

## МЕТОДЫ И МОДЕЛИ ПРОГНОЗИРОВАНИЯ СПОРТИВНЫХ СОБЫТИЙ

Константинов А. В., Букреев Д. А.

ФГБОУ ВО «Мелитопольский государственный  
университет», Мелитополь,  
e-mail: dmitriy.bukreev@mel-su.ru

Научный руководитель: Букреев Д. А.

### Введение

Современное развитие цифровых технологий привело к значительному увеличению объёмов спортивной статистики, доступной для анализа, что существенно изменило характер спортивной аналитики и расширило возможности прогнозирования игровых событий. В условиях, когда практически каждый матч сопровождается подробными числовыми и видеоданными, а информационные платформы регулярно обновляют показатели команд и игроков, возрастает необходимость применения интеллектуальных методов для анализа этих данных и построения достоверных прогнозов. В частности, футбол как один из наиболее популярных и статистически насыщенных видов спорта предоставляет исследователям уникальные возможности для выявления скрытых закономерностей и моделирования вероятности спортивных исходов. Традиционные экспертные подходы постепенно уступают место методам статистического анализа, машинного обучения и нейросетевого моделирования. Эти подходы позволяют учитывать широкий спектр факторов – от индивидуальных игровых показателей до трендовых характеристик команд, событий матча и контекстных параметров. Развитие методов искусственного интеллекта, доступность инструментов обработки данных и рост вычислительных мощностей способствуют тому, что прогнозирование спортивных событий приобретает точный, формализованный и воспроизводимый характер. При этом особую значимость приобретает разработка программных решений, способных объединять сбор данных, их подготовку, обучение моделей и визуализацию результатов в единой функциональной системе.

Актуальность данного исследования определяется необходимостью повышения точности прогнозирования и упрощения процедуры аналитической обработки для широкого круга пользователей – от спортивных аналитиков и преподавателей до любителей спорта. Несмотря на существование зарубежных сервисов спортивной аналитики, большинство из них ориентированы на крупные коммерческие платформы и не учитывают специфики локальных чемпионатов или учебно-исследовательских задач. В этой связи представляется важным создание программного инструмента, реализующего современные математические и алгоритмические

подходы, но при этом доступного, прозрачного и адаптируемого под конкретные требования.

Цель исследования – анализ методов и моделей прогнозирования спортивных событий и разработка программного средства, реализующего интеллектуальный подход к предсказанию результатов футбольных матчей.

### Материалы и методы исследования

Материалом исследования выступили статистические данные о футбольных матчах, включающие как базовые показатели команд, так и расширенные характеристики, отражающие их игровую форму, результативность и поведенческие особенности в конкретных соревнованиях. Используемый набор данных был сформирован на основе открытых спортивных ресурсов, агрегаторов статистики и архивов матчей, что позволило получить достаточно репрезентативную выборку для построения моделей прогнозирования.

Первоначальный этап исследования был связан с предварительной обработкой данных. Статистические записи зачастую содержали пропуски, дубли и некорректные значения, поэтому была проведена очистка выборки, унификация форматов и нормализация числовых признаков.

Для выбора наиболее значимых факторов использовались методы корреляционного анализа, статистической значимости признаков и оценка важности атрибутов с помощью алгоритмов машинного обучения, таких как Random Forest и Gradient Boosting. Обучение моделей проводилось на разделённых обучающих и тестовых выборках, что обеспечивало объективную оценку качества прогнозирования. Для оценки эффективности применялись стандартные метрики классификации – accuracy, precision, recall, F1-score, а также показатели качества вероятностных предсказаний, включая log-loss и ROC-AUC. Подход с перекрёстной валидацией позволил минимизировать эффект случайных флуктуаций и повысить достоверность результатов.

### Результаты исследования и их обсуждение

Прогнозирование спортивных событий представляет собой сложную аналитическую задачу, требующую учёта большого количества факторов, различающихся по природе, стабильности и степени влияния на итоговый результат. В отличие от классических областей статистического моделирования, где параметры могут быть относительно устойчивыми во времени, спортивные данные характеризуются высокой динамичностью: состояние команды, тактическая схема, психологические аспекты, мотивация и даже внешние условия способны существенно изменять исход события. Именно поэтому выбор методов прогнозирования требует

комплексного подхода, основанного на сочетании математических, статистических и интеллектуальных моделей.

Статистические методы традиционно являются основой спортивной аналитики благодаря своей интерпретируемости и простоте применения. Наиболее распространённой моделью является распределение Пуассона, которое применяется для прогнозирования количества голов в футбольном матче [1]. Модель исходит из предположения, что количество забитых голов можно описать как редкое событие, происходящее в рамках определённого среднего значения. Это позволяет оценивать вероятности различных счётов и, следовательно, вероятность победы, ничьей или поражения. Другим важным статистическим инструментом является логистическая регрессия, применяемая для оценки вероятности наступления бинарных или многоклассовых исходов [2]. Её преимущество заключается в возможности учёта большого количества независимых переменных и получении вероятностной интерпретации результатов. Однако статистические модели имеют ограниченную способность к работе со сложными нелинейными зависимостями и могут терять точность при наличии значительных шумов в данных или высоком уровне вариативности спортивных показателей.

Байесовские подходы основываются на теории вероятностей и позволяют учитывать априорные знания, постепенно корректируя прогнозы по мере появления новых данных. В спортивной аналитике байесовские сети применяются для моделирования взаимодействий между факторами и построения гибких прогнозов, опирающихся на вероятностное распределение параметров. Преимущество этого подхода проявляется при анализе малых выборок или в условиях, когда параметры системы меняются неравномерно. Однако высокая вычислительная сложность и потребность в точной настройке структуры моделей ограничивают их практическое использование в оперативных системах прогнозирования.

Методы машинного обучения получили широкое распространение благодаря способности выявлять сложные взаимосвязи в данных, которые могут быть недоступны классическим статистическим моделям. Среди наиболее эффективных алгоритмов, используемых в спортивной аналитике, можно выделить:

- деревья решений, позволяющие выявлять правила, определяющие исход матча на основе пороговых разделений данных;
- Random Forest, представляющий собой ансамбль деревьев решений и обеспечивающий высокую устойчивость к шумам и переобучению;
- градиентный бустинг, который последовательно улучшает прогнозы, концентрируясь на наиболее сложных для классификации случаях;

- метод опорных векторов (SVM), эффективно работающий на данных высокой размерности и обеспечивающий чёткие разделяющие поверхности между классами.

Эти методы обладают высокой обобщающей способностью и позволяют формировать более стабильные модели, особенно при правильной настройке параметров и достаточном объёме тренировочной выборки.

Применение нейронных сетей в спортивной аналитике стало возможным благодаря росту вычислительных мощностей и доступности больших массивов данных. В отличие от статистических методов, нейросети способны автоматически выявлять скрытые закономерности, используя многослойные представления данных. В рамках исследования рассматривались полносвязные нейронные сети, а также более сложные архитектуры, применяемые для анализа временных рядов, такие как LSTM и GRU. Последние позволяют учитывать динамику формы команды, последовательность предыдущих результатов и ряд других временных факторов, которые существенно влияют на вероятность победы.

Нейросетевые модели часто демонстрируют более высокую точность по сравнению с классическими методами, однако требуют тщательно подготовленных данных и значительных вычислительных ресурсов для обучения.

Особое место среди современных подходов занимают ансамблевые и гибридные модели, комбинирующие преимущества различных методов. Их эффективность объясняется тем, что каждый алгоритм выделяет свои характерные закономерности, а объединение результатов позволяет компенсировать индивидуальные недостатки.

В спортивной аналитике наибольшее распространение получили ансамбли моделей, объединяющие:

- статистические методы и ML;
- ML и нейросетевые модели;
- несколько моделей машинного обучения с разной структурой.

Такие ансамбли позволяют получать более устойчивые прогнозы, особенно на выборках с высокой вариативностью.

Каждый подход имеет свои преимущества и ограничения, а их выбор определяется целями исследования, качеством данных и особенностями конкретного спортивного соревнования.

#### *Модели машинного обучения в прогнозировании результатов матчей*

Модели машинного обучения занимают центральное место в современном прогнозировании спортивных событий благодаря способности выявлять сложные и неочевидные взаимосвязи между игровыми показателями. В отличие от классических статистических методов, ориентированных преимущественно на линейные зависимости, алгоритмы машинного обучения

позволяют работать с высокоразмерными, шумными и неоднородными данными, что делает их особенно полезными для анализа футбольных матчей, где результат определяется сочетанием множества факторов. Применение таких моделей требует тщательной подготовки данных, корректного выбора признаков и взвешенного подхода к организации процесса обучения. Качество обучающей выборки является ключевым фактором, определяющим итоговую эффективность моделей машинного обучения. Рабочий датасет содержал информацию о большом количестве футбольных матчей, включающую показатели результативности, форму команд, статистику домашних и выездных игр, а также различные контекстные характеристики: турнир, этап чемпионата, наличие значимых игровых событий и кадровых изменений в составе. Поскольку данные имели различную природу, важным этапом стала их унификация: числовые признаки нормировались, категориальные подвергались кодированию, а показатели, имеющие временную динамику, дополнялись производными характеристиками. Это позволило усилить информативность исходных данных и придать моделям способность улавливать устойчивые тренды. Для предотвращения смещения результатов выборка была разделена на тренировочную и тестовую части, а при необходимости применялась перекрёстная валидация.

Одной из ключевых задач при построении моделей стало определение наиболее значимых факторов, влияющих на исход матча. Для решения этой задачи применялись методы оценки важности признаков, основанные на ансамблевых моделях, таких как Random Forest и Gradient Boosting. Эти алгоритмы позволяют анализировать вклад каждого параметра в процесс классификации и строить ранжированные списки факторов.

По результатам анализа были выделены наиболее значимые признаки:

- текущая форма команды (результаты последних матчей);
- разница забитых и пропущенных голов;
- статистика домашних и выездных встреч;
- характеристики соперника;
- средняя результативность команды за несколько предыдущих туров;
- очковый баланс и турнирная мотивация.

Эти показатели отражают реальную игровую силу команды и обладают высокой прогностической ценностью, что подтверждается их регулярным появлением среди значимых факторов при обучении моделей [5].

В качестве основных моделей прогнозирования рассматривались алгоритмы, традиционно демонстрирующие высокую устойчивость при работе с табличными данными:

1. Деревья решений – обеспечивают интерпретируемость и быстрый процесс обучения,

но склонны к переобучению. Применялись как базовый ориентир.

2. Random Forest – представляет собой ансамбль деревьев решений и отличается способностью обобщать данные при наличии шумов. Модель показала высокую устойчивость и хорошие значения точности.

3. Gradient Boosting – один из наиболее эффективных алгоритмов для задач классификации спортивных исходов. Позволяет уточнять прогнозы путём последовательного исправления ошибок предыдущих моделей. Демонстрировал наиболее высокие значения точности и F1-метрики.

4. Метод опорных векторов – использовался для построения разделяющих поверхностей между классами, обеспечивая высокую точность, но требуя тщательной параметризации.

Все модели обучались на подготовленном датасете с использованием оптимальных процедур настройки гиперпараметров: перебора по сетке, кросс-валидации и контроля переобучения. Для повышения эффективности обучения применялись:

- функции активации ReLU;
- алгоритм оптимизации Adam;
- механизмы регуляризации (Dropout);
- нормализация входных данных.

Применение нейронной сети позволило улучшить качество прогнозов в ситуациях, когда между признаками наблюдалась выраженная нелинейная зависимость. В частности, модель лучше справлялась с прогнозированием исходов матчей между командами сопоставимого уровня, где традиционные методы часто демонстрируют погрешности.

Важно подчеркнуть, что каждая модель продемонстрировала свои преимущества в зависимости от структуры данных. Статистические методы обеспечивали хорошую интерпретируемость, модели машинного обучения – высокую точность и устойчивость, а нейронные сети – способность выявлять скрытые зависимости. Сравнение результатов показало, что наиболее эффективными оказались ансамблевые модели и нейронные сети, обладающие лучшим балансом метрик accuracy и F1-score. При этом итоговая точность предсказаний превышала показатели, характерные для классических статистических методов.

#### *Сравнительный анализ точности различных моделей*

Сравнительный анализ различных моделей прогнозирования является ключевым элементом исследования, поскольку позволяет объективно оценить, насколько успешно тот или иной алгоритм справляется с задачей классификации исходов футбольных матчей. Поскольку данные обладают сложной структурой, включают в себя как числовые, так и категориальные признаки, а также отличаются высокой вариативностью, вы-

бор наиболее эффективной модели не может быть основан на формальном сравнении отдельных метрик [3]. Требуется комплексная оценка, учитывающая устойчивость модели, точность прогнозов, способность обобщать результаты и адекватно реагировать на изменения данных [4].

Для обеспечения объективности сравнения применялся единый набор входных данных и идентичные условия разбиения выборки на тренировочную и тестовую части. Каждая модель проходила обучение с использованием одинаковых признаков, что позволяло минимизировать влияние сторонних факторов [6]. Для оценки качества использовались несколько метрик, каждая из которых отражала свою сторону эффективности:

- Accuracy – общая доля правильных предсказаний;
- Precision – точность распознавания конкретного исхода;
- Recall – полнота выявления данного класса;
- F1-score – гармоническое среднее между precision и recall;
- ROC-AUC – способность модели различать классы при разных порогах решения.

Использование нескольких метрик позволило избежать односторонней оценки и получить более полное представление о работе каждой модели. Такой подход особенно важен для прогнозирования исходов спортивных матчей, где распределение классов часто бывает несбалансированным.

1. Статистические модели показали ожидаемо стабильные, но умеренные результаты.

2. Логистическая регрессия продемонстрировала достаточно высокую точность на выборках с выраженной разницей в игровых показателях команд, однако её эффективность снижалась при анализе матчей сопоставимого уровня. Точность модели составила около 65–68 %, что соответствует средним значениям для традиционных методов спортивной аналитики.

3. Модель Пуассона, использующая вероятностный подход для оценки количества голов, хорошо справлялась с прогнозированием вероятностей определённых счётов, однако при сведении задачи к классификации исходов (победа/ничья/поражение) давала более ограниченный результат из-за предположений о независимости событий и постоянстве среднего значения интенсивности.

В целом статистические методы подтвердили свою базовую полезность, но уступили более современным алгоритмам, особенно на сложных участках выборки.

Алгоритмы машинного обучения показали более высокие результаты по ключевым метрикам.

1. Деревья решений продемонстрировали адекватное качество прогнозов (accuracy около 70 %), но склонность к переобучению ограничила их практическую применимость.

2. Random Forest значительно улучшил ситуацию: ансамбль деревьев показал точность порядка 73–75 %, устойчиво работал с шумными данными и обеспечил более высокий уровень обобщения. Особенно полезным оказался механизм определения важности признаков, что позволило уточнить модель и улучшить её интерпретируемость.

3. Градиентный бустинг (в вариантах XGBoost и CatBoost) стал одним из наиболее эффективных алгоритмов. Его точность достигала 77–79 %, а F1-score показывал высокие значения для всех классов. Преимущество бустинга проявилось в способности выявлять тонкие нелинейные закономерности и эффективно работать с большим количеством зависимых признаков.

4. Нейросетевые модели продемонстрировали наиболее высокий уровень прогностической точности, особенно при наличии достаточного объёма данных для обучения. Полносвязная нейронная сеть, оптимизированная под структуру данных и дополненная механизмами регуляризации, показала точность в диапазоне 78–82 %.

Помимо абсолютных значений точности важно учитывать устойчивость прогнозов, анализ показал:

- статистические модели демонстрируют стабильность, но низкую адаптивность;
- модели машинного обучения устойчивы к шумам и дают высокие значения F1-score;
- нейронные сети показывают лучшую устойчивость при сложных взаимосвязях признаков, но требуют более тщательной подготовки данных.

Особое преимущество ансамблевых методов заключалось в минимизации ошибок: разброс значений по метрикам был значительно ниже, чем у отдельных моделей.

В совокупности именно ансамблевые алгоритмы и нейросетевые модели показали наиболее высокую эффективность и обоснованно могут рассматриваться как основа автоматизированных систем прогнозирования.

Разработка интеллектуальной системы прогнозирования спортивных событий явилась завершающим этапом исследования и позволила объединить отдельные аналитические компоненты в единую функциональную среду. Программный комплекс был спроектирован таким образом, чтобы обеспечить полный цикл работы с данными – от их загрузки и предобработки до построения прогноза и визуализации результатов. В отличие от традиционных статистических инструментов, ориентированных преимущественно на анализ отдельных показателей, интеллектуальная система реализует целостный подход, позволяющий пользователю работать с большими массивами данных и получать аналитически обоснованные предсказания в удобной интерактивной форме.

### Заключение

Проведённое исследование позволило всесторонне рассмотреть современные подходы к прогнозированию спортивных событий и выявить наиболее эффективные методы анализа футбольных матчей. Особое внимание было уделено сравнительному изучению статистических моделей, алгоритмов машинного обучения и нейронных сетей, что позволило сформировать целостное представление о возможностях каждого класса методов и определить области их практического применения. Анализ показал, что традиционные статистические подходы сохраняют ценность благодаря своей интерпретируемости, однако в ряде случаев их предсказательная сила ограничивается линейностью используемых моделей и устойчивостью параметров.

Наиболее высокие результаты продемонстрировали алгоритмы машинного обучения – прежде всего ансамблевые методы градиентного бустинга и Random Forest, которые обладают способностью выявлять сложные зависимости и обеспечивать устойчивость при наличии шумов в данных. Применение нейронных сетей позволило дополнить картину: модели глубокого обучения продемонстрировали высокую точность при анализе нелинейных и слабо структурированных признаков, особенно в тех случаях, когда необходимо учитывать динамику формы команды и временные закономерности.

Практическим результатом исследования стало создание интеллектуальной программной системы, включающей инструменты обработки данных, построения моделей и визуализации результатов. Такая система обеспечивает удобный и интуитивно понятный механизм работы с массивами спортивной статистики и может использоваться как в учебных, так и в прикладных задачах. Возможность интеграции различных моделей, адаптивность архитектуры и реализованный механизм сравнения прогнозов делают разработанную систему гибким инструментом для анализа широкого спектра спортивных данных.

Таким образом, представленный в исследовании подход подтверждает высокую эффективность использования современных методов машинного обучения и нейросетевого моделирования при прогнозировании спортивных событий. Перспективы дальнейшего развития включают расширение набора данных, интеграцию текстовой аналитики для учёта новостного контекста, применение рекуррентных и трансформерных моделей для анализа временных последовательностей, а также создание гибридных систем, объединяющих несколько типов алгоритмов. Всё это будет способствовать повышению точности прогнозов и укреплению роли интеллектуальных технологий в области спортивной аналитики.

### Список литературы

1. Андреев К. С. Компьютерное зрение и машинное обучение. СПб.: Питер, 2020.
2. Брук Т. Статистические методы в спортивной аналитике. М.: Спорт и наука, 2020.
3. Гудфеллоу Я., Бенджио И., Курвиль А. Глубокое обучение. М.: ДМК Пресс, 2018.
4. Половников Е. П. Технологии машинного обучения в автоматизированных системах. М.: Академия, 2021.
5. Bukreiev D. O. et al. Features of the use of software and hardware of the educational process in the conditions of blended learning // AET 2020-Symposium on Advances in Educational Technology. Technology (AET 2020). SCITEPRESS, 2022. №. 2. С. 236-244.
6. Bukreiev D. Neuro-network technologies as a mean for creating individualization conditions for students learning // SHS Web of Conferences. EDP Sciences, 2020. Т. 75. С. 04013.

### СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ YOLOV8–YOLO11 В ЗАДАЧАХ ОБНАРУЖЕНИЯ ОБЪЕКТОВ В РЕАЛЬНОМ ВРЕМЕНИ

Корж А. А., Олейник Н. П.

ФГБОУ ВО «Мелитопольский государственный университет», Мелитополь,  
e-mail: alex.korhz13@mail.ru,  
nata.oleynik.2014@mail.ru,

*Постановка проблемы.* Отсутствие сопоставимых по методике сравнений YOLOv8–YOLO11 приводит к выбору моделей «по привычке» или по разрозненным публикациям, где различаются датасеты, размер входа и правила измерения задержки. Для практических систем это означает либо избыточные требования к вычислительным ресурсам, либо неприемлемое снижение качества детекции при переносе на целевую сцену.

#### Материалы и методы исследования

История семейства YOLO начинается с работы J. Redmon и соавт. (2015) [1], где предложен одноэтапный подход к детекции объектов. В последующих версиях развивались идеи ускорения и повышения точности; например, для YOLOv7 показано, что переработка backbone/neck и обучение с учётом задач реального времени позволяют сохранить высокую скорость при конкурентной точности [2].

Ultralytics в YOLOv8 перешли к anchor-free предсказаниям и унифицированному фреймворку задач, оставив NMS как обязательную часть конвейера [3].

В YOLOv9 введены GELAN и механизм PGI, рассчитанный на улучшение градиентного потока на этапе обучения [4]. Работа по YOLOv10 (Tsinghua University) предложила end-to-end детекцию без NMS через согласованное двойное назначение (one-to-many и one-to-one) и более лёгкие вычислительные блоки [5].

В YOLO11 Ultralytics сфокусировались на снижении параметрической сложности и уско-