



**МИНОБРНАУКИ РОССИИ**  
**федеральное государственное бюджетное образовательное учреждение**  
**высшего образования**  
**«Московский государственный технологический университет «СТАНКИН»**  
**(ФГБОУ ВО «МГТУ «СТАНКИН»)**

**Институт  
информационных  
технологий**

**Кафедра  
информационных технологий  
и вычислительных систем**

**КУРСОВОЙ ПРОЕКТ**  
**ПО ДИСЦИПЛИНЕ**  
**«ОСНОВЫ НОВЫХ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ»**

СТУДЕНТА 4 КУРСА бакалавриата ГРУППЫ ИДБ-20-02

**ЕРДОГАНА ДЕНИЗА ЕРДАЛОВИЧА**

**НА ТЕМУ**  
Исследование инструментов data science и разработка веб-сервиса анализа успеваемости  
студента

Направление: 09.03.01 Информатика и вычислительная техника  
Профиль подготовки: Программное обеспечение средств вычислительной техники и  
автоматизированных систем

Отчет сдан «\_\_\_\_\_» \_\_\_\_\_ 2024 г.

Оценка \_\_\_\_\_

Преподаватель Бибиков О.Д., преподаватель

\_\_\_\_\_  
(подпись)

**МОСКВА 2024**

**ОГЛАВЛЕНИЕ**

ГЛАВА 1. Формирование требований к программному продукту .....	3
1.1. Анализ потребностей .....	5
1.2. Определение функциональных требований .....	7
1.3. Определение нефункциональных требований .....	8
1.4. Сбор и анализ требований .....	8
1.5. Документирование требований .....	15
ГЛАВА 2. Проектирование .....	17
2.1. Формирование общих требований .....	17
2.2. Формирование структуры приложения .....	18
2.3. Организация модели для проведения экспериментов .....	18
ГЛАВА 3. Реализация .....	21
ГЛАВА 4. Тестирование .....	25
ГЛАВА 5. Разработка и развёртывание продукта .....	31
ГЛАВА 6. Эксплуатация и сопровождение .....	33

## **ГЛАВА 1. ФОРМИРОВАНИЕ ТРЕБОВАНИЙ К ПРОГРАММНОМУ ПРОДУКТУ**

Государству необходимо давать образование населению, преимущественно молодому, так как оно более обучаемое и является важной частью будущего, что необходимо для дальнейшего процветания.

Совершенствование качества образования и подготовка высококвалифицированных специалистов – это факторы повышения благосостояния во многих странах.

Исполнение вышеупомянутых функций осуществляется учебным учреждением при помощи следующих инструментов: мониторинг качества образования, обеспечение информационным материалом, проведение практических занятий, оценивание результатов обучения студента для поддержания мотивации последнего, социализация студента. Чтобы студенту иметь хороший рейтинг, ему нужно: иметь дисциплину для получения дополнительных баллов или автомата, выполнять различного рода работы (что развивает когнитивные способности студента), а главное сдавать итоговые аттестационные работы, которые эмулируют реальную стрессовую ситуацию, которая встречается в реальной жизни на работе.

Мониторинг – это контроль образовательного процесса, а также его результатов согласно определенным стандартам. Метрика уровня образования при мониторинге - успеваемость студента.

Успеваемость студента - это одна из важнейших характеристик образовательной деятельности учебного заведения, по которой судят о достигнутых результатах или об имеющихся проблемах в образовании. Хорошая успеваемость - свидетельство усвоения содержания образовательной программы гарантирующий подготовку востребованного обществом специалиста, который владеет современными знаниями и умением работать с передовыми технологиями.

Так образом критерий успеваемости важен:

- высшему учебному заведению как: показатель качества оказываемого образования, показатель качественного исполнения своих функций перед государством (если мы говорим про бюджетные учреждения);
- государству как: степень качества спонсируемых кадров для будущего страны, именно, образованные кадры имеют наибольшую ценность (в последнее время их процент падает);
- студентам как: характеристика для самоконтроля своей учебной деятельности.

Мониторинг качества образования имеет комплексный характер. Первый подразумевает не только оценку текущего состояния качества образования, но и прогнозирование будущего состояния.

Прогнозирование результатов обучения позволяет эффективно управлять процессом обучения, профессиональным развитием студентов и, одновременно, принимать меры по снижению риска неудовлетворительных результатов в обучении.

Прогнозирование может осуществляться множеством методов. Стандартные статистические методы известны и применимы издавна, они трудоёмки в вычислениях и ограничены в силу своей специфики.

В контексте реального времени, можно подчеркнуть, что компьютерная техника сделала прорыв за последние десятилетия. Этот прорыв предоставляет набор передовых инструментов, часть из которых являются инструментами специалиста по данным (Data Scientist-a): электронно-вычислительная техника для обработки и анализа больших массивов данных, полученных в ходе исследований, построение статистических моделей, которые позволяют оценить текущее состояние учебных достижений студентов и осуществить прогноз их будущего.

Создание программного модуля прогнозирующего успеваемость позволяет выявлять студентов, которые с большой вероятностью будут иметь академические задолженности, и применять управляющее воздействие

раньше, чем появиться реальная проблема с успеваемостью. Это приводит к более эффективной учебно-воспитательной работе единого деканата. Так анализ результатов ЕГЭ может быть также использован для выявления среди первокурсников «групп риска» по успеваемости.

Для того, чтобы сформировать требования к ПО, необходимо выделить ключевые моменты, которые следует учесть при формировании:

1. Анализ потребностей.
2. Определение функциональных требований.
3. Определение нефункциональных требований.
4. Сбор и анализ требований.
5. Документирование требований.

### **1.1. АНАЛИЗ ПОТРЕБНОСТЕЙ**

Прогнозирование успеваемости может быть интересно следующим группам людей и организаций:

- учащимся и их родителям. Они могут использовать прогнозы для: определения сильных и слабых сторон в обучении, планирования дополнительных занятий и подготовки к экзаменам;
- преподавателям и администрации учебных заведений. Прогнозы помогают им оценить эффективность учебных программ и методов обучения, а также выявить студентов, нуждающихся в дополнительной поддержке;
- исследователям в области образования. Они могут использовать данные о прогнозировании успеваемости для разработки новых методов обучения и оценки их эффективности;
- работодателям. Прогнозы могут помочь им определить потенциал будущих сотрудников и принять решение о найме или продвижении по службе;
- государственным органам. Они могут использовать прогнозы для оценки качества образования в стране и разработки стратегий его

улучшения;

- психологам и консультантам. Прогнозы позволяют им выявить возможные проблемы в обучении и предложить пути их решения.

После того, как были выделены потенциальные потребители, можно определить основные потребности:

- получения статистического прогноза: пользователи заинтересованы в возможности получения прогноза их успеваемости по введенным ими данным;
- визуализация и анализ данных: программный модуль должен обеспечить визуализацию основных показателей студента по отношению всех студентов, дабы студенту стало понятно, в каких аспектах он объективно силен, а в каких нужно приложить усилия. Также стоит продемонстрировать корреляционную матрицу признаков, чтобы пользователь понимал, что влияет на его результаты. Это позволит лучше понимать и интерпретировать полученные результаты;
- доступность и удобство использования: важным аспектом для пользователей является удобство использования и доступность программного модуля с любого устройства с почти любым уровнем соединения интернета. Также модуль должен обладать хорошим показателем юзабилити, чтобы интерфейс решения привлекал пользователя;
- точность и достоверность прогнозов: пользователи ожидают, что прогнозы будут точными и достоверными, основанными на объективных данных и научных методах;
- поддержка и помощь: пользователи могут нуждаться в поддержке помощи при использовании сайта, например, при интерпретации результатов прогнозирования или настройке параметров;
- конфиденциальность данных: пользователи хотят быть уверены, что их личные данные защищены и не будут переданы третьим

лицам без их согласия.

## **1.2. ОПРЕДЕЛЕНИЕ ФУНКЦИОНАЛЬНЫХ ТРЕБОВАНИЙ**

Функциональные требования представляют собой описание того, что должен делать программный модуль или система. Они определяют функции, сервисы или возможности, которые будут предоставлены пользователям, а также способы их взаимодействия с системой.

Основной целью определения функциональных требований является установление четкого набора задач и функций, которые должны быть реализованы в программном продукте. Это позволяет разработчикам и заказчикам иметь общее представление о том, что именно должно быть включено в проект, и ориентироваться в процессе его разработки и тестирования.

### **1. Интерпретируемость прогноза:**

- a. Программный модуль должен обеспечивать понятный пользователю результат прогноза.
- b. Сайт должен предоставлять аналитическую выжимку с графической интерпретацией пользователю для удобства восприятия информации.

### **2. Удобное управление:**

- a. Интерфейс должен последовательно ввести пользователя по основным этапам прогнозирования результатов.
- b. Форма опроса должна раскрывать смысл необходимых для прогнозирования полей путём комментирования соответствующих элементов.
- c. Функционал сайта не должен быть перегружен.

### **3. Обратная связь:**

- a. Программный модуль должен предоставлять краткую информацию об этапах взаимодействия пользователю.
- b. Программный модуль должен на каждом этапе предоставлять

обратную связь пользователю.

- с. У пользователя должна быть возможность написать сообщение на сайте разработчикам программного модуля.

### **1.3. ОПРЕДЕЛЕНИЕ НЕФУНКЦИОНАЛЬНЫХ ТРЕБОВАНИЙ**

Нефункциональные требования определяют условия, в которых должно функционировать программное обеспечение, и обеспечивают качество и удобство использования для конечных пользователей.

1. **Безопасность:** пользовательские данные должны быть защищены от несанкционированного доступа и храниться в безопасном режиме. Это означает использование надежных методов аутентификации, шифрования данных и механизмов контроля доступа.
2. **Совместимость:** сайт должен быть совместим с множеством браузеров и их версиями на различных устройствах.
3. **Эффективность:** сайт должен быть эффективен в своём прогнозировании и соответствовать реальному будущему, а также должен быть оптимизирован в своих вычислениях, дабы пользователь не ждал результатов долгое время.

### **1.4. СБОР И АНАЛИЗ ТРЕБОВАНИЙ**

Актуальность и целесообразность разработки программного модуля по выбранной теме зависят от имеющихся продуктов. Если найдётся продукт, который имеет: возможность к масштабируемости, финансирование, спрос у пользователей, и т. д., то разработка аналогичного продукта (модуля) не будет иметь смысла, так как соревноваться на уровне ресурсов невозможно, а решение перехватить аудиторию у конкурентов без наличия ноу-хау (и то не факт, что это поможет) невыполнимо.

В связи с этим было принято решение проанализировать рынок продуктов связанных с прогнозированием успеваемости студентов, так как без спроса - предложения не будет.



Нужно учитывать, что проблема успеваемости имеет комплексный характер, поэтому стоит учитывать, как психологический портрет студента, так и его профессиональные навыки. То есть для создания продукта анализирующего успеваемость студента нужна команда из: психологов, социологов, программистов, аналитиков, и т. д. Для создания такого продукта у владельца должна быть мотивация (обычно выражающаяся в денежном эквиваленте). Однако на уровне рынка, данное решение вряд ли найдёт спонсирование, что доказывается отсутствием аналогов на рынке.

В процессе сравнения решений нужно находить как сильные, так и слабые стороны имеющихся продуктов. Однако оценить объективно продукты невозможно без критериев оценивания, которые дают качественную или количественную характеристику продукту.

В связи с этим был создан следующий список критериев с надлежащей аргументацией оценивания продуктов:

- цена - это критерий для пользователей, поскольку она напрямую влияет на их способность и возможность пользоваться продуктом;
- простота использования (юзабилити) – это критерий приложения, который отражает насколько легко и интуитивно понятно пользователь может взаимодействовать с интерфейсом приложения;
- эффективность - это критерий того, насколько точно приложение может предсказывать будущие значения успеваемости студента;
- анонимность – это критерий, который характеризует степень защиты информации об пользователе;
- универсальность – это критерий, который показывает, насколько хорошо предсказывает успеваемость студента вне зависимости от его параметров, которые не учитываются в исследовании.

Данные критерии ключевые (но единственные) при сравнении решений связанных с успеваемостью студентов.

Масштабных приложений анализа успеваемости студентов по опросу нет. Большинство похожих проектов и решений рассмотрено в научных работах. Однако рассмотреть отдалённо похожие продукты стоит:

Gradescope это американская компания, которая предлагает онлайн-инструменты оценивания успеваемости обучающихся используя искусственный интеллект.

Одноимённая бесплатная нейросеть - инструмент для всех, кто занимается выставлением оценок и аттестацией.

У инструмента существует обёртка в виде сайта с таким же названием. Сайт предоставляет возможность следить за своими оценками, а также получать уведомления о сроках сдачи работ. Кроме того, GradeScore позволяет преподавателям создавать курсы и управлять ими, а также общаться со студентами.

Сервис имеет инструменты для анализа данных успеваемости студентов, такие как графики и таблицы. Платформа позволяет прогнозировать успеваемость студентов на основе различных факторов: прошлые оценки, посещаемость, участие в проектах и т.д.

Оценивание решения:

- сайт доступен только на английском языке (-);
- интерфейс понятен на первых этапах работы, но сложен для освоения специальных возможностей (-);
- ограниченная версия доступна бесплатно, продвинутая за 10 \$ в месяц. Бесплатная версия не содержит даже половину функций платной версии(+);
- сайт на сайте не учитывает личные качества обучающегося, а смотрит только на предметную успеваемость (-);
- так как основной инструмент – это тяжёлая нейросеть с огромным количеством параметров и данных, то эффективность прогнозированияобеспечена на высоком уровне (+);
- анонимности нет, студент предоставляет свои личные данные при

участии на сайте (-);

- универсальность на слабом уровне. Сайт предназначен на узкую область предметов в определённом формате. Если у студента имеется десяток дисциплин в высшем учебном заведении, то на сайте учитываются только доступные, причём только в тех темах, которые рассматриваются преподавателем (-).

Пример работы сайта представлен на рисунке (см. Рис. 1.1).

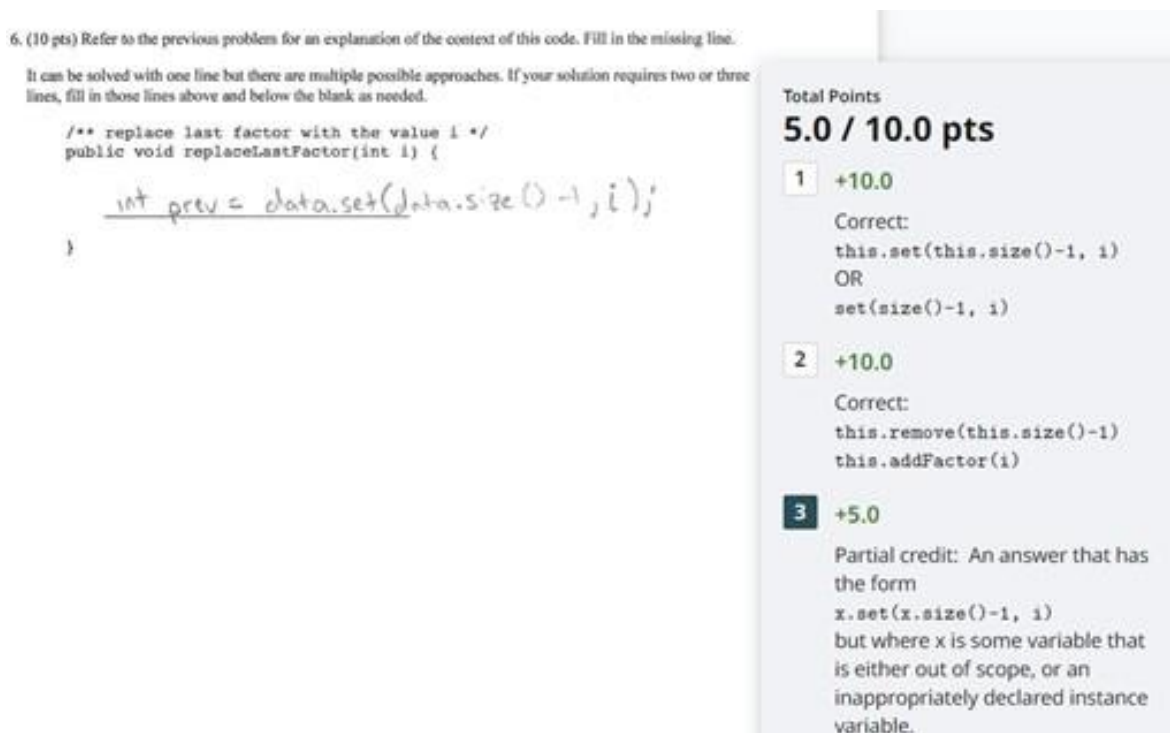


Рис. 1.1. Пример работы сайта GradeScope

Knewton - это компания, которая разрабатывает персонализированные образовательные системы на основе искусственного интеллекта. Компания использует алгоритмы машинного обучения для анализа данных учащихся и создания индивидуальных планов обучения.

Knewton также предлагает платформу для создания и управления онлайн-курсами. Одним из главных преимуществ является возможность создания персонализированных учебных программ для каждого студента. Это

позволяет ученикам быстрее достигать своих образовательных целей и повышает их мотивацию к обучению.

Оценивание решения:

- веб-сервис полностью на английском языке (-);
- интерфейс запутанный и сложный для нового пользователя (-);
- knewton предлагает платные услуги. Стоимость зависит от выбранного пакета и начинается от \$399 в год (-);
- сайт учитывает личные характеристики студента при прогнозировании его успеваемости: возраст, пол, уровень образования, средний балл аттестата, интересы, предпочитаемый стиль обучения, цели в образовании, язык обучения, страна проживания. Однако слабо рассматривается психологическая составляющая студента и внешние факторы (+);
- эффективность обеспечена на среднем уровне, так как нейросеть лежащая в основе инструмента слабее, чем у аналогичных конкурентов(-);
- анонимности нет, так как студент отдаёт свои личные данные при участии на сайте (-);
- универсальность жалеет оставлять лучшего, так как многие внешние факторы не учитываются при прогнозировании (-).

Пример работы сайта представлен на рисунке (см. Рис. 1.2).

<input type="checkbox"/>	First Name ↓	Last Name	Exceptions	Assignment Mastery	Review Questions	Review Time	Predicted Grade	Attempt	Test Grade
<input type="checkbox"/>	Alfredo	Westervelt		95%	25	22m	● 90 – 98%	0 of 2	Not Started
<input type="checkbox"/>	Allison	Aminoff		85%	25	22m	● 75 – 83%	1 of 2	70%
<input type="checkbox"/>	Dulce	Mango		96%	25	22m	● 89 – 96%	1 of 2	92%
<input type="checkbox"/>	Jakob	Vaccaro	ⓘ	75%	25	22m	● 77 – 85%	1 of 2	In Progress
<input type="checkbox"/>	Jaxson	Stanion		65%	25	22m	● 78 – 86%	0 of 2	Not Started
<input type="checkbox"/>	Jaylon	Bator		76%	25	22m	● 43 – 63%	1 of 2	50%

Рис. 1.2. Пример работы сервиса Knewton

Civitas Learning - это платформа для онлайн-обучения, которая предлагает курсы по различным направлениям, таким как предпринимательство, финансы, маркетинг и другое.

Одной из главных особенностей Civitas Learning является использование искусственного интеллекта и машинного обучения для персонализации обучения. Платформа анализирует данные об учащихся, их интересах и целях, чтобы создать индивидуальные планы обучения.

Civitas Learning также предлагает различные инструменты для совместной работы и общения между студентами и преподавателями. Платформа позволяет создавать группы, обсуждать темы, задавать вопросы и получать обратную связь от преподавателей. Платформа также предоставляет возможность отслеживать прогресс обучения и оценивать результаты. Это помогает учащимся видеть свои достижения и мотивирует их продолжать обучение.

Оценивание решения:

- анонимность не обеспечивается, так как для действия на платформе необходимо предоставить основную информацию о себе (-);
- стоимость использования Civitas Learning зависит от выбранного пакета. Базовый пакет стоит от 99 долларов в месяц, а премиум-пакет - от 199 долларов (-);
- спектр предметов довольно широк, но качество их проверки и преподавания неизвестен (+);
- платформа полностью на английском языке (-);
- эффективность исследования приемлема, подробное рассмотрение стоит отдельного исследования (+);
- универсальность хорошая, так как учитывается большое количество факторов разных категорий (личностных, учебных, и т. д.) (+);
- интерфейс не удобен. Большое количество вкладок, сильная

нагруженность страниц (-).

Пример работы приложения представлен (см. Рис. 1.3).

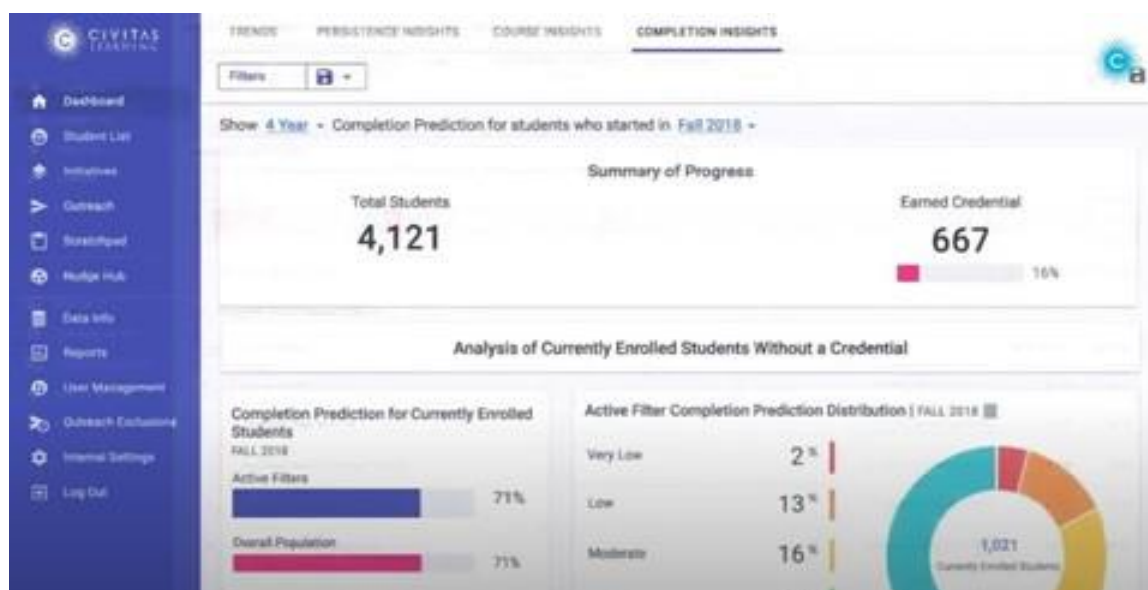


Рис. 1.3. Пример работы сервиса Civitas Learning

Для наглядного сравнения упомянутые решения при помощи таблицы (см. Таблица № 1.1).

Таблица 1.1.

Сравнение имеющихся решений

Название Криетерий	GradeScope	Knewton	Civitas Learning
<b>Цена</b>	+	-	-
<b>Юзабилити</b>	-	-	-
<b>Эффективность</b>	+	-	+
<b>Анонимность</b>	-	-	-
<b>Универсальность</b>	-	-	+

Готовых решений в широком масштабе немного, все они представляют идентичные модели, которые отслеживают учебные успехи и прогнозируют на будущую успеваемость.

Отсутствие анонимности в каждом решении сильный минус, так для оценки своей успеваемости, пользователь обязан заполнять подробную

информацию о себе, которая не влияет на сам прогноз. Сами системы нагруженные, и представляют собой огромные модели с необходимостью кормления им большого количества данных на протяжении долгого времени. У пользователя нет возможности зайти, заполнить требуемую информацию о себе, как о студенте и получить прогноз. Также из-за масштаба готовых решений сильно нагружается интерфейс. Пользователю скорее станет лень во всём разбираться, чем часами тратить свои силы на изучение продукта. Все решения представлены на английском языке без возможности включения альтернативных языков, что сужает круг пользователей.

Всё выше озвученное означает, что создание программного модуля по данной теме имеет смысл, с учётом того, что недостатки конкурентов будут исправлены.

## **1.5. ДОКУМЕНТИРОВАНИЕ ТРЕБОВАНИЙ**

Документирование требований является важным этапом в процессе разработки программного обеспечения. Этот этап включает в себя создание и утверждение технического задания, которое является основным документом, содержащим все функциональные и нефункциональные требования к разрабатываемой системе. Процесс утверждения требований важен для обеспечения понимания и согласия всех заинтересованных сторон с поставленными целями проекта.

Создание технического задания:

На основе собранных и проанализированных требований мы разрабатываем документ, который четко определяет все функциональные возможности, интерфейсы, данные и ограничения системы.

Техническое задание должно быть структурированным и понятным: оно является основой для дальнейшей работы разработчиков, тестировщиков и других специалистов.

Вывод:

В ходе процесса формирования требований к программному продукту, в данном случае прогнозирующего модуля в обёртке веб-сайта, были проведены следующие этапы: анализ бизнес-потребностей, определение функциональных и нефункциональных требований, сбор и анализ требований, а также документирование требований.

Анализ бизнес-потребностей позволил идентифицировать потребителей и определить основные потребности, которые должно удовлетворять разрабатываемое приложение.

Определение функциональных требований позволило четко сформулировать, какие функции и возможности должны быть реализованы в сайте.

Определение нефункциональных требований выявило важные аспекты.

Сбор и анализ требований включал в себя изучение аналогичных прогнозирующих решений для выявления их сильных и слабых сторон, что помогло определить наиболее востребованные функции и возможности для включения в наше приложение.

Документирование требований включало создание технического задания, что помогло обеспечить понимание и согласие с поставленными целями проекта.

Таким образом, в результате проведенных этапов формирования Требования была создана база для дальнейшей разработки прогнозирующего модуля успеваемости студента, которая будет удовлетворять потребности пользователей и соответствовать стандартам качества и безопасности.



## ГЛАВА 2. ПРОЕКТИРОВАНИЕ

### 2.1. ФОРМИРОВАНИЕ ОБЩИХ ТРЕБОВАНИЙ

Используя информацию, полученную во время формирования требований к программному продукту, получим общие требования к приложению.

Программный модуль, разработан для прогнозирования успеваемости студентов, представляет из себя веб-сайт, направленный на улучшение успеваемости студентов путём улучшения мониторинга обучения.

Пользователи программного модуля получают возможность прогнозировать свою успеваемость с любого устройства, предоставляю информацию о себе как о студенте, посредством заполнения обязательных данных на сайте при прогнозировании. Это позволит им проанализировать им будущее состояние их успеваемости, оценить влияние факторов на успеваемость, а также сравнить себя с остальными студентами.

Основные функциональные возможности программного модуля включают в себя следующее:

- выбор возможного решения модуля (посмотреть к какому классу студентов относится пользователь, спрогнозировать будущую успеваемость в виде рейтинга);
- заполнение данными обязательной формы на сайте;
- получение результатов от программного модуля с подробной интерпретацией данных;
- анализ полученных результатов.

Уровень доступа пользователей не будет регулироваться в соответствии с их статусом. Все пользователи будут иметь неограниченный доступ к функциональным возможностям сайта.

Таким образом, программный модуль будет предоставлять доступ к разнообразным решениям, обеспечивая удобство использования и эффективность работы как для студентов, так и для преподавателей.

## **2.2. ФОРМИРОВАНИЕ СТРУКТУРЫ ПРИЛОЖЕНИЯ**

Главное меню веб-сайта будет представлено в форме удобного интерфейса, где пользователь сможет выбрать тип решаемой задачи.

После, в зависимости от выбранной задачи, пользователю нужно будет ввести необходимые данные, на основе которых программный модуль предоставит решения пользователю. Все данные нужно будет вводить поэтапно. В каждом этапе будет подробное описание требуемых данных.

После всех заполненных данных у пользователю будет возможность получить решение.

Решение будет выводиться в отдельном окне. Там будет находиться основная информация относящаяся к требуемой задаче, а также визуальная информация помогающая интерпретировать данные.

Таким образом, структура и функциональности веб-сайта для проведения прогнозов будут ориентированы на обеспечение пользователей доступом к удобному и инновационному инструменту для мониторинга своей успеваемости.

## **2.3. ОРГАНИЗАЦИЯ МОДЕЛИ ДЛЯ ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТОВ**

Всё решение под обёрткой сайта будет основываться на модели машинного обучения – случайный лес.

Случайный лес (Random Forest) — это один из методов машинного обучения, который основан на использовании ансамбля решающих деревьев. Случайный лес представляет собой композицию из нескольких деревьев решений, каждое из которых построено по-своему.

Принцип работы случайного леса:

- Создание выборки. Случайный лес работает с обучающей выборкой, которая состоит из объектов с известными ответами;
- построение деревьев. Из обучающей выборки случайным образом

выбираются признаки и на их основе строятся деревья решений. При этом для каждого дерева выбирается случайный набор признаков, который используется для разделения узлов дерева;

- голосование. После построения всех деревьев решений происходит голосование. Для каждого объекта в выборке определяется класс, за который проголосовало большинство деревьев. Этот класс и будет считаться предсказанием случайного леса;
- оценка качества. Качество случайного леса оценивается по точности предсказаний. Чем больше деревьев в ансамбле, тем выше точность.

Случайный лес имеет несколько преимуществ перед другими методами машинного обучения:

- высокая точность. Случайный лес может достигать высокой точности предсказаний благодаря использованию ансамбля деревьев решений;
- устойчивость к переобучению. Случайный лес устойчив к переобучению благодаря использованию случайных признаков при построении деревьев;
- простота интерпретации. Случайный лес легко интерпретировать, так как он представляет собой ансамбль деревьев решений.

В целом, случайный лес является мощным и гибким методом машинного обучения, который может быть использован для решения различных задач.

Данная модель была выбрана в связи с тем, что при небольшом количестве данных при работе показывала лучшие показатели по метрикам данных.

Основной метрикой при оценки модели была MAE (mean absolute error). MAE (Mean Absolute Error) — это показатель, используемый для оценки качества прогнозов в машинном обучении и статистике. Он представляет собой среднее абсолютное отклонение между фактическими и

предсказанными значениями. MAE рассчитывается по следующей формуле (см. Рис. 2.1):

$$MAE(y^{true}, y^{pred}) = \frac{1}{N} \sum_{i=1}^N |y_i - f(x_i)|$$

Рис. 2.1. Формула MAE

где:  $N$  — количество наблюдений,  $y_i$  — фактическое значение,  $f(x_i)$  — предсказанное значение. Чем меньше значение MAE, тем лучше качество прогноза. Данная метрика используется из-за ненормального распределения большинства признаков. Таким образом, разработка программного модуля для прогнозирования успеваемости студентов, объединяющую техническую организацию и образовательные аспекты. Он позволяет взглянуть по-новому на старое.

### ГЛАВА 3. РЕАЛИЗАЦИЯ

Реализация представляет собой развёрнутый на сервере сайт, внешний вид которого выглядит следующим образом (см. Рис. 3.1):

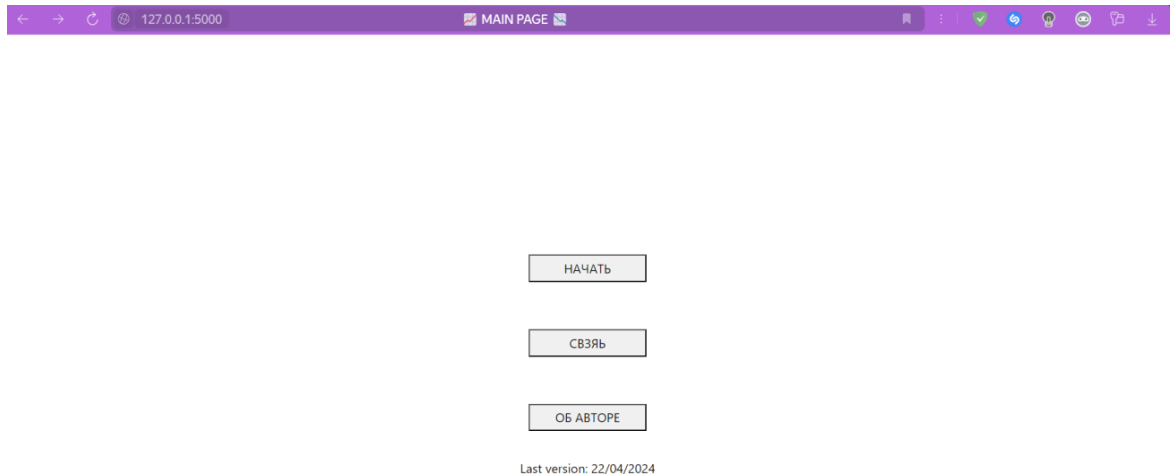


Рис. 3.1. Главная страница сайта

На главной странице сайта имеется три кнопки со следующим назначением:

- START – кнопка для перехода заполнения формы для дальнейшего прогнозирования;
- FEEDBACK – кнопка для отправки отзыва или пожелания автору сайта;
- ABOUT US – кнопка для перехода на страницу подробной информации об авторе.

Перейдём по каждой кнопке и рассмотрим подробнее каждый раздел. Начнём с раздела START.

Данный раздел представляет собой форму для сбора данных от пользователя для дальнейшего прогнозирования по ним (данным) моделью машинного обучения (см. Рис. 3.2).

**USER FORM**

**COLLECTING BASIC DATA**  
This section is dedicated to collecting basic information about the student.

**YOUR GENDER:**  
Only traditional genders are considered. This feature affects the actions of the subject throughout his life (mainly due to the hormonal background).

☐ MALE  
☐ FEMALE

**YOUR AGE:**  
The number of full years.

**COLLECTING INTERNAL FEATURES**  
The collection of internal signs is necessary to take into account the factors that characterize the student as a person.

**HOW MANY HOURS A DAY DO**

Рис. 3.2. Страница заполнения формы

После завершения ввода обязательных данных появится возможность спрогнозировать будущую успеваемость нажав на кнопку PREDICT. Или вернуться на главную страницу нажав на кнопку BACK (см. Рис. 3.3).

**HOW CONVENIENT IS IT FOR YOU TO MONITOR YOUR PROGRESS:**  
It means [website/modular magazine/electronic educational system, etc.].

1-0 2-0 3-0 4-0 5-0

**YOUR FAMILY'S INCOME (IN RUBLES):**  
For all family members combined per month.

UP TO 150,000-0  
FROM 150,000 TO 350,000-0  
FROM 350,000+0 REQUIRED

**THE OPINION OF YOUR RELATIVES ABOUT THE NEED FOR HIGHER EDUCATION:**

1-0 2-0 3-0 4-0 5-0

**WHETHER YOUR RELATIVES HAVE HIGHER EDUCATION:**  
Mostly from parents.

HAVE-0  
ABSENT-0

**BACK** **PREDICT**

Last version 22/04/2024

Рис. 3.3. Конец страницы заполнения формы

При нажатии кнопки PREDICT откроется новая страница с прогнозируемым результатом (см. Рис. 3.4):



Рис. 3.4. Страница прогноза

С этой страницы можно вернуться на главную нажав на кнопку HOME. Рассмотрим нажатие кнопки FEEDBACK на главной странице. При её нажатии пользователь попадёт на следующую страницу (см. Рис. 3.5):



Рис. 3.5. Страница с обратной связью

На этой странице пользователь может оставить сообщение создателю сайта, чтобы сделать это необходимо нажать на кнопку SEND. Если пользователь захочет вернуться назад, то следует нажать кнопку BACK.

Рассмотрим нажатие кнопки ABOUT US на главной странице. При нажатии на этой кнопки появится следующая форма (см. Рис. 3.6):



Рис. 3.6. Страница об авторе



## ГЛАВА 4. ТЕСТИРОВАНИЕ

Тестировать сам алгоритм случайный лес вручную не предстоит возможности, так как алгоритм выбора признаков довольно-таки сложен и не представлен широкой аудитории, авторы метода оставляют анонимность реализации за собой. Однако рассмотреть более простой алгоритм, который является главной и большой частью алгоритма случайного леса – возможно.

Таким образом мы рассмотрим работу модели решающего дерева, которая используется в случайном лесе.

Данные для обучения студентов необходимые модели (облегчены для возможности ручных расчётов).

Листинг 4.1.

Пример данных

```
students_dict = {
    "Рейтинг": [75, 69, 90, 100, 45],
    "Средний балл в школе": [4, 3, 5, 5, 3],
    "Пол": [1, 1, 0, 1, 1]
}
```

Данные в табличной форме выглядят следующим образом (см. Рис. 4.1):

	Рейтинг	Средний балл в школе	Пол
0	75	4	1
1	69	3	1
2	90	5	0
3	100	5	1
4	45	3	1

Рис. 4.1. Табличные данные студентов

Прогнозируемый признак – Рейтинг.

Листинг 4.2.

Прогнозируемый признак

```
Y = dataframe["Рейтинг"]
```

Все остальные признаки (пол, средний балл в школе) используются для обучения дерева.

Листинг 4.3.

Признаки студентов

```
X = dataframe.drop(columns=["Рейтинг"])
```

Сам подбор параметров модели долгий и аналитический процесс, который не столь относится к тестированию, как к аналитике, так что рассматривать их подбор мы не будем.

Разбиение данных происходит поэтапно пополам по всем переменным.

Основной формулой для разбиения используется MAE (см. Рис. 4.2).

$$MAE(y^{true}, y^{pred}) = \frac{1}{N} \sum_{i=1}^N |y_i - f(x_i)|$$

Рис. 4.2. Формула MAE

в нашем случае где:

- $f(x_i)$  – среднее значение рейтинга по выборке;
- $y_i$  – значение рейтинга студента;
- $N$  – количество студентов.

Отсюда, чем меньше абсолютная ошибка, тем точнее работает модель, относительно этого подхода алгоритм решающего дерева разбивает выборку

на подвыборки. Также отметим, что разбиение числовых признаков происходит между соседними значениями признака, то есть если у нас есть оценки 4 и 5, то разбиение будет проходить по 4.5.

Суть алгоритма заключается в том, чтобы отделить друг от друга объекты (показатели студентов) прямыми в многомерном пространстве (равное количеству изучаемых признаков) так, чтобы наиболее похожие признаки были в одной области.

Решение (первая итерация):

- среднее значение признака  $f_{(xi)} = \frac{(75+69+90+100+45)}{5} = 75,8$ ;
- подсчитаем начальное значение MAE  $f_{(xi)} = \frac{abs(75-75,8)+abs(69-75,8)+abs(90-75,8)+abs(100-75,8)+abs(45-75,8)}{5} = \frac{76,8}{5} = 15,36$ .

Теперь нужно разделить записи так, чтобы суммарное значение MAE уменьшилось.

Для этого разделим данные по среднему баллу в школе следующим образом: “Средний балл в школе  $< 4.5$ ”. Тогда в левой подвыборке у нас останутся три записи, а в правой две.

Подсчитаем для каждой подвыборки своё MAE.

Вторая итерация.

Левая ветвь:

- среднее значение признака  $f_{(xi)} = \frac{(75+69+45)}{3} = 63$ ;
- подсчитаем начальное значение MAE  $\frac{abs(75-63)+abs(69-63)+abs(45-63)}{3} = \frac{36}{3} = 12$ .

Правая ветвь:

- среднее значение признака  $f_{(xi)} = \frac{(100+90)}{2} = 95$ ;
- подсчитаем начальное значение MAE  $\frac{abs(90-95)+abs(100-95)}{2} = \frac{10}{2} = 5$ .

Суммарный MAE:  $12 + 5 = 17$

Возьмём другое разделение, чтобы сравнить, насколько эффективное разделение было выбрано изначально. Для этого возьмём разделение “Средний балл в школе  $< 3.5$ ”.

Левая ветвь:

- среднее значение признака  $f_{(xi)} = \frac{(69+45)}{2} = 57$ ;
- подсчитаем начальное значение MAE  $\frac{abs(69-57)+abs(45-63)}{2} = \frac{30}{2} = 15$ .

Правая ветвь:

- среднее значение признака  $f_{(xi)} = \frac{(75+100+90)}{3} = 88,3$ ;
- подсчитаем начальное значение MAE:

$$\frac{abs(75-88,3)+abs(100-88,3)+abs(90-88,3)}{3} = \frac{26,7}{3} = 8,9.$$

Суммарный MAE:  $15 + 8,9 = 29,3$ .

Отсюда понятно, что разделение при первом подходе лучше, так как оно разделяет пространство признаков так, что более похожие объекты находятся рядом. Теперь возьмём для левой ветки разделение вида “Средний балл в школе  $< 3.5$ ”.

Третья итерация.

Левая ветвь:

- среднее значение признака  $f_{(xi)} = \frac{(69+45)}{2} = 57$ ;
- подсчитаем начальное значение MAE  $\frac{abs(69-57)+abs(45-57)}{3} = \frac{24}{2} = 12$ .

Правая ветвь:

Остался один элемент со значением рейтинга 75.

Суммарный MAE:  $12 + 0 = 12$ .

Для другой же ветки (правой) возьмём разделение вида “Пол  $\leq 0.5$ ”.

Левая ветвь:

Остался один элемент со значением 90.

Правая ветвь:

Остался один элемент со значением 100.

Суммарный MAE:  $0 + 0 = 0$ .

В итоге у нас получилась готовая модель, которая может прогнозировать данные.

Код аналогичной модели:

Листинг 4.4.

Модель решающего дерева

```
model = DecisionTreeRegressor(criterion='absolute_error')
model.fit(X, Y)
```

Если мы визуализируем наше дерево, то получится следующий график (см. Рис. 4.3):

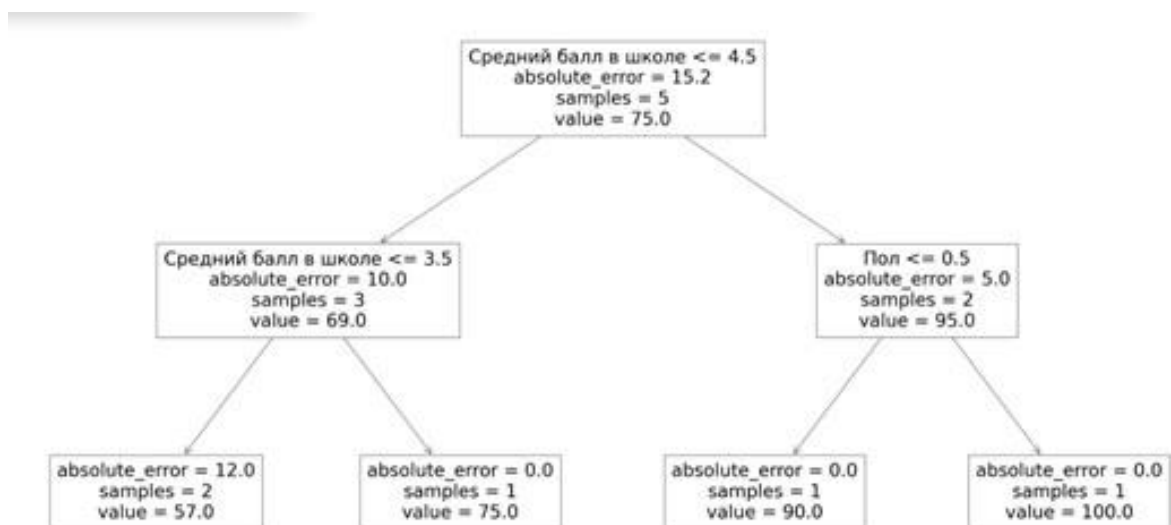


Рис. 4.3. Решающие дерево.

Теперь спрогнозируем будущий рейтинг студента со следующими данными: средний балл в школе — 5, пол — 0 (мужской). Если смотреть по дереву, то прогнозируемый рейтинг студента будет составлять 90.

Код прогнозирования:

Код прогнозирования

```
new_student = pd.DataFrame({"Средний балл в школе": [5], "Пол": [0]})  
res = model.predict(new_student)
```

Что же покажет нам модель (см. Рис. 4.4):

```
1 print("Прогнозируемый балл студента:", *res)
```

Прогнозируемый балл студента: 90.0

Рис. 4.4. Прогнозируемый моделью рейтинг

Как показано прогнозируемый моделью балл соответствует подсчитанному баллу нами вручную, что показывает, что тестирование было успешно пройдено. Следовательно, программный модуль работает правильно.

## ГЛАВА 5. РАЗРАБОТКА И РАЗВЁРТОВАНИЕ ПРОДУКТА

В процессе разработки программного продукта клиентская и серверная часть, а также база данных запускается локально, на персональном компьютере программиста:

Таблица 5.1.

Информация о портах, на которых локально запускается веб-приложение

Часть приложения	Порт размещения
Клиентская	5173
Серверная	7000
База данных	5432

Все изменения, сделанные в процессе разработки фиксируются с помощью системы контроля версий git, а репозиторий веб-приложения размещается на сервисе GitLab.

Для развёртывания готового программного продукта был арендован VPS (Virtual Private Server) на хостинге Beget (см. Рис. 5.1).

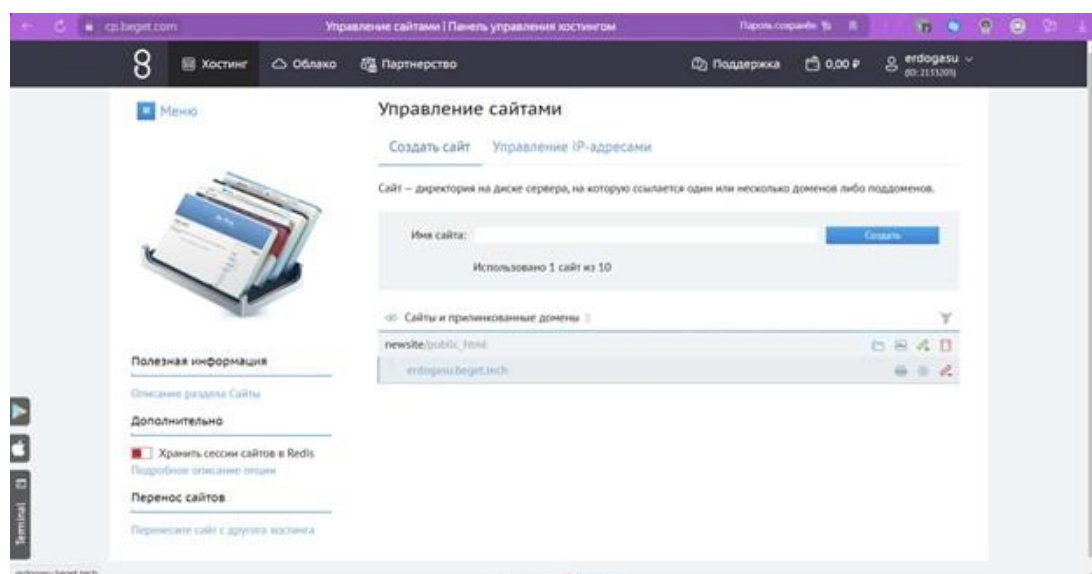


Рис. 5.1. Сайт хостинга

Для удобства, клиентская часть и серверная часть поделены на разные репозитории. Сбора веб-приложения осуществляется с помощью Docker. Сначала запускается контейнер с серверной частью, а после неё запускается контейнер с клиентской частью.

С помощью сервиса Gitlab CI/CD к веб приложению подключена возможность быстро развёртывать приложение в случае необходимости внесения изменения в код. Система настроена таким образом, чтобы после любого изменения в репозитории (если оно было добавлено в основную ветку) автоматически запускать сборку веб-приложения.



## ГЛАВА 6. ЭКСПЛУАТАЦИЯ И СОПРОВОЖДЕНИЕ

После внедрения в эксплуатацию веб-приложения предусмотрено систематическое обнаружение и анализ ошибок и неполадок.

Для обеспечения эффективной связи между пользователями и разработчиками предусмотрена отдельная страничка, где пользователь прямо на сайте может оставить сообщение администратору (см. Рис. 6.1).



The image shows a web form for providing feedback. At the top, the title "FEEDBACK MESSAGE" is displayed in a large, bold, black, monospace-style font. Below the title, the instruction "WRITE YOUR MESSAGE TO THE AUTHOR:" is written in a smaller, bold, black, monospace-style font. Underneath this instruction is a large, empty rectangular text input field with a thin gray border. At the bottom left of the form is a button labeled "BACK" in a black, monospace-style font, and at the bottom right is a button labeled "SEND" in the same style. Centered below the buttons is the text "Last version: 22/04/2024" in a small, black, monospace-style font.

Рис. 6.1. Форма для обратной связи

Или же пользователь можете написать автору по следующим предоставленным контактам (см. Рис. 6.2):

# ABOUT AUTHOR

NAME SURNAME: DENIZ ERDOGAN

E-MAIL: ERDOGAN33@MAIL.RU

GIT REPOSITORY: [\\*CLICK HERE\\*](#)

BIOGRAPHY: 22 YEARS OLD. FROM RUSSIA.  
DATA SCIENTIST. STUDENT OF  
UNIVERSITY. STUDENT OF  
SCHOOL 21.

BACK

Last version: 22/04/2024

Рис. 6.2. Контакты автора

Также в конце каждой страницы написано последнее обновление сайта.