

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра социальной информатики

**АЛГОРИТМЫ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА  
В СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ С ИСПОЛЬЗОВАНИЕМ  
SPSS**

(автореферат бакалаврской работы)

студента 4 курса 451 группы  
направления 09.03.03 - Прикладная информатика  
профиль Прикладная информатика в социологии  
Социологического факультета  
Беднова Данилы Витальевича

Научный руководитель

кандидат физико-математических наук, доцент \_\_\_\_\_ Л.Б. Тяпаев  
подпись, дата

Зав. кафедрой

кандидат социологических наук, доцент \_\_\_\_\_ И.Г. Малинский  
подпись, дата

Саратов 2024

## **ВВЕДЕНИЕ**

**Актуальность проблемы.** В настоящее время анализ данных стал неотъемлемой частью многих сфер деятельности, будь то наука, бизнес или общественные исследования. С ростом объема данных и сложности задач анализа, возникает необходимость в использовании интеллектуальных систем, способных обрабатывать и анализировать данные более эффективно и точно.

Актуальность заключается и в том, что в современном мире объем данных, которые необходимо обрабатывать и анализировать, постоянно растет. Традиционные методы статистического анализа могут быть недостаточными для обработки таких больших объемов данных. В этом контексте использование искусственного интеллекта в анализе данных становится все более актуальным. Исследование на примере SPSS Statistics дополнительно актуализирует проблему и позволяет оценить эффективность искусственного интеллекта в обработке и анализе данных, а также выявить его преимущества и ограничения. Это исследование может быть полезным для различных областей социальной жизни, таких как бизнес, наука, медицина и других, где требуется обработка больших объемов данных для принятия важных решений.

**Степень научной разработанности** проблемы велика, так как изучение возможностей эффективного применения искусственного интеллекта в анализе данных коррелирует с широкими разработками методов и алгоритмов для улучшения процесса анализа и принятия решений на основе полученных результатов. Описанию научной разработанности проблемы автор посвятил отдельную главу.

**Объект** исследования – интегрированный искусственный интеллект для анализа данных.

**Предмет** исследования – опыт использования интегрированного искусственного интеллекта для анализа данных с помощью программного обеспечения SPSS Statistics.

**Эмпирическую базу** исследования составили вторичная социологическая информация (первая глава исследования), авторское социологическое

исследование «Влияние многопользовательских онлайн игр на студентов города Саратова». Эмпирическую базу исследования дополняет первичный анализ вторичных данных с использованием логистической регрессии, дерева решений и кластеризации методом k-средних.

**Цель:** выявление возможностей интегрированного ИИ для анализа социологических данных с помощью SPSS STATISTICS.

**Задачи:**

- изучить и описать традиционные методы анализа данных;
- выявить особенности применения ИИ в анализе данных;
- выявить особенности использования SPSS STATISTICS в процессе анализа данных;
- применить на практике методы искусственного интеллекта для анализа социологических данных.

**Структура** бакалаврской работы: введение, четыре главы (по три и два параграфа соответственно), заключение, список использованных источников, четыре приложения.

## **ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ**

**Первая глава «Традиционные методы анализа данных»** посвящена основным принципам анализа данных. Рассматриваются различные методы сбора и предварительной обработки данных, такие как опросы, эксперименты, наблюдения и т.д. Также в этой главе представлено социологическое исследование на тему влияния многопользовательских онлайн игр на студентов города Саратов. Данные могут быть получены из различных источников, таких как опросы, наблюдения, эксперименты и существующие базы данных. Важно выбирать надежные и достоверные источники, чтобы результаты исследования были валидными.

Методы сбора данных

**Опросы** – это один из наиболее распространенных методов сбора данных. Они могут быть проведены с использованием структурированных или полуструктурированных вопросов, что позволяет собрать качественные и количественные данные от респондентов.

**Наблюдения** – этот метод включает в себя систематическое наблюдение и запись явлений, как они происходят в реальной жизни. Наблюдения могут быть структурированными и неструктурированными.

**Эксперименты** – предполагают контроль переменных и наблюдение за их влиянием на исследуемый объект. Это один из наиболее надежных методов, так как позволяет установить причинно-следственные связи.

**Анализ существующих данных** – использование уже собранных данных из архивов, баз данных и предыдущих исследований.

### **Принципы обработки данных**

После сбора данных важно провести их обработку. Обработка данных включает этапы очистки, преобразования и агрегации данных.

**Очистка данных** – удаление дубликатов, исправление ошибок, заполнение пропущенных значений.

**Преобразование данных** – перевод данных в удобный формат для анализа. Это может включать нормализацию, стандартизацию, категоризацию данных.

**Агрегация данных** – объединение данных для получения обобщенной информации, например, суммирование, вычисление средних значений и медиан.

Методы дескриптивной статистики.

Дескриптивная статистика включает методы, которые помогают описывать и визуализировать данные. К основным показателям дескриптивной статистики относятся:

**Медиана** – значение, которое находится посередине распределения.

**Мода** – значение, которое наиболее часто встречается в выборке.

**Среднее арифметическое** – сумма всех значений, деленная на их количество.

**Дисперсия** – мера вариативности данных.

**Графическое представление данных** – использование графиков, диаграмм и гистограмм для визуализации распределения данных.

В рамках первой главы я также провел социологическое исследование на тему влияния многопользовательских онлайн игр на студентов города Саратов.

Предметом исследования являются особенности социализации молодежи через многопользовательские онлайн игры.

Объектом исследования является социологический феномен онлайн игр.

Целью исследования было выяснить, как часто студенты играют в онлайн игры, какие жанры предпочитают, и как игры влияют на их успеваемость и социальную жизнь.

**Во второй главе «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ЕГО ПРИМЕНЕНИЕ В АНАЛИЗЕ ДАННЫХ»** рассмотрен искусственный интеллект и его применение как в анализе данных в целом, так и в социологии в частности. Искусственный интеллект — это область науки, изучающая создание интеллектуальных систем, способных воспринимать, обрабатывать и анализировать информацию, а также принимать решения на основе этих данных. Мы рассмотрели различные методы и технологии искусственного интеллекта, такие как машинное обучение, нейронные сети, генетические алгоритмы и т.д. Также — методы искусственного интеллекта, применяемые в социологии. Например, логистическая регрессия или дерево решений. Особое внимание было уделено генеративным алгоритмам, которые позволяют создавать новые наборы данных на основе уже имеющихся. В исследовании использовались такие подходы, как генеративные состязательные сети (GAN) и вариационные автоэнкодеры (VAE). Эти методы позволяют генерировать данные, которые сложно отличить от реальных, что существенно повышает качество и точность анализа. В 3 разделе данной главы была проведена

генерация социологических данных с помощью генеративного искусственного интеллекта.

Выделяют следующие шаги генеративного ИИ:

1. Сбор данных.
2. Предобработка данных. Происходит очистка данных и при необходимости их преобразование для дальнейшего понимания.
3. Обучение модели. Основываясь на трансформере, модель обучается, делает прогнозы и рассчитывает разницу между фактическим и требуемым.
4. Обратное распространение и оптимизация. На этом этапе происходит завершение обучения и генератор уже может сам воссоздать новый набор данных, с схожими статистическими характеристиками.

Проанализировав результирующую таблицу, можно сделать выводы, что, во-первых, наблюдается высокая точность модели, во-вторых, это дает надежны на качественное проведение более глубокого и комплексного исследования.

**Третья глава «ОСНОВЫ SPSS STATISTICS И ЕГО ИСПОЛЬЗОВАНИЕ ДЛЯ АНАЛИЗА ДАННЫХ»** посвящена SPSS Statistics. SPSS Statistics — это программное обеспечение для статистического анализа данных, разработанное компанией IBM. Оно предоставляет широкий спектр инструментов и методов для анализа данных, включая статистические тесты, регрессионный анализ, факторный анализ, многомерный анализ и т.д. Мы изучили основные функции и возможности SPSS Statistics и рассмотрели, как интегрированный искусственный интеллект может улучшить процесс анализа данных в этой программе. Были рассмотрены такие инструменты и методы для анализа данных как:

Описательная статистика - одна из наиболее часто используемых функций SPSS Statistics. Она позволяет исследователям получить сводную информацию о распределении данных, включая среднее значение, медиану,

моду, минимальное и максимальное значения, стандартное отклонение и многое другое.

Корреляционный анализ - еще одна важная функция SPSS Statistics. Она позволяет исследователям определить степень взаимосвязи между двумя или более переменными. SPSS Statistics предоставляет различные методы для расчета коэффициентов корреляции, таких как Пирсона, Спирмена и Кендалла.

Регрессионный анализ - еще одна мощная функция SPSS Statistics. Она позволяет исследователям определить связь между зависимой переменной и одной или несколькими независимыми переменными. SPSS Statistics предоставляет различные методы для проведения регрессионного анализа, включая линейную регрессию, логистическую регрессию, множественную регрессию и многое другое.

Анализ дисперсии - еще одна важная функция SPSS Statistics. Она позволяет исследователям определить наличие статистически значимых различий между группами. SPSS Statistics предоставляет различные методы для проведения анализа дисперсии, включая однофакторный анализ дисперсии (ANOVA) и многофакторный анализ дисперсии (MANOVA).

Факторный анализ - еще одна мощная функция SPSS Statistics. Она позволяет исследователям определить скрытые факторы, которые объясняют наблюдаемые взаимосвязи между переменными. SPSS Statistics предоставляет различные методы для проведения факторного анализа, включая главные компоненты анализа, анализ главных факторов и факторный анализ с использованием метода максимального правдоподобия.

Также мы рассмотрели, как происходит импорт данных в SPSS, так как это одна из первоочередных и важных задач. Импорт данных может быть осуществлен из различных источников, это и электронные таблицы и базы данных, текстовые файлы и другие. Стоит отметить, что загрузка данных из различных источников происходит в удобном формате. Самый распространенный, безусловно, текстовый файл.

Также в данной главе были рассмотрены примеры применения интегрированного искусственного интеллекта в SPSS Statistics, такие как автоматический выбор модели, оптимизация параметров и т.д.

**В четвертой главе «АНАЛИЗ СОЦИОЛОГИЧЕСКИХ ДАННЫХ С ПРИМЕНЕНИЕМ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»** был проведен анализ данных социологического исследования из первой главы, но используя три различных метода искусственного интеллекта. В частности, были использованы: логистическая регрессия, дерево решений и кластеризация методом k-средних.

### **Примеры использования методов ИИ в социологии**

Методы искусственного интеллекта (ИИ) находят все большее применение в социологических исследованиях. Они позволяют обрабатывать большие объемы данных, выявлять скрытые закономерности и делать более точные прогнозы. Рассмотрим несколько реальных примеров использования ИИ в анализе социологических данных.

#### **Пример 1: Прогнозирование политических предпочтений**

Целью исследования было прогнозирование политических предпочтений избирателей на основе демографических и поведенческих данных. Были использованы регрессионный анализ и классификация, такие как логистическая регрессия и деревья решений. Модель случайного леса показала наилучшие результаты с точностью прогнозирования около 85%.

#### **Пример 2: Анализ настроений в социальных сетях**

Исследование общественного мнения и настроений по различным социальным вопросам на основе данных из социальных сетей было выполнено с использованием методов обработки естественного языка (NLP). Анализ показал, что большинство пользователей высказываются негативно по вопросам экономической нестабильности и позитивно по вопросам социальной справедливости.

#### **Пример 3: Оценка эффективности образовательных программ**

Для оценки влияния образовательных программ на академическую успеваемость и социальное поведение студентов были использованы рекуррентные нейронные сети (RNN). Исследование показало, что образовательные программы значительно улучшают успеваемость студентов и положительно влияют на их социальное поведение.

#### **Пример 4: Прогнозирование уровня преступности**

Использование нейронных сетей и геопространственного анализа для прогнозирования уровня преступности в городских районах на основе социально-экономических данных показало высокую точность прогнозирования. Геопространственный анализ помог выделить районы, требующие повышенного внимания правоохранительных органов.

#### **Пример 5: Анализ данных опросов с помощью кластерного анализа**

Для выявления скрытых сегментов в данных опросов населения применялись алгоритмы кластеризации, такие как K-средних и иерархическая кластеризация. Это позволило выделить несколько ключевых сегментов населения, различающихся по своим предпочтениям, поведению и социальным установкам.

#### **Применение методов ИИ для анализа социологических данных**

В данной главе я рассмотрел применение методов искусственного интеллекта для анализа социологических данных. Были использованы три метода: логистическая регрессия, дерево решений и кластеризация методом k-средних. Целью являлось сравнение результатов, полученных с помощью этих методов, с результатами, полученными в ходе построения частотных таблиц.

#### **Логистическая регрессия**

Логистическая регрессия используется для предсказания вероятности принадлежности наблюдения к одному из нескольких классов. В данном случае целью являлось предсказание влияния игр на жизнь респондентов (положительное, нейтральное или отрицательное).

#### **Дерево решений**

Дерево решений используется для классификации данных на основе правил, сформированных из признаков. Оно позволяет визуализировать процесс принятия решений и выявлять ключевые факторы, влияющие на результаты.

### **Кластеризация методом k-средних**

Кластеризация методом k-средних используется для разделения данных на группы, в которых наблюдения внутри группы имеют больше сходств друг с другом, чем с наблюдениями из других групп. Этот метод оказался самым информативным для определения групп респондентов с различными характеристиками.

## **ЗАКЛЮЧЕНИЕ**

В заключение данной работы можно сделать следующие выводы.

В ходе исследования были рассмотрены основные принципы анализа данных, искусственный интеллект и его применение в анализе данных, а также проведен анализ социологического опроса как методами искусственного интеллекта так и традиционными методами анализа такими как таблицы частот, корреляционный анализ и т.д.

Основными принципами анализа данных являются сбор, обработка, интерпретация и представление данных. Сбор данных включает в себя различные методы, такие как опросы, эксперименты и наблюдения. Обработка данных включает в себя их очистку, трансформацию и агрегацию. Интерпретация данных позволяет выявить закономерности, тенденции и взаимосвязи между переменными. Представление данных позволяет визуализировать результаты анализа и делать выводы.

Искусственный интеллект (ИИ) является областью компьютерных наук, которая изучает создание интеллектуальных систем, способных выполнять задачи, требующие интеллектуальных способностей человека. ИИ имеет широкий спектр применений в анализе данных, включая машинное обучение, обработку естественного языка, компьютерное зрение и автоматическое

планирование. Он позволяет автоматизировать и ускорить процесс анализа данных, а также обнаруживать скрытые закономерности и делать прогнозы.

Методы искусственного интеллекта показали, что они могут быть полезными для углубленного анализа социологических данных. K-Means кластеризация оказалась наиболее информативной для выявления групп респондентов с различными характеристиками. Логистическая регрессия и дерево решений показали средние результаты точности, однако их можно улучшить с помощью дальнейшей настройки моделей и использования дополнительных данных.

В заключение, интегрированный искусственный интеллект для анализа на примере сравнения с SPSS Statistics является мощным инструментом, который позволяет автоматизировать и ускорить процесс анализа данных, делать более точные выводы и прогнозы. Он также позволяет пользователям сосредоточиться на сущности данных и принимать обоснованные решения на основе анализа. Искусственный интеллект и его применение в анализе данных играют все более важную роль в современном обществе, и их развитие будет продолжаться в будущем.