

АНАЛИЗ ВОСТРЕБОВАННОСТИ АНАЛИТИКОВ ДАННЫХ В АГРОПРОМЫШЛЕННЫХ ПРЕПРИЯТИЯХ

Рыбалкин Богдан Александрович, бакалавр 1 курса института экономики и управления АПК ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, br3027645@gmail.com

Ульянкин Александр Евгеньевич, ассистент кафедры статистики и кибернетики института экономики и управления АПК ФГБОУ ВО РГАУ-МСХА имени К.А. Тимирязева, aeulianckin@rgau-msha.ru

Аннотация. В статье описан процесс анализа востребованности аналитиков данных в сфере агропромышленности, данные взяты с сайта <https://hh.ru>, а также приведены примеры использования библиотеки Request для работы с API. Использование этого метода позволяет получать большие объемы информации для ее дальнейшей обработки

Ключевые слова: Request, sqlite3, анализ данных, аналитики данных, сельскохозяйственные предприятия

ANALYSIS OF THE DEMAND FOR DATA ANALYSTS IN AGRO-INDUSTRIAL ENTERPRISES

Rybalkin Bogdan Aleksandrovich, 1st year Bachelor of the Institute of Economics and Management of the Agro-Industrial Complex of the Russian State Agrarian University-Moscow Timiryazev Agricultural Academy, br3027645@gmail.com

Ulianckin Alexander Evgenyevich, Assistant at the Department of Statistics and Cybernetics of the Institute of Economics and Management of the Agroindustrial Complex of the Federal State Budgetary Educational Institution of the Russian State Agrarian University-Moscow State Agricultural Academy named after K.A. Timiryazev, aeulianckin@rgau-msha.ru

Annotation. The article describes the process of analyzing the demand for data analysts in the agro-industry, the data is taken from the <https://hh.ru> website, and also provides examples of using the Request library to work with the API. The use of this method allows you to obtain large amounts of information for its further processing

Keywords: Request, data analysis, data analytics, agricultural enterprises

В современном мире, данные становятся новым золотом, появилась востребованная и перспективная профессия — Data Scientist. Это специалист по

анализу данных, который может превратить огромные массивы информации в полезные знания.

Data Scientist занимается сбором, обработкой и анализом данных. Он использует математические методы, статистический анализ и машинное обучение для выявления закономерностей и тенденций. На основе полученных результатов Data Scientist создает модели и алгоритмы, которые помогают принимать обоснованные решения и прогнозировать будущие события.

Для работы Data Scientist должен обладать глубокими знаниями в области математики, статистики и информатики. Также ему необходимо владеть языками программирования, такими как Python, R или SQL. Кроме того, Data Scientist должен уметь работать с большими данными (Big Data) и понимать принципы машинного обучения и искусственного интеллекта.

В агропромышленных компаниях Data Scientist занимается анализом данных о производстве продукта и его востребованности. Например он может исследовать данные о росте пшеницы и множестве факторов, что бы понять как повысить урожай, или например исследовать востребованность молокопродуктов на рынке, что бы отдать информацию о возможности увелечения прибыли при увелечении производства или наоборот сокращения производства при низком спросе. Исследуя это он помогает компании принять верные решения на пути к развитию.

На данный момент Data Scientistc является очень востребованной профессией о чем говорит количество вакансий в размере 853. Все данные я получал с помощью открытого API сайта hh.ru, используя библиотеку request. Для работы с базой данных я использовал библиотеку sqlite3.

Requests — это библиотека, которая создана для простой и быстрой работы с запросами. Стандартные HTTP-библиотеки Python, например Urllib3, часто требуют значительно больше кода для одного и того же действия.

API — это механизм, который позволяют двум программам взаимодействовать друг с другом, используя набор определений и протоколов.

Sqlite3 — это библиотека, которая предоставляет легкую дисковую базу данных, не требующую отдельного серверного процесса и позволяющую получать доступ к базе данных с использованием нестандартного варианта языка запросов SQL.

Matplotlib — библиотека на языке программирования Python для визуализации данных двумерной и трёхмерной графикой. Получаемые изображения могут быть использованы в качестве иллюстраций в публикациях.

Анализ собранных мной данных включал следующие этапы:

1. **Сбор данных.** Сбор основных показателей: название вакансии, зарплату, описание вакансии, название компании и описание компании и еще несколько других параметров. Обработка и замена (при надобности) параметров которые могут быть пустыми. Занесение всего в базу данных.

2. **Фильтрация данных с дальнейшим занесением в другую таблицу.** Обход всех данных полученных ранее и их фильтрация. Фильтрация осуществляется по ключевым словам: «Агро», «Сель», «АПК», «РСХБ».

Подходящие данные заносятся в новую таблицу.

```
import sqlite3

con = sqlite3.connect("data.db")
cur = con.cursor()

words_list = ["агро", "сель", "анк", "рехб"]
id_set = set()
for data in cur.execute("SELECT * FROM DataScientists").fetchall():
    data = list(data)
    id = data.pop(0)

    if any([word in str(d).lower() for d in data for word in words_list if d]):
        cur.execute( sql: "INSERT INTO AgroDataScientists (name, area, requirement, responsibility, experience, schedule, "
                        "employment, salary, description, key_skills, employer_name, employer_description, employer_area, "
                        "employer_industries) "
                        "VALUES (?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?, ?)",
                    tuple(data)).fetchone()
        id_set.add(id)

print()
print(len(id_set))

con.commit()
```

Рисунок 1 – Фильтрация данных

3. Анализ полученных данных. Анализ заключается в получении разных показателей которые могут нам сказать о востребованности Data Scientist специалистов в агропромышленности: например такие как: отношения количества вакансий от агро-предприятий к общему количеству, средняя зарплата в агро-компаниях относительно средней зарплаты у остальных и.д. Так же интерес могут представлять такие показатели как: города в которых предлагаются вакансии и ключевые навыки.

4. Визуализация данных. . Для создания графиков, диаграмм и других визуальных представлений используем библиотеки Matplotlib.

После анализа данных можно сказать о следующем:

- 26 вакансий из 853 относятся к агропромышленности. То есть около 3% от общего числа, это достаточно хороший результат.
- Зарплата среди всех вакансий Data Scientist равна 213 861 рублей, среди вакансий связанных с агропромышленностью средняя зарплата равна 151 250 рублей

Согласно этим данным можно сказать что специалисты Data Scientist достаточно востребованы в агропромышленности.

Библиографический список

1. Библиотеки Python для анализа данных и машинного обучения: использование Python в Data Science URL: <https://practicum.yandex.ru/>

2. Ульянов, А. Е. Автоматизация анализа международных эколого-экономических систем на основе технологии машинного обучения / А. Е. Ульянов // Электронный сетевой политематический журнал "Научные труды

КубГТУ". – 2020. – № 6. – С. 74-83. – EDN MLTKLP.

3. Что такое Big Data и почему их называют «новой нефтью» URL:
<https://trends.rbc.ru/>

4. SQLite3 что это за библиотека - описание модуля URL:
<https://proglib.io/p/samouchitel-po-python-dlya-nachinayushchih-chast-22-osnovy-raboty-s-sqlite-2023-06-15>