

## АВТОМАТИЗАЦИИ АНАЛИЗА БОЛЬШИХ ОБЪЕМОВ ДАННЫХ С ПРИМЕНЕНИЕМ НЕЙРОСЕТЕВОЙ ТЕХНОЛОГИИ

**Садикова Шахноза Шукуруллаевна**

доцент, ТГТУ, *Sshakhnoza@yandex.ru*,

**Кимизбаева Орзигуль Эшоновна**

доцент, ТГТУ,

**Садикова Ферузахон Ботирхон кизи**

докторант, ТГТУ.

### Аннотация

*Mazkur maqolada har kuni ishlab chiqariladigan, to'planadigan va qayta ishlanadigan katta hajmdagi ma'lumotlar yuzasidan fikrlar bildirilgan bo'lib, katta hajmdagi ma'lumotlar (Big Data)ni tahlil qilishda foydalaniladigan texnologiyalarni qo'llash usullari keltirilgan. Big Data strategiya tanlashda, mahsulotlarni ishlab chiqishda yoki mijozlarga ularning xatti-harakatlaridan kelib chiqqan holda aniq tavsiyalar berishda biznes qarorlarini qabul qilishda asosiy vosita sifatida ko'rib chiqilgan. Ma'lumotlarni tahlil qilish, model yaratish va ma'lumotlarni vizualizatsiya qilish algoritmi ishlab chiqilgan. Ishlab chiqilgan algoritm asosida Python dasturlash tilidan foydalanib turli xil tasvirlarni vizuallashtirish imkonini beruvchi neyron tarmog' ishlab chiqilgan. Bu esa katta hajmdagi ma'lumotlarni tahlil qilish orqali kompaniyalar uchun yangi bilimlarga ega bo'lish va samarali boshqaruv imkoniyatini yaratishda yordam beradi.*

### Аннотация

*В данной статье выражены мнения о больших объемах производимой, хранимой и обрабатываемой информации, а также приводятся методы применения технологий, используемых при анализе больших объемов данных (Big Data). Большие данные рассматриваются как ключевой инструмент принятия бизнес-решений при выборе стратегии, разработке продуктов или выработке конкретных рекомендаций клиентам на основе их поведения. Разработан алгоритм анализа данных, создания моделей и визуализации данных. На основе разработанного алгоритма была разработана нейронная сеть, позволяющая визуализировать различные изображения с помощью языка программирования Python. Это поможет компаниям получить новые знания и создать эффективные возможности управления посредством анализа больших объемов данных.*

### Annotation

*This article expresses opinions about the large volumes of information produced, stored and processed, and also provides methods for using technologies used in the analysis of large amounts of data (Big Data). Big Data is seen as a key tool for making business decisions when choosing a strategy, developing products, or making specific recommendations to customers based on their behavior. An algorithm for analyzing data, creating models and visualizing data has been developed. Based on the developed algorithm,*

*a neural network was developed that allows you to visualize various images using the Python programming language. This will help companies gain new knowledge and create effective management capabilities by analyzing large amounts of data.*

#### **Kalit so‘zlar**

*Katta hajmdagi ma'lumotlar, avtomatlashtirish texnologiyalari, mashinali o'qitish algoritmlari, ma'lumotlarni vizualizatsiya qilish, neyron tarmoq, kataloglash jarayoni, avtomatik tasniflash.*

#### **Ключевые слова**

*Большие объемы данных, технологии автоматизации, алгоритмы машинного обучения, визуализация данных, нейросеть, процессе каталогизации, автоматическая классификация.*

#### **Keywords**

*Big Data, automation technologies, machine learning algorithms, data visualization, neural network, cataloging process, automatic classification.*

Сегодняшний мир охватывается огромным количеством данных, которые производятся, собираются и обрабатываются каждый день. Как следствие, анализ больших объемов данных (Big Data) является ключевым инструментом при принятии бизнес-решений в выборе стратегии, разработке продуктов или конкретных рекомендаций клиентам на основе их поведения. Таким образом, проанализировав огромное количество данных, компании могут получить новые знания и помочь себе в эффективном управлении.

Однако, анализ больших объемов данных без автоматизации является невозможным. Вручную обработать огромный объем информации потребует значительного количества времени и человеческих ресурсов, что скажется на эффективности и точности результатов. Кроме того, человеческий фактор может привести к ошибкам при обработке и интерпретации данных. Именно поэтому автоматизация анализа больших данных становится все более востребованной.

В настоящее время технологии автоматизации анализа больших данных стали очень продвинутыми, благодаря чему можно найти быстрое и эффективное решение для обработки данных. Большинство инструментов автоматизации анализа данных используют технологии и алгоритмы машинного обучения, которые позволяют системам автоматически находить связи и закономерности в данных, что в свою очередь дает возможность принимать более точные и обоснованные решения.

Автоматизация анализа больших объемов данных (Big Data) - это процесс применения различных методов и инструментов для автоматизации обработки, анализа и интерпретации больших объемов данных. Целью автоматизации анализа больших данных является улучшение производительности, точности и эффективности анализа данных за счет использования алгоритмов машинного обучения, искусственного интеллекта, автоматизированных платформ и инструментов для обработки и анализа больших объемов данных. Автоматизация

анализа больших данных также позволяет быстрее выявлять тенденции, прогнозировать поведение клиентов и принимать более обоснованные решения на основе данных.

Существует много инструментов для обработки и анализа больших объемов данных. Некоторые из них:

1. Hadoop - это платформа для обработки больших объемов данных, которая позволяет работать с данными, хранящимися на множестве серверов.
2. Apache Spark - это распределенная вычислительная система, которая позволяет обрабатывать данные в режиме реального времени.
3. Apache Cassandra - это децентрализованная база данных, которая способна обрабатывать большие объемы данных на многих серверах.
4. Tableau - это инструмент для визуализации данных, который позволяет создавать инфографику, графики и дашборды для анализа данных.
5. Power BI - это набор инструментов для анализа данных, который позволяет создавать интерактивные дашборды, отчеты и визуализации данных.
6. SAS - это программное обеспечение для анализа данных, которое используется для статистического анализа, моделирования данных и машинного обучения.
7. Python и R - это языки программирования, которые позволяют проводить анализ данных, строить модели и визуализировать данные.

Это только несколько примеров. Существует множество других инструментов, в зависимости от нужд и целей анализа данных.

Алгоритм проведения анализа данных, построения модели и визуализации данных может быть следующий:

1. Сбор данных: Начните с сбора данных, необходимых для анализа. Это может включать в себя данные из базы данных, онлайн-источников, опросов и т. д.
2. Предварительная обработка данных: Очистка данных от пропусков, выбросов, дубликатов и других неправильных значений. Преобразование данных в подходящий формат (например, даты, числа, текст) и выполнение стандартизацию и нормализацию данных.
3. Исследовательский анализ данных (EDA): Проведение исследовательского анализа данных, для понятия основных характеристик данных, включая распределение, корреляции и важные особенности.
4. Построение модели: Выбор подходящего метода машинного обучения или статистического моделирования для создания модели на основе имеющихся данных. Это может быть линейная регрессия, деревья решений, нейронные сети или другие алгоритмы.
5. Оценка модели: Оценка производительности модели с помощью метрик, таких как средняя абсолютная ошибка (MAE), среднеквадратическая ошибка (MSE)

или коэффициент детерминации ( $R^2$ ). В случае классификации используется метрики, такие как точность, полнота, F1-мера и т. д.

6. Визуализация данных: Построение график и диаграмм, для визуализации результатов анализа данных и моделирования. Это может включать в себя изображения распределений, корреляций, важности признаков, кривые обучения и т. д.

7. Интерпретация результатов: анализ и интерпретация результатов анализа данных и моделирования, для принятия решения на основе полученных знаний.

8. Подготовка отчетности: Создание отчета, обобщающего процесс анализа данных, методологии моделирования, результатов и выводов для дальнейшего использования.

По приведенному алгоритму можно визуализировать данные на языке программирования Python.

1. Импорт необходимой библиотеки:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Activation, Dropout
from tensorflow.keras.utils import plot_model
```

2. Загрузка датасет:

```
df = pd.read_csv('dataset.csv')
```

3. Проверка данных:

```
print(df.head())
print(df.describe())
print(df.isnull().sum())
```

4. Визуализация данных:

```
sns.pairplot(df, hue='target')
plt.show()
```

5. Разделение данных на тренировочный и тестовый датасеты:

```
X = df.drop('target', axis=1)
y = df['target']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

6. Создание модели нейросети:

```
model = Sequential()
model.add(Dense(64, input_dim=X_train.shape[1], activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dropout(0.5))
model.add(Dense(1, activation='sigmoid'))
```

```
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```

7. Обучение модели:

```
history = model.fit(X_train, y_train, epochs=100, batch_size=32, validation_split=0.1)
```

8. Визуализация процесса обучения:

```
plt.plot(history.history['accuracy'], label='accuracy')  
plt.plot(history.history['val_accuracy'], label='val_accuracy')  
plt.title('Accuracy')  
plt.xlabel('epoch')  
plt.ylabel('accuracy')  
plt.legend(loc='lower right')  
plt.show()
```

9. Оценка модели:

```
score = model.evaluate(X_test, y_test, verbose=0)  
print("Test loss:", score[0])  
print("Test accuracy:", score[1])
```

10. Строение графика предсказаний и реальных значений:

```
y_pred = model.predict(X_test)  
fig, ax = plt.subplots()  
ax.scatter(y_test, y_pred)  
ax.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], 'k--', lw=4)  
ax.set_xlabel('Measured')  
ax.set_ylabel('Predicted')  
plt.show()
```

Разработанную нейронную сеть можно использовать для визуализации изображений. Например, можно обучить нейронную сеть на совокупности изображений, представляющих различные виды цветовой гаммы, и использовать ее для генерации новых картинок в разных цветах и оттенках. Кроме этого можно использовать для классификации и группировки товаров на изображениях. Алгоритм может основываться на определенных признаках, чертах или параметрах, которые помогут разделить товары на группы.

Для использования нейронной сети для классификации и группировки товаров на изображениях можно создать модель глубокого обучения, например, сверточную нейронную сеть (CNN). В процессе обучения такой модели можно подать большой набор размеченных изображений товаров, чтобы модель научился различать их различные характеристики и особенности.

После обучения модели можно использовать ее для классификации новых изображений товаров. Например, в онлайн-магазинах, можно загрузить изображение товара, а затем использовать обученную нейронную сеть, чтобы определить, к какой категории товаров он относится. Таким образом, нейронная сеть поможет автоматизировать процесс классификации и группировки товаров на изображениях.

Кроме того, нейронная сеть может быть использована для визуализации данных о товарах, например, путем отображения графиков или диаграмм, которые отображают различные характеристики и особенности товаров, такие как цвет, размер, форма и другие параметры.

Такая модель сможет автоматически классифицировать товары на изображениях на основе их внешнего вида, цвета, формы и других характеристик. После обучения модель можно применять для автоматической классификации новых изображений товаров, что упростит процесс каталогизации и организации товаров для розничных и интернет-магазинов.

Кроме того, с помощью нейронной сети также можно провести группировку товаров на изображениях на основе их сходства и внешних характеристик. Это позволит улучшить точность анализа ассортимента товаров, упростить поиск и навигацию по каталогам, а также улучшить персонализацию рекомендаций.

Таким образом, использование нейронной сети для классификации и группировки объектов на изображениях может значительно улучшить эффективность управления ассортиментом, улучшить опыт покупателей и повысить конкурентоспособность розничных и интернет-магазинов.

Проведенный анализ, позволил нам предположить, как технологии автоматизации анализа больших данных влияют на вероятность участия компаний в процессе каталогизации и организации товаров для розничных и интернет-магазинов.

## ЛИТЕРАТУРА:

1. Балашов М.К. Будущее маркетинга: нейронные сети как инструмент обслуживания клиентов // Наука и образование сегодня. 2017. № 11. С. 52-55.
2. Киселев А.В. Влияние технологий обработки больших данных на развитие сферы электронной торговли // StudNet. 2020. URL: [https:// cyberleninka.ru/ article/n/ vliyanietehnologiy-obrabotki-bolshih-dannyh-na-razvitie-sferyelektronnoy-torgovli](https://cyberleninka.ru/article/n/vliyanietehnologiy-obrabotki-bolshih-dannyh-na-razvitie-sferyelektronnoy-torgovli)
3. Соколов М.П. Обработка больших объемов параметрической информации с помощью нейронной сети Кохонена // Научный вестник МГТУ ГА. 2006. №100. URL: <https://cyberleninka.ru/article/n/obrabotka-bolshih-obemovparametricheskoy-informatsii-s-romoschyu-neuronnoy-setikohonena> (дата обращения: 28.05.2023).