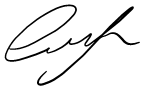
**ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ**

**НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ**

**«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»**

Факультет компьютерных наук Департамент программной инженерии

|  |  |
| --- | --- |
| СОГЛАСОВАНО  Научный руководитель,  сотрудник факультета компьютерных наук, ассистент департамента анализа данных и искусственного интеллекта  А. О. Боревский  « » 2024 г. | УТВЕРЖДАЮ  Академический руководитель образовательной программы  «Программная инженерия»    Н. А. Павлочев  «  » 2024 г. |

 **СТАТИСТИКА ФУТБОЛИСТОВ И ТРАНСФЕРНЫЙ РЫНОК:   
МАТЕМАТИЧЕСКИЙ АНАЛИЗ**

|  |  |
| --- | --- |
| Подп. и дата |  |
| Инв. № дубл. |  |
| Взам. Инв. № |  |
| Подп. и дата |  |
| Инв. № подл. |  |

**Программа и методика испытаний   
ЛИСТ УТВЕРЖДЕНИЯ**

**RU.17701729.11.04-01 ПМИ 01-1-ЛУ**

Исполнители:

Студент группы БПИ-224

/ А. Г. Алексеев /

Студент группы БПИ-221

/ С. Ю. Панкратов /

Студент группы БПИ-216

/ Р. Р. Сайфутдинов /

«24» марта2024 г.

УТВЕРЖДЕН

RU.17701729.11.04-01 ПМИ 01-1-ЛУ

**СТАТИСТИКА ФУТБОЛИСТОВ И ТРАНСФЕРНЫЙ РЫНОК:   
МАТЕМАТИЧЕСКИЙ АНАЛИЗ**   
**Программа и методика испытаний**

|  |  |
| --- | --- |
| Подп. и дата |  |
| Инв. № дубл. |  |
| Взам. Инв. № |  |
| Подп. и дата |  |
| Инв. № подл. |  |

**RU.17701729.11.04-01 ПМИ 01-1**

**Листов 21**

**АННОТАЦИЯ**

Программа и методика испытаний — это документ, в котором содержится информация о программном продукте, а также полное описание приемочных испытаний для данного программного продукта.

Настоящая Программа и методика испытаний для « Статистика футболистов и трансферный рынок: математический анализ » содержит следующие разделы:  
«Объект испытаний», «Цель испытаний», «Требования к программе», «Требования к программной документации», «Средства и порядок испытаний», «Методы испытаний», «Приложения».

В разделе «Объект испытаний» указано наименование, краткая характеристика и назначение программы.

В разделе «Цель испытаний» указана цель проведения испытаний.

Раздел «Требования к программе» содержит основные требования к программе, которые подлежат проверке во время испытаний (требования к функционалу).

Раздел «Требования к программным документам» содержит состав программной документации, которая представляется на испытания.

Раздел «Средства и порядок испытаний» содержит информацию о технических и программных средствах, которые следует использовать во время испытаний, а также порядок этих испытаний.

Раздел «Методы испытаний» содержит информацию об используемых методах испытаний.

Настоящий документ разработан в соответствии с требованиями:

1) ГОСТ 19.101-77 Виды программ и программных документов [1];  
2) ГОСТ 19.102-77 Стадии разработки [2];  
3) ГОСТ 19.103-77 Обозначения программ и программных документов [3]; 4) ГОСТ 19.104-78 Основные надписи [4];  
5) ГОСТ 19.105-78 Общие требования к программным документам [5];   
6) ГОСТ 19.106-78 Требования к програм1мным документам, выполненным печатным способом [6];  
7) ГОСТ 19.301-79 Программа и методика испытаний. Требования к содержанию и оформлению [10].

Изменения к данному документу оформляются согласно ГОСТ 19.603-78 [8], ГОСТ 19.604-78 [9].

**СОДЕРЖАНИЕ**

[ГЛОССАРИЙ 5](#_Toc162298977)

[1. ОБЪЕКТ ИСПЫТАНИЙ 6](#_Toc162298978)

[1.1. Наименование программы 6](#_Toc162298979)

[1.2. Краткая характеристика области применения 6](#_Toc162298980)

[2. ЦЕЛЬ ИСПЫТАНИЙ 7](#_Toc162298981)

[2.1. Документы, на основании которых ведется разработка 7](#_Toc162298982)

[3. ТРЕБОВАНИЯ К ПРОГРАММЕ 8](#_Toc162298983)

[3.1.1. Состав выполняемых функций 8](#_Toc162298984)

[3.1.2. Организация хранения данных 11](#_Toc162298985)

[3.1.3. Организация выходных данных 12](#_Toc162298986)

[3.1.4 Требования к интерфейсу 13](#_Toc162298987)

[4. ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ 15](#_Toc162298988)

[4.1. Состав программной документации 15](#_Toc162298989)

[4.2. Специальные требования к программной документации 15](#_Toc162298990)

[5. СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ 16](#_Toc162298991)

[5.1. Технические средства, используемые во время испытаний 16](#_Toc162298992)

[5.2. Программные средства, используемые во время испытаний 16](#_Toc162298993)

[5.3. Порядок проведения испытаний 16](#_Toc162298994)

[6. СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ 17](#_Toc162298995)

[6.1. Проверка требований к технической документации 17](#_Toc162298996)

[6.2. Проверка требований к функциональным характеристикам 17](#_Toc162298997)

[6.3. Проверка требований к надёжности 18](#_Toc162298998)

[СПИСОК ИСТОЧНИКОВ 19](#_Toc162298999)

[ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ 21](#_Toc162299000)

# ГЛОССАРИЙ

1. Пользователь – человек, который использует продукт «Статистика футболистов и трансферный рынок: математический анализ».
2. СУБД– система управления базами данных.
3. Датасет - коллекция данных, которые обрабатываются и анализируются в рамках проекта.
4. Признаки (features) - отдельные характеристики или параметры, которые используются в модели машинного обучения для предсказания стоимости футболиста.
5. ML-алгоритм (machine learning algorithm) - систематическая процедура, которую компьютер использует для создания модели из данных. В этом проекте, ML-алгоритм будет использоваться для предсказания трансферной стоимости футболиста.
6. Feature importance - мера, которая определяет, какой существенный вклад признаки вносят в предсказания модели.
7. Трансферный рынок - сфера в спортивной индустрии, где футбольные клубы могут купить или продать права на игроков.
8. Предсказание - результат работы модели машинного обучения, которое указывает на предполагаемую трансферную стоимость футболиста.

# ОБЪЕКТ ИСПЫТАНИЙ

# Наименование программы

Наименование программы – «Статистика футболистов и трансферный рынок: математический анализ».

Наименование программы на английском языке – «Football player statistics and the transfer market: a mathematical analysis».

Наименование программы для пользователя – «FootPricing».

# Краткая характеристика области применения

Целевая аудитория программы для предсказания трансферной стоимости футболистов – это спортивные менеджеры и агенты, работающие в футбольных клубах, специалисты отдела аналитики, а также любые заинтересованные стороны, включая футбольные академии, футболистов и их скаутинговые агентства, студентов и исследователей в области спортивной аналитики и любителей футбола, которые желают иметь представление о ценообразовании на трансферном рынке и оценку футбольных талантов на основе отслеживаемых показателей и статистических данных игрока.

Благодаря этой программе менеджеры могут получить прогноз справедливой трансферной стоимости конкретного футболиста по определённым статистическим и физическим показателям, собранным на основе его выступлений за клуб и сборную, что поможет им в проведении трансферной политики их команды. Так, например, при проведении переговоров между клубами о продаже/покупке игрока команда может ссылаться на прогноз для установления справедливой цены трансфера, чтобы не занизить/завысить стоимость футболиста, тем самым сэкономив свои ресурсы. Помимо этого, программа предполагает возможность предоставления обоснования прогнозируемой стоимости на основе показателей, которые использовались в алгоритме, что поможет в проведении аналитических исследований. Так, например, можно будет определить, какие факторы сильнее всего влияют на трансферную стоимость футболиста, чтобы отбирать потенциальных игроков для последующей продажи.

# ЦЕЛЬ ИСПЫТАНИЙ

# Документы, на основании которых ведется разработка

Целью испытаний является проверка корректности выполнения программой функций, изложенных в п. 4 «Требования к программе» настоящего Технического задания из комплекта документации в соответствии с ЕСПД (Единой системой программной документации).

# ТРЕБОВАНИЯ К ПРОГРАММЕ

* 1. **Требования к функциональным характеристикам**

**3.1.1. Состав выполняемых функций**

**3.1.1.1. Сбор данных**

**Цель:** сбор необходимых данных о футболистах, включая статистическую информацию с сайта «Transfermarkt» и количество подписчиков в инстаграм через сайт «Pixwox», для последующего анализа и предсказания трансферной стоимости игрока.

**Задачи:**

1. Адаптация существующего парсера для сбора данных с сайта «Transfermarkt».

2. Разработка механизма для сбора информации о количестве подписчиков в инстаграм через альтернативные сервисы, такие как «Pixwox».

3. Сбор данных, даже если они содержат отсутствующие значения, с последующей их обработкой на более позднем этапе.

**Требования к данным:**

- Актуальность: данные должны быть собраны на момент последнего доступного обновления источников.

- Полнота: данные могут содержать отсутствующие значения, которые будут обрабатываться на последующем этапе.

- Точность: при сборе данных необходимо стремиться к максимальной точности в рамках доступных источников.

**Методы сбора данных:**

- Использование скриптов автоматизированного парсинга для сбора данных с «Transfermarkt».

- Запросы к «Pixwox» для получения данных о подписчиках в инстаграм в связи с ограничениями доступа к «Instagram» на территории РФ.

**Формат данных:**

- Данные будут сохраняться в «.xlsx» формате для последующего удобства обработки и анализа.

**Процедура:**

1. Запуск парсера для сбора необходимых данных о футболистах с сайта «Transfermarkt».

2. Сбор данных о подписчиках футболистов через сервис Pixwox ввиду текущих ограничений доступа к «Instagram».

3. Сохранение собранных данных с допусками на отсутствующие значения в соответствующих полях таблицы для их последующей обработки.

**Контрольные точки:**

- Получение исходных данных о футболистах в соответствии с заданными параметрами и разрешенными отсутствующими значениями.

- Проверка наличия данных о количестве подписчиков через альтернативные источники информации.

- Подготовка комплекта собранных данных к следующему этапу обработки.

**Критерии приемки:**

- Данные успешно собраны, и количественные показатели соответствуют ожиданиям проекта.

- Присутствие отсутствующих значений допускается, при условии их успешной обработки на последующих этапах.

* + - 1. **Предобработка данных**

**Цель**: подготовить собранные данные к этапу их анализа и обучения модели, обеспечив их чистоту, однородность и пригодность для использования в предсказательных моделях.

**Задачи:**

1. Идентификация и обработка выбросов и аномальных значений в наборе данных.

2. Заполнение или удаление пропущенных значений в соответствии с лучшими практиками обработки данных.

3. (Опционально) Кодирование категориальных данных для перевода их в числовой формат, пригодный для обработки моделями машинного обучения.

4. (Опционально) Масштабирование или нормализация числовых признаков для выравнивания диапазонов значений.

5. Автоматизация процедур предобработки для обеспечения воспроизводимости результатов.

**Требования к данным:**

- Корректность: данные должны соответствовать реалиям футбольного мира и быть проверенными на отсутствие очевидных ошибок.

- Согласованность: данные должны быть представлены в едином формате, чтобы обеспечить их совместимость на этапе анализа.

- Нормализованность: данные подвергнуты нормализации для обеспечения справедливости последующего анализа.

**Методы предобработки данных:**

- Использование статистических методов для выявления и удаления выбросов и аномалий.

- Кодирование категориальных признаков при помощи one-hot encoding, label encoding или прочих подходящих методов.

- Выделение новых признаков из исходных при помощи преобразований.

**Контрольные точки:**

- Готовность набора данных без выбросов и аномалий к следующему этапу анализа.

- Отсутствие пропущенных значений или их корректная обработка в исходных данных.

- Преобразование необходимых категориальных признаков в формат, пригодный для машинного обучения.

- Подтверждение эффективности проведенной нормализации данных.

**Критерии приемки:**

- Исходные данные очищены, предобработаны и унифицированы для обеспечения их готовности к использованию в математических моделях.

- Проведенная предобработка не искажает первоначальное содержание данных и не ухудшает их качества для прогнозирования.

**3.1.1.3. Реализация моделей машинного обучения**

**Цель:** создание и оптимизация регрессионной модели CatBoost для предсказания трансферной стоимости футболиста на основе обработанных данных о его характеристиках и спортивной деятельности.

**Задачи:**

1. Подготовка данных для обучения, валидации и тестирования модели.

2. Реализация модели CatBoost для выполнения регрессионных задач.

3. Подбор оптимальных гиперпараметров модели с использованием Optuna для достижения лучших результатов предсказания.

4. Тестирование различных функций потерь для оптимизации качества модели.

5. Реализация возможности логарифмирования целевой переменной для улучшения производительности модели в случае её нелинейных зависимостей.

**Требования к модели:**

- Точность: модель должна обеспечивать высокую степень точности предсказаний трансферной стоимости.

- Обобщающая способность: модель должна показывать хорошие результаты как на обучающем наборе данных, так и на валидационном и тестовом.

- Стабильность: модель должна быть устойчива к малым изменениям в входных данных.

**Методы реализации:**

- Использование библиотеки CatBoost для построения регрессионной модели.

- Применение библиотеки Optuna для нахождения наилучших гиперпараметров модели.

- Анализ различных функций потерь и выбор наиболее подходящей для конкретной задачи и данных.

- Эксперименты с логарифмированием целевой переменной при наличии сильного перекоса в распределении.

**Процедура:**

1. Подготовка обучающего, валидационного и тестового наборов данных.

2. Инициализация базовой модели CatBoost с предварительно выбранными гиперпараметрами.

3. Проведение оптимизации гиперпараметров с помощью Optuna для улучшения результатов модели.

4. Выбор и применение оптимальной функции потерь на основе полученных результатов.

5. При необходимости, трансформация целевой переменной посредством логарифмического преобразования.

**Контрольные точки:**

- Валидность и готовность исходных данных для обучения модели.

- Успешный отбор и оптимизация гиперпараметров модели.

- Выбор и подтверждение адекватности функции потерь.

- Проверка качества и точности предсказаний на тестовых данных.

**Критерии приемки:**

- Оптимизация модели приводит к улучшению её обобщающей способности без переобучения.

- Результаты предсказаний модели соответствуют реалистичным ожиданиям относительно трансферной стоимости футболистов.

- Модель способна корректно работать с различными вариантами представления целевой переменной (например, с логарифмированными значениями).

# Организация хранения данных

**Цель**: структурировать и систематизировать хранение обязательных данных о футболистах в файлах «.xlsx» для их последующего использования в аналитических и моделирования целях.

**Задачи:**

1. Выбрать оптимальный формат организации данных в табличном файле «.xlsx» для удобства доступа и манипуляций.

2. Организовать данные таким образом, чтобы они были легко интегрированы с инструментами анализа данных и машинного обучения.

3. Обеспечить легкость обновления данных в «.xlsx» файлах вручную или с помощью автоматизированных инструментов.

**Требования к характеристикам данных:**

- Полнота: включение всех необходимых и обязательных данных для каждого футболиста в таблицу.

- Актуальность: информация должна быть актуальной и регулярно обновляемой.

- Доступность: данные должны легко извлекаться для анализа и моделирования.

**Обязательные характеристики данных:**

- Добавление в файл всех ключевых атрибутов, описанных в предыдущем этапе работы над проектом.

**Методы организации хранения:**

- Размещение данных в табличных файлах формата «.xlsx», используя механизмы, предоставляемые программой «Microsoft Excel» или совместимыми приложениями.

- Определение структурированного формата таблицы для удобного отображения данных и последующего их анализа.

**Контрольные точки:**

- Создание структурированных «.xlsx» файлов с обязательными данными для всех футболистов.

- Проверка файлов на наличие ошибок и актуальность информации.

- Успешная интеграция «.xlsx» файлов с другими этапами проекта.

**Критерии приемки:**

- Удобство и эффективность системы хранения данных в «.xlsx» файлах, включая возможность простого анализа и обновления.

- Все ключевые данные о футболистах успешно структурированы по соответствующим категориям, доступны для просмотра и готовы к использованию в модели.

# Организация выходных данных

**Цель**: Обеспечить представление результатов предсказаний трансферной стоимости футболистов и сопутствующей аналитики в доступных и понятных форматах для конечного пользователя.

**Задачи:**

1. Представление результатов предсказаний в виде ясных и однозначных числовых значений.

2. Визуализация ключевых данных для облегчения их интерпретации и улучшения восприятия.

3. Сохранение результатов анализа и SHAP-значений в удобной форме для последующего доступа и анализа.

**Требования к выходным данным:**

- Доступность: данные должны быть предоставлены в форме, легко интерпретируемой конечным пользователем.

- Содержательность: наличие минимально необходимого набора метрик и характеристик для объяснения модели.

**Методы реализации:**

1. Передача числовых результатов предсказаний через таблицы «.xlsx», применяя четкое и однозначное форматирование.

2. Применение базовых визуализаций для представления статистических сводок по трансферной стоимости и важности признаков.

3. Выгрузка данных из SHAP для иллюстрации влияния каждого признака на прогнозируемое значение.

**Процедура:**

1. Собрать результаты предсказаний и сохранить их в «.xlsx» файле со всеми соответствующими метриками.

2. Подготовить графики и диаграммы с использованием доступных программных средств, таких как Microsoft Excel или Python с библиотеками Matplotlib или Seaborn.

3. Внедрить процесс получения SHAP-значений для объяснения предсказаний в контексте вклада каждого признака.

**Контрольные точки:**

1. Ясность и точность представления числовых прогнозов.

2. Адекватность выбранных типов визуализаций для представляемых данных.

3. Включение SHAP значений для подтверждения надежности прогнозов.

**Критерии приемки:**

- Числовые предсказания четко сформулированы и легко интерпретируемы.

- Визуализации корректно отражают результаты анализа и прогнозы, поддерживая понимание.

- Данные о влиянии признаков на прогнозы предоставлены и доступны для осмысления и дальнейшего анализа.

# Требования к интерфейсу

**Цель**: упростить процесс работы с программой для предсказания трансферной стоимости футболиста путём создания удобной в использовании программной среды без разработки отдельного пользовательского интерфейса.

**Задачи:**

1. Организовать набор «.ipynb» файлов для управления работой модели и её настройками.

2. Упростить ввод данных и параметров моделирования для пользователя за счёт чёткой и понятной структуры ноутбуков Jupyter.

3. Обеспечить в «.ipynb» файлах удобное введение и изменение данных, необходимых для запуска и тестирования модели.

**Требования к «конфигурации»**:

- Понятность: переменные и параметры должны быть названы таким образом, чтобы их назначение было интуитивно понятным.

- Доступность: управление параметрами должно производиться максимально просто, без необходимости изменения исходного кода модели.

- Переносимость: файлы должны легко запускаться на различных устройствах и платформах поддерживающих «Jupyter Notebook».

**Методы реализации:**

1. Использование «Jupyter Notebook» для создания интерактивных «.ipynb» файлов.

2. Структурирование ноутбуков с ясным разделением на блоки: подготовка данных, настройка параметров модели, тренировка модели, валидация результатов и вывод предсказаний.

3. Вставка комментариев и пояснений в ноутбуки для облегчения понимания логики работы и управления процессами.

**Процедура:**

1. Разработать шаблоны «Jupyter Notebook» с предустановленными ячейками для ввода данных и параметров модели.

2. Создать чёткую инструкцию по использованию ноутбуков, чтобы пользователи могли самостоятельно менять нужные «конфигурации».

3. Убедиться, что изменение параметров модели не требует глубокого погружения в код и доступно пользователям с базовыми навыками программирования.

**Контрольные точки:**

1. Созданы ноутбуки Jupyter с чёткой структурой и инструкциями.

2. Безошибочное управление параметрами модели через изменение значения переменных в ноутбуках.

3. Успешное выполнение процесса обучения и предсказания модели с использованием изменённых «конфигураций».

**Критерии приемки:**

- Ноутбуки Jupyter обеспечивают централизованное и удобное управление всеми процессами обработки и анализа данных.

- Пользователь может легко модифицировать вводные данные и параметры моделирования без вмешательства в программный код.

- Программная среда работает стабильно и надёжно на разных устройствах и позволяет получить точные результаты предсказаний модели.

# ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

# Состав программной документации

* + 1. «Статистика футболистов и трансферный рынок: математический анализ». Техническое задание (ГОСТ 19.201-78) [7];
    2. «Статистика футболистов и трансферный рынок: математический анализ». Программа и методика испытаний (ГОСТ 19.301-78) [10];
    3. «Статистика футболистов и трансферный рынок: математический анализ». Пояснительная записка (ГОСТ 19.404-79) [11];
    4. «Статистика футболистов и трансферный рынок: математический анализ». Руководство программиста (ГОСТ 19.504-79) [12];
    5. «Статистика футболистов и трансферный рынок: математический анализ». Текст программы (ГОСТ 19.401-78) [13];

# Специальные требования к программной документации

1. Документы к программе должны быть выполнены в соответствии с ГОСТ 19.106-78 и ГОСТами к каждому виду документа (см. п. 5.1).
2. Пояснительная записка должна быть загружена в систему Антиплагиат через LMS «НИУ ВШЭ». Лист, подтверждающий загрузку пояснительной записки, сдается в учебный офис вместе со всеми материалами не позже, чем за день до защиты курсовой работы.
3. Вся документация также воспроизводится в печатном виде, она должна быть подписана академическим руководителем образовательной программы 09.03.04 «Программная инженерия», руководителем разработки и исполнителем перед сдачей курсовой работы в учебный офис, не позже одного дня до защиты.
4. Документация также сдается в электронном виде в формате .pdf или .docx, а программа – в архиве формата .zip или .rar.
5. Все документы перед защитой курсовой работы должны быть загружены в информационно-образовательную среду НИУ ВШЭ LMS (Learning Management System) в личном кабинете, дисциплина – «Курсовой проект, 3 курс ПИ», одним архивом.

# СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ

# Технические средства, используемые во время испытаний

В рамках тестирования программы для предсказания трансферной стоимости футболиста использовался следующий конфигурационный набор оборудования:

1) Операционная система: Windows 10, для обеспечения совместимости с используемыми аналитическими инструментами и Jupyter Notebook.

2) Процессор: Intel Core i5-9400F, предоставляющий достаточную мощность для эффективной работы моделей машинного обучения.

3) Оперативная память: 8 Гб, что позволяет уверенно обрабатывать большие объемы данных и избегать задержек в выполнении операций.

4) Жесткий диск: 256 Гб SSD, с 150 Гб свободной памяти, гарантирующий быстрый доступ к данным и программному обеспечению, ускоряя процесс тестирования и анализа.

5) Подключение к сети Интернет, что является необходимым компонентом для доступа к актуальным данным, обновлению программных пакетов и взаимодействия с потенциальными онлайн-сервисами.

# Программные средства, используемые во время испытаний

Во время испытаний использовались следующие программные средства:  
1) Jupyter Notebook  
2) Дистрибутив Anaconda версии 2023.3

# Порядок проведения испытаний

Испытания должны проводиться в следующем порядке:  
1) Установить «.ipynb» файлы программы из [репозитория](https://github.com/rafaeldks/Football-Player-Statistics-and-Transfer-Market-Math-Perspective);  
2) Последовательно запустить «.ipynb» файлы программы;  
3) Провести испытания, описанные в разделе «Методика испытаний»;  
4) Завершить работу «.ipynb» файлов и закрыть их.

# СРЕДСТВА И ПОРЯДОК ИСПЫТАНИЙ

# Проверка требований к технической документации

Состав программной документации проверяется наличием полного комплекта документов программной документации в системе SmartLMS и наличием всех требуемых подписей. Также проверяется соответствие документации требованиям ГОСТ.

Комплект документов полный. Все документы удовлетворяют представленным требованиям.

# Проверка требований к функциональным характеристикам

**Процесс сбора данных:**

1. Корректность отправки get-запросов:

- Проверить, что get-запросы корректно отправляются на страницы игроков и клубов платформы Transfermarkt.

- Убедиться, что сервер не блокирует множество запросов от одного источника как подозрительную активность, принимая во внимание количество посылаемых запросов и интервалы времени между ними.

2. Аккуратность извлечения данных:

- Установить, что библиотека «beautiful soup» правильно выделяет требуемую статистику из HTML страницы: базовую информацию об игроке, информацию о клубной статистике и т.д.

- Удостовериться, что собранная информация соответствует последним обновлениям данных на платформе.

3. Объединение и сохранение данных:

- Проверить, что собранные данные правильно объединяются в один датафрейм Pandas.

- Подтвердить, что датафрейм сохраняется в формате Excel без потери данных и искажения структуры.

**Процесс предобработки данных:**

1. Проверка корректности конвертации данных:

- Убедиться в точности перевода столбца возраста, стоимости и дат в числовые типы данных с проверкой на адекватность распознанных значений.

2. Проверка правильности разделения и формирования новых столбцов:

- Подтвердить, что статистика выступлений клуба и игрока, а также все остальные выделенные признаки соответствуют заявленной структуре и используют актуальные данные.

3.Применение One-Hot кодирования:

- Проверить, что One-Hot кодирование для переменных, таких как сильная нога игрока, выполняется корректно без потери информации.

4. Замена пропусков данных:

- Убедиться, что замены пропущенных значений на средние или нули осуществляется правильно с сохранением целостности набора данных.

**Процесс обучения модели:**

1. Проверка конфигураций обучения:

- Убедиться, что конфигурации обучения (выбор функции потерь, опция обучения только на топ-5 лигах, опция логарифмирования таргета) задаются корректно и приводят к изменениям в процессе обучения модели.

- Проверить правильность загрузки датасета по заданному пути и его соответствие требуемой структуре.

2. Проверка наличия пропусков в данных:

- Удостовериться в том, что перед обучением модели производится проверка датасета на наличие пропусков и в случае их обнаружения предпринимаются соответствующие действия.

3. Проверка выделения целевой переменной и деления данных:

- Проверить корректность выделения целевой переменной и корректное разделение данных на обучающую и тестовую выборки.

4. Проверка оптимизации гиперпараметров:

- Проверить эффективность подбора гиперпараметров модели при помощи библиотеки Optuna и соответствие полученных параметров оптимальному пространству параметров, определенному в исследовании.

5. Проверка качества модели:

- Удостовериться в том, что после обучения модели выводятся все заявленные метрики качества (RMSE, MAE, R^2, MEDAE, EVS, MAPE), и они осмысленны в контексте поставленной задачи.

- Проверить адекватность переобучения модели на всем наборе данных перед предсказанием на новых данных.

6. Проверка сохранения обученной модели:

- Подтвердить, что процесс обучения завершается сохранением модели в файл «best\_model.cbm» для последующего использования.

# Проверка требований к надёжности

Произведя проверку функциональных требований в п. 6.2, можно сделать вывод, что требования к надежности удовлетворены

# СПИСОК ИСТОЧНИКОВ

1. ГОСТ 19.101-77 Виды программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
2. ГОСТ 19.102-77 Стадии разработки. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
3. ГОСТ 19.103-77 Обозначения программ и программных документов. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
4. ГОСТ 19.104-78 Основные надписи. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
5. ГОСТ 19.105-78 Общие требования к программным документам. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
6. ГОСТ 19.106-78 Требования к программным документам, выполненным печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
7. ГОСТ 19.201-78 Техническое задание. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
8. ГОСТ 19.603-78 Общие правила внесения изменений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
9. ГОСТ 19.604-78 Правила внесения изменений в программные документы, выполненные печатным способом. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
10. ГОСТ 19.404-79 Программа и методика испытаний. Требования к содержанию и оформлений. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
11. ГОСТ 19.404-79 Пояснительная записка. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
12. ГОСТ 19.504-79 Руководство программиста. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
13. ГОСТ 19.401-78 Текст программы. Требования к содержанию и оформлению. // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
14. ГОСТ 19.301-79 Программа и методика испытаний. Требования к содержанию и оформлению // Единая система программной документации. – М.: ИПК Издательство стандартов, 2001.
15. «Transfermarkt» [Электронный ресурс] – Сайт – Режим доступа: свободный, ссылка:

* https://www.transfermarkt.com/

# ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Лист регистрации изменений | | | | | | | | | |
| Номера листов (страниц) | | | | | Всего листов (страниц в докум.) | № документа | Входящий  № сопроводит ельного докум. и дата | Подп. | Дата |
| Изм. | Измененных | Замененных | Новых | Аннулированн ых |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |