

# Курс “Интеллектуальный анализ данных в бизнесе на платформе SAS”

к. ф.м.н Нурутдинова Алсу Рафаиловна

---

- Лекции (13): основы анализа данных
- Семинары (14): знакомство с SAS

## Семинары: практические аспекты

Задачи по пособию + домашнее задание + доклады

## Система оценивания

### Теория

---

Проверочные работы на семинарах

Коллоквиум (устный или письменный)

### Практика:

---

- Практические задания

### Зачет

---

Допуск к зачету: работа на практике(20б) +домашнее задание(20б) +доклады(10б)

Зачет-50 баллов (тест+задача)

Тест в google формах (<https://support.google.com/docs/answer/7032287?hl=ru>)

В курсе изучаются задачи анализа данных и методы их решения на программной платформе SAS.

Основное внимание уделяется статистическим методам анализа данных.

Потребность в специалистах по интеллектуальному анализу данных, машинному обучению и бизнес-аналитике растёт ежегодно, а кандидатов на вакансии найти всё труднее. Уровень конкуренции на рынке труда цифровой экономики за последние 2 года снизился в 1,5 раза. Денежное вознаграждение специалистов растёт на 10% в год. Это создаёт отличные условия для начала активной работы в этой области.

*(по данным hh.ru)*



**~110 000 ₽**

Средняя заработная плата  
специалиста по анализу данных

В последние годы разворачивается очередная волна трансформации моделей деятельности в бизнесе и социальной сфере, вызванная появлением цифровых технологий нового поколения, которые в силу масштабов и глубины влияния получили наименование «сквозных», — искусственного интеллекта, робототехники, Интернета вещей, технологий беспроводной связи и ряда других. Их внедрение, по оценкам, способно повысить производительность труда в компаниях на 40% [WEF, 2018a]. В ближайшем будущем именно эффективное использование новых цифровых технологий будет определять международную конкурентоспособность как отдельных компаний, так и целых стран<sup>1</sup>, формирующих инфраструктуру и правовую среду для цифровизации.

- Data scientist
  - Работа с данными
  - Знание инструментов и методов
  - Опыт решения задач
- Менеджер
  - Понимание, как работает машинное обучение
  - Понимание узких мест, оценивание сроков
- Заказчик
  - Метрики качества
  - Требования к данным
  - Ограничения современных подходов

Тема 1.

## Интеллектуальный анализ данных и машинное обучение в условиях цифровой экономики.

---

Интеллектуальный анализ данных и машинное обучение в условиях цифровой экономики.

Основные понятия data mining и машинного обучения.

Виды данных.

Принятие решений и конкуренция на основе аналитики в цифровой экономике.

Аналитическая экономика.

Аналитика и результаты бизнеса. Анализ данных в маркетинге. Прогноз привлечения и оттока клиентов. Кластеризация клиентов и анализ профилей кластера. Задачи выявления ассоциаций и последовательностей. Применение прогнозного



моделирования в банковском скоринге и в задачах выявления мошенников.  
Текстовая аналитика и сферы ее применения.

Как перевести часы в минуты?



# Как перевести часы в минуты?

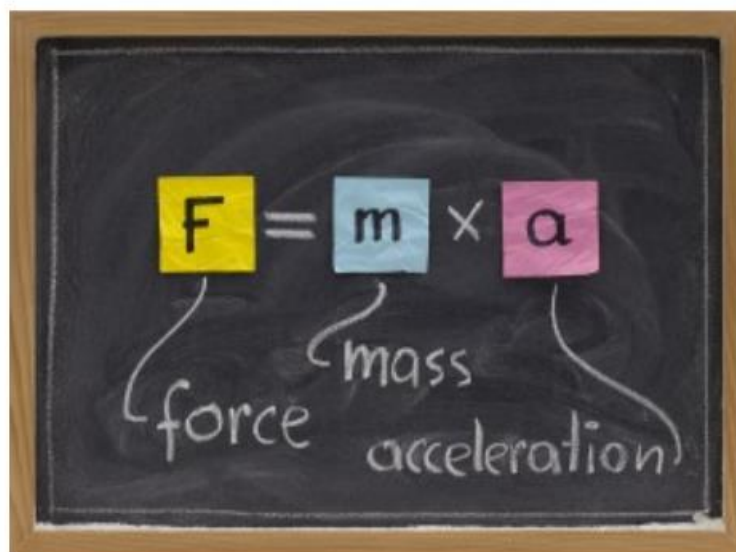
- $x$  — часы
- $f(x) = 60x$  — преобразование в минуты, функция

# Какая сила приложена к телу?

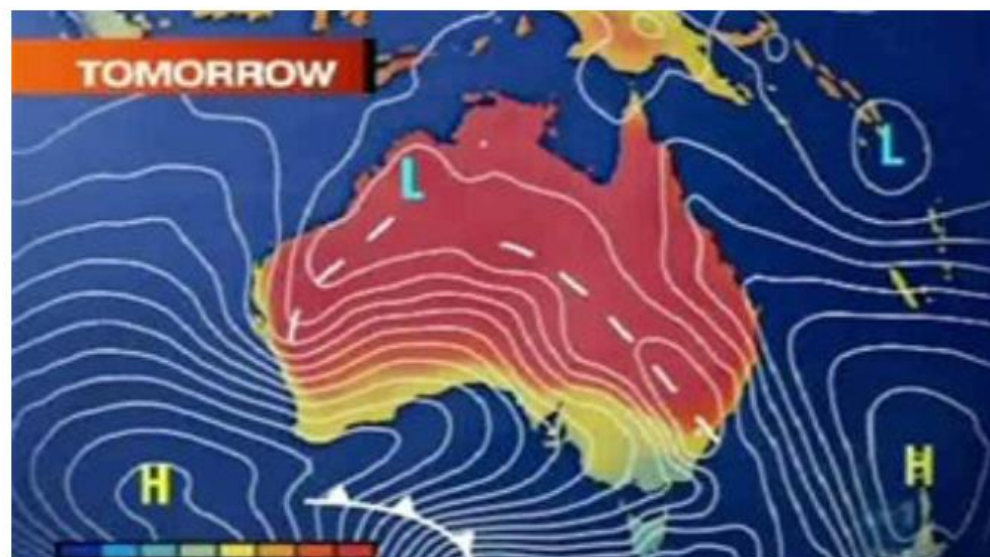
- Известны масса тела  $m$  и его ускорение  $a$
- Чему равна сила  $F$ ?

# Какая сила приложена к телу?

- Известны масса тела  $m$  и его ускорение  $a$
- Чему равна сила  $F$ ?
- Второй закон Ньютона:  $F = ma$



Как предсказать погоду?



## Уравнения Навье-Стокса

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} = -\frac{\partial P}{\partial x} + Re \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right),$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} = -\frac{\partial P}{\partial y} + Re \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right),$$

$$\frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} = -\frac{\partial P}{\partial z} + Re \left( \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right),$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0.$$

# Уравнения Навье-Стокса

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} = -\frac{\partial p}{\partial x} + Re \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} \right),$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} = -\frac{\partial p}{\partial y} + Re \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right),$$

$$\frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} = -\frac{\partial p}{\partial z} + Re \left( \frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2} \right),$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0.$$



# Анализ тональности текста

- Какой эмоциональный окрас имеет текст?
- Варианты: позитивный, нейтральный, негативный
- Применение: автоматический анализ отзывов от пользователей

# Анализ тональности текста

*«Я вижу очень большой минус, что курс будет на готовой библиотеке sci-kit. Курс от Andrew лучше тем, что ученик сам пишет алгоритм и видит изнутри, как он работает.»*

**Какой окрас?**

# Анализ тональности текста

*«Большое спасибо! Судя по всему, это как раз то, чего не хватает всем зарубежным курсам по Machine Learning и Knowledge Discovery. Это теория, математика, объяснение того, как оно устроено “в кишках”.»*

**Какой окрас?**

# Анализ тональности текста

- $x$  — текст на русском языке
  - $f(x)$  — его окрас (принимает значения -1, 0, 1)
  - Можно ли выписать формулу для  $f(x)$ ?
- 
- На входе — вовсе не числа
  - Точная зависимость может не существовать

# Больше сложных задач!

- Какой будет спрос на товар в следующем месяце?
- Сколько денег заработает магазин за год?
- Вернет ли клиент кредит?
- Заболеет ли пациент раком?
- Сдаст ли студент следующую сессию?
- На фотографии гуманитарий или технарь?
- Кто выиграет битву в онлайн-игре?

# Больше сложных задач!

- Везде — очень сложные неявные зависимости
- Нельзя выразить их формулой
- Но есть некоторое число примеров
  - Тексты с известным окрасом
- Будем приближать зависимости, используя примеры

# Анализ данных и машинное обучение

— про то, как восстановить сложные зависимости  
по конечному числу примеров

Многие задачи, возникающие в практических приложениях, не могут быть решены заранее известными методами или алгоритмами. Это происходит по той причине, что нам заранее не известны механизмы порождения исходных данных или же известная нам информация недостаточна для построения модели источника, генерирующего поступающие к нам данные. Как говорят, мы получаем данные из черного ящика. В этих условиях ничего не остается, как только изучать доступную нам последовательность исходных данных и пытаться строить предсказания совершенствуя нашу схему в процессе

предсказания. Подход, при котором прошлые данные или примеры используются для первоначального формирования и совершенствования схемы предсказания, называется методом машинного обучения (Machine Learning)