

УДК 004.8

ИССЛЕДОВАНИЕ И АНАЛИЗ ДАННЫХ С ПОМОЩЬЮ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА О СПОРТИВНЫХ РЕЗУЛЬТАТАХ ДЛЯ ВЫЯВЛЕНИЯ КЛЮЧЕВЫХ ФАКТОРОВ УСПЕХА КОМАНД

Маматкасымова Алийма Төрөжановна¹, Атамкулова Мушарап Тешеевна²,
Кудайберди кызы Майрам³, Мойдунов Расулбек Жалилович⁴

¹*Ошский технологический университет им. М. Адышева, кандидат физико-математических наук, доцент,*

²*Ошский технологический университет им. М. Адышева, преподаватель*

³*Ошский технологический университет им. М. Адышева, магистрант,*

⁴*Кыргызско-Узбекский международный университет им. Б. Сыдыкова, учитель*

¹E-mail: mamatkasymova1973@mail.ru

²E-mail: musharap_74@bk.ru

³E-mail: kudaiberdieva9316@gmail.com

Аннотация. В рамках данной статьи с помощью ИИ исследовано и проведено комплексный анализ данных о женщинах-волейболистках на основе информации, полученной из открытой базы данных Volleybox. Исследование продемонстрировало возможности применения современных методов анализа данных для изучения физических и игровых характеристик спортсменок, а также их взаимосвязи с игровыми позициями и ролью в команде. Результаты данной работы могут быть полезны в образовательной деятельности, спортивной аналитике, а также при разработке систем поддержки принятия решений в профессиональном спорте.

Annotation. In this article, AI was used to research and conduct a comprehensive analysis of data on female volleyball players, based on information obtained from the Volleybox open database. The study demonstrated the potential of applying modern data analysis methods to study the physical and playing characteristics of athletes, as well as their relationships with their positions and team roles. The results of this work can be useful in educational activities, sports analytics, and in the development of decision support systems in professional sports

Ключевые слова. Анализ данных, цифровые платформы, базы данных, Volleybox, спортивная аналитика, языка программирования, Google Colab, Python, рост, вес, возраст, игровая позиция

Key words. Data analysis, digital platforms, databases, Volleybox, sports analytics, programming languages, Google Colab, Python, height, weight, age, playing position

Введение

В условиях стремительного развития цифровых технологий и увеличения объёмов доступной информации методы анализа и обработки больших данных занимают ключевое место в современных научных исследованиях и практических приложениях. Одной из сфер, где аналитика данных приобретает всё большее значение, является спорт. Использование методов анализа данных и машинного обучения способствует более глубокому пониманию факторов, влияющих на спортивные достижения, и повышает конкурентоспособность команд и отдельных игроков.

Женский волейбол представляет собой динамично развивающийся и высококонкурентный вид спорта, в котором физические, антропометрические и игровые характеристики спортсменок играют решающую роль. Современные цифровые платформы и открытые источники данных аккумулируют значительные объёмы информации о волейболистках, включая данные о росте, весе, возрасте, игровой позиции, высоте нападающего удара и блока, а также информацию о национальности и статусе активности игрока. Анализ таких данных позволяет выявлять закономерности распределения физических параметров по игровым амплуа, определять типичные профили спортсменок для различных позиций и исследовать общие тенденции развития женского волейбола на глобальном уровне.

Актуальность данной исследования обусловлена необходимостью применения современных методов анализа больших данных для исследования спортивных показателей и выявления ключевых факторов, влияющих на успешность команд и игроков.

Целью исследования является проведение комплексного разведочного анализа базы данных женщин-волейболисток, собранной с платформы Volleybox, с целью изучения распределения антропометрических и игровых характеристик, выявления взаимосвязей между ними и оценки пригодности данных для дальнейшего аналитического и прогнозного использования. Результаты данной работы могут быть полезны в образовательной деятельности, спортивной аналитике, а также при разработке систем поддержки принятия решений в профессиональном спорте.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Изучить существующие подходы к сбору и анализу спортивных данных, а также возможности применения технологий анализа больших данных в спортивной аналитике.
2. Описать структуру и основные характеристики выбранного набора данных женщин-волейболисток, включая объем, состав признаков и степень их полноты.
3. Выполнить разведочный анализ данных (EDA) с целью изучения распределений

основных антропометрических и игровых показателей спортсменок.

4. Проанализировать взаимосвязь физических характеристик игроков с их игровыми позициями и ролью в команде.

5. Выявить ключевые закономерности и тенденции в распределении игроков по возрасту, росту и весу.

Объектом исследования в данной исследовании являются данные о профессиональных женщинах-волейболистках, содержащие информацию об их антропометрических характеристиках, игровых позициях, показателях результативности и статусе активности, собранные из открытого источника — базы данных Volleybox. Рассматриваемый объект представляет собой совокупность структурированных статистических данных, отражающих физические и игровые особенности спортсменок, а также общие тенденции развития женского волейбола на международном уровне.

Предметом исследования исследовании являются платформа Google Colab, методы работы и результаты анализа спортивных данных, отражающие взаимосвязь антропометрических и игровых характеристик женщин-волейболисток с их игровыми позициями и ролью в команде. В рамках исследования рассматриваются закономерности распределения физических показателей, таких как рост, вес, высота нападающего удара и блока, а также их влияние на специализацию спортсменок и общие тенденции формирования профессиональных волейбольных команд.

Методы исследования. В ходе исследования использовались методы сбора, очистки и предобработки данных, разведочный и описательный статистический анализ, а также методы визуального и сравнительного анализа, реализованные с применением инструментов анализа данных на основе языка программирования Python, что позволило выявить основные закономерности, тенденции и особенности структуры исследуемого набора спортивных данных[2],[7].

Описание выбранного набора данных

Набор данных Volleyball Women Players Database – Volleybox представляет собой крупную и систематизированную коллекцию информации о профессиональных и полупрофессиональных женщинах-волейболистках со всего мира, собранную путём веб-скрапинга с открытой базы данных Volleybox (<https://women.volleybox.net/>). Датасет содержит 72 875 записей и представлен одним CSV-файлом объёмом около 7 МБ. Он служит обобщённым срезом мирового женского волейбола и включает ключевые биографические и антропометрические характеристики игроков, что делает его ценным ресурсом для исследовательских, статистических и учебных задач.

Позиция важна для анализа зависимости антропометрических характеристик от игровой роли: например, рост и высота прыжка для центральных блокирующих, мобильность и вес для либеро. Также могут быть представлены параметры, связанные с текущим статусом игрока: активна ли она, завершила ли карьеру, ищет ли клуб, состоит ли в конкретной команде. В некоторых случаях в данных содержатся дополнительные ссылки, метаданные или признаки, отражающие структурированную информацию, имеющуюся на сайте Volleybox.

Исследовательский анализ данных (EDA)

В ходе исследования получены следующие характеристики в виде диаграммы:

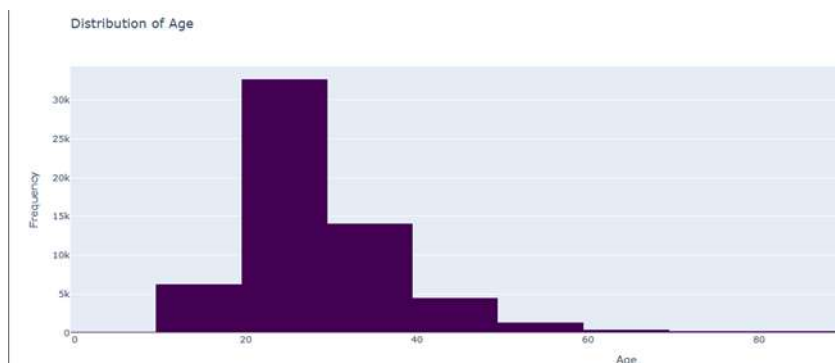


рис. 1. Визуализация распределения возраста игроков

В целом гистограмма показывает распределение активных игроков по возрастным группам и ясно отражает, в каком возрасте сосредоточена основная часть действующих спортсменов.

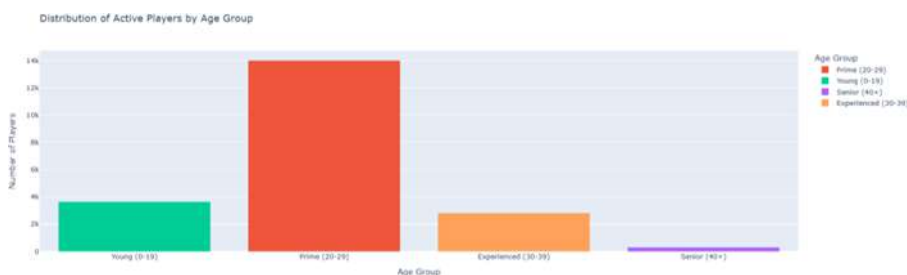


рис. 2. Визуализация распределения игроков по возрастным группам

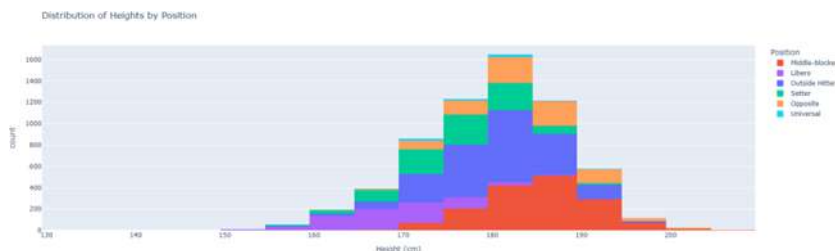


рис. 3. Визуализация распределения высоты, веса, нападающего удара и блоков по положению

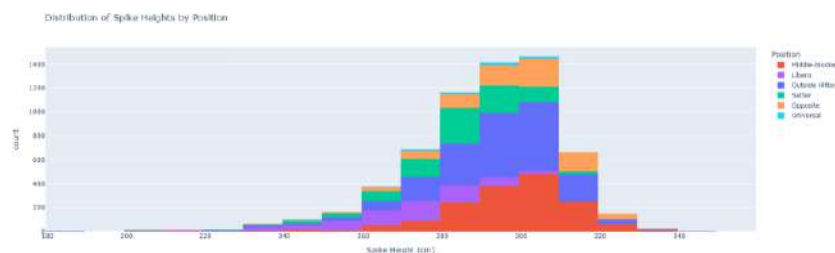


рис. 4. Визуализация распределения высоты нападающего удара по положению

Аналогичную закономерность можно наблюдать на графике распределения высоты нападающего удара.

Такой анализ позволяет выявить структурные различия между амплуа и может быть использован для оценки соответствия физических данных требованиям позиции.

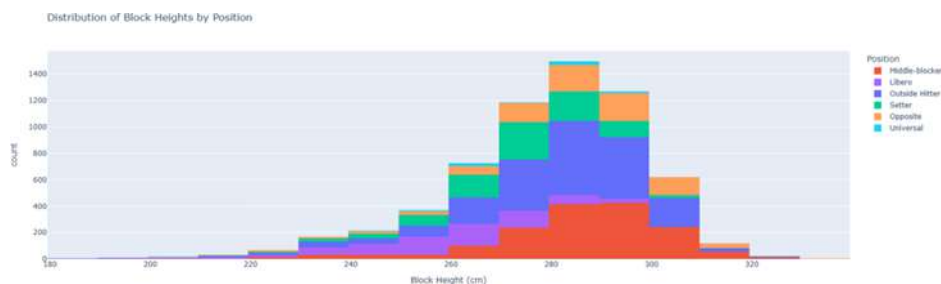


рис. 5. Визуализация распределения высоты блоков по положению

Из графика видно, что центральные блокирующие обладают наиболее высокими значениями block reach — их данные смещены в правую часть распределения, что соответствует их роли в защите и атаке у сетки. физических требований к их роли.

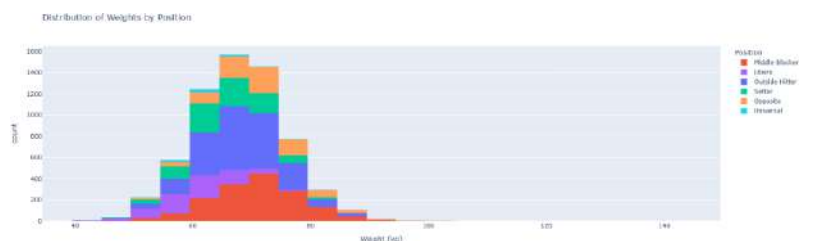


рис. 6. Визуализация веса по позиции

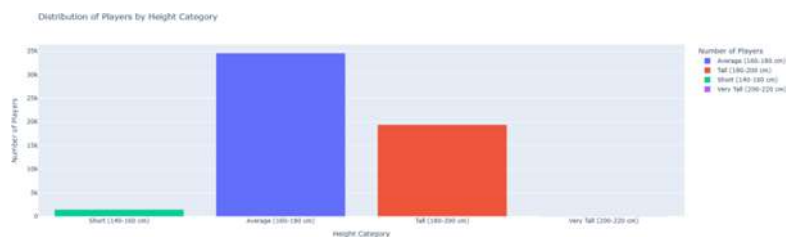


рис. 7. Визуализация распределения игроков по росту

В целом распределение показывает явный сдвиг в сторону средних и высоких значений роста, что соответствует физическим требованиям волейбола, где рост является одним из ключевых факторов для успеха, особенно в атаке и блоке.

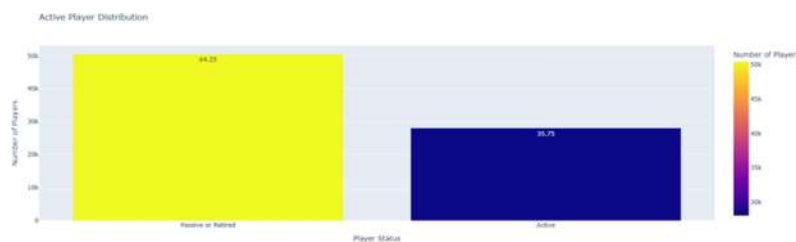


рис. 9. Визуализация распределения активных игроков

Таким образом, диаграмма отражает здоровую и динамичную ситуацию, где абсолютное большинство спортсменов находятся в активной фазе карьеры. Это характерно для базы данных действующей лиги, актуального состава сборных или активного игрового сезона.

Наблюдается четкая положительная корреляция между ростом и весом игроков. Это закономерно и ожидаемо, так как оба параметра связаны с телосложением спортсмена.

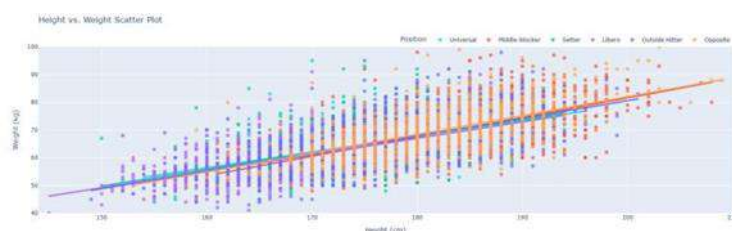


рис. 10. Диаграмма рассеивания роста и веса

Такое распределение демонстрирует рациональный подход к формированию команды, где каждый игрок максимально эффективно использует свои природные данные для решения конкретных тактических задач.

Заключение

В результате сравнительного анализа было установлено, что антропометрические характеристики спортсменок имеют выраженную зависимость от их игровой специализации. Выявленные закономерности подтверждают обоснованность существующей специализации в женском волейболе и подчёркивают значимость объективных данных при подборе и оценке игроков. Кроме того, в работе были выявлены общие тенденции распределения спортсменок по возрасту и географическому признаку, что позволяет получить более целостное представление о структуре профессионального женского волейбола.

Установлено, что после соответствующей предобработки и фильтрации данных набор может служить надёжной основой для более сложных задач, таких как кластеризация игроков, классификация по игровым позициям или прогнозирование спортивных показателей [1],[7].

Таким образом, выполненная исследование подтверждает эффективность применения методов анализа данных в спортивной аналитике и демонстрирует их практическую ценность. Полученные результаты могут быть использованы в образовательных целях, при проведении дальнейших научных исследований, а также в практической деятельности тренеров и аналитиков для повышения качества принятия решений в профессиональном волейболе.

Список использованных литератур

- [1]. Хастие Т., Тибширани Р., Фридман Дж. Статистическое обучение: основы и методы.
- [2]. М.: Вильямс, 2020.
- [3]. Mitchell R. Web Scraping with Python: Collecting Data from the Modern Web. — 2nd ed.
- [4]. Sebastopol: O'Reilly Media, 2018. — 488 p.
- [5]. Гудфеллоу Я., Бенджио Й., Курвилль А. Глубокое обучение. — М.: ДМК Пресс, 2021.
- [6]. Маккини У. Python и анализ данных. — СПб.: Питер, 2022.
- [7]. Вандер Плас Д. Python для науки о данных. — М.: ДМК Пресс, 2020.
- [8]. Хант Дж. Практическое машинное обучение на Python. — М.: ДМК Пресс, 2022.
- [9]. Gulnarida Zhalilova, Aliyma Mamatkasymova, Elnura Zhusupova, Kunduz Zhalzhaeva. Forecasting data science professionals' salaries using machine learning methods based on real data// *International conference on «Modern problems of applied science and engineering»*, On May 2-3, 2024. Том 3244, Выпуск 1/ 27 ноября, 2024г. <https://pubs.aip.org/aip/acp/article-abstract/3244/1/030034/3322836/Forecasting-data-science-professionals-salaries>