

УДК 159.9.075:004.891

**ИССЛЕДОВАНИЕ ПСИХОЛОГИЧЕСКИХ ПРОБЛЕМ БОЛЬШИХ  
КОЛЛЕКТИВОВ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИЙ  
ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ**

**Мирянова В.Н.**

*Кандидат технических наук, доцент кафедры «Информатика и управление в технических системах»*

*ФГАОУ ВО «Севастопольский государственный университет»,  
Севастополь, Россия*

**Дементьев К.В.**

*Аспирант кафедры «Информатика и управление в технических системах»*

*ФГАОУ ВО «Севастопольский государственный университет»,  
Севастополь, Россия*

**Мирянова А.Д.**

*Ассистент кафедры «Психология»*

*ФГАОУ ВО «Севастопольский государственный университет»,  
Севастополь, Россия*

**Аннотация**

Работа посвящена изучению психологических проблем с использованием современных технологий анализа данных. Предложен подход, который сочетает психологические теории с возможностями машинного обучения и анализа больших данных для достижения более глубокого понимания сложных взаимосвязей внутри больших коллективов. Исследование включает в себя анализ эмоциональных состояний, полученных из открытых данных, социальных сетей, прогнозирование поведения с помощью машинного обучения. Результаты исследования могут быть полезны для разработки инновационных подходов к решению психологических проблем в обществе.

**Ключевые слова:** технологии анализа данных, психологические проблемы, машинное обучение, большие данные, социальные сети, искусственный интеллект, психическое здоровье.

***ANALYSIS OF LARGE COLLECTIVES PSYCHOLOGICAL PROBLEMS  
USING DATA MINING TECHNOLOGIES***

***Mirianova V.N.***

*Ph.D., Associate Professor of the Department «Computer Science and Management in Technical Systems»,*

*Sevastopol State University,*

*Sevastopol, Russia*

***Dementiev K.V.***

*Postgraduate at the Department «Computer Science and Management in Technical Systems»,*

*Sevastopol State University,*

*Sevastopol, Russia*

***Mirianova A.D.***

*Assistant of the Department «Psychology»,*

*Sevastopol State University,*

*Sevastopol, Russia*

**Abstract**

The work is devoted to the study of psychological problems using modern data analysis technologies. An approach is proposed that combines psychological theories with the capabilities of machine learning and big data analysis to achieve a deeper understanding of complex relationships within large groups. The study includes the analysis of emotional states obtained from open social network data, behavior prediction using machine learning. The results of the study can be useful for

developing innovative approaches to solving psychological problems in society.

**Keywords:** Data Analysis Technologies, Psychological Issues, Machine Learning, Big Data, Social Media, Artificial Intelligence, Mental Health.

Технологии в современном мире развиваются с беспрецедентной скоростью. Их использование становится все более важным в различных областях жизни, включая психологию. Наука о поведении и мышлении людей теперь не ограничивается традиционными методами исследования, но все больше использует новые подходы, основанные на данных и машинном обучении.

В этой связи данная работа посвящена изучению методики исследования психологических проблем большого коллектива с использованием технологий интеллектуального анализа данных. Предлагается подход, который объединяет психологические теории и методы с мощностью машинного обучения и анализа данных для более глубокого и точного понимания сложных взаимосвязей внутри больших коллективов.

Технологии интеллектуального анализа данных предоставляют нам новые возможности для изучения психологических явлений, так как они позволяют обрабатывать огромные объемы информации и выявлять сложные закономерности, которые могут оставаться незамеченными при использовании более традиционных методов.

Машинное обучение, как ключевой компонент интеллектуального анализа данных, предлагает эффективные алгоритмы для обработки и интерпретации данных, что может значительно упростить профессиональную деятельность психологов и помочь в решении актуальных проблем.

Важность этого исследования подчеркивается необходимостью внедрения новых технологий в психологическую практику и развития инновационных подходов к решению психологических проблем больших коллективов.

Задачи данного исследования представляются следующими:

1. Изучение существующих психологических теорий и методик, применяемых для изучения проблем больших коллективов.
2. Анализ возможностей и методов интеллектуального анализа данных и машинного обучения в контексте психологического исследования.
3. Разработка подхода, который бы объединял психологические теории и методы с машинным обучением и анализом данных для более глубокого и точного понимания проблем больших коллективов.
4. Проверка эффективности предложенного подхода на реальных данных, с целью выявления сложных взаимосвязей и закономерностей внутри больших коллективов.
5. Оценка потенциала данного подхода для внедрения в практику психологического консультирования и решения актуальных проблем больших коллективов.

Исследования, в которых использовались технологии интеллектуального анализа данных для изучения психологических проблем, представлены в широком диапазоне источников. Так в статье [1] авторы использовали метод машинного обучения и анализ данных из социальных сетей для прогнозирования поведения людей и их реакций на различные события. В исследовании [2] использован анализ данных для изучения психоэмоциональных состояний на основе анализа данных с сенсоров мобильных устройств. В работах [3] – [7] большие данные применены для анализа общественной безопасности и прогнозирования социального поведения в городских условиях, предложены модели исследований в области педагогики, психологии и здравоохранения. Все эти исследования демонстрируют потенциал интеллектуального анализа данных для изучения психологических проблем и предоставления новых инсайтов.

Интеллектуальный анализ данных (Data Mining) – это процесс обнаружения полезной информации в больших объемах данных. Существует

множество методов и техник, которые используются для этого. Приведём некоторые из них.

Метод, именуемый классификацией, используется для систематизации каждого элемента в наборе данных в одну из predetermined групп или классов. Например, можно классифицировать электронные письма как «спам» или «не спам».

Кластеризация используется для группировки похожих элементов вместе. В отличие от классификации, группы или кластеры не определяются заранее.

Прогнозирование будущих значений на основе прошлых и текущих данных. Например, прогнозирование продаж на основе прошлых продаж и текущих тенденций.

Метод ассоциативных правил применяется для выявления некоторых закономерностей в больших наборах данных. Например, «покупатели, которые покупают хлеб, часто покупают и молоко».

Метод выявления аномалий может обнаружить аномальные или необычные паттерны в данных. Это может быть полезно для обнаружения мошенничества или других аномальных активностей.

Регрессионный анализ используется для моделирования и прогнозирования непрерывных переменных на основе других переменных в наборе данных.

Сетевой анализ применяется для анализа социальных сетей, веб-ссылок и других сетевых структур.

Также имеют место быть методы текстового анализа, пространственного анализа, визуализации данных, которые используются для извлечения полезной информации из текстовых, пространственных данных, а также помогают в понимании и интерпретации данных.

Рассмотрим приложение интеллектуального анализа данных при рассмотрении выборки данных, содержащей информацию (смоделирована) о психологическом состоянии студентов некоторого учебного заведения.

В качестве инструмента для анализа данных задействуем набор из библиотеки Pandas для Python.

Скрипт Python для подключения необходимых библиотек приведен ниже:

```
import numpy as np # линейная алгебра
import pandas as pd # обработка данных
%matplotlib inline
import matplotlib as mpl
import matplotlib.pyplot as plt
from matplotlib.ticker import StrMethodFormatter
import seaborn as sns
import datetime
import random
import os
```

Скрипт для прочтения выборки из csv-файла:

```
df = pd.read_csv('Student Mental health.csv')
```

Скрипт для вывода первых пяти записей из выборки:

```
df.head()
```

	Время	Пол	Возраст	Факультет	Курс	Средний балл	Семейное положение	Депрессия	Тревожность	Панические атаки	Обращения
0	08/07/2023 12:02	Ж	18	Вычислительной математики и кибернетики	Курс 1	3.00 - 3.49	Нет	Да	Нет	Да	Нет
1	08/07/2023 12:04	М	21	Философский	Курс 2	3.00 - 3.49	Нет	Нет	Да	Нет	Нет
2	08/07/2023 12:05	М	19	ИТ	Курс 1	3.00 - 3.49	Нет	Да	Да	Да	Нет
3	08/07/2023 12:06	Ж	22	юридический	Курс 3	3.00 - 3.49	Да	Да	Нет	Нет	Нет
4	08/07/2023 12:13	М	23	Механико-математический	Курс 4	3.00 - 3.49	Нет	Нет	Нет	Нет	Нет

Рисунок 1 – Фрагмент выборки (составлено авторами)

Выборка (датасет), представленная на рис. 1, содержит 404 записи со следующими полями:

- «Время» – дата проведения анкетирования студента;
- «Пол» – пол анкетизируемого;
- «Возраст» – возраст анкетизируемого;
- «Факультет» – именование факультета университета;
- «Курс» – текущий учебный год;
- «Средний балл» – совокупный средний балл успеваемости;
- «Семейное положение» – семейное положение;
- «Депрессия» – факт наличия депрессии;
- «Тревожность» – факт наличия состояния тревожности;
- «Панические атаки» – факт наличия панических атак;
- «Обращения» – факт обращения к специалисту для лечения.

Предобработка данных – это важный этап в процессе анализа данных, который включает в себя различные действия по очистке, трансформации и структурированию данных для дальнейшего использования. Этот процесс является критически важным, поскольку качество данных напрямую влияет на качество выводов, которые можно сделать на основе этих данных. Он может включать несколько основных этапов: очистка, трансформация и структурирование данных.

Предобработка данных важна по нескольким причинам. Во-первых, она помогает улучшить качество данных, что может привести к более точным и надежным результатам анализа. Во-вторых, она может помочь упростить процесс анализа данных, делая данные более легкими для понимания и обработки. Наконец, она может помочь выявить и исправить потенциальные проблемы в данных, которые могут повлиять на результаты анализа.

Скрипт для предобработки выборки приведен ниже:

```
df.nunique()  
df.isnull().sum()
```

```
df = df.dropna(how='any',axis=0)  
df.isnull().sum()
```

Для определения доминирующей возрастной категории выборки возможно выполнить построение гистограммы:

```
plt.figure(figsize=(10,10))  
plt.hist(df['Age'])  
plt.title("Возрастное распределение")  
plt.xlabel("Возраст")  
plt.ylabel("Кол-во анкетированных студентов")
```

Из рис. 2, а видно, что большинство из выборки в возрасте 18, 19 и 24 лет.

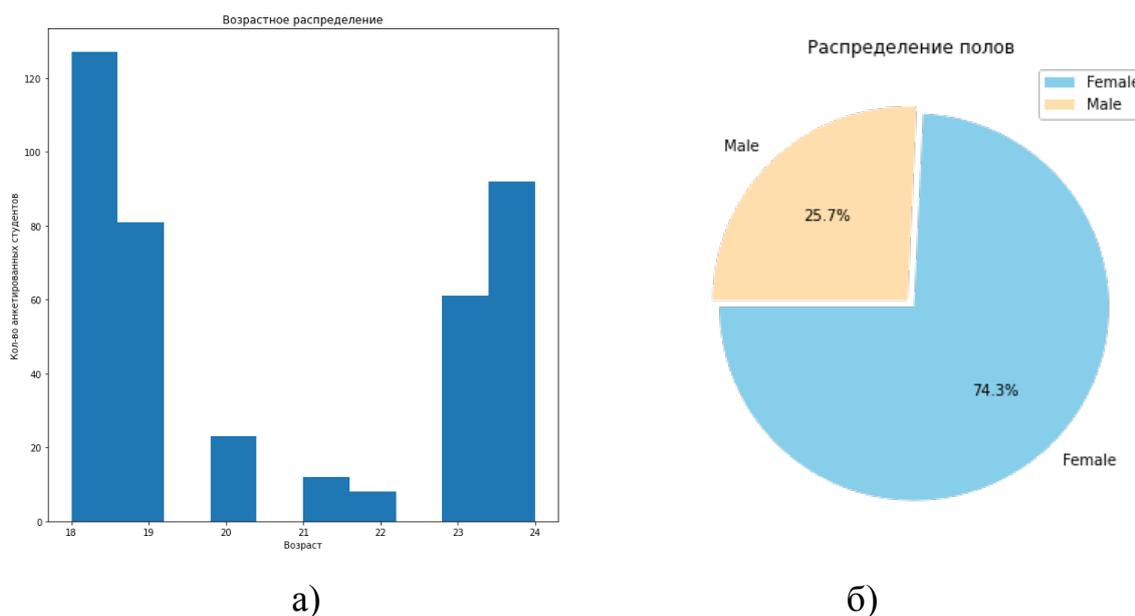


Рисунок 2 – Гистограмма возраста студентов (а) и распределение полов анкетлируемых студентов (б) (составлено авторами)

Распределение полов, анкетлируемых в процентном соотношении, позволяет оценить pie-диаграмма (рис. 2, б).

По гистограмме на рис. 3, а можно сделать вывод о том, что одновременно депрессия и тревожность не встречается у преобладающего Дневник науки | [www.dnevniknauki.ru](http://www.dnevniknauki.ru) | СМИ ЭЛ № ФС 77-68405 ISSN 2541-8327

большинства студентов.

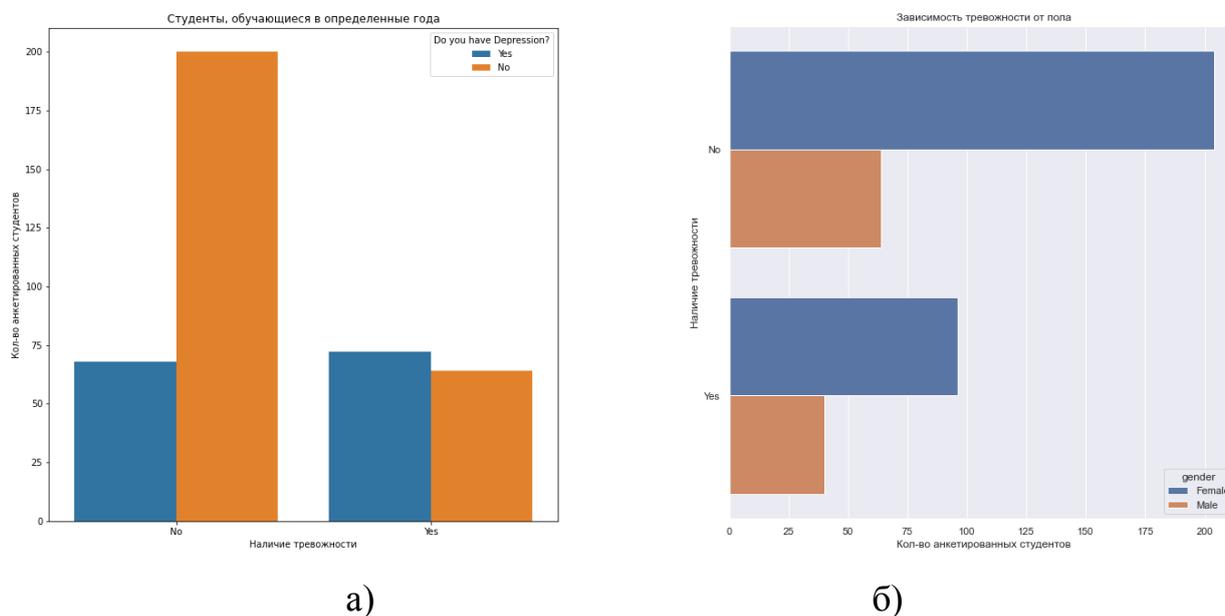


Рисунок 3 – Гистограмма тревожности и депрессии выборки (а) и зависимость тревожности от полов (б) (составлено авторами)

Гистограмма на рис. 3, б демонстрирует большую склонность к тревожности у студентов мужского пола.

С помощью аналогичных построений было определено, что, например, мужчины менее склонны к депрессии; что хорошисты и отличники более подвержены паническим атакам. Также были выявлены зависимости показателей печали, истощения, эйфории, расстройства сна, перепадов настроения, тревоги, оправданий, нервных срывов, признания ошибок, чрезмерных мыслей, агрессии, оптимизма, сексуальной активности и концентрации на профессиональный диагноз.

Таким образом, методы интеллектуального анализа данных позволяют выявлять психологические состояние коллективов, что в свою очередь может являться инструментом, например, для психологической коррекции или подбора персонала.

**Библиографический список**

1. Попова Е.П. Прогнозирование реакции пользователей в социальных сетях методами машинного обучения/Е.П. Попова, В.Н. Леоненко // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 1. № 1. С. 118–124.
2. Чернятьев А.Л. Цифровая аналитика эмоций: пилотажное исследование распознавания эмоций человека с помощью сенсоров мобильных устройств/ А.Л. Чернятьев, А.П. Лебедев // Вестник Костромского государственного университета. Серия: Педагогика. Психология. Социокинетика. 2021. Т. 27, № 4. С. 200–207. ISSN 2073-1426. <https://doi.org/10.34216/2073-1426-2021-27-4-200-207>.
3. Hashem I. A. T. и др. The role of big data in smart city 11 // International Journal of Information Management. 2016. Т. 36. № 5. С. 748-758.
4. Панов А.И. Анализ применения искусственного интеллекта в сфере безопасности // Экономика и качество систем связи. 2022. №4 (26). С. 46–53.
5. Ушаков Д.С. Искусственный интеллект как инструмент психологического исследования // Сибирский психологический журнал. 2023. № 90. С. 188–200. DOI: 10.17223/17267080/90/11.
6. Васильева Т.В. Эффективное применение алгоритмов машинного обучения для ранней диагностики нарушений обучаемости у детей дошкольного возраста/ Т.В. Васильева, Г.В. Васильев // Russian Journal of Education and Psychology. 2023. №3. С.30–44.
7. Использование методов машинного обучения при коррекции поведенческих факторов риска в профилактике сердечно-сосудистых заболеваний / Г.Г. Рапаков, Г.Т. Банщиков, В.А. Горбунов, А.В. Ударатин // Вестник Череповецкого государственного университета. 2020. № 4(97). С. 54-68.

*Оригинальность 88%*