalog\\avtosignal\\dx1-5000-new;http://\\pandora-auto.ru\\;50;13000;0;Fignya;1456412556851858379;[16222585];[\\\"2001-07-13 15:51:13\\\"]};google

        headers: "Content-Type: application/json\n"
        pathVariables: []
```



Результаты проекта



Для бизнеса:

- Более точная настройка рекламы, повышение эффективности продвижения сайта
- Реализация «пилотного» проекта бизнес-аналитики за минимальный бюджет
- Повышение интереса к возможностям и инструментам бизнес-аналитике
- Понимание возможностей и ограничений методов и инструментов

Результаты проекта



Для участников:

- Полученный опыт: более адекватная оценка своих возможностей, готовность к получению **отрицательного** результата
- Понимание сложности различных этапов проекта и реальных затрат на предобработку данных
- Интересный проект в портфолио

Прогнозирование целевых действий на сайте

для вас выступал

Лагерев Дмитрий Григорьевич

LagerevDG@mail.ru

+7 910 337 47 50

