# Введение

Последние десятилетия в результате стремительного развития вычислительной техники

Каждый день в мире совершаются миллионы финансовых транзакций. Хоть деньги и частично утратили свой физический, ощущаемый облик, мошенники, желающие их похитить, остались.

В связи с огромным количеством совершаемых транзакций, человеку будет практически нереально анализировать весь этот поток. На помощь ему приходят вычислительные машины, которые в последние годы стремительно развиваются и способны выполнять невероятные по объему и скорости вычисления.

В связи с этим набирает закономерную популярность такая наука как машинное обучение. С помощью предлагаемых ею методов возможно предсказыватьрезультат тех или иных задач.

В данной работе я постараюсь обучить модель, способную отличать мошеннические транзакции от немошеннических.

По соображениям конфиденциальности личных данных, в открытом доступе не имеется наборов реальных финансовых транзакций. Поэтому был выбран набор, созданный синтетически на основе реальных.

## Постановка задачи

## Теоретическая часть

## Графы

## Классификаторы

Классификация  — один из разделов [машинного обучения](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5), посвященный решению следующей задачи. Имеется множество объектов (ситуаций), разделённых некоторым образом на классы. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Это множество называется [обучающей выборкой](http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0). Классовая принадлежность остальных объектов неизвестна. Требуется построить [алгоритм](http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC), способный классифицировать произвольный объект из исходного множества.

Классифицировать объект — значит, указать номер (или наименование класса), к которому относится данный объект.

Классификация объекта — номер или наименование класса, выдаваемый [алгоритмом](http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC) классификации в результате его применения к данному конкретному объекту.

Пусть  — множество описаний объектов,  — конечное множество номеров (имён, меток) классов. Существует неизвестная целевая зависимость — отображение , значения которой известны только на объектах конечной [обучающей выборки](http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0) . Требуется построить [алгоритм](http://www.machinelearning.ru/wiki/index.php?title=%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC) , способный классифицировать произвольный объект .

### Метод k-ближайших соседей

### Градиентный бустинг

### Дерево решений

Решающими деревьями называется семейство моделей, которые позволяют восстанавливать нелинейные зависимости произвольной сложности. Они воспроизводят логические схемы, позволяющие получить окончательное решение о классификации объекта с помощью ответов на иерархически организованную систему вопросов. Причем вопрос, задаваемый на последующем иерархическом уровне, зависит от ответа, полученного на предыдущем уровне.

Каждой из вершин дерева за исключением листьев соответствует некоторый вопрос, подразумевающий несколько вариантов ответов, соответствующих выходящим ребрам. В зависимости от выбранного варианта ответа осуществляется переход к вершине следующего уровня. Листьям поставлены в соответствие метки, указывающие на отнесение распознаваемого объекта к одному из классов.

Решающее дерево называется бинарным, если каждая внутренняя или корневая вершина инцидентна только двум выходящим рёбрам.

#### Определение решающего дерева

Рассмотрим бинарное дерево, в котором:

* каждой внутренней вершине ν приписана функция (или предикат) 
* каждый листовой вершине ν приписан прогноз . В случае классификации листу может быть приписан вектор вероятностей.

### Случайный лес

## Практическая часть

## Описание данных

## Предварительный анализ

## Исследование признаков (feature engineering)

## Эксперименты на моделях

## Результаты

## Заключение