Этап 1. Описание предметной области

**Введение в предметную область**

Футбол является одним из самых популярных видов спорта, привлекающим внимание миллионов болельщиков. Современные технологии и доступность данных позволяют использовать нейронные сети для анализа футбольных матчей и прогнозирования результатов.

Существует множество факторов, связанных с футбольными матчами, которые могут быть прогнозируемыми и которые могут учитываться при предсказании исхода игры.

Эти факторы могут включать в себя различные статистические данные, такие как процент владения мячом, количество ударов по воротам, количество фолов или угловых и т.д. Некоторые из этих факторов могут быть более предсказуемыми, так как они больше зависят от стиля игры команды и меньше подвержены влиянию случайных факторов. К таким факторам можно отнести, например, количество фолов или количество угловых. Их значения часто связаны с общим стилем и тактикой команды.

Однако есть и другие факторы, которые хуже поддаются прогнозированию. Например, количество желтых карточек или количество забитых голов и сам результат матча. Эти факторы больше зависят от случайных событий, таких как погода, настрой команды, удача, решения судьи. Их прогнозирование может быть более сложным и подверженным большим колебаниям.

В контексте данного проекта было принято решение сосредоточиться на разработке модели, которая будет прогнозировать результат футбольного матча на основе предсказания количества голов. Это решение обусловлено популярностью данной метрики.

При прогнозировании результатов футбольных матчей широко используются различные статистические метрики, которые помогают анализировать и оценивать шансы команд на победу. Две из наиболее популярных метрик - это xG (Expected Goals) и рейтинг ELO.

**xG (Expected Goals):** Эта метрика представляет собой статистический показатель, который оценивает вероятность забить гол в конкретной ситуации на основе множества факторов (расстояние до ворот, угол атаки, тип удара и т.д). xG позволяет определить, насколько опасной была атака и каковы шансы на гол. Чем выше xG, тем больше вероятность, что команда забьет гол. В данном проекте будут использоваться значения xG, предоставленные сторонними источниками, такими как api-football.com, которые предоставляют статистические данные о футбольных матчах. Эти значения xG будут важными компонентами для прогнозирования количества забитых голов в текущем матче.

**Рейтинг ELO:** Рейтинг ELO — это система рейтингования команд и игроков, которая была разработана для шахмат и впоследствии стала популярной в футболе и других видов спорта. Каждая команда или игрок имеет свой начальный рейтинг, и после каждого матча этот рейтинг обновляется в зависимости от результата игры и рейтинга соперника. Рейтинг ELO позволяет определить относительную силу команд и прогнозировать исходы матчей на основе их рейтингов. В данном проекте для расчета рейтинга ELO будет использоваться собственный алгоритм, учитывающий результаты команд с сезона 2021/2022 и до текущего сезона. Это необходимо, так как отсутствует сторонний ресурс, который предоставлял бы данные о рейтинге ELO для футбольных команд.

**Цель проекта**

Целью данного проекта является создание веб-приложения для предсказания результатов футбольных матчей с использованием нейронной сети. Нейронная сеть будет обучена на основе статистических данных о предыдущих матчах команд и вычислит прогнозируемое количество голов в текущем матче для каждой команды. Это позволит пользователям получать информативные прогнозы и делать обоснованные решения при анализе футбольных событий.

**Задачи для достижения цели**

1. Проанализировать существующие методы прогнозирования и их применимость к проекту.
2. Исследовать доступные источники данных о футбольных матчах и собрать необходимую статистику.
3. Рассмотреть алгоритм вычисления рейтинга ELO для футбольных команд и разработать собственный алгоритм для проекта.
4. Спроектировать базу данных для хранения статистических данных о матчах и рейтингах ELO.
5. Определить параметры и конфигурацию нейронной сети для прогнозирования результатов матчей.
6. Изучить методики тестирования нейронных сетей и провести тестирование разработанной модели.
7. Разработать структуру и интерфейс веб-приложения для предоставления прогнозов пользователю.

**Дополнительная теория**

Для оценки работы прогнозов нейронной сети будут использоваться следующие три метрики:

**Точность прогноза (Accuracy):** Это метрика, которая измеряет долю правильных прогнозов модели относительно общего числа прогнозов. Точность прогноза выражается в процентах и помогает оценить, насколько модель верно классифицирует данные. Высокая точность указывает на хорошую производительность модели.

**Продвинутая метрика F1 (F1-Score):** F1-мера является средним гармоническим между точностью (Precision) и полнотой (Recall) модели. F1-мера помогает оценить способность модели правильно классифицировать данные при учете как ложных положительных, так и ложных отрицательных прогнозов. Формула для расчета F1 метрики представлена на рисунке 1.

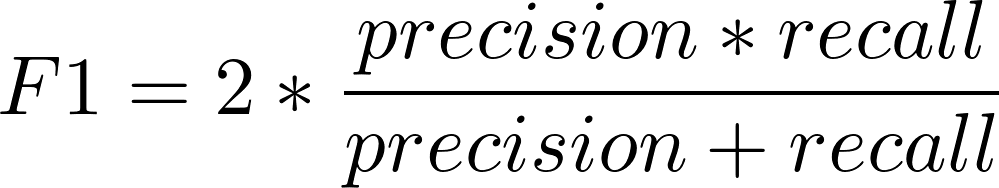


Рисунок 1 – Формула для расчета F1 метрики

**ROI (Return on Investment):** ROI — это метрика, которая оценивает доходность инвестиций. В контексте данного проекта ROI может быть использован для измерения прибыльности ставок, сделанных на основе прогнозов модели. ROI учитывает как доход, так и затраты и позволяет определить, насколько использование прогнозов модели приводит к прибыли или убыткам.

Проблема

Футбольные матчи известны своей непредсказуемостью, вызванной различными факторами, такими как состояние игроков, погодные условия и случайные события. Эта непредсказуемость создает проблему для болельщиков и аналитиков, которые хотели бы иметь инструмент, способный более точно прогнозировать результаты матчей.

Актуальность

В настоящее время можно наблюдать рост интереса к футболу и к ставкам на спорт. Этот интерес поддерживается не только болельщиками, но и привлекает внимание аналитиков и инвесторов.

Поэтому становится актуальным развитие инструментов, которые могли бы более точно анализировать и прогнозировать результаты матчей. Рост доступности статистических данных о футбольных событиях и развитие технологий машинного обучения создают возможности для создания таких инструментов.

Этап 2. Обзор существующих программных средств

Одной из самых популярных платформ, которая используется для прогнозирования результатов футбольных матчей на основе алгоритмов нейронной сети, является xGscore (<https://xgscore.ru/)>. На рисунке 2 представлена главная страница платформы xGscore.

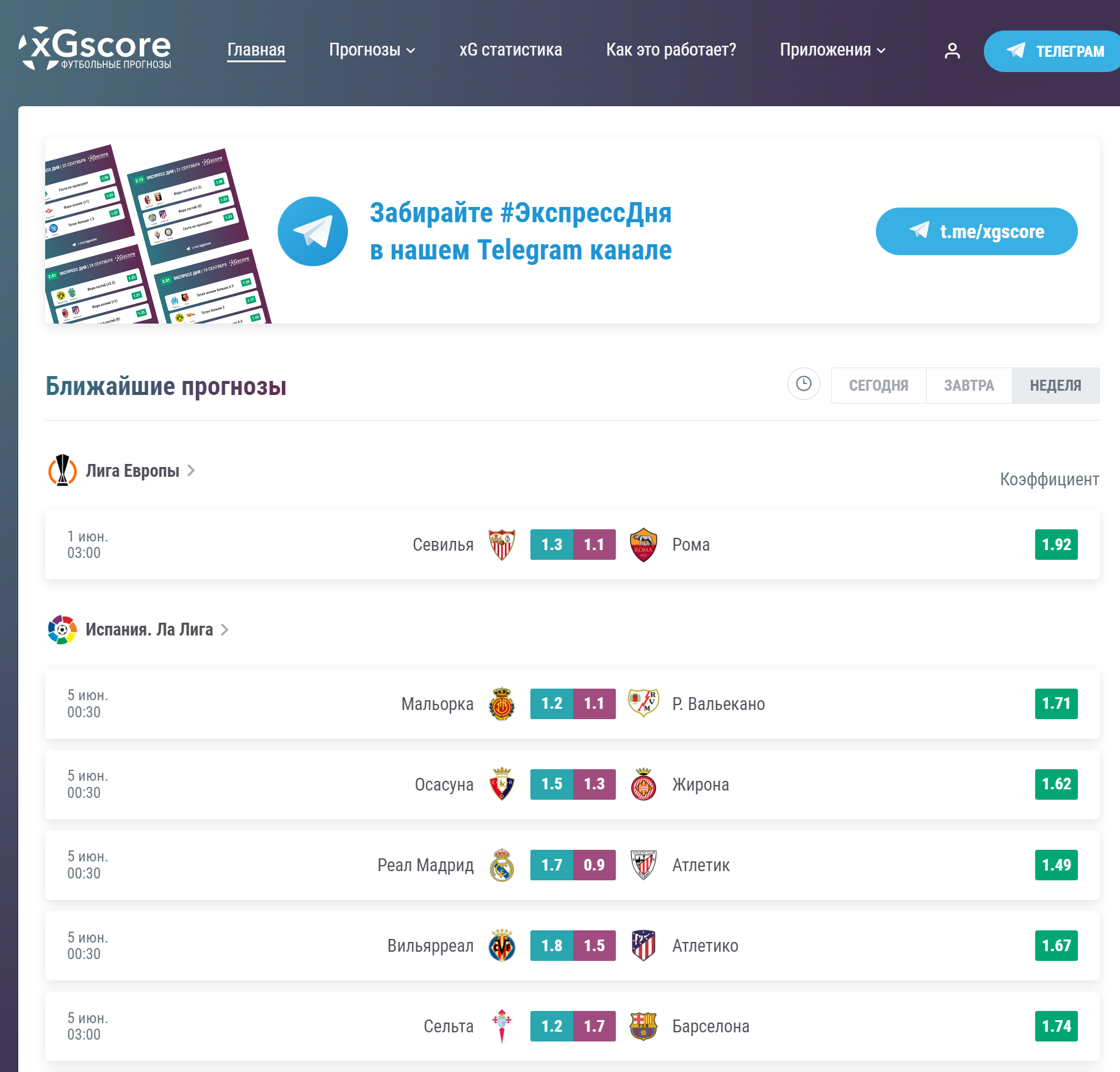


Рисунок 2 – Главная страницы платформы xGscore

Основным фокусом модели прогнозирования xGscore является определение веса каждого прошедшего матча для расчета ожидаемого количества голов. Алгоритм учитывает различные факторы, включая голы и ожидаемые голы (xG), дату матча, место проведения (домашнее или выездное), силу соперника и уровень турнира. Путем суммирования показателей всех матчей формируются данные о забитых и пропущенных голах для хозяев и гостей. Объединение этих показателей позволяет определить ожидаемый счет матча.

Прогнозы на платформе xGscore предоставляются для топ-7 чемпионатов (Англия, Испания, Италия, Германия, Франция, Португалия и Нидерланды), а также для двух главных еврокубков (Лига чемпионов и Лига Европы). Важно отметить, что система не предоставляет прогнозы для Чемпионата России и вероятность включения его в платформу невысока, так как она была разработана украинскими разработчиками.

Исходя из вышесказанного, принято решение разработать систему, сосредоточенную на прогнозировании результатов футбольных матчей для Чемпионата России, а также для Чемпионата Англии (это позволит сравнить результаты прогнозирования, полученные от данной модели, с прогнозами, предоставляемыми платформой xGscore).

**Этап 3. Процесс AS IS vs TO BE**

На рисунке 3 представлена контекстная диаграмма верхнего уровня для процесса футбольного прогнозирования.

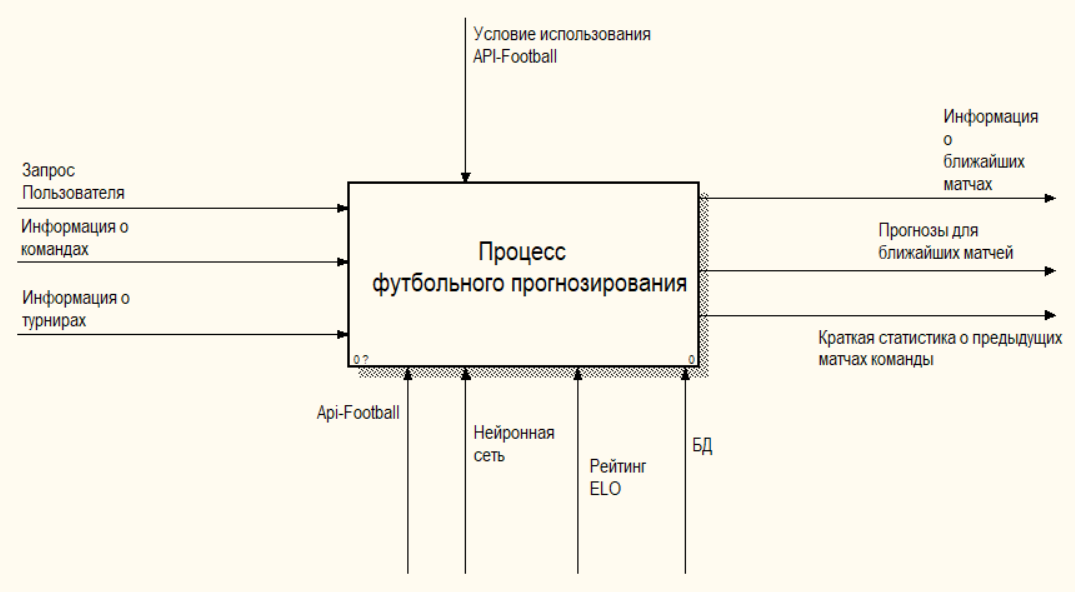


Рисунок 3 – Диаграмма вариантов использования

На рисунке 4 представлена декомпозиция процесса футбольного прогнозирования.

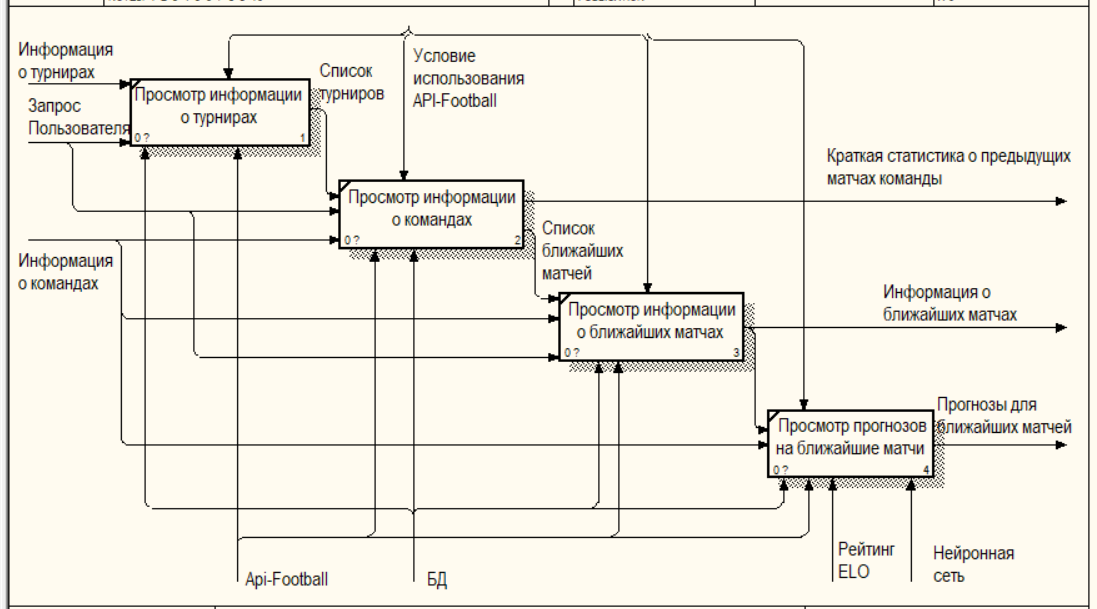


Рисунок 4 – Декомпозиция «Процесс футбольного прогнозирования»

Пользователям доступен просмотр информации о футбольных турнирах, для которых система предоставляет прогнозы. Эта информация включает в себя подробности о турнирах, такие как турнирная таблица, а также информацию об участвующих командах, которая берется из API-Football. Для каждого турнира пользователь может посмотреть информацию по командам и их краткую статистику.

Пользователи также могут узнать о ближайших матчах для каждого из турниров, включая информацию о дате и времени проведения матчей. Кроме того, система предоставляет прогнозы для этих ближайших матчей (следующий тур). Эти прогнозы формируются с использованием нейронной сети, которая анализирует данные о предыдущих результатов команд (полученных из API-Football и хранящихся в БД), а также рейтинг ELO для команд, который рассчитывается внутри системы на основе данных о результатах сезона.

# Этап 4. Описание вариантов использования

На рисунке 5 представлена диаграмма вариантов использования



Рисунок 5 – Диаграмма вариантов использования

Пользователь может просматривать доступные турниры, а также информацию о них, например турнирную таблицу. Также пользователь может просматривать информация о ближайших матчах (дата и время, задействованные команды), а также получать прогнозы по этим матчам (прогнозируемое количество мячей, процент победы хозяев, ничьи и победы гостей). Кроме того, пользователю будет доступна краткая статистика о последних проведенных матчах команды (количество забитых мячей и процент правильности полученного прогноза).

# Этап 6. Выработка требований и постановка задачи

Требования:

1. Вычисление прогнозируемого количества голов для каждой команды в матче с помощью нейронной сети;
2. Своевременное обновление рейтинга ELО;
3. Обеспечение доступа пользователям к прогнозам на ближайшие матчи, информации о турнирах и командах с помощью веб-приложения;
4. Совместимость с данными о сезоне 2022/2023 и последующими данными.

Постановка задачи:

Разработать веб-приложение, основанное на нейронной сети, для предсказания результатов футбольных матчей с целью предоставления пользователям информативных прогнозов.