



**ИТ-класс**  
В МОСКОВСКОЙ ШКОЛЕ

# **Особенности подготовки обучающихся к чемпионату профессионального мастерства «Профессионалы» по компетенции «Машинное обучение и большие данные»**

- 1 Чемпионаты профессионального мастерства
- 2 Компетенция «Машинное обучение и большие данные»
- 3 Особенности реализации курса в школьной программе

# КОНЦЕПЦИЯ ВСЕРОССИЙСКОГО ЧЕМПИОНАТНОГО ДВИЖЕНИЯ ПО ПРОФЕССИОНАЛЬНОМУ МАСТЕРСТВУ:



ИТ-класс  
в московской школе

Интегрирование и тиражирование лучших и перспективных практик по всем профессиональным направлениям

Создание условий и системы мотивации, способствующих повышению значимости и престижа рабочих профессий

Содействие оперативному и эффективному кадровому обеспечению различных отраслей экономик

Мониторинг и анализ траектории карьерного роста молодежи

Создание и развитие профессионального экспертного сообщества




Реализация программ импортозамещения, содействие развитию, популяризации отечественных производителей

# ЧЕМПИОНАТ ПРОФЕССИОНАЛЬНОГО МАСТЕРСТВА

## Цель чемпионата

Создание условий и системы мотивации, способствующих повышению значимости и престижа рабочих профессий, профессиональному росту молодежи путем внедрения лучших практик и профессиональных навыков, а также содействие оперативному и эффективному кадровому обеспечению различных отраслей экономики

## Задачи чемпионата

-  Выявление, развитие и поддержка талантливой и перспективной молодежи;
-  Развитие актуальных компетенций, необходимых экономике региона;
-  Развитие системы образования

## Целевая аудитория:

### Эксперты



- 1 Учителя и преподаватели
- 2 Мастера ПО
- 3 Работодатели

### Участники



- 1 Студенты СПО
- 2 Школьники от 14 лет

# Этапы Чемпионата

- ВНУТРЕННИЙ ОТБОР В ОРГАНИЗАЦИЯХ;
- КВАЛИФИКАЦИОННЫЙ ОТБОР НА ЧЕМПИОНАТ И ФИНАЛ ЧЕМПИОНАТА МОСКОВСКИЕ МАСТЕРА

01


02


ОТБОРОЧНЫЙ ЭТАП НА ФИНАЛ ЧЕМПИОНАТА ПО ПРОФЕССИОНАЛЬНОМУ МАСТЕРСТВУ «ПРОФЕССИОНАЛЫ»

- ФИНАЛ ЧЕМПИОНАТА ПО ПРОФЕССИОНАЛЬНОМУ МАСТЕРСТВУ «ПРОФЕССИОНАЛЫ»;
- ФИНАЛ ЧЕМПИОНАТА ВЫСОКИХ ТЕХНОЛОГИЙ

03

# Регламентирующая документация

- ПОРЯДОК ПРОВЕДЕНИЯ ЧЕМПИОНАТА
  - МЕТОДИКА ПРОВЕДЕНИЯ ЧЕМПИОНАТА (РЕГЛАМЕНТ)
  - ПОРЯДОК ПРОВЕДЕНИЯ КВАЛИФИКАЦИОННОГО ОТБОРА
  - ПОЛОЖЕНИЕ О СБОРНОЙ КОМАНДЕ МОСКВЫ
  - ФОРМА СОГЛАСИЙ НА ОПД И НА УЧАСТИЕ
  - ПРОГРАММА ПОДГОТОВКИ ЮНИОРОВ
  - МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ
  - ПРОГРАММА ЧЕМПИОНАТА
  - ДЕЛОВАЯ ПРОГРАММА
- 

- ОПИСАНИЕ КОМПЕТЕНЦИИ
  - КОНКУРСНОЕ ЗАДАНИЕ
  - ПЛАН ЗАСТРОЙКИ ПЛОЩАДКИ
  - ПРОГРАММА СОРЕВНОВАНИЙ
  - ПЕРЕЧЕНЬ ЛИЧНЫХ ИНСТРУМЕНТОВ
  - ИНФРАСТРУКТУРНЫЙ ЛИСТ
- 

- КОНЦЕПЦИЯ ЧЕМПИОНАТНОГО ДВИЖЕНИЯ
- ПОЛОЖЕНИЕ ЧЕМПИОНАТНОГО ДВИЖЕНИЯ
- ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ ЧЕМПИОНАТНОГО ДВИЖЕНИЯ
- БРЕНДБУК ЧЕМПИОНАТНОГО ДВИЖЕНИЯ

РЕГЛАМЕНТИРУЮЩАЯ ДОКУМЕНТАЦИЯ  
ФЕДЕРАЛЬНОГО ОПЕРАТОРА

КОНКУРСНАЯ ДОКУМЕНТАЦИЯ ЧЕМПИОНАТА «МОСКОВСКИЕ МАСТЕРА»

НОРМАТИВНАЯ И ИНАЯ ДОКУМЕНТАЦИЯ ЧЕМПИОНАТА «МОСКОВСКИЕ МАСТЕРА»

# ПРЕИМУЩЕСТВА УЧАСТИЯ В ЧЕМПИОНАТАХ ПО ПРОФЕССИОНАЛЬНОМУ МАСТЕРСТВУ «ПРОФЕССИОНАЛЫ»:



ИТ-класс  
в московской школе

Обмен лучшим практиками

Победители и призеры имеют возможность пройти стажировку

Выявление и поддержка талантливой молодежи

Мотивация и поддержка интереса школьников к изучению современных языков программирования, к решению задач машинного обучения

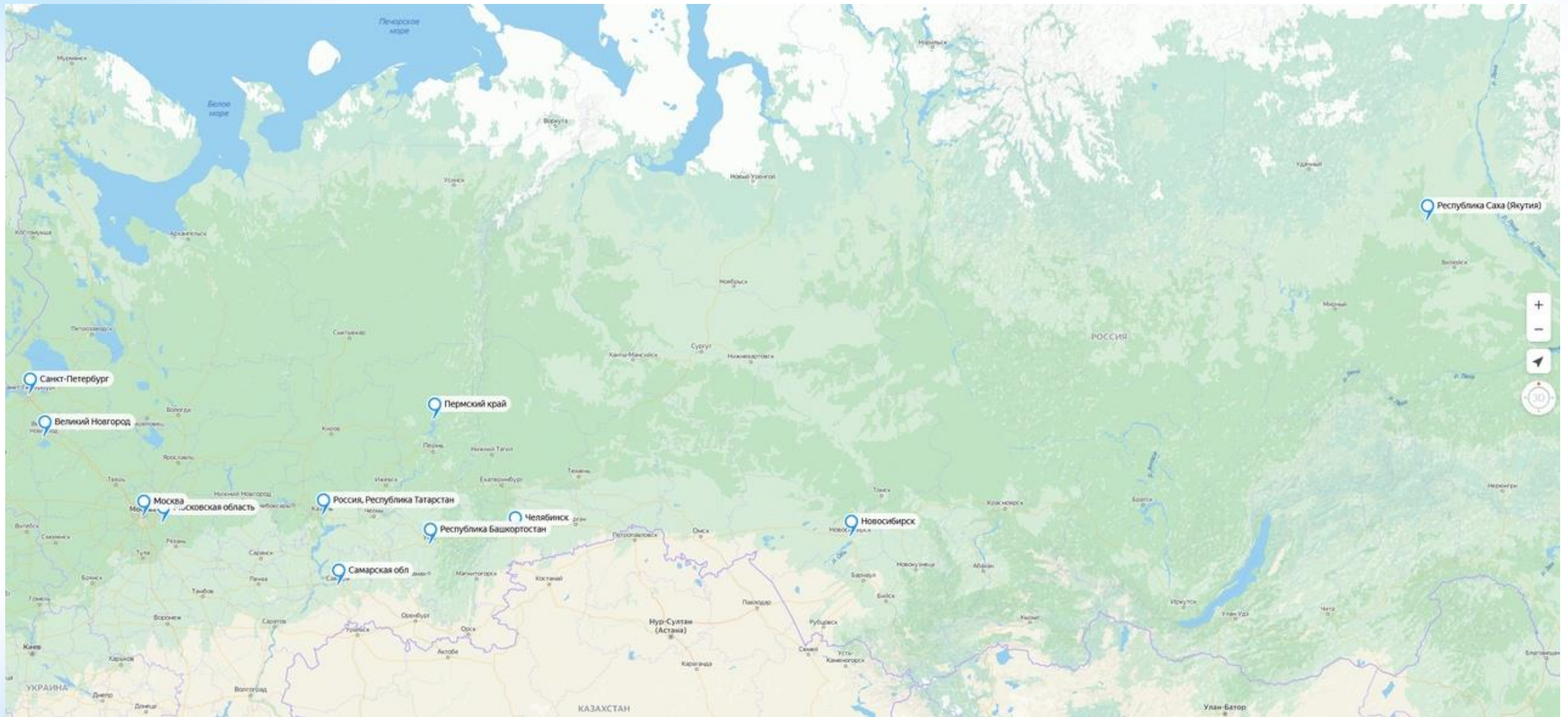
Популяризация информационных технологий

Рейтинг проектных классов

# Сборная Москвы



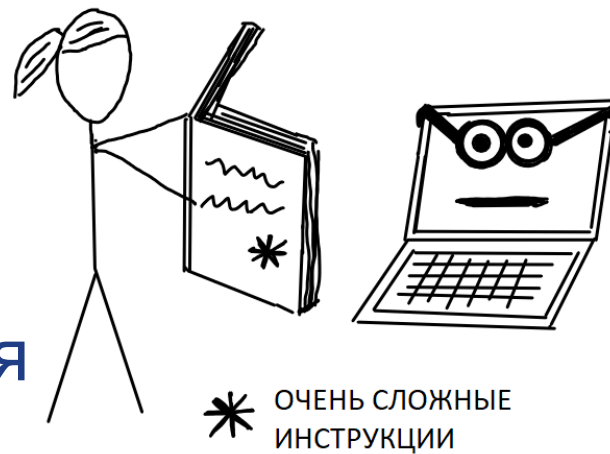
# История развития компетенции



# Machine Learning & Big Data

**Машинное обучение — это** раздел искусственного интеллекта (ИИ), предметом которого является поиск методов решения задач путем обучения в процессе решения сходных задач

Без машинного обучения



С машинным обучением



# Применение машинного обучения



ИТ-класс  
в московской школе



- Рекомендательные системы



- Ранжирование в поиске



- Распознавание и тегирование изображений



- Чат-боты



- Написание текстов



- Оптимизация маркетинговых компаний



- Прогнозирование трафика, продаж



- Персонализация коммуникаций с клиентом



- Рекомендации товаров (контента) на сайте



- Сегментация клиентов

## Плюсы:

- Высокая зарплата
- Влияние на бизнес
- Востребованность

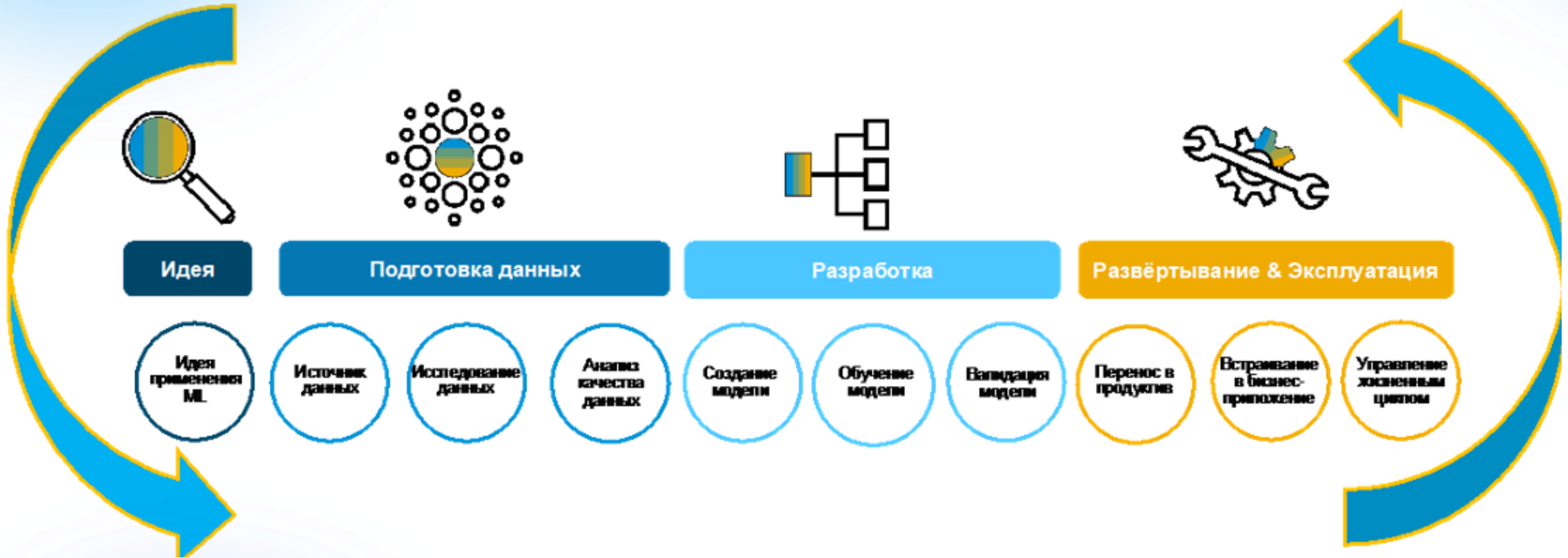
## Минусы:

- Непредсказуемые результаты

Средний уровень окладов –  
170 000Р/месяц



# Этапы решения задач



- 1 Чемпионаты профессионального мастерства
- 2 Компетенция «Машинное обучение и большие данные»
- 3 Особенности реализации курса в школьной программе

# Оцениваемые умения



ИТ-класс  
в московской школе

Сбора и подготовки  
данных;

Владения методами  
анализа данных, в том  
числе и визуального;

Поиска структуры в  
данных, выявления  
взаимосвязей и  
закономерностей;

Выдвижения гипотез и  
построение выводов по  
данным

Выбирать, обучать и  
разворачивать модель;

Оформлять техническую  
документацию;

Разрабатывать  
программное решение;

Презентовать результаты

# Пример задания

## Задача:

Необходимо решить задачу определения злоумышленника по его поведению в сети Интернет. По последовательности из десяти веб-сайтов, посещенных подряд одним и тем же человеком, требуется идентифицировать этого человека.

## Идея:

Пользователи Интернета по-разному переходят по ссылкам, и это может помочь их идентифицировать (кто-то сначала в почту, потом про футбол почитать, затем новости, социальная сеть, потом, наконец, – работать, кто-то – сразу работать, если это возможно).

Набор данных содержит информацию о сеансах просмотра пользователями, в которых:

- ✓ **site<sub>i</sub>** - это идентификаторы сайтов в этом сеансе (в соответствии с словарем site.pkl);
- ✓ **time<sub>j</sub>** - это отметки времени посещения соответствующего сайта;
- ✓ **target** - принадлежит ли эта сессия злоумышленнику.

|   | timestamp           | site                       |
|---|---------------------|----------------------------|
| 0 | 2013-11-15 08:12:07 | fpdownload2.macromedia.com |
| 1 | 2013-11-15 08:12:17 | laposte.net                |
| 2 | 2013-11-15 08:12:17 | www.laposte.net            |
| 3 | 2013-11-15 08:12:17 | www.google.com             |
| 4 | 2013-11-15 08:12:18 | www.laposte.net            |

|            | site1 | time1               | site2   | time2               | site3   | time3               | site4   | time4               | site5   | time5               | ... | time6               | site7   |
|------------|-------|---------------------|---------|---------------------|---------|---------------------|---------|---------------------|---------|---------------------|-----|---------------------|---------|
| session_id |       |                     |         |                     |         |                     |         |                     |         |                     |     |                     |         |
| 1          | 23713 | 2014-03-24 15:22:40 | 23720.0 | 2014-03-24 15:22:48 | 23713.0 | 2014-03-24 15:22:48 | 23713.0 | 2014-03-24 15:22:54 | 23720.0 | 2014-03-24 15:22:54 | ... | 2014-03-24 15:22:55 | 23713.0 |
| 2          | 8726  | 2014-04-17 14:25:58 | 8725.0  | 2014-04-17 14:25:59 | 665.0   | 2014-04-17 14:25:59 | 8727.0  | 2014-04-17 14:25:59 | 45.0    | 2014-04-17 14:25:59 | ... | 2014-04-17 14:26:01 | 45.0    |
| 3          | 303   | 2014-03-21 10:12:24 | 19.0    | 2014-03-21 10:12:36 | 303.0   | 2014-03-21 10:12:54 | 303.0   | 2014-03-21 10:13:01 | 303.0   | 2014-03-21 10:13:24 | ... | 2014-03-21 10:13:36 | 303.0   |
| 4          | 1359  | 2013-12-13 09:52:28 | 925.0   | 2013-12-13 09:54:34 | 1240.0  | 2013-12-13 09:54:34 | 1360.0  | 2013-12-13 09:54:34 | 1344.0  | 2013-12-13 09:54:34 | ... | 2013-12-13 09:54:34 | 1346.0  |
| 5          | 11    | 2013-11-26 12:35:29 | 85.0    | 2013-11-26 12:35:31 | 52.0    | 2013-11-26 12:35:31 | 85.0    | 2013-11-26 12:35:32 | 11.0    | 2013-11-26 12:35:32 | ... | 2013-11-26 12:35:32 | 11.0    |

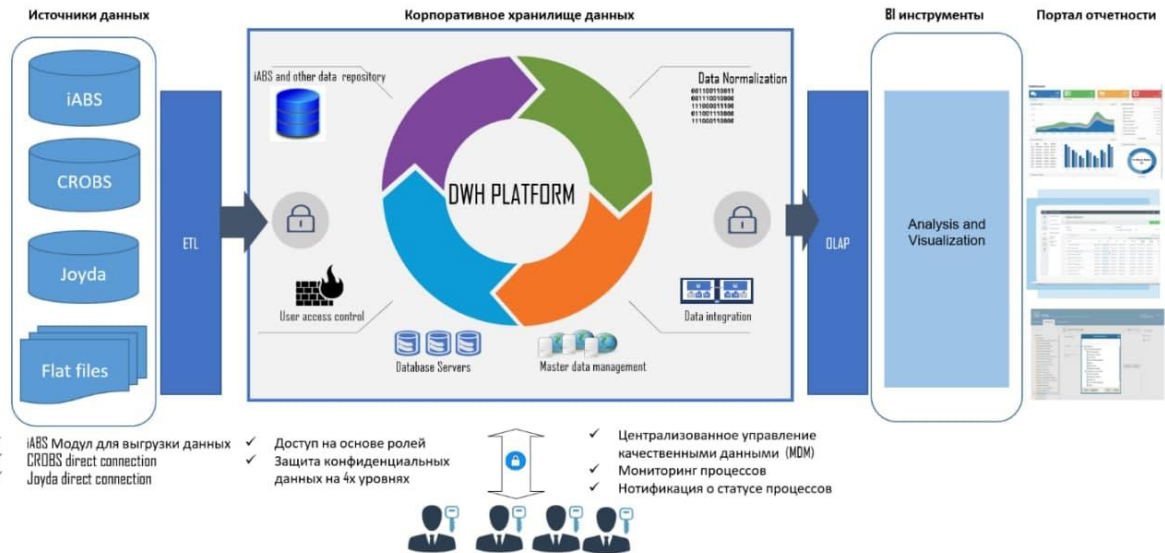
# Парсинг и предобработка данных

| Плановые почасовые объемы для режимных генерирующих единиц |  |                       |                    |                                  |                                   |   |                               |                                   |   |          |
|--|--|-----------------------|--------------------|----------------------------------|-----------------------------------|---|-------------------------------|-----------------------------------|---|----------|
| Субъект РФ   |  | ОАО "Алтайэнергобыт"  |                    |                                  |                                   |   |                               |                                   |   |          |
| Дата поставки  |  | 2015-01-01            |                    |                                  |                                   |   |                               |                                   |   |          |
| Режимная генерирующая единица                              |  | Узел расчетной модели |                    | Почасовые плановые объемы, МВтч. |                                   |   |                               |                                   |   |          |
|  |  |                       |                    | 00-01                            |                                   |   | 01-02                         |                                   |   |          |
| Номер  | Наименование                           | Номер                 | Наименование       | Технологический минимум, МВтч    | Плановый объем производства, МВтч | Максимально допустимое значение производств активной мощности, МВтч | Технологический минимум, МВтч | Плановый объем производства, МВтч | Максимально допустимое значение производств активной мощности, МВтч | Тех. мил |
| 10703  | Волгоградская ГТЭС (ГТЭС Иновация)     | 1000510               | Смоленская (ВГТЭС) | 0,000                            | 0,000                             | 0,000   | 0,000                         | 0,000                             | 0,000   |          |
| 10702  | МУП "Волгоградский теплоэлектролизный" | 1000514               | Курганская         | 11,000                           | 11,000                            | 11,000  | 11,000                        | 11,000                            | 11,000  |          |
| 10100  | ТЭЦ АЛТА                               | 1000576               | Южная              | 7,000                            | 7,000                             | 7,000   | 7,000                         | 7,000                             | 7,000   |          |
| 10701  | ТЭЦ "Тункусуйфа"                       | 1000507               | Владимирская       | 8,000                            | 8,000                             | 8,000   | 8,000                         | 8,000                             | 8,000   |          |
|  | Итого                                  |                       |                    | 26,000                           | 26,000                            | 26,000  | 26,000                        | 26,000                            | 26,000  |          |

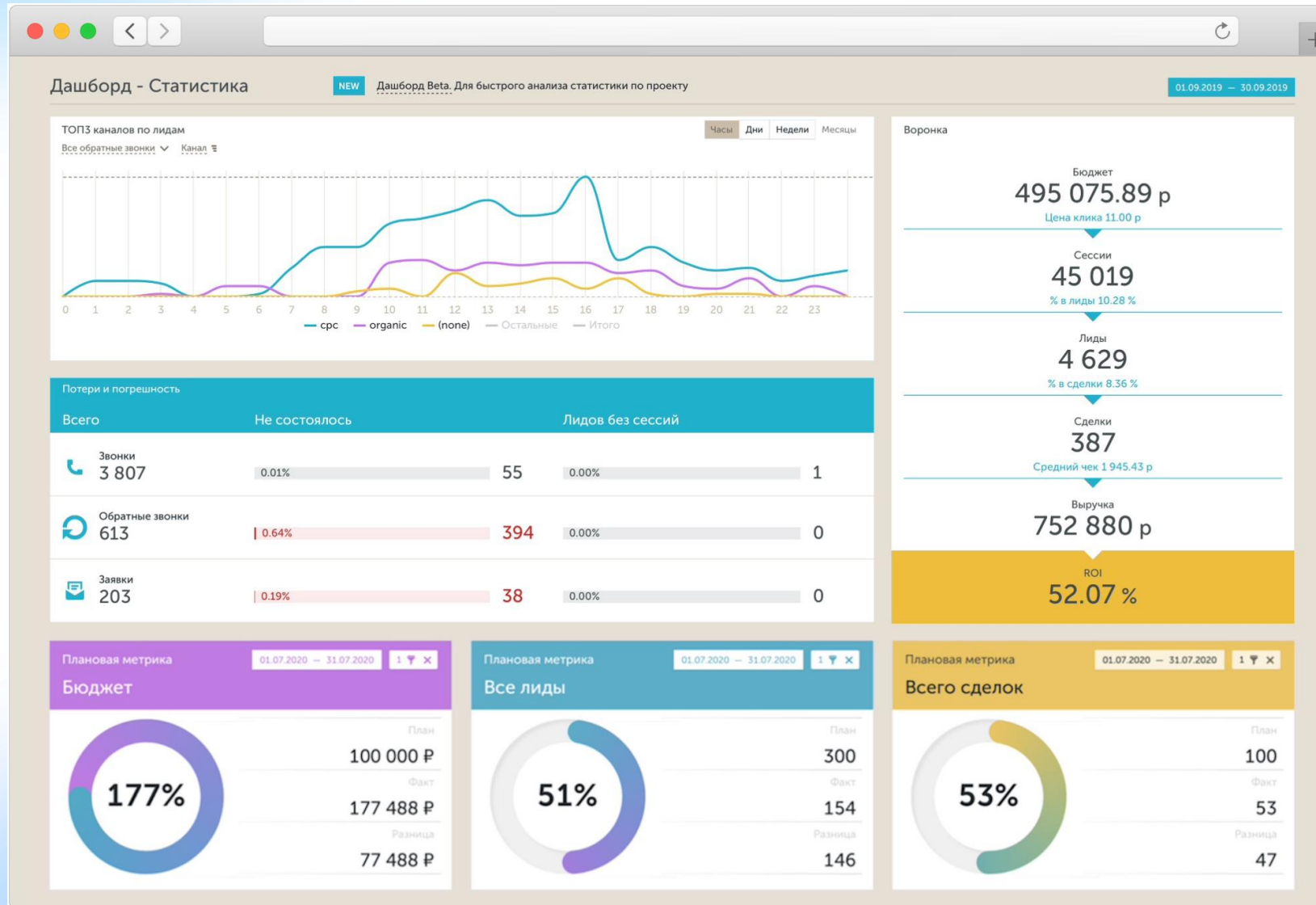
```

start = time.time()
cnxn = pyodbc.connect(
    "DRIVER= {ODBC Driver 13 for SQL Server}; SERVER=DESKTOP-1NF3KQV\\SQLEXPRESS; DATABASE=
)
production_extracted = pd.read_sql("SELECT * FROM Production", cnxn)
prices_extracted = pd.read_sql("SELECT * FROM Prices", cnxn)
data_joined = pd.merge(
    production_extracted, prices_extracted, how="left", on=["ID_node", "_datetime"]
)
time_extract_join = time.time() - start
print("Time of extract and join = ", round(time_extract_join, 1))
cnxn.close()
    
```

Time of extract and join = 10.1



# Разведочный анализ данных



# Построение, обучение и оптимизация модели



# Построение, обучение и оптимизация модели



# Построение, обучение и оптимизация модели

```
In [20]: from sklearn.base import BaseEstimator
class Never5Classifier(BaseEstimator):
    def fit(self, X, y=None):
        pass
    def predict(self, X):
        return np.zeros((len(X), 1), dtype=bool)
```

```
In [21]: never_5_clf = Never5Classifier()
cross_val_score(never_5_clf, X_train, y_train_5, cv=3, scoring="accuracy")
```

```
Out[21]: array([0.909 , 0.90715, 0.9128 ])
```

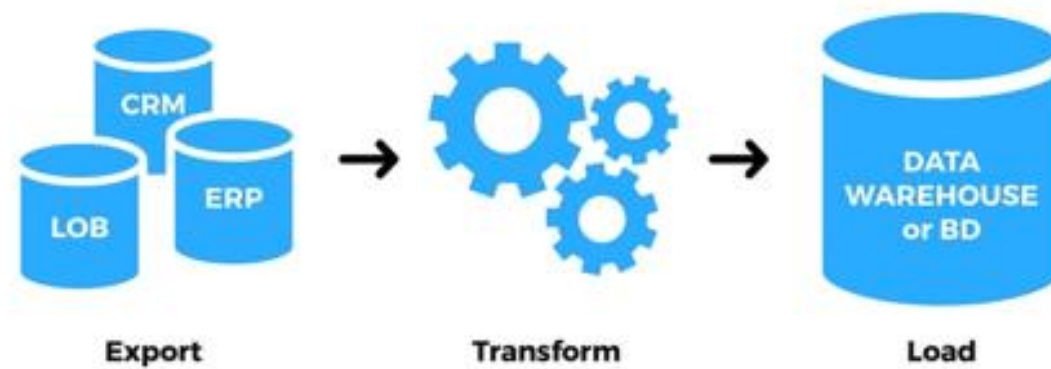


**Вывод:** Правильность (accuracy) обычно не считается предпочтительным показателем производительности для классификаторов, особенно при работе с несбалансированной выборкой

# Разработка средств интеграции и поддержки готового решения



ИТ-класс  
в московской школе



# Роль и место изучения Data Science



## Образовательный маршрут

1. Программирование (обработка структурированных данных, алгоритмы оптимизации)
2. Введение в ИТ-специальность (основы технологий искусственного интеллекта)
3. Проектная и исследовательская деятельность (кейсовые проекты по разработке моделей машинного обучения)
4. Кружок «Глубокое машинное обучение» (нейросети)



## Олимпиады и мероприятия

- ➔ Всероссийская олимпиада по искусственному интеллекту (сентябрь-ноябрь)
- ➔ Московская предпрофессиональная олимпиада, продуктовый сектор, ИТ (ноябрь-апрель)
- ➔ Чемпионат профессионального мастерства города Москвы «Московские мастера» (декабрь-март)
- ➔ Проектные конкурсы «Инженеры будущего» и «Наука для жизни» (февраль-март)
- ➔ Национальная технологическая олимпиада (НТО)

## Основные преимущества



Структурированный интерактивный документ, позволяющий запускать отдельные блоки кода в любом порядке.



Отображение и сохранение результатов вывода (текст, графика, структурные объекты, в т.ч. интерактивные)



Подходит для оформления отчетов проектных и исследовательских работ, хранения и распространения учебных и методических материалов

Тема 10. Визуализация данных. Табличные данные.ipynb

Файл Изменить Вид Вставка Среда выполнения Инструменты Справка Последнее изменение: 28 января

Содержание

Визуализация табличных данных

- Библиотека seaborn
- Библиотека plotly
- Обсуждение, задания.
- Примечание

+ Раздел

```
[ ] df.describe()
```

|       | Year_of_Release | NA_Sales    | EU_Sales    | JP_Sales    | Other_Sales | Global_Sales | Critic_Score | Critic_Count | User_Score  | User_Count   |
|-------|-----------------|-------------|-------------|-------------|-------------|--------------|--------------|--------------|-------------|--------------|
| count | 6825.000000     | 6825.000000 | 6825.000000 | 6825.000000 | 6825.000000 | 6825.000000  | 6825.000000  | 6825.000000  | 6825.000000 | 6825.000000  |
| mean  | 2007.436777     | 0.394484    | 0.236089    | 0.064158    | 0.062677    | 0.777590     | 70.272088    | 28.931136    | 7.185626    | 174.722344   |
| std   | 4.211248        | 0.967385    | 0.687330    | 0.287570    | 0.269871    | 1.963443     | 13.868572    | 19.224165    | 1.439942    | 587.428538   |
| min   | 1985.000000     | 0.000000    | 0.000000    | 0.000000    | 0.000000    | 0.010000     | 13.000000    | 3.000000     | 0.500000    | 4.000000     |
| 25%   | 2004.000000     | 0.060000    | 0.020000    | 0.000000    | 0.010000    | 0.110000     | 62.000000    | 14.000000    | 6.500000    | 11.000000    |
| 50%   | 2007.000000     | 0.150000    | 0.060000    | 0.000000    | 0.020000    | 0.290000     | 72.000000    | 25.000000    | 7.500000    | 27.000000    |
| 75%   | 2011.000000     | 0.390000    | 0.210000    | 0.010000    | 0.070000    | 0.750000     | 80.000000    | 39.000000    | 8.200000    | 89.000000    |
| max   | 2016.000000     | 41.360000   | 28.960000   | 6.500000    | 10.570000   | 82.530000    | 98.000000    | 113.000000   | 9.600000    | 10665.000000 |

```
[ ] top_platforms = df.Platform.value_counts().sort_values(ascending = False).head(5).index.values
sns.boxplot(y="Platform", x="Critic_Score", data=df[df.Platform.isin(top_platforms)], orient="h")
```

```
<Axes: xlabel='Critic_Score', ylabel='Platform'>
```

Другой полезный график для отображения таблиц - само изображение таблицы с подсветкой значений цветом, соответствующих их величине. Выполняется командой `heatmap()`, которой указываем таблицу и дополнительные параметры отображения, например:

- `annot=True`, - подписывать значение клеточек таблицы
- `fmt=".1f"` - в каком виде выводить числа (здесь число с плавающей точкой, один знак после запятой,
- `linewidths=.5` - толщина линий, разделяющих клеточки таблицы. По умолчанию будет показана и цветовая шкала.

# IDE для Data Scientist'a

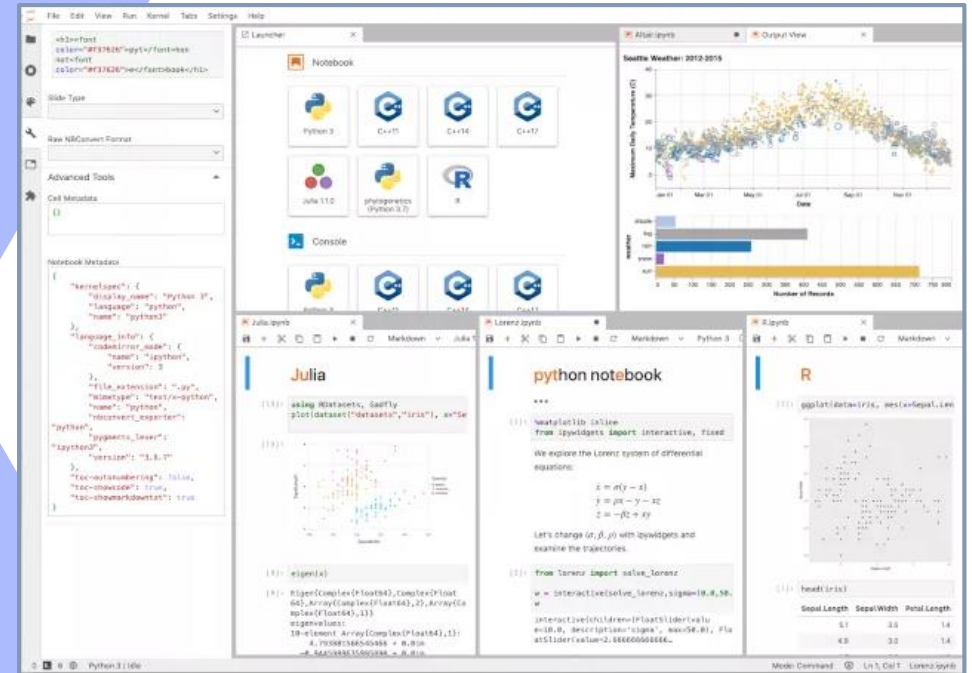


PyCharm

Spyder

Visual Studio Code (VS Code)

JupyterLab



# Библиотеки и модули Python



**Numpy, Pandas** – числовые и табличные данные

**Pillow, OpenCV, pyAudio** – медиа-данные  
**Folium** – картография

**Matplotlib, Seaborn, Plotly** – графический анализ

**Scikit-Learn** – модели машинного обучения и вспомогательные инструменты

**TensorFlow.Keras** – упрощенное конструирование и обучение нейросетей  
**PyTorch** – библиотека глубокого обучения

1. Концепция всероссийского чемпионатного движения «Профессионалы». Все документы и новости <https://pro.firpo.ru/>
2. Концепция чемпионатного движения на странице <https://pro.firpo.ru/o-nas/dokumenty/>
3. Этапы чемпионата: Региональные чемпионаты, Отборочные (межрегиональные), финал или чемпионат высоких технологий, информация о том, где и когда проходили финальные мероприятия в 2023 году <https://pro.firpo.ru/meropriyatiya/grafik-finalnogo-etapa/>
4. Сайт регионального этапа чемпионата, на котором размещена вся информация и документы <https://prof.mcrpo.ru/>
5. Презентация чемпионата:  
<https://prof.mcrpo.ru/insertfiles/2023/%D0%A0%D0%A7%202023/%D0%9F%D1%80%D0%B5%D0%B7%D0%B5%D0%BD%D1%82%D0%B0%D1%86%D0%B8%D1%8F%20%D0%A0%D0%A7.pdf>
6. Методика проведения чемпионата:  
<https://prof23.mcrpo.ru/insertfiles/%D0%BF%D1%80%D0%BE%D0%B2%D0%B5%D0%B4%D0%B5%D0%BD%D0%B8%D1%8F%20%D1%87%D0%B5%D0%BC%D0%BF%D0%B8%D0%BE%D0%BD%D0%B0%D1%82%D0%B0.pdf> (здесь обязательно изучить пункты 11-18,20 о правах и обязанностях участников и экспертов)
7. Комплекты конкурсной документации <https://pro.firpo.ru/kompetentsi/>

## Контакты:



Киселева Светлана Владимировна



Mksvet.vl@mail.ru



+7(977)4797925