

УДК 550.82

А.А. Иванов¹, lepri4dw@gmail.com

А.А. Бакиров¹, amir.bakirov@gmail.com

А.Т. Асанбеков¹, ask.net.ua@gmail.com

Ю.С. Корякина², jk169@list.ru

¹Кыргызско-Германский Институт Прикладной Информатики

²Институт Машиноведения Автоматики и Геомеханики НАН КР

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ПРИ РАЗРАБОТКЕ ЦИФРОВЫХ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО РАЗМЕЩЕНИЮ ШКОЛ В ГОРОДСКОЙ СРЕДЕ

В данной статье рассматривается разработка цифровой системы поддержки принятия решений по размещению новых школ в городской среде. Рассмотрены источники открытых данных о населении и образовательных учреждениях, методы анализа плотности населения и загруженности школ, а также подходы к пространственной визуализации данных на карте. Проанализированы существующие методы определения потребности в образовательной инфраструктуре. На основе проведенного анализа предложена модель определения приоритетных зон для строительства новых школ с использованием методов анализа данных и элементов машинного обучения. Разработанная система позволяет формировать рекомендации по размещению образовательных учреждений на основе анализа городской среды.

Ключевые слова: машинное обучение, СППР, городские данные, размещение школ, образовательная инфраструктура, ИСУО, анализ загруженности школ, плотность населения, цифровая карта.

Введение

В условиях стремительной урбанизации и пространственного расширения современных городов вопрос равномерного и эффективного размещения объектов социальной инфраструктуры становится одной из ключевых задач градостроительного управления. Особенно остро данная проблема проявляется в сфере школьного образования, поскольку несбалансированное размещение общеобразовательных учреждений приводит к перегруженности отдельных школ, снижению качества образовательного процесса, увеличению расстояния между местом проживания учащихся и школой, а также росту ежедневной маятниковой миграции, что дополнительно влияет на транспортную систему города [1].

Однако в условиях динамичного роста населения и активного изменения городской застройки традиционные подходы к территориальному планированию не всегда позволяют своевременно реагировать на возникающие потребности. Методы, основанные на статических генеральных планах, нормативных радиусах доступности и экспертной оценке, зачастую не учитывают в полной мере демографические изменения, внутреннюю миграцию и точечную застройку. В результате решения о строительстве новых школ могут приниматься на основе разрозненных или устаревших данных, без достаточного учета будущего роста населения и фактической загруженности существующих образовательных учреждений [2].

Следствием этого является то, что процесс выбора мест для строительства новых школ остается длительным и трудоемким. На практике градостроителям и сотрудникам муниципальных органов необходимо вручную сопоставлять различные источники информации: демографическую статистику, карты жилой застройки, сведения о проектной вместимости школ, фактическом количестве учащихся и пространственной доступности образовательных учреждений. Такой подход требует значительных временных затрат и усложняет подготовку объективного аналитического обоснования для принятия решений.

С практической точки зрения данная проблема проявляется в том, что планирование и обоснование строительства новых общеобразовательных учреждений в городской среде может занимать от 3 до 12 месяцев в зависимости от доступности данных, уровня цифровизации и сложности рассматриваемой территории.

Учитывая указанные ограничения, возникает необходимость перехода от преимущественно ручных и экспертных методов планирования к цифровым системам поддержки принятия решений. Такие системы должны обеспечивать интеграцию разнородных данных, выполнение пространственного анализа и формирование аналитических рекомендаций для специалистов. В отличие от традиционного подхода, цифровая система поддержки принятия решений (СППР) позволяет рассматривать городскую среду не только с точки зрения текущего состояния образовательной инфраструктуры, но и с учетом прогнозируемого роста населения, изменения плотности застройки и возможного дефицита учебных мест.

Соответственно, ожидаемым эффектом от применения предлагаемой методологии является существенное сокращение времени, необходимого для проведения градостроительного анализа. За счет автоматизации ключевых этапов - агрегации демографических показателей, оценки загруженности существующих школ, анализа пространственной доступности и расчета локального дефицита мест - становится возможным переход от длительного ручного анализа к более оперативной и воспроизводимой процедуре подготовки решений. В перспективе использование проектируемой СППР может сократить аналитический цикл, традиционно занимающий несколько месяцев, до 10–12 минут.

Технологической основой для решения данной задачи выступает развитие так называемой концепции «Умного города» и доступность больших объемов городских данных. Эти факторы открывают новые возможности для применения геоинформационных технологий, методов машинного обучения и элементов искусственного интеллекта в задачах градостроительного планирования. Использование прогнозных моделей машинного обучения в сочетании с демографическими расчетами позволяет не только фиксировать текущую нехватку школьных мест, но и предиктивно оценивать будущий спрос на образовательную инфраструктуру в разрезе отдельных районов и микрорайонов города.

Таким образом, актуальность исследования определяется необходимостью формирования научно обоснованной методологии построения цифровой системы поддержки принятия решений для анализа загруженности школ, прогнозирования дефицита учебных мест и определения приоритетных зон размещения новых общеобразовательных учреждений в городской среде. В рамках данного подхода акцент делается не на описании уже полностью разработанной системы, а на обосновании предлагаемой методологии и архитектуры будущей СППР, которая может быть использована для повышения оперативности и качества градостроительных решений.

Цель данного исследования описание и оценка эффективности разрабатываемой цифровой системы поддержки принятия решений для предиктивного анализа дефицита школьных мест и определения приоритетных зон размещения новых школ, а также ее сравнительный анализ с существующими подходами к градостроительному планированию.

Объектом исследования выступают цифровые системы поддержки принятия решений (СППР) в сфере градостроительного планирования и развития инфраструктуры.

Предметом исследования являются методы машинного обучения и алгоритмы пространственного моделирования, интегрированные в архитектуру данных систем для оценки и прогнозирования дефицита школьных мест.

В качестве тестового полигона для апробации разработанной системы выбран город Бишкек (Кыргызская Республика). Выбор обусловлен высокими темпами внутренней миграции, интенсивным жилищным строительством и наличием ярко выраженных зон дефицита образовательной инфраструктуры. Интеграция официальных данных Национального статистического комитета КР, данных информационных систем управления образованием (ИСУО) [5] и открытых геоданных (OpenStreetMap) позволяет сформировать релевантную базу для проведения исследования.

Для достижения поставленной цели в рамках работы решаются следующие задачи:

1. Анализ современных подходов к геопространственному прогнозированию потребности в объектах социальной инфраструктуры.

2. Сбор, интеграция и предобработка демографических и картографических данных города Бишкек.
3. Разработка архитектуры СППР, включающей ML-модели (в частности, алгоритмы градиентного бустинга) для прогнозирования локального дефицита школьных мест.
4. Формирование рекомендательной подсистемы генерации координат для новых школ с использованием аналитических слоев и технологий искусственного интеллекта.
5. Апробация разработанного веб-ориентированного комплекса и оценка полученных результатов.

Анализ существующих систем планирования образовательной инфраструктуры

Для анализа образовательной инфраструктуры и планирования размещения школ используются различные информационные системы, основанные на обработке статистических данных и геоинформационных технологиях. Рассмотрим несколько распространённых решений: OpenEMIS, School Mapping System и Frontline Location Analytics.

OpenEMIS представленная на рисунке 1 - система управления образовательной информацией (Education Management Information System), [6] разработанная при поддержке UNESCO. Система используется в различных странах для централизованного сбора и анализа статистических данных в сфере образования. Основная задача системы - хранение и обработка информации о школах, учащихся, преподавателях и образовательных ресурсах.

Основные функции:

- сбор и хранение данных об образовательных учреждениях;
- анализ образовательной статистики;
- формирование аналитических отчетов и панелей мониторинга.

Недостатки:

- отсутствует пространственный анализ образовательной инфраструктуры;
- система не предназначена для определения мест строительства новых школ.

Таким образом, OpenEMIS используется преимущественно для анализа образовательной статистики и мониторинга состояния образовательной системы.

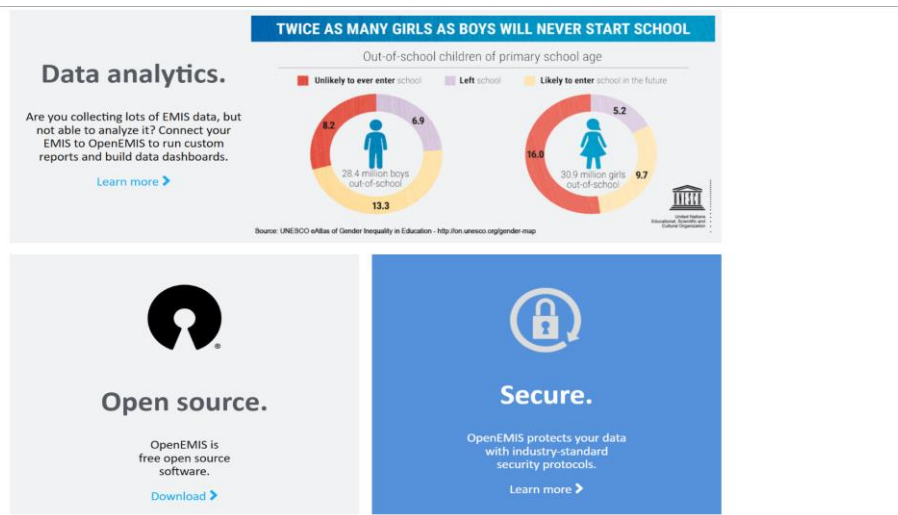


Рисунок 1 – Информационный блок с сайта системы OpenEMIS, демонстрирующий возможности аналитики данных

School Mapping System представленная на рисунке 2 представляет собой геоинформационную систему (GIS) [7], используемую для анализа расположения образовательных учреждений и планирования образовательной инфраструктуры. Такие системы отображают школы на карте и позволяют анализировать пространственное распределение образовательных учреждений.

Основные функции:

- отображение школ на интерактивной карте;
- анализ расстояния до ближайших школ;

- определение зон обслуживания школ;
- выявление районов с недостаточным количеством школ.

Недостатки:

- ограниченные аналитические возможности;
- решения о размещении новых школ принимаются пользователями вручную.

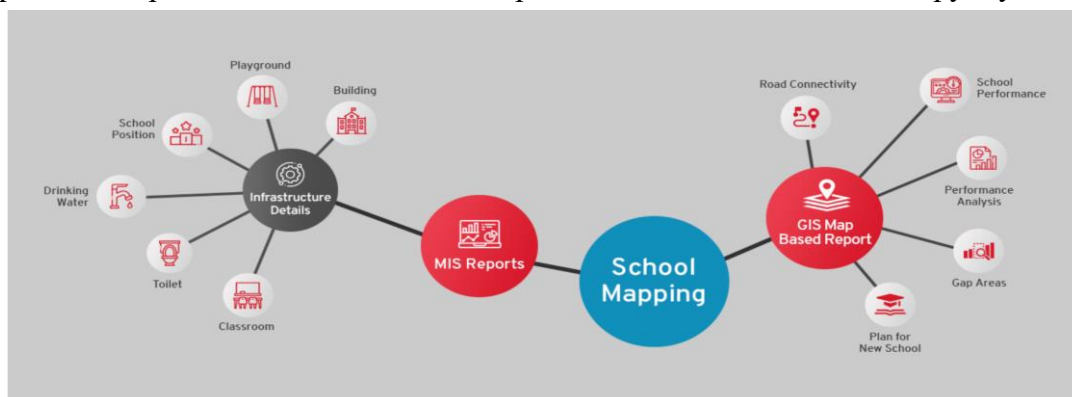


Рисунок 2 – Концептуальная схема работы School Mapping System

Frontline Location Analytics представленная на рисунке 3 - коммерческая аналитическая система, используемая школьными округами США для анализа образовательной инфраструктуры. Система объединяет демографические данные и данные о школах и позволяет проводить пространственный анализ распределения учащихся.

Основные функции:

- анализ демографических данных районов;
- прогноз количества учащихся;
- моделирование сценариев изменения школьных границ;
- анализ загрузки школ.

Недостатки:

- высокая стоимость внедрения;
- решения о строительстве школ принимаются администраторами на основе анализа данных.

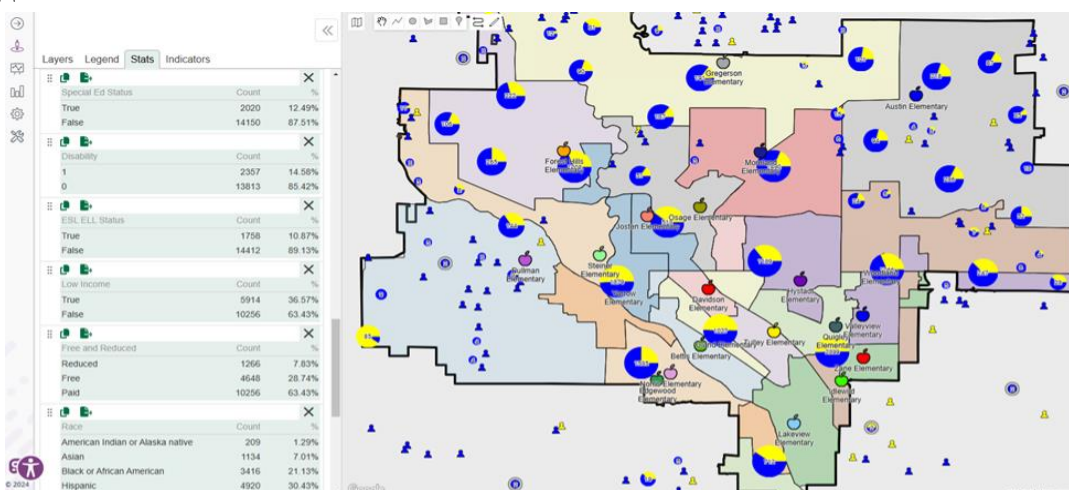


Рисунок 3 – Пример интерфейса Frontline Location Analytics с картографической визуализацией

Для более наглядного сопоставления рассмотренных решений целесообразно обобщить их основные функции и ограничения в сравнительной таблице 1. Такое сравнение позволяет определить, какие задачи уже решаются существующими системами, а какие остаются недостаточно автоматизированными. Особое внимание уделяется наличию пространственного анализа, прогнозирования и автоматического формирования рекомендаций по размещению новых образовательных учреждений.

Таблица 1 - Сравнение существующих систем анализа образовательной инфраструктуры

Система	Основные функции	Ограничения
OpenEMIS	Сбор и анализ образовательной статистики	Нет пространственного анализа инфраструктуры
School Mapping System	Отображение школ на карте, анализ доступности	Нет интеллектуального анализа
Frontline Location Analytics	Демографический анализ и прогнозирование загрузки школ	Решения принимаются пользователями, ограничен рынком США
Разрабатываемая система	Анализ плотности населения и загруженности школ, формирование рекомендаций по размещению новых школ	Требуется расширение данных и дальнейшее развитие

Проведённый анализ существующих систем показал, что большинство решений в мире ориентировано на сбор и анализ образовательной статистики или визуализацию инфраструктуры на карте. Такие системы, как **OpenEMIS** представленная на рисунке 1, используются для управления образовательными данными на уровне государства, а системы класса **School Mapping** представленная на рисунке 2 позволяют анализировать территориальное распределение школ с использованием геоинформационных технологий. Коммерческие решения, например **Frontline Location Analytics** представленная на рисунке 3, предоставляют расширенные инструменты аналитики и прогнозирования демографических показателей.

Однако рассмотренные системы в основном выполняют функции анализа и визуализации данных представленная на таблице 1, а решения о размещении новых школ принимаются пользователями на основе полученной информации. Автоматическое формирование рекомендаций по размещению образовательной инфраструктуры в таких системах, как правило, отсутствует.

Следует отметить, что в **Кыргызской Республике на данный момент отсутствуют цифровые системы**, предназначенные для анализа плотности населения и загруженности школ с целью определения приоритетных зон строительства новых образовательных учреждений. В связи с этим разработка цифровой системы формирования рекомендаций по размещению школ на основе анализа городских данных является актуальной задачей и может повысить эффективность планирования образовательной инфраструктуры.

Материалы и методы

Общая архитектура системы и источники данных

Разрабатываемая система поддержки принятия решений (СППР) представленная на рисунке 5 представляет собой веб-ориентированный программно-аналитический комплекс, архитектура которого основана на интеграции трех типов данных: геопро пространственных, образовательных и демографических.

В качестве источника картографической информации использовалась база OpenStreetMap, из которой с помощью API Nominatim и Overpass были получены административные границы Бишкека, полигоны застройки и координаты школ.

Оценка загрузки образовательной инфраструктуры проводилась на основе данных ИСУО в формате XML, включающих проектную вместимость школ, фактическое число учащихся и структуру классов. Демографическая база сформирована на основе данных Национального статистического комитета Кыргызской Республики, включая численность

населения, показатели прироста и возрастную структуру. Расчет плотности населения был откалиброван с учетом данных переписи, что повысило точность пространственного анализа.

Существующие методологии разработки систем поддержки принятия решений

Разработка цифровых систем поддержки принятия решений (СППР), ориентированных на анализ городской инфраструктуры и прогнозирование, основывается на сочетании методологий системного проектирования, обработки данных и машинного обучения. В современной практике выделяется несколько ключевых подходов, применяемых при создании подобных систем.

Одним из фундаментальных подходов является системная инженерия (Systems Engineering) представленная на рисунке 4, предполагающая рассмотрение системы как совокупности взаимосвязанных компонентов: источников данных, аналитических модулей, пользовательского интерфейса и механизмов интеграции. Этот подход широко применяется при разработке геоинформационных и аналитических систем, обеспечивая целостность и согласованность архитектуры.

Важную роль играет концепция Data-Driven Development, в рамках которой процесс разработки строится вокруг сбора, очистки, интеграции и анализа данных. Это особенно актуально для систем, работающих с демографической и геопространственной информацией.

Для построения аналитических моделей широко применяется методология CRISP-DM (Cross-Industry Standard Process for Data Mining), включающая этапы понимания задачи, подготовки данных, моделирования, оценки и внедрения. Она обеспечивает структурированность и воспроизводимость результатов.

В контексте геоинформационных систем используется пространственно-ориентированный подход (Spatial Data Analysis), направленный на обработку геоданных, анализ плотности и оценку доступности объектов, что важно для задач территориального планирования.

Таким образом, разработка систем поддержки принятия решений в сфере городского планирования требует применения комбинированного методологического подхода, включающего:

- системную инженерию для проектирования архитектуры;
- data-driven подход для работы с данными;
- CRISP-DM для построения моделей машинного обучения;
- пространственный анализ для обработки геоданных.

Использование совокупности данных методологий позволяет обеспечить комплексность анализа, повысить точность прогнозирования и создать эффективный инструмент поддержки принятия решений.

Для обоснования разрабатываемой системы поддержки принятия решений необходимо сначала рассмотреть различие между существующей методологией анализа образовательной инфраструктуры и предлагаемой методологией СППР. Существующий подход, как правило, ориентирован на сбор данных, их базовую обработку, построение отчетов и визуализацию текущего состояния. Однако такой подход в меньшей степени учитывает прогнозирование, автоматизированное формирование рекомендаций и сценарный анализ.

В предлагаемой методологии акцент сделан на более глубокой интеграции геопространственных, образовательных и демографических данных, а также на применении методов машинного обучения для прогнозирования дефицита ученических мест. Сравнение существующей методологии и предлагаемого подхода представлено на рисунке 4.

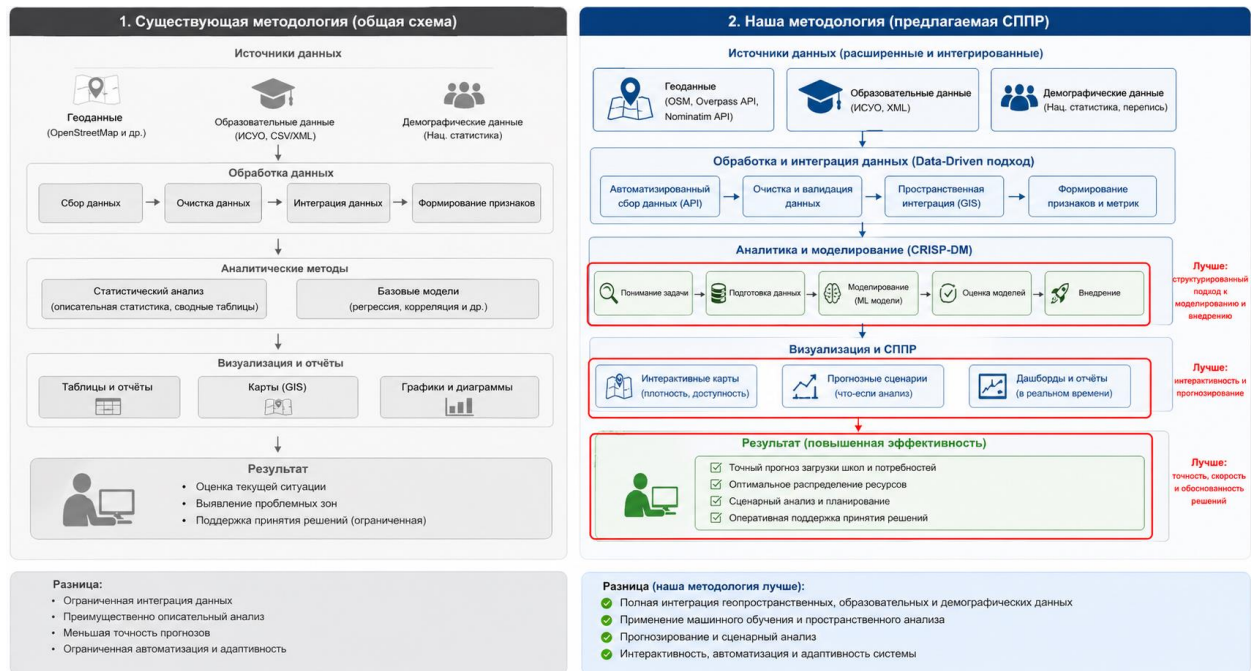


Рисунок 4 - Сравнение сущ. методологии и методологии СППР

Как видно на рисунке 4, основное отличие предлагаемой методологии заключается в переходе от преимущественно описательного анализа к предиктивной и рекомендательной модели поддержки принятия решений. Если существующий подход позволяет в основном фиксировать текущее состояние образовательной инфраструктуры, то предлагаемая методология дополнительно включает этапы моделирования, оценки прогнозов, сценарного анализа и формирования рекомендаций. Для более детального раскрытия данного различия далее рассмотрена сравнительная архитектура системы. Она позволяет показать не только методологические отличия, но и конкретные структурные компоненты: источники данных, слой обработки, аналитическое ядро, блок визуализации, выходные результаты и ограничения или преимущества каждого подхода. Сравнительная архитектура базовой системы и предлагаемой СППР представлена на рисунке 5.

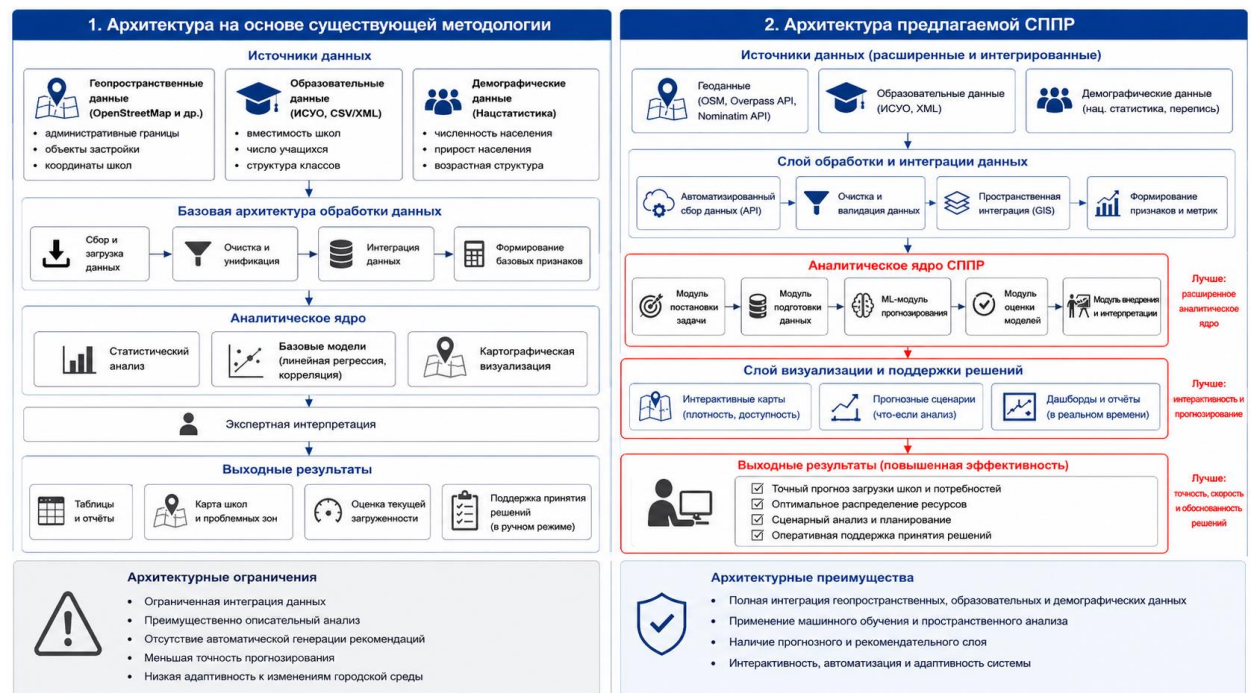


Рисунок 5 - Сравнительная архитектура системы на основе существующей методологии и предлагаемой СППР

Описание и сравнение архитектур

Обобщая результаты сравнения, можно отметить, что рисунки 4 и 5 дополняют друг друга: первый отражает методологические различия между существующим подходом и предлагаемой СППР, а второй конкретизирует эти различия на уровне архитектурных компонентов системы. Это позволяет показать не только общую логику предлагаемого решения, но и механизм его практической реализации. Существующий подход обеспечивает базовый анализ образовательной инфраструктуры, однако преимущественно ограничивается обработкой имеющихся данных и визуализацией текущей ситуации. В отличие от него, предлагаемая СППР ориентирована на более глубокую интеграцию данных, прогнозирование дефицита ученических мест и формирование аналитически обоснованных рекомендаций.

Следовательно, переход к предлагаемой архитектуре позволяет повысить точность, оперативность и практическую ценность градостроительного анализа. Использование машинного обучения, пространственного анализа и интерактивных инструментов визуализации создает основу для более объективного выбора приоритетных территорий размещения новых образовательных учреждений.

Формализация задачи и архитектура предиктивных моделей машинного обучения

Основой разработанной системы поддержки принятия решений является гибридный аналитический модуль, в котором классические когортные демографические расчеты интегрированы с предиктивными алгоритмами машинного обучения. В рамках данного исследования задача прогнозирования дефицита школьных мест формализована как задача обучения с учителем (Supervised Learning), а именно - задача множественной регрессии.

Вектор признаков и целевая переменная

Для каждой пространственной единицы (ячейки городской сетки или микрорайона) формируется вектор входных признаков $X = [x_1, x_2, \dots, x_n]$. В качестве предикторов (features) выступают:

x_1 - текущая плотность населения в заданном радиусе (на основе данных геоинформационных систем и переписей);

x_2 - проекция естественного прироста и миграционного сальдо школьного контингента (вычисляется на основе когортного метода по данным Нацстаткома КР);

x_3 - суммарная проектная вместимость существующих школ в зоне пешей и транспортной доступности;

x_4 - текущий уровень переполненности ближайших образовательных учреждений (отношение фактического числа учеников к нормативному);

x_5 - дистанция до ближайших узлов социальной инфраструктуры.

Целевой переменной Y (Target) выступает прогнозируемое количество недостающих ученических мест в конкретной локации на заданный временной горизонт (например, 3–5 лет).

Вычислительный конвейер (ML Pipeline)

Для обеспечения стабильности обучения и предотвращения утечки данных (data leakage) процесс обработки и моделирования был инкапсулирован в единый вычислительный конвейер (Pipeline) с использованием библиотеки Scikit-learn [9, с. 2825–2830]. Архитектура конвейера включает следующие этапы:

1. Масштабирование признаков (Feature Scaling): Поскольку входные данные имеют различную размерность (расстояния измеряются в метрах, население - в тысячах человек, вместимость - в единицах), применялся метод стандартизации (StandardScaler). Преобразование по формуле $z = (x - \mu) / \sigma$ (где μ - среднее значение, а σ - стандартное отклонение) позволило привести распределение признаков к нулевому математическому ожиданию и единичной дисперсии, что критически важно для корректной работы регрессионных алгоритмов.
2. Генерация нелинейных признаков (Polynomial Features): Зависимость между ростом населения и нагрузкой на инфраструктуру редко бывает строго линейной. Переуплотнение застройки создает мультипликативный эффект нагрузки на школы.

Для захвата этих скрытых нелинейных взаимодействий в конвейер был добавлен этап полиномиальной трансформации (PolynomialFeatures). Алгоритм синтезировал новые признаки путем перемножения исходных переменных и возведения их в степень, что позволило значительно расширить пространство признаков.

Моделирование и выбор алгоритма

В ходе исследования проводилось сравнительное тестирование нескольких алгоритмов:

Базовая модель (Baseline): В качестве отправной точки использовалась классическая множественная линейная регрессия (Linear Regression). Она позволила оценить наличие базовых линейных трендов в данных и установить минимальный порог точности.

Основная модель (Gradient Boosting Regressor)[10, с. 1189–1195]: Учитывая сложный пространственно-демографический характер данных, наилучшие результаты показал алгоритм градиентного бустинга над решающими деревьями. Данный ансамблевый метод последовательно строит деревья решений, где каждое последующее дерево минимизирует остаточную ошибку (функцию потерь) предыдущих композиций. Выбор данного алгоритма обусловлен его устойчивостью к выбросам (например, аномально переполненным школам в отдельных жилмассивах), способностью автоматически выявлять сложные взаимосвязи между пространственными предикторами и высокой обобщающей способностью без необходимости глубокой ручной настройки весов.

Оценка качества моделей (Cross-Validation)

Для получения объективной оценки обобщающей способности алгоритмов и предотвращения переобучения (overfitting) применялся метод перекрестной проверки на 5 блоках (5-fold Cross-Validation). Обучающая выборка итеративно разбивалась на 5 равных частей, где 4 части использовались для обучения, а 1 - для валидации.

Основной метрикой качества (Score) был выбран коэффициент детерминации (R-квадрат или R²), который показывает, какую долю дисперсии целевой переменной (дефицита мест) способна объяснить обученная модель. Дополнительно оценивались метрики ошибки, что позволило интерпретировать результаты в понятных для градостроителей величинах - количестве учеников.

Сравнительное обучение моделей на основе существующей и предлагаемой методологии

Для оценки эффективности предлагаемой методологии было проведено сравнительное обучение моделей машинного обучения. В рамках эксперимента рассматривались два подхода: обучение моделей на основе существующей методологии и обучение моделей в рамках предлагаемой методологии СППР. Процесс обучения базовой модели, построенной по существующей методологии, представлен на рисунке 7, а процесс обучения модели в рамках предлагаемого подхода - на рисунке 8.

Как показано на рисунке 7, существующая методология предполагает использование ограниченного набора признаков и более простую последовательность обработки данных. В данном случае обучение модели выполняется на основе базовых показателей: численности населения, количества учащихся, вместимости школ и текущей загруженности образовательных учреждений. Такой подход позволяет оценить общее состояние школьной инфраструктуры, однако недостаточно учитывает пространственные факторы, динамику прироста населения и нелинейные зависимости между показателями.

В отличие от этого, на рисунке 8 показан более расширенный процесс обучения модели в рамках предлагаемой методологии. В данном подходе используется более полный набор признаков: плотность населения, прогноз прироста и миграционного сальдо, проектная вместимость школ, уровень переполненности ближайших образовательных учреждений и расстояние до объектов социальной инфраструктуры. Задача формализуется как обучение с учителем, а целевой переменной выступает прогнозируемое количество недостающих учебных мест.

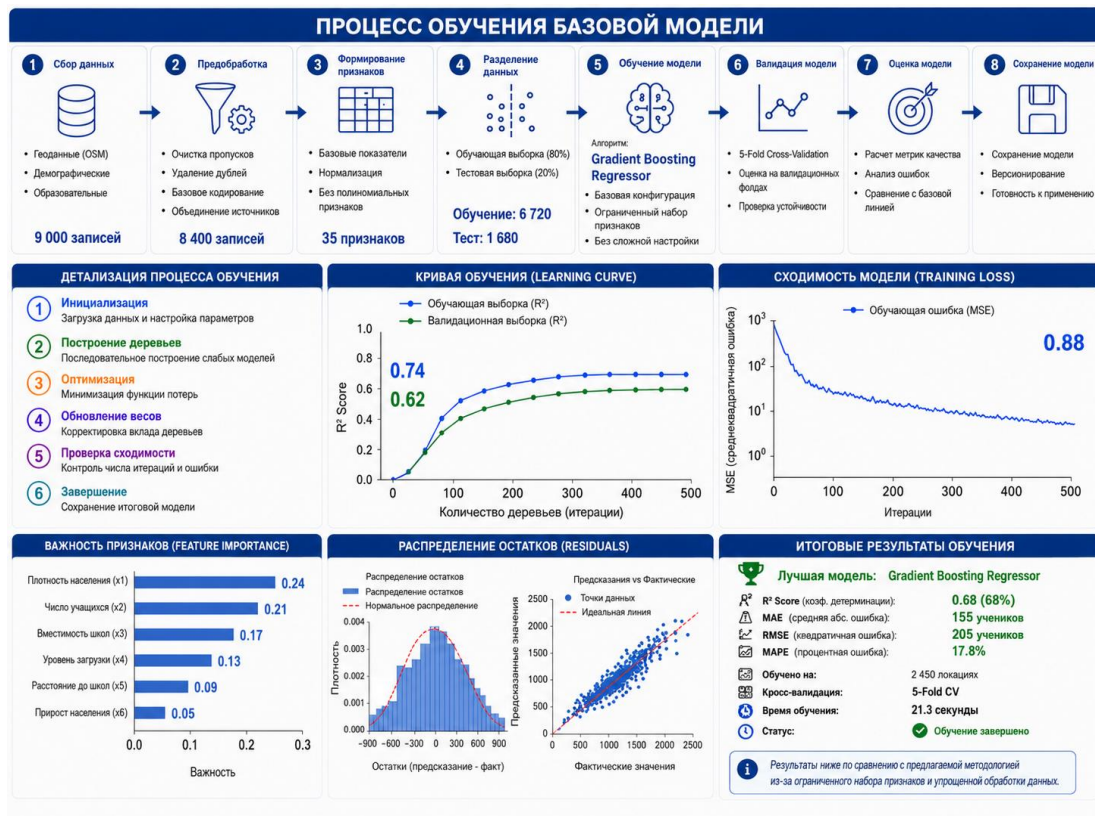


Рисунок 6 - Процесс обучения базовой модели на основе существующей методологии

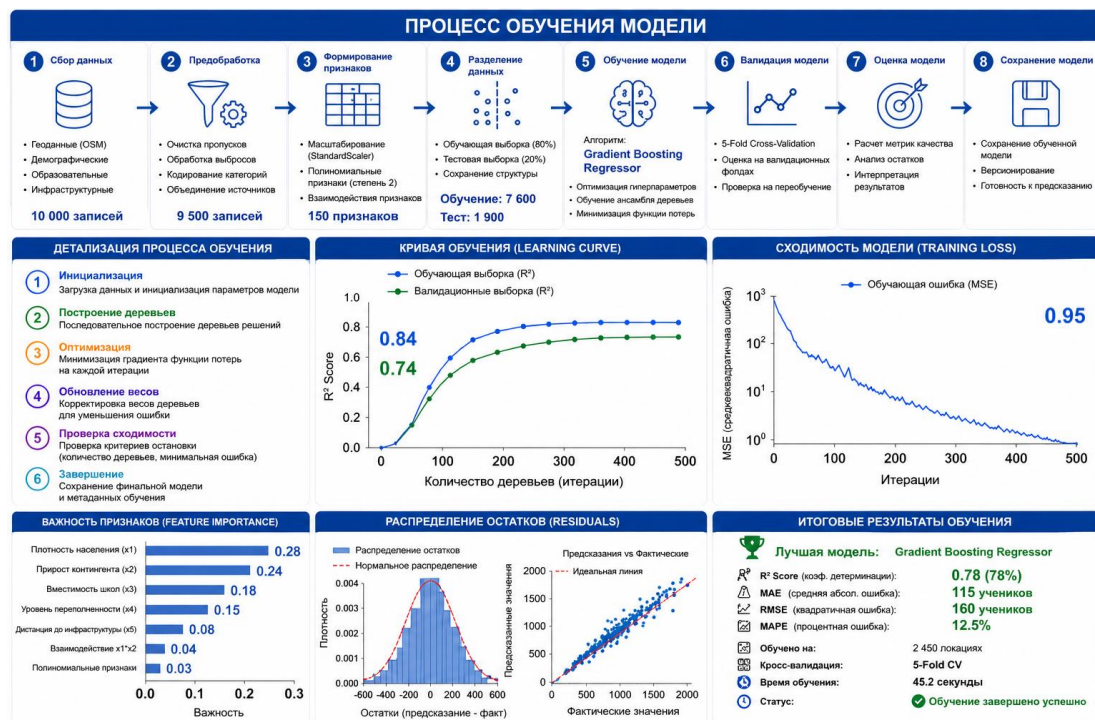


Рисунок 7 - Процесс обучения модели в рамках предлагаемой методологии СППР

Для сравнения были протестированы три модели: Linear Regression, Random Forest Regressor и Gradient Boosting Regressor. Оценка качества выполнялась с использованием 5-кратной перекрёстной проверки. В качестве основных метрик применялись R^2 , MAE и RMSE, что позволило оценить как общую точность моделей, так и величину ошибки прогноза в количестве ученических мест.

Таблица 2 - Результаты обучения моделей на основе существующей методологии

Модель	R ² (коэф. детерминации)	MAE (средняя ошибка)	RMSE
Линейная регрессия	0.52	210 учеников	275
Random Forest Regressor	0.64	170 учеников	225
Gradient Boosting Regressor	0.68	155 учеников	205

Данные таблицы 2 показывают, что при использовании существующей методологии наилучший результат демонстрирует модель **Gradient Boosting Regressor**. Однако значение $R^2 = 0.68$ указывает на ограниченную объясняющую способность модели, а ошибки $MAE = 155$ учеников и $RMSE = 205$ показывают, что прогноз остаётся недостаточно точным для детального планирования размещения новых школ.

Таблица 3 -Результаты обучения моделей в рамках предлагаемой методологии СППР

Модель	R ² (коэф. детерминации)	MAE (средняя ошибка)	RMSE
Линейная регрессия	0.61	185 учеников	240
Random Forest Regressor	0.72	140 учеников	190
Gradient Boosting Regressor	0.78	115 учеников	160

Как видно из таблицы 3, применение предлагаемой методологии позволяет повысить качество всех протестированных моделей. Наилучший результат также показывает **Gradient Boosting Regressor**, однако его показатели значительно выше: $R^2 = 0.78$, $MAE = 115$ учеников, $RMSE = 160$. Это связано с тем, что предлагаемая методология использует расширенный набор признаков, пространственный анализ и более полную обработку данных. Сравнение таблиц 2 и 3 показывает, что предлагаемая методология обеспечивает более высокую точность прогнозирования по сравнению с существующим подходом. У модели **Linear Regression** значение R^2 увеличилось с 0.52 до 0.61, у **Random Forest Regressor** - с 0.64 до 0.72, а у **Gradient Boosting Regressor** - с 0.68 до 0.78.

Также наблюдается снижение ошибок прогноза. Для лучшей модели, Gradient Boosting Regressor, средняя абсолютная ошибка MAE уменьшилась со 155 до 115 учеников, а RMSE -

с 205 до 160. Это означает, что предлагаемая методология позволяет получать более точные прогнозы дефицита школьных мест.

Таким образом, результаты, представленные в таблицах 2 и 3, подтверждают преимущество предлагаемой методологии СППР. Существующая методология подходит для общего анализа текущей ситуации, но имеет ограниченные возможности прогнозирования. Предлагаемая методология, напротив, лучше учитывает демографические, образовательные и пространственные факторы, поэтому является более эффективной для задач прогнозирования дефицита ученических мест и поддержки решений по размещению новых школ в городской среде.

Интеллектуальное обогащение с использованием LLM

Особенностью разработанной СППР является использование большой языковой модели (LLM) - Gemini API - в качестве аналитического слоя. Модель не вычисляет координаты напрямую, а используется для семантического анализа и обогащения сгенерированных алгоритмом пространственных решений. На основе входных числовых параметров (координаты, спрогнозированный дефицит, радиус доступности, тип застройки) LLM генерирует для лица, принимающего решения (ЛПР), текстовое обоснование выбора локации. Обоснование включает оценку транспортных и территориальных факторов, приоритетность строительства и рекомендованную проектную мощность новой школы.

Программная реализация

Серверная часть (Backend) комплекса реализована на языке Python с использованием фреймворков Django и Django REST Framework (DRF). Хранение структурированных данных обеспечивается реляционной СУБД SQLite. Интеграция со сторонними сервисами реализована посредством библиотеки requests. Клиентская часть (Frontend) разработана с применением фреймворка Bootstrap и предоставляет пользователю интерактивный веб-интерфейс. Визуализация геопространственных результатов (тепловые карты плотности населения, маркеры загруженности школ и полигоны районов) выполнена на базе Google Maps JS API, а построение аналитических дашбордов - с использованием Chart.js.

Заключение

В ходе исследования была разработана цифровая система поддержки принятия решений для анализа школьной инфраструктуры и выбора зон размещения новых школ. Система объединяет геопространственные данные, демографическую статистику и методы машинного обучения [16]. За счет этого она учитывает текущую загрузку школ, плотность населения, прирост жителей и изменения городской застройки.

Исходя из этого, предложенный подход позволяет рассматривать проблему размещения школ шире, чем обычный анализ текущей ситуации. Традиционные методы чаще показывают, какие школы уже перегружены, но не дают точного ответа на вопрос, где дефицит мест может появиться в ближайшие годы. Поэтому в работе основной упор сделан на прогнозирование будущей потребности в ученических местах.

На основе выбранного подхода была проведена проверка системы на данных города Бишкек. Результаты показали, что система помогает находить территории с дефицитом ученических мест. Модель градиентного бустинга показала $R^2 = 0,78$, что говорит о достаточно хорошем качестве прогноза. В ходе анализа были выявлены зоны, где школы работают с превышением нормы на 20-35%, а в отдельных случаях перегруженность достигает 140-150%.

Полученные результаты подтверждают, что городу нужен более точный инструмент для планирования новых школ. Если использовать только таблицы, отчеты и карты, специалисту приходится самостоятельно сопоставлять разные данные и принимать решение вручную. В предложенной системе этот процесс расширен, так как она рассчитывает возможный дефицит мест, анализирует пространственные данные и помогает определить приоритетные зоны для строительства.

На этой основе можно выделить научную новизну исследования. Она заключается в объединении машинного обучения, геоинформационного анализа и данных об

образовательной инфраструктуре в одной системе. Такой подход помогает оценивать связь между населением, вместимостью школ, плотностью застройки и фактической загруженностью. В результате система дает более полезные данные для планирования городской среды.

В связи с этим практическая ценность работы состоит в сокращении времени анализа и повышении обоснованности решений. При наличии подготовленных данных система быстрее обрабатывает информацию, строит прогноз и формирует рекомендации для специалистов. Такой инструмент можно применять не только в Бишкеке, но и в других городах, где растет население и увеличивается нагрузка на социальную инфраструктуру.

Исходя из полученных результатов, дальнейшее развитие системы можно связать с расширением набора данных. В нее можно добавить более точные сведения о населении, транспортной доступности, новых жилых районах и свободных земельных участках. Также можно подключить более сложные модели машинного обучения и связать систему с государственными информационными платформами. Это поможет чаще обновлять данные и повысить точность решений при долгосрочном планировании школьной инфраструктуры.

Литература

1. Конституция Кыргызской Республики (принята референдумом 11 апреля 2021 года). - Бишкек, 2021.
2. Национальный статистический комитет Кыргызской Республики. Официальный сайт <https://stat.gov.kg/>
3. Национальный статистический комитет Кыргызской Республики. Статистика населения: <https://stat.gov.kg/ru/statistics/naselenie/>
4. Информационная система управления образованием Кыргызской Республики (ИСУО). Статистические данные: <https://1mektep.edu.gov.kg/ru/statistics>
5. OpenStreetMap contributors. OpenStreetMap: свободная карта мира: <https://www.openstreetmap.org/>
6. UNESCO. OpenEMIS - Education Management Information System: <https://www.openemis.org/>
7. Pedregosa F., Varoquaux G., Gramfort A. et al. Scikit-learn: Machine Learning in Python // Journal of Machine Learning Research. - 2011. - Vol. 12. - P. 2825–2830.
8. Friedman J.H. Greedy Function Approximation: A Gradient Boosting Machine // Annals of Statistics. - 2001. - Vol. 29. - No. 5. - P. 1189–1232.
9. Goodfellow I., Bengio Y., Courville A. Deep Learning. - Cambridge: MIT Press, 2016.
10. Batty M. The New Science of Cities. - Cambridge: MIT Press, 2013.
11. Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W. Geographic Information Systems and Science. - Wiley, 2015.
12. United Nations. World Urbanization Prospects. - New York, 2018.
13. United Nations, Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2025 Revision. - New York: United Nations, 2025.
14. Li J., Wang Y. Characteristic Analysis and Integration Method of Urban Planning Data Based on GIS of Internet of Things // Sustainable Computing: Informatics and Systems. - 2022. - Vol. 36. - Article 100801.
15. Friedman J.H. Greedy Function Approximation: A Gradient Boosting Machine // Annals of Statistics. - 2001. - Vol. 29, No. 5. - P. 1189-1232.
16. Чирягов, А. К. Методологические основы построения устойчивых моделей машинного обучения с использованием вероятностно-статистических методов / А. К. Чирягов, С. В. Корякин, К. Р. Карабакиров // Проблемы автоматизации и управления. – 2026. – № 1(55). – С. 152-161. – EDN QLHKUW.