

# **Прогнозирование благонадёжности компании в процессе проверки потенциальных контрагентов**

Одинцов Сергей Сергеевич  
Факультет аналитики Big Data

## Описание проблемы

### Текущая ситуация:

В состав группы компаний входят 5 организаций. В каждой из них есть подразделение, отвечающее за проверку потенциальных контрагентов. Особенности организации процесса:

- Проверки осуществляются вручную
- Методы проверок различаются от компании к компании, несмотря на наличие общих внутренних документов
- Отличается подготовка персонала, осуществляющего указанные проверки в разных компаниях группы.

### Проблема:

Периодически заключаются договоры с компаниями, которые не выполняют взятые на себя обязательства. В ходе последующих расследований обычно выясняется, что данные контрагенты и не собирались выполнять условия договоров. В итоге, компании группы несут финансовые потери.

## Задача

**Минимизировать финансовые потери группы компаний, связанные с заключением договоров с недобросовестными контрагентами**

## Решение

Единый метод проверки потенциальных контрагентов для всех компаний группы на основе модели машинного обучения. Алгоритм позволит выявлять недобросовестных контрагентов на основании стандартного набора данных. Это позволит решить поставленную задачу за счёт:

- Автоматизированная проверка контрагента с заранее заданной точностью
- Минимизация влияния человеческого фактора, т.е. уровня подготовки персонала, осуществляющего проверку, т.к. от линейных работников будет требоваться только внести стандартный набор исходных данных в модель
- Повышение точности выявления неблагонадёжных контрагентов за счёт процесса постоянного дополнительного обучения модели на новых данных

## Как это реализовано



## Стандартный набор данных для проверки

|                     |   |
|---------------------|---|
| <b>Risk</b>         | признак добросовестности/не добросовестности контрагента. Целевая переменная.           |
| <b>Sector_score</b> | исторический уровень финансового риска.   |
| <b>Score_A</b>      | объём собственных средств в тыс. руб.   |
| <b>PARA_A</b>       | несоответствие в расходах, отраженных в пунктах бух. отчётности организации в тыс. руб. |
| <b>Score_B</b>      | объём кредиторской задолженности в тыс. руб.  |
| <b>PARA_B</b>       | несоответствие в затратах, отраженных в пунктах бух. отчётности организации в тыс. руб. |
| <b>Numbers</b>      | оценка нарушений бухгалтерской отчётности за 5 последних лет                            |
| <b>Score_B.1</b>    | объём просроченной кредиторской задолженности в тыс. руб.                               |



|                     |   |
|---------------------|---|
| <b>Risk_C</b>       | объём дебиторской задолженности в тыс. руб.                       |
| <b>Money_Value</b>  | объём нарушений, выявленный в ходе предыдущего аудита в тыс. руб. |
| <b>Score_MV</b>     | объём оборотных активов организации в тыс. руб.                   |
| <b>PROB</b>         | объём проблемных активов на балансе предприятия в тыс. руб.       |
| <b>RiSk_E</b>       | соотношение собственных и заемных средств                         |
| <b>Risk_D</b>       | соотношение собственных и заемных средств                         |
| <b>Prob</b>         | объём проблемной дебиторской задолженности в тыс. руб.            |
| <b>Risk_F</b>       | коэффициент автономии   |
| <b>Score</b>        | коэффициент финансовой устойчивости                               |
| <b>CONTROL_RISK</b> | степень платежеспособности  |
| <b>Audit_Risk</b>   | уровень финансового риска на основании аудиторского заключения    |

## Обработка данных

1. Анализ полноты данных.
2. Проверка типов данных.
3. Проверка признаков на корреляцию.
4. Анализ распределения данных для каждого признака.
5. Нормализация признаков.

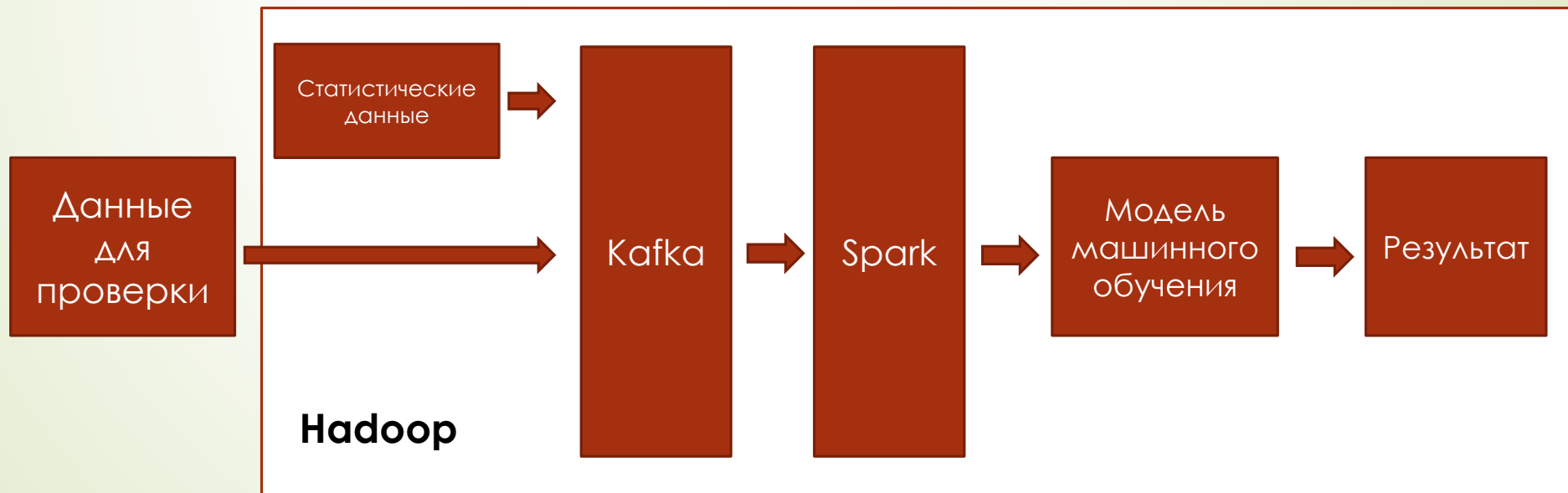


## Построение модели

1. Отбор признаков.
2. Нормализация признаков.
3. Оценка качества.

## Архитектура

На начальном этапе, для модели будет достаточно одного виртуального сервера в корпоративном облаке. Но в дальнейшем, с накоплением данных о проверках, порядка 2500 проверок в год на группу компаний и планируемым увеличением числа компаний в группе до 10, необходимо будет развернуть, распределённое хранилище на базе Hadoop Spark



## Финансы

**Исходные данные:** за прошлый год: всего группа компаний столкнулась с 10 недобросовестными контрагентами. Из них были выявлены только 5. Средний убыток от одного недобросовестного контрагента составил 500 тыс. руб. Таким образом суммарный убыток составил порядка 2,5 млн. руб.

**Предлагаемое решение** обладает следующими метриками точности:

- Процент добросовестных компаний, определённых по результатам проверки как добросовестные (TPR) = 95%
- Процент недобросовестных компаний, определённых в результате проверки как недобросовестные (TNR) = 95%, т.е. ошибка составляет всего 5%.

### **Результат применения модели.**

Таким образом, применение данной модели в прошлом году позволило бы выявить 9 из 10 недобросовестных компаний. Что позволило бы избежать потерь в размере 2 млн. руб.