МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ КЫРГЫЗСКОЙ РЕСПУБЛИКИ

КЫРГЫЗСКО-ГЕРМАНСКИЙ ИНСТИТУТ ПРИКЛАДНОЙ ИНФОРМАТИКИ

**КУРСОВОЙ ПРОЕКТ**

на тему “Разработка системы прогнозирования риска семейного насилия в регионах Кыргызстана с использованием методов машинного обучения”

Выпонили: Студенты группы AIN-1-22

Кубанычбек уулу А.

Тахирова К. ­­­­­­­­­­­­­­­­

Проверили:

Старший преподаватель

Токтоналы А.

Консультант по немецкому языку:

Доцент КГИПИ

Ланге А.

Содержание

[Введение 3](#_Toc194770332)

[Глава 1. Теоретические основы прогнозирования семейного насилия 5](#_Toc194770333)

[1.1 Анализ предметной области: семейное насилие и его социальные последствия 5](#_Toc194770334)

[1.2 Методы прогнозирования рисков в социальных исследованиях 5](#_Toc194770335)

[1.3 Обзор существующих моделей и алгоритмов машинного обучения для прогнозирования рисков 6](#_Toc194770336)

[1.4 Анализ существующих информационных систем прогнозирования рисков семейного насилия 7](#_Toc194770337)

[Глава 2. Разработка системы прогнозирования риска семейного насилия 8](#_Toc194770338)

[2.1 Анализ существующих данных: источники, формат, предобработка 8](#_Toc194770339)

[2.2 Выбор инструментов и методов машинного обучения 8](#_Toc194770340)

[2.3 Разработка модели прогнозирования: обучение, тестирование и оценка качества 8](#_Toc194770341)

[2.4 Визуализация результатов и интерпретация прогнозов 8](#_Toc194770342)

[2.5 Оценка эффективности системы и возможные 8](#_Toc194770343)

[Заключение 9](#_Toc194770344)

[Список использованных источников 10](#_Toc194770345)

[Приложение 1. Код программы 12](#_Toc194770346)

# **Введение**

Семейное насилие остается острой проблемой, затрагивающей тысячи семей в Кыргызстане. Несмотря на существующие законы и меры поддержки, число случаев домашнего насилия остается высоким. Одной из причин является отсутствие инструментов для раннего выявления потенциальных угроз. Использование методов машинного обучения для прогнозирования риска семейного насилия позволит более эффективно анализировать факторы риска и разрабатывать превентивные меры.

Актуальность данной темы обусловлена необходимостью создания интеллектуальных систем, способных анализировать большие объемы данных и выявлять паттерны, указывающие на высокий риск насилия. Современные технологии, включая алгоритмы машинного обучения, позволяют автоматизировать процесс анализа и прогнозирования, что делает систему точной и надежной. Для достижения цели проекта необходимо выполнить следующие задачи:

1. Провести анализ предметной области [13, 16, 20], изучить существующие подходы к прогнозированию риска семейного насилия.
2. Исследовать методы обработки данных, включая сбор, предобработку и анализ факторов риска.
3. Разработать архитектуру системы прогнозирования, определить функциональные требования и выбрать подходящий стек технологий.
4. Реализовать модель машинного обучения для прогнозирования риска семейного насилия.
5. Провести тестирование модели и оценить ее точность, надежность и применимость в реальных условиях.

Современные технологии, такие как Python и Jupyter Notebook [1, 3], предоставляют эффективные инструменты для анализа данных, включая обработку, прогнозирование и визуализацию. Объект исследования данной работы являются проблема прогнозирования риска семейного насилия с использованием методов машинного обучения, а предметом – методы машинного обучения и их применение для прогнозирования риска семейного насилия в регионах Кыргызстана.

Содержание работы структурировано следующим образом:

* Введение. Содержит обоснование актуальности темы курсового проекта, формулировку цели и задач исследования, а также вклад каждого участника команды
* Глава 1. Теоретические основы прогнозирования семейного насилия. Анализируется предметная область, методы прогнозирования рисков, обработки данных и существующие инструменты, модели.
* Глава 2. Разработка системы прогнозирования риска семейного насилия. Описывается проектирование системы, архитектура, функциональные требования и выбранный стек технологий. Представлены этапы создания продукта, подготовка данных, разработка алгоритмов и результаты тестирования.
* Заключение. Подводятся итоги, формулируются выводы и рекомендации.
* Список использованных источников. Содержит перечень научных, технических и литературных источников, которые использовались в процессе выполнения курсового проекта.
* Приложение. Представлены дополнительные материалы: исходный код, визуализации и результаты тестирования.

Работа над проектом была организована в команде из двух человек, каждый из которых выполнял конкретные задачи и вносил свой вклад:

* Кубанычбек уулу Азиз руководил командой, принимал важные технические решения, вел проектную документацию, а также разрабатывал код, как видно из разделов 2.3 и 2.4.
* Тахирова Кыз Сайкал отвечала за качество и надежность программы, вела проектную документацию, писала код, как показано в разделах 2.3 и 2.4.

# **Глава 1. Теоретические основы прогнозирования семейного насилия**

## **Анализ предметной области: семейное насилие и его социальные последствия**

Семейное насилие является одной из наиболее острых социальных проблем, оказывающих влияние на физическое и психологическое состояние жертв. Оно включает в себя различные формы насилия, такие как физическое, психологическое, экономическое и сексуальное.

Факторы риска семейного насилия разнообразны и могут включать:

* **Экономические**: бедность, безработица, финансовая зависимость жертвы.
* **Социальные**: низкий уровень образования, алкогольная или наркотическая зависимость агрессора, культурные традиции, поддерживающие насилие.
* **Психологические**: детские травмы, перенесенный опыт насилия, психические расстройства.
* **Юридические**: слабая правоприменительная практика, недостаточная защита жертв со стороны государства.

Семейное насилие в Кыргызстане остается серьезной проблемой, несмотря на предпринимаемые меры. Согласно данным Министерства внутренних дел [13], в 2023 году за 7 месяцев было зарегистрировано 8956 случаев насилия в семье. Для борьбы с насилием были приняты законодательные меры, включая **Закон «О защите от семейного насилия»**, однако их эффективность остается недостаточной. Основные проблемы включают:

* Недостаточное финансирование кризисных центров.
* Низкий уровень доверия жертв к правоохранительным органам.
* Ограниченные возможности раннего выявления угроз.

## **Методы прогнозирования рисков в социальных исследованиях**

Прогнозирование рисков семейного насилия требует применения современных методов анализа данных. Классические статистические методы, такие как линейная и логистическая регрессия, позволяют выявить зависимости между факторами и вероятностью наступления события. Однако для работы со сложными и многомерными данными более подходящими являются методы машинного обучения. Деревья решений, ансамбли моделей (бэггинг, бустинг, стекинг) и нейронные сети способны обрабатывать большие объемы данных и выявлять скрытые закономерности. Эти методы особенно полезны при работе с неструктурированными данными, такими как текстовые заявления или медицинские записи. Они способны обнаруживать сложные нелинейные зависимости, однако требуют значительных вычислительных ресурсов и больших объемов данных для обучения.

## **Обзор существующих моделей и алгоритмов машинного обучения для прогнозирования рисков**

Среди алгоритмов машинного обучения наиболее популярными для прогнозирования рисков являются случайный лес (Random Forest), градиентный бустинг (XGBoost) и ансамблевые методы, такие как стекинг. Случайный лес хорошо справляется с задачами классификации и регрессии благодаря своей простоте и интерпретируемости. Он строит множество деревьев решений на основе случайных подвыборок данных и усредняет их результаты, что снижает риск переобучения. Этот метод особенно полезен при работе с категориальными данными, такими как тип насилия или пол агрессора.

XGBoost отличается высокой точностью и скоростью обучения, что делает его одним из лучших выборов для работы с большими датасетами. Алгоритм использует градиентный бустинг, последовательно улучшая качество модели за счет корректировки ошибок предыдущих итераций. XGBoost также позволяет настраивать множество гиперпараметров, таких как скорость обучения, глубина деревьев и количество деревьев, что обеспечивает гибкость в решении различных задач.

Стекинг представляет собой ансамблевый метод, который комбинирует прогнозы нескольких моделей с использованием метамодели для финального решения. Этот подход позволяет объединить преимущества разных алгоритмов, повышая общее качество прогнозов.

## **Анализ существующих информационных систем прогнозирования рисков семейного насилия**

Традиционные методы прогнозирования риска семейного насилия, такие как социологические исследования, опросы и экспертные оценки, обладают рядом ограничений. Они часто субъективны, требуют значительных временных затрат и не позволяют оперативно обрабатывать большие объемы данных. В мировой практике уже существуют успешные примеры подобных систем. Например, в США применяется система прогнозирования риска на основе полицейских отчетов, анализирующая повторные вызовы, жалобы и другие индикаторы насилия. В Великобритании используются аналитические платформы, которые идентифицируют домохозяйства с повышенным риском семейного насилия на основе социальных, экономических и правовых данных.

Применение машинного обучения в прогнозировании риска семейного насилия имеет значительные преимущества, такие как:

* **Высокая точность прогнозирования** за счет анализа множества факторов и выявления скрытых закономерностей.
* **Автоматизация процесса выявления угроз**, что позволяет оперативно реагировать на потенциальные случаи насилия.
* **Обработка больших объемов данных** из различных источников (социальные службы, правоохранительные органы, медицинские учреждения), что невозможно при традиционных методах.

Использование современных технологий в прогнозировании семейного насилия позволит создать эффективную систему раннего выявления угроз, что в перспективе приведет к снижению уровня насилия, повышению защиты жертв и улучшению работы социальных служб.

# **Глава 2. Разработка системы прогнозирования риска семейного насилия**

* 1. **Анализ существующих данных: источники, формат, предобработка**
  2. **Выбор инструментов и методов машинного обучения**
  3. **Разработка модели прогнозирования: обучение, тестирование и оценка качества**
  4. **Визуализация результатов и интерпретация прогнозов**
  5. **Оценка эффективности системы и возможные направления её улучшения**

# **Заключение**

3стр

# **Список использованных источников**

Учебная и научная литература

1. Project Jupyter Documentation. Официальная документация инструмента Jupyter. Ссылка: <https://docs.jupyter.org/en/latest/>
2. Microsoft. Train a model in Python with automated machine learning. Ссылка: [https://learn.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-azure-machine-learning-tutorial](https://learn.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-azure-machine-learning-tutorial%20). Опубликовано 3 июля 2024г.
3. Python Documentation. Ссылка: [https://docs.python.org/3/](https://docs.python.org/3/%20)
4. Deep Learning. An MIT Pressbook / Ian Goodfellow, Yoshua Bengio, Aaron Courville. 773с. Ссылка: <https://www.deeplearningbook.org/>
5. Машинное обучение без лишних слов / Андрей Бурков. 2020 г. – 192с.

Интернет-ресурсы

1. «В Кыргызстане увеличилось количество случаев домашнего насилия» [Электронный ресурс]. – Режим доступа: [www.vb.kg/doc/437760\_v\_kyrgyzstane\_yvelichilos\_kolichestvo\_slychaev\_domashnego\_nasiliia.html](http://www.vb.kg/doc/437760_v_kyrgyzstane_yvelichilos_kolichestvo_slychaev_domashnego_nasiliia.html). Опубликовано 18 июня 2024 8:57
2. «Модели машинного обучения» [Электронный ресурс]. – Режим доступа: [yandex.cloud/ru/docs/glossary/ml-models?utm\_referrer=https%3A%2F%2Fwww.google.com%2F](https://yandex.cloud/ru/docs/glossary/ml-models?utm_referrer=https%3A%2F%2Fwww.google.com%2F)
3. «В Кыргызстане выросло количество случаев семейного насилия» [Электронный ресурс]. – Режим доступа: [kaktus.media/doc/508966\_v\_kyrgyzstane\_vyroslo\_kolichestvo\_slychaev\_semeynogo\_nasiliia.html](https://kaktus.media/doc/508966_v_kyrgyzstane_vyroslo_kolichestvo_slychaev_semeynogo_nasiliia.html). Опубликовано 10 сентября 2024 10:11
4. Курс «Машинное обучение» [Электронный ресурс]. – Режим доступа: [www.dmitrymakarov.ru/](http://www.dmitrymakarov.ru/)
5. Open Data Kyrgyzstan [Электронный ресурс]. – Режим доступа: [data.gov.kg/en/dataset/othet-no-jinuam-cobepwnbwnm-cemenhoe-hacnjine-3a-2022-rod](https://data.gov.kg/en/dataset/othet-no-jinuam-cobepwnbwnm-cemenhoe-hacnjine-3a-2022-rod)
6. «Что такое «закон Салтанат» и как наказывают за семейное насилие в Кыргызстане» [Электронный ресурс]. – Режим доступа: [factcheck.kg/chto-takoe-zakon-saltanat-i-kak-nakazyvayut-za-semejnoe-nasilie-v-kyrgyzstane/](https://factcheck.kg/chto-takoe-zakon-saltanat-i-kak-nakazyvayut-za-semejnoe-nasilie-v-kyrgyzstane/)
7. Информация по семейному насилию [Электронный ресурс]. – Режим доступа: [mvd.gov.kg/rus/domesticViolence/reports/34](https://mvd.gov.kg/rus/domesticViolence/reports/34)
8. Верховный суд КР [Электронный ресурс]. – Режим доступа: [sot.kg/statistics/main/index.php](https://sot.kg/statistics/main/index.php)
9. Национальный статический комитет КР [Электронный ресурс]. – Режим доступа: [stat.gov.kg/ru/opendata/category/13/](https://stat.gov.kg/ru/opendata/category/13/)
10. «Кыргызстан: Семейное насилие в отношении женщин и девочек с инвалидностью» [Электронный ресурс]. – Режим доступа: [www.hrw.org/ru/news/2023/12/14/kyrgyzstan-domestic-violence-against-women-girls-disabilities](http://www.hrw.org/ru/news/2023/12/14/kyrgyzstan-domestic-violence-against-women-girls-disabilities)

# **Приложение 1. Код программы**