

УДК 004.5, 004.415

## РЕКОМЕНДАТЕЛЬНАЯ СИСТЕМА ПОДБОРА ИГРОКОВ В КОМАНДНЫХ ВИДАХ СПОРТА, ПОСТРОЕННАЯ НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Р. Р. Шигапов<sup>1</sup>[0000-0002-3163-0959], А. А. Ференец<sup>2</sup>[0000-0002-7859-9901]

Институт информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета  
ул. Кремлевская, 35, г. Казань, 420008

<sup>1</sup> shigapov.rinat.2000@gmail.com, <sup>2</sup> ist.kazan@gmail.com

### **Аннотация**

Описана разработка на основе машинного обучения рекомендательной системы подбора игроков на примере хоккея с возможностью расширения ее использования в различных командных видах спорта. Для каждого вида спорта рассмотрены амплуа и характеристики игроков, которые были структурированы и разделены на общие группы. Проанализирована информация о хоккее, футболе, баскетболе и волейболе. Для каждого из рассмотренных параметров выведены коэффициенты, показывающие их влияние на результат матча. Протестированы модели, построенные на основе различных алгоритмов машинного обучения. Создан веб-интерфейс приложения.

**Ключевые слова:** спорт, хоккей, подбор игроков, рекомендательная система, машинное обучение

### **ВВЕДЕНИЕ**

Переход игрока из одного спортивного клуба в другой (трансфер) является неотъемлемой частью командных игр. За последние несколько лет количество трансферов неуклонно росло, как и общая сумма денег, потраченная клубами на покупку новых игроков [1].

Спортивные клубы заинтересованы в новых игроках для улучшения результатов команды. Но нередко случается, что игрок переоценён и не приносит ожи-

даемую от него пользу [2]. Из-за неудачной трансферной политики команда может проигрывать матчи, клуб — потерять большие деньги, а карьеры купленных игроков — пойти на спад. В связи с этим сейчас в основных спортивных лигах как мира, так и России автоматизируется сбор данных о статистике определённых игроков, матчей или же команд, а их последующий анализ является актуальной задачей [3]. Так появилась гипотеза, что можно на основе машинного обучения реализовать рекомендательную систему для подбора игроков в командных видах спорта.

Для реализации подобной системы требуется решить следующие задачи: изучить влияние различных доступных статистических параметров игроков на результаты команды, подготовить набор данных (датасет), обучить модели машинного обучения и выбрать наилучшую из них, а также создать алгоритм рекомендации и реализовать пользовательский интерфейс.

### **ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ**

Проанализированы списки наиболее популярных в мире видов спорта [4, 5] и выделены четыре: футбол, хоккей, волейбол, баскетбол.

Для поиска новых талантливых игроков клубы создают специальные скаутинговые отделы. Как правило, в этих отделах работают люди с опытом в игре, например, игроки, которые закончили свою спортивную карьеру. Раньше в скаутских отделах полагались только на мнение авторитетных лиц. В таких отделах при проведении анализа игроков основное влияние на результат имел человеческий фактор, который не всегда объективен: для анализа требуется очень качественное представление информации, так как объём статистических данных весьма большой. Сейчас же с каждым годом новые технологии все чаще стали внедряться в ведение статистики и аналитики в спорте. Появляются новые системы, которые предлагают различные функции для усовершенствования команды:

- InStat — веб-приложение, направленное на получение статистики матчей, создание специальных тренировок и отчётов по командам соперников. Основным видом спорта является футбол, также некоторые функции развиты для баскетбола и хоккея. Отметим, что для сбора статистики каждый матч просматривают два человека, что делает данный способ трудозатратным. Особенностью является то, что статистика привязана к видео [6]. Главная

особенность платформы — возможность просматривать статистику вместе с эпизодами игры.

- IceBerg — средство для сбора статистики с помощью машинного обучения и специального оборудования. На основе полученных данных система может создавать отчеты по играм и специальные тренировки как для определенного игрока, так и для всей команды [7].
- DataVolley4 — программное обеспечение, с помощью которого команды создают отчёты по командам соперника и изменяют тактики своей игры [8].

Исходный код названных продуктов — коммерческая тайна, а результаты точности реализуемых ими функций неизвестны. Отметим также, что в данных системах отсутствует функция рекомендаций.

Помимо общепринятых характеристик выступления команды, в каждом виде спорта у игроков есть свои характеристики, которые показывают, насколько успешно выступает данный игрок. Важность этих характеристик зависит не только от вида спорта, но и от позиции, на которой выступает игрок (амплуа).

Для футбола (ф), хоккея (х), баскетбола (б) и волейбола (в) характерны свои позиции игроков:

- нападающий (х, ф),
- полузащитник (ф),
- защитник (х, ф),
- вратарь (х, ф),
- либеро (в),
- связующий (в),
- доигровщик (в),
- диагональный (в),
- центральный связующий (в),
- игрок передней линии (б),
- игрок задней линии (б).

В данных видах спорта прослеживаются закономерности в позициях игроков, которых можно разделить на игроков нападения и обороны, а также вратарей.

В хоккее для полевых игроков выделяют 10 основных параметров, а для вратарей – 7 [9]. В футболе для полевых игроков выделяют 13 основных параметров, а для вратарей – 8 [10]. В баскетболе используют 9 параметров для игроков [11], а в волейболе — 7 [12]. Отметим, что для каждого вида спорта и каждого амплуа игрока есть свои характерные параметры.

Если сравнить статистику по разным видам спорта, то можно найти много общих характеристик, однако среди них есть и свойственные каждому отдельному спорту. Можно разделить все признаки на несколько групп:

- игровые (количество проведённых матчей, время на площадке),
- вратарские (для футбола и хоккея),
- атакующие (голы, передачи, набранные очки),
- оборонительные (перехваты, блок-шоты, отборы, блокированные удары, эффективность приёма),
- штрафные (штрафные минуты, фолы, красные и желтые карточки).

Во всех выбранных видах спорта для победы необходимо набрать больше очков, чем соперник. Наиболее отличается волейбол: игра ограничивается не временем, а количеством очков — для победы необходимо выиграть 3 сета, в каждом из которых набрать 25 очков, но с разницей в 2 очка. В остальных играх матч ограничивается временем:

- футбол — 2 тайма по 45 минут;
- хоккей — 3 периода по 20 минут;
- баскетбол — 4 периода по 10 минут.

Другим важным аспектом командных игр являются замены. В этом компоненте наиболее жесткие правила в футболе. С 2020 года из-за COVID-19 разрешены пять замен за весь матч, до этого было всего три. В волейболе разрешены 6 замен за партию. В баскетболе и хоккее ограничений на замены нет. Правила замен влияют на количество игроков в команде:

- В футболе за команду играют 10 полевых игроков плюс вратарь, и на замене сидят от 3 до 7 человек в зависимости от лиги [13].
- В волейболе в команде может быть максимум 14 человек, в игре участвуют 6 человек и 8 запасных [14].
- В баскетболе на площадке находятся 5 человек и от 5 до 7 – на замене [15].

- В хоккее на площадке находятся 5 игроков и вратарь. На замене сидят примерно еще 3 пятёрки (игроки, играющие вместе на площадке) и запасной вратарь, то есть всего в хоккейный состав входят примерно 22 игрока [16].

Если сравнивать игры по скорости изменения ситуации на площадке, то безоговорочно впереди будет хоккей, так как в отличие от остальных видов спорта игроки тут катаются на коньках, что в разы быстрее бега [17].

Мы можем сделать вывод, что несмотря на то, что в каждом виде спорта существуют свои особенности, в целом структура игр имеет общую базу и строится на единых принципах. Благодаря этому, создав алгоритм рекомендательной системы, можно будет расширить его и на другие виды спорта.

В данной работе показана реализация рекомендательной системы на основе хоккея с шайбой.

### **ПРОЕКТИРОВАНИЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ**

Игровые характеристики — это основные характеристики для любого ампула. Для более релевантных данных был сделан выбор игроков, которые сыграли определённое количества игр в сезоне, в каждой из которых провели минимальное среднее время на площадке. Эти данные зависят от выбранного вида спорта.

В Континентальной хоккейной лиге (КХЛ) в регулярном сезоне 56 игр. При этом в команде больше игроков, чем может быть заявлено на матч. Из-за постоянных ротаций в составе или же, например, травм игрок может сыграть не в каждом матче. Для корректных рекомендаций игрок, сыгравший в 25% матчей сезона, может учитываться.

Хоккейный матч длится 60 минут, как правило в команде 4 пятёрки (5 человек) полевых игроков. Так как время между игроками делится не поровну (лидеры обычно играют больше), то можно считать минимальным необходимым

временем, проведённым на площадке, 15% общего времени всех игр. Таким образом, если игрок сыграл в 14 матчах сезона более 9 минут, то он может быть рекомендован системой.

Для полевых игроков (нападающих и защитников) параметры почти одинаковы, но оказывают разное влияние. Для нападающих более существенны атакующие характеристики, так как их основная задача забивать голы в ворота противников. В настоящее время, если игрок не умеет или не хочет отрабатывать в защите, существует практика не ставить его в состав. Также последние несколько лет стали популярны такие термины, как «нападающий оборонительного плана» и «атакующий защитник» [18]. Эти термины означают, что игроки выполняют функции, не связанные напрямую со своим амплуа. Например, «нападающий оборонительного плана» может забивать мало голов и не набирать большого количества очков, но зато такой игрок может выходить в меньшинстве и блокировать собой броски. Как правило, такие игроки «таскают рояль», именно поэтому их действия могут оставаться незамеченными, и их часто недооценивают. Чтобы в разрабатываемой системе не было таких случаев, для полевых игроков необходимо рассматривать все доступные характеристики, как атакующие, так и оборонительные.

Если предыдущие параметры добавляли баллы в оценку, то штрафные характеристики вычитают их, так как это действия, которые отрицательно сказываются на игре команды. Зачастую в игре двух равных команд основным фактором успеха является именно игра в неравных составах, и если игроки не дисциплинированы, заработав удаление, они могут подвести команду.

Для оценки нападающего необходимо понять, с какими весами брать каждый параметр.

На официальном сайте КХЛ [19] представлены 25 параметров полевых игроков.

Так как некоторые характеристики могут повторять друг друга, например, заброшенные шайбы плюс передачи – это очки, а в других лигах и видах спорта может не быть такой подробной статистики, приведем основные параметры:

- Количество проведенных игр — игровая, сколько игрок отыграл матчей в сезоне;
- Заброшенные шайбы — атакующая, количество голов за сезон или матч;
- Передачи — атакующая, количество результативных передач, после которых были забиты голы;
- Плюс/Минус — игровая, коэффициент полезности игрока: при участии в атаке, которая завершилась голом, прибавляется балл, при нахождении на поле и взятии ворот командой соперника балл вычитается;
- Штрафное время — штрафная, удаление игрока с поля за нарушение, в зависимости от вида нарушения игрок может получить 2, 5, 10 или 20 минут штрафного времени;
- Время на площадке — игровая, время, которое игрок непосредственно играл;
- Броски по воротам — атакующая, количество бросков, произведенное игроком в створ ворот;
- Вбрасывания — атакующая, характеристика, как правило, относящаяся к центральному нападающему; при вводе шайбы судьей в игру происходит борьба центральных нападающих за шайбу;
- Выигранные вбрасывания — атакующая характеристика;
- Силовые приемы — оборонительная, количество силовых приёмов в рамках правил против соперника;

- Фолы против игрока — атакующая, количество заработанных игроком штрафов соперника на себе;
- Блокированные броски — оборонительная, броски, которые игрок заблокировал, не дав дойти шайбе до ворот.

Чтобы доказать предположение об атакующих, оборонительных и штрафных характеристиках, сделаем следующие действия:

1. Для каждой команды и каждого игрока выводится среднее значение выбранных параметров, кроме количества проведенных игр и времени на площадке из-за того, что они нужны для составления выборки, а не рекомендации игроков. Возьмем для примера казанский «Ак Барс» (см. таблицу 1).

Таблица 1 — Параметры нападающих казанского «Ак Барса» в среднем за матч

| голы | передачи | плюс/минус | Штрафы | Броски | вбрасыв. | вбрасыв. | выиг. прием | силов. прием | фолы против | блок. броски |
|------|----------|------------|--------|--------|----------|----------|-------------|--------------|-------------|--------------|
| 2.5  | 2.7      | 1.9        | 6.7    | 22.3   | 38.5     | 31.4     | 9.7         | 2.8          | 5.6         |              |

Таблица 2 — Параметры нападающих в среднем для игрока

| голы | передачи | плюс/минус | Штрафы | Броски | вбрасыв. | вбрасыв. | выиг. прием | силов. прием | фолы против | блок. броски |
|------|----------|------------|--------|--------|----------|----------|-------------|--------------|-------------|--------------|
| 0.14 | 0.17     | 0.32       | 0.53   | 1.23   | 0.43     | 3.7      | 0.57        | 0.13         | 0.35        |              |



2. Проверяется, как коррелируют данные команд из предыдущего пункта с результатами команды. В таблице 3 можно заметить верность рассуждений по поводу параметров: забитые шайбы, передачи, плюс/минус, штрафы, броски, выигранные вбрасывания. Параметр вбрасывания не должен учитываться, так как количество вбрасываний никак не может повлиять на игру, в отличие от выигранных. Хотя параметры «блокированные броски» и «фолы против» слабо коррелируют с результатами, их не учитывать нельзя. Можно предположить, что слабая корреляция связана со сложностью выявления заблокированных бросков — бросок заблокирован, гола нет. С «фолами против» объяснить сложнее, но, если вспомнить о важности большинства/меньшинства, этот параметр нельзя не учитывать. Возможно, влияет подготовка «спецбригад» для большинства и меньшинства, и этот параметр теряется.

Таблица 3 — Корреляция параметров нападающих в среднем по клубам и результатами матчей

| голы | передачи | плюс/минус | Штрафы | броски | вбрасыв. | выигр. вбрасыв. | сил. прием | фолы против | блок. броски |
|------|----------|------------|--------|--------|----------|-----------------|------------|-------------|--------------|
| 0.75 | 0.71     | 0.74       | -0.66  | 0.53   | -0.11    | 0.30            | 0.10       | 0.16        | 0.37         |

Для оценки игрока воспользуемся данными, полученными в доказательствах. Разделим данные таблицы 3 на данные таблицы 2, убрав вбрасывания (см. таблицу 4).

Таблица 4 — Весовые коэффициенты игровых характеристик для нападающих

| голы | передачи | плюс/минус | Штрафы | Броски | вбрасыв. | выигр. вбрасыв. | сил. прием | фолы против |
|------|----------|------------|--------|--------|----------|-----------------|------------|-------------|
| 5.35 | 4.18     | 2.31       | -0.65  | 0.43   | 0.08     | 0.17            | 1.23       | 0.54        |

В отличие от нападающих у защитников меньше действий, которые можно зафиксировать. В хоккейной среде есть фраза: «Лучший защитник тот, которого не видно», то есть главная задача игрока обороны – не пропустить шайбу. Именно поэтому у защитников обычно смотрят на параметр плюс/минус, который показывает, насколько успешно защитник сыграл в обороне. Среди игроков обороны также есть свои исключения – «атакующие защитники», которые иногда могут набрать очков больше, чем нападающие. Эта тенденция пришла из-за океана из Национальной хоккейной лиги (НХЛ) [20]. Поэтому, как и у нападающих, при выводе оценки берутся оборонительные, атакующие и штрафные характеристики. Метод оценки игроков защиты схож с нападающими, за исключением параметра «выигранные вбрасывания», так как защитники редко принимают в них участие. В таблицах 5 и 6 мы получили ожидаемые параметры, учитывая результаты нападающих. Для более релевантных оценок параметры «фолы против» и «броски» берутся с другими весами.

Таблица 5 — Параметры защитников в среднем для игрока

| голы | пере-<br>дачи | плюс/<br>минус | Штра-<br>фы | Брос-<br>ки | выиг.<br>вбра-<br>сыв. | силов.<br>прием | фолы<br>против |
|------|---------------|----------------|-------------|-------------|------------------------|-----------------|----------------|
| 0.08 | 0.14          | 0.31           | 0.49        | 0.87        | 0.67                   | 0.09            | 0.62           |

Таблица 6 — Корреляция параметров защитников в среднем по клубам и результатами матчей

| голы | пере-<br>дачи | плюс/<br>минус | Штра-<br>фы | Брос-<br>ки | выиг.<br>вбра-<br>сыв. | силов.<br>прием | фолы<br>против |
|------|---------------|----------------|-------------|-------------|------------------------|-----------------|----------------|
| 0.23 | 0.57          | 0.74           | -0.24       | 0.25        | 0.18                   | 0.04            | 0.35           |

Таблица 7 — Весовые коэффициенты игровых характеристик для защитников

| ГОЛЫ | пере-<br>дачи | плюс/<br>минус | Штра-<br>фы | Брос-<br>ки | выиг.<br>вбра-<br>сыв. | силов.<br>прием | фолы<br>против |
|------|---------------|----------------|-------------|-------------|------------------------|-----------------|----------------|
| 2.87 | 4.07          | 2.39           | -0.49       | 0.28        | 0.27                   | 0.44            | 0.56           |

Позиция вратаря есть не во всех видах спорта. В выбранных видах спорта позиция вратаря встречается только в футболе и хоккее. Вратарь защищает ворота, чтобы соперник не забил гол. На успешности защиты основана статистика вратарей. У них совсем другие параметры, из-за чего их характеристики были вынесены в отдельную вратарскую категорию. В неё входят:

- отражённые броски,
- штрафное время,
- выигрыши,
- броски в створ ворот,
- процент отражённых бросков,
- время на площадке.
- проигрыши,
- пропущено шайб,
- коэффициент надежности (60 минут × пропущенные шайбы / время на площадке),
- «сухие» игры (игры, в которых вратарь пропустил 0 голов).

Также в хоккее, в отличие от полевых игроков, как правило, замена вратаря встречается крайне редко. Это может произойти в случае травмы или если у основного вратаря не пошла игра.

Другой важной особенностью в построении рекомендательной системе стало количество вратарей. Если в каждой хоккейной команде может быть около 15 нападающих, 10 защитников, то вратарей обычно максимум 3. К тому же третий вратарь, обычно только номинальный, например, заканчивающий карьеру или, наоборот, очень молодой. Это происходит из-за того, что третий вратарь крайне редко играет, на матч заявляются только два вратаря. Существуют разные тактики использования вратарей, но зачастую в клубе есть один основной вратарь, который играет 90% матчей, и второй вратарь, который хуже по классу и выходит на игру лишь иногда. Поэтому некоторые тренеры дают резервным вратарям дополнительные задания, такие как ведение статистики [21].

Основываясь на вышесказанном, можно сделать вывод, что система для вратарей не будет пользоваться спросом, а количество статистических параметров не даст её развить, как этого хотелось бы.

Рекомендации для команд строятся на основе игроков, входящих в состав этой команды. С помощью машинного обучения предполагается оценка игроков на будущий сезон, матч или временной интервал. Далее для каждой команды рассчитывается средний рейтинг для каждого амплуа. Затем формируется список из определённого количества игроков, рейтинг которых выше среднего по команде в этом амплуа. Тем самым выбранные игроки будут положительно влиять на команду. Также к рекомендации прикрепляется ссылка на статистику каждого игрока.

Исходя из наблюдений за трансферами за прошлые года, перед заключением основного контракта клуб заключает с игроком «пробный» или «просмотровый» контракт [22]. Обычно это контракт со спортсменом, который ни разу не работал с тренерским штабом и не знаменит. Это происходит из-за того, что часто на игрока могут влиять неспортивные причины, например, межличностные. Поэтому при использовании клубами созданной системы, если игрок не знаком руководству, то рекомендуется подписывать игрока изначально на «пробные» контракты.

Из выявленных проблем можно указать, что на рынке свободных агентов очень мало вратарей, и, как правило, команды укомплектованы этим амплуа:

вряд ли какая-то команда захочет отдавать вратаря при успешных результатах, ведь считается, что «вратарь — это половина команды».

## РЕАЛИЗАЦИЯ И ТЕСТИРОВАНИЕ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ

Для создания оценок и работы системы необходимы были данные игроков со статистикой за каждый сыгранный матч. В интернете необходимого набора данных (датасета) в открытом доступе найдено не было.

Для создания датасета были взяты данные с официального сайта Континентальной хоккейной лиги [19]. КХЛ — это международная лига, которая считается сильнейшей в Европе и второй в мире. Также на основе результатов российских команд в этой лиге выбирается победитель чемпионата России.

Для решения этой задачи был реализован парсер на языке программирования Python с использованием библиотек BeautifulSoup4 и requests [23–25].

Для вариации возможных улучшений точности модели машинного обучения данные были собраны для матчей по отдельности и для всего сезона вместе. В общем было получено: 451 107 записей об игроках в матчах; 19 675 записей об игроках в сезонах; 3 578 игроков.

Для предобработки датасета были проведены следующие действия:

1. Время, проведённое на площадке, для более удобного расчёта было переведено в секунды из минут плюс секунд.
2. Из-за переименования некоторых клубов была проделана работа для унифицирования названий в датасете. Например, в 2010 году клубы «ХК МВД» и «ХК Динамо» объединились в «ОХК Динамо», для каждого игрока с одним из этих клубов решено записывать клуб «Динамо Москва» [26].
3. Были удалены некоторые данные, которые могут испортить статистику, например, матч между Авангардом и Витязем от 9 января 2010 года, закончившийся массовой дракой. Команды не доиграли матч и заработали 14 штрафных часов [27]. В подобных случаях в датасете вместе результата матча прописано «-:-». Такие матчи были проанализированы и удалены.

4. Были выбраны данные, начиная с сезона 2014/2015, из-за того, что до этого параметры «фолы против», «силовые приёмы» и «блокированные броски» не считались.
5. Были выбраны игроки, которые минимум в 14 матчах сыграли больше 9 минут (540 секунд).
6. Датасет был разделён по амплуа игроков.
7. Набор данных для нападающих и защитников был разделён на данные, включающие игры плей-офф и не включающие эти игры.

Для обучения модели использовались стандартные для этой задачи технологии: язык программирования Python и библиотека scikit-learn [23, 28]. Были рассмотрены следующие алгоритмы машинного обучения, решающие задачу регрессии:

- Линейная регрессия [29] — самый простой алгоритм машинного обучения, который строит модель зависимости одного параметра от одного или нескольких других.
- Случайный лес [30] — алгоритм, реализованный с помощью решающих деревьев.
- XGBoost [31] — алгоритм, построенный на ансамбле слабых предсказывающих моделей
- Регрессия LASSO [32] — линейная регрессия с добавленной регуляризацией, которая заключается во введении дополнительного параметра. В этом алгоритме используется регуляризация L1, которая взвешивает ошибки по абсолютному значению.
- Ридж-регрессия [33] — алгоритм, аналогичный предыдущему, но он использует регуляризацию L2, то есть взвешивает ошибки по их квадрату.

Для оценки моделей была выбрана характеристика MAPE (средняя абсолютная ошибка в процентах) [34]. Этот коэффициент легко интерпретируется и часто используется для сравнения моделей (см. рис.3).

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

Рисунок 3. Формула подсчёта MAPE

Таблица 8 — Средняя абсолютная ошибка в процентах для моделей нападающих

| Алгоритм/<br>Параметр     | Линейная<br>регрессия | Случайный<br>лес | XGBoost    | Регрессия<br>Lasso | Ридж-ре-<br>грессия |
|---------------------------|-----------------------|------------------|------------|--------------------|---------------------|
| За сезон                  | 0.43128452            | 0.43101564       | 0.42478811 | 0.41351516         | 0.49690804          |
| По времени<br>на площадке | 0.27168860            | 0.27341709       | 0.27544888 | 0.26920587         | 0.27161855          |
| Без игр<br>плей-офф       | 0.30768135            | 0.31707778       | 0.30556947 | 0.30856364         | 0.30743493          |

Таблица 9 — Средняя абсолютная ошибка в процентах для моделей защитников

| Алгоритм/<br>Параметр     | Линейная<br>регрессия | Случайный<br>лес | XGBoost    | Регрессия<br>Lasso | Ридж-ре-<br>грессия |
|---------------------------|-----------------------|------------------|------------|--------------------|---------------------|
| За сезон                  | 0.45104690            | 0.46619587       | 0.47591978 | 0.46723438         | 0.44972414          |
| По времени<br>на площадке | 0.36275719            | 0.35827988       | 0.39512220 | 0.35803927         | 0.36216254          |
| Без игр<br>плей-офф       | 0.31173705            | 0.26844788       | 0.26821248 | 0.26116363         | 0.31264784          |

Из результатов таблиц 8 и 9 можно сделать вывод, что наилучшим способом является регрессия LASSO, причём для защитников лучше использовать датасет без игр плей-офф. Это объясняется тем, что в плей-офф тактики команд сильно изменяются и становятся более оборонительными, защитники реже рискуют и идут в нападение, это влияет на оценку защитников. Для нападающих наилучшим образом обучается модель на датасете по времени, проведённому на льду. Из-за того, что игроки проводят разное время на льду, не всегда объективно сравнивать

игроков по матчам. Поэтому мы смотрим параметры за определённое время. Клуб может узнать, сколько в среднем проводит игрок времени на льду в текущей команде, и оценить необходимость его покупки.

Для создания программы, предоставляющей пользовательский интерфейс были выбраны современные популярные технологии: Java, Spring, Hibernate, PostgreSQL.

Для рекомендации игроков пользователь должен выбрать команду и амплуа игрока. Поэтому было реализовано веб-приложение, в котором пользователю на выбор предоставляются команды и позиции игроков. После выбора игроков система предоставляет список рекомендованных игроков. Также для окончательного выбора игроков есть ссылки на статистику игроков на сайте КХЛ.

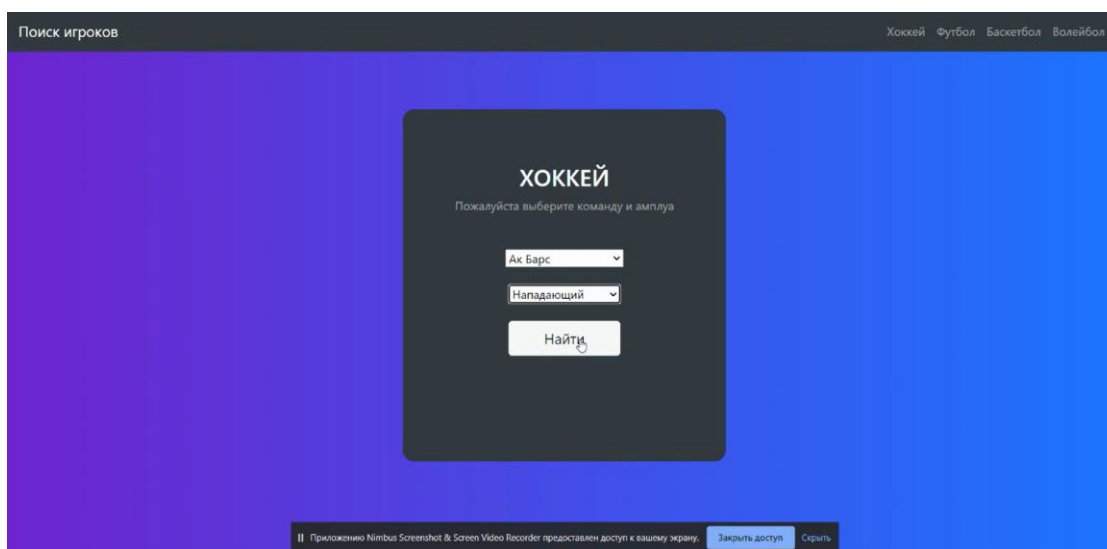


Рисунок 4. Выбор параметров для рекомендательной системы



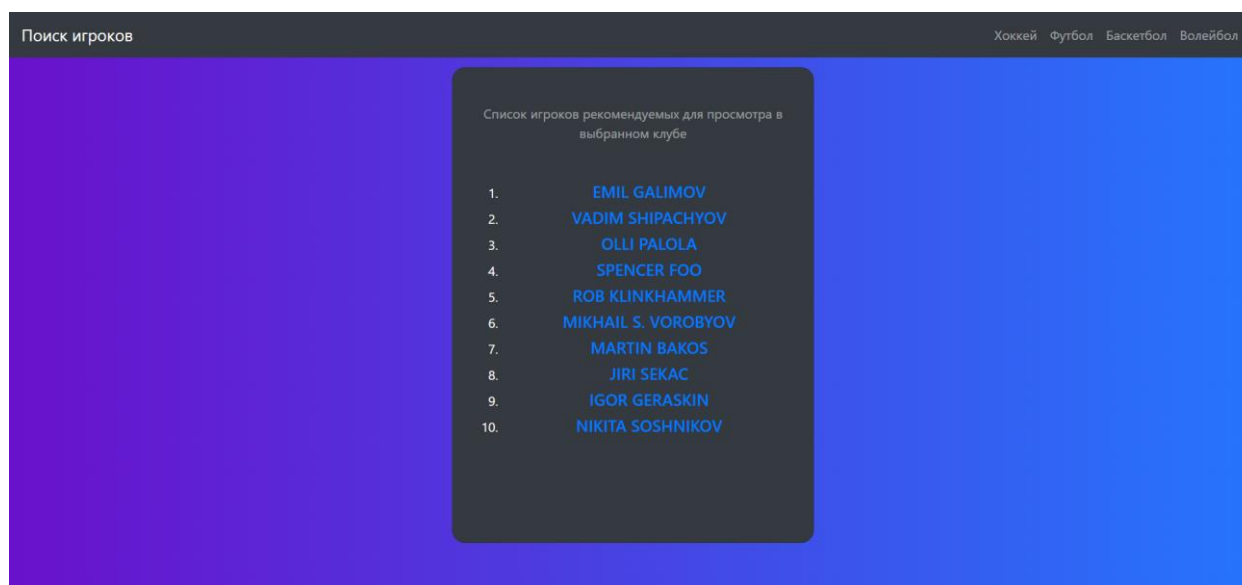


Рисунок 5. Рекомендация игроков нападения для хоккейного клуба «Ак Барс»

Для примера рассмотрим рекомендацию нападающих для казанского хоккейного клуба «Ак Барс».

Среди предложенных игроков есть Иржи Секач и Роб Клинкхаммер. Эти игроки уже играли за выбранный клуб и занимали ведущие роли в нём. Более того, в сезоне 2017/2018 оба игрока выступали за казанский клуб и выиграли вместе с ним кубок Гагарина. Можно предположить, что система выбрала игроков, которые достаточно хорошо вписывались в систему клуба и приносили большую пользу в игре команды.

Также среди предложенных игроков есть Вадим Шипачёв. В сезоне 2021/2022 он был признан самым ценным игроком лиги. В начале мая 2022 года стало известно, что игрок переходит в «Ак Барс». Значит, рекомендации системы совпали с необходимыми параметрами нападающего для казанской команды и ожидаемыми результатами от Вадима Шипачёва.

Среди остальных игроков есть как опытные игроки (как Никита Сошников), так и молодые и перспективные, например, Игорь Гераськин и Михаил Воробьёв.

Также для тестирования были просмотрены предложения рекомендательной системы за прошлый сезон. В списке предложенных был Дмитрий Кагарлицкий, который в этом сезоне стал одним из наиболее удачных приобретений ка-

занского клуба. Можно отметить, что в списке не было, например, Даниила Тарасова, который также стал игроком «Ак Барса» в этом сезоне, но не оправдал ожиданий казанской команды.

### **ЗАКЛЮЧЕНИЕ**

Изучено влияние различных доступных статистических параметров игроков на результаты команды, собран датасет и обучены модели машинного обучения, а также созданы алгоритм рекомендации и пользовательский интерфейс.

Результатом работы является рекомендательная система на основе игроков КХЛ, полученная с применением технологий машинного обучения. Системой могут пользоваться спортивные клубы, агенты, тренеры для поиска игроков, а также болельщики для изучения возможностей трансферов клубов.

В будущем планируется использовать систему для других видов спорта и расширить источники данных для обучения системы.

### **СПИСОК ЛИТЕРАТУРЫ**

1. Гореликов В.А. Трансфер игрока в профессиональном спорте как маркетинговый продукт спортивного клуба // Наука и спорт: современные тенденции. 2021. Т. 9, № 3. С. 115–124. <https://doi.org/10.36028/2308-8826-2021-9-3-115-124>
2. Зачем переходили? Летние трансферы КХЛ, которые пока не принесли никакой пользы. URL: <https://www.championat.com/hockey/article-4497349-igroki-neudachno-vystupayuschie-za-novye-kluby-v-khl-lindholm-tarasov-bartoshak-bryukvin-shalunov-dansk-rejdeborn.html>
3. Уровень зрелости российского спорттеха. URL: <https://rb.ru/analytics/russia-sporttech/>
4. Топ-10 самых популярных видов спорта в мире в 2021 году. URL: <https://themoney.co/ru/top-10-des-sports-les-plus-populaires-au-monde-en-2021/>
5. Россияне назвали самые популярные виды спорта. URL: [https://www.magram.ru/news/rossiyane\\_nazvali\\_samye\\_populyarnye\\_vidy\\_sporta.html](https://www.magram.ru/news/rossiyane_nazvali_samye_populyarnye_vidy_sporta.html)
6. Официальный сайт InStat. URL: <https://instatsport.com/>
7. Официальный сайт IceBerg. URL: <https://www.icebergsports.com/>
8. Официальный сайт DataVolley4. URL: <https://www.dataproject.com/Products/GLOBAL/en/Volleyball/DataVolley4>

9. Хоккейная статистика. URL: <https://www.eliteprospects.com/>
10. Футбольная статистика. URL: <https://ru.whoscored.com/Statistics>
11. Статистика на баскетбол. Где искать?  
URL: <https://bukmekerov.net/stati/cool/statistika-na-basketbol-gde-iskat/>
12. Волейбольная статистика. URL: <http://www.volleyservice.ru/>
13. Международный совет футбольных ассоциаций (IFAB).  
URL: <https://www.theifab.com/>
14. Правила волейбола.  
URL: <https://www.sportzone.ru/sport/rules.html?sport=volleyball>
15. Официальные правила баскетбола 2020. URL:  
<https://russiabasket.ru/Files/Documents/%d0%9e%d1%84%d0%b8%d1%86%d0%b8%d0%b0%d0%bb%d1%8c%d0%bd%d1%8b%d0%b5%20%d0%9f%d1%80%d0%b0%d0%b2%d0%b8%d0%bb%d0%b0%20%d0%91%d0%b0%d1%81%d0%ba%d0%b5%d1%82%d0%b1%d0%be%d0%bb%d0%b0%202020%20v.1.1.pdf>
16. Официальные правила хоккея 2021–2022.  
URL: <https://fhr.ru/upload/iblock/0a2/Ofitsialnaya-kniga-Pravil-IKHF.pdf>
17. Самые быстрые хоккеисты.  
URL: <https://www.sport-express.ru/hockey/nhl/reviews/bure-ili-makdevid-samye-bystrye-hokkeisty-v-istorii-nhl-1344222/>
18. Позиции игроков в хоккее и их функции.  
URL: <https://dcntrsport.com/poziczii-igrokov-v-hokkee-i-ih-funkczii/>
19. Официальный сайт Континентальной хоккейной лиги (КХЛ).  
URL: <https://www.khl.ru/>
20. Официальный сайт Национальной хоккейной лиги.  
URL: <https://www.nhl.com/>
21. Резервные вратари по-разному убивают время.  
URL: <https://www.nhl.com/ru/news/rezervnye-vratari-po-raznomu-ubivaiut-vremia/c-874329>
22. Хоккеисты на просмотровых контрактах в КХЛ. URL: <https://www.championat.com/hockey/article-3808815-hokkeisty-na-prosmotrovyyh-kontraktah-v-khl-v-sezone-201920.html>
23. Официальный сайт языка программирования Python.

URL: <https://www.python.org/>

24. Документация библиотеки requests.

URL: <https://requests.readthedocs.io/en/latest/>

25. Документация библиотеки Beautiful Soup.

URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

26. ХК «Динамо» и ХК МВД прекратили свое существование.

URL: <https://sportrbc.ru/news/5755069a9a79474acad120d9>

27. Самая безумная бойня в истории российского хоккея.

URL: <https://www.sport-express.ru/hockey/khl/reviews/10-let-massovym-drakam-v-matche-vityaz-avangard-1629757/>

28. Официальный сайт библиотеки scikit-learn.

URL: <https://scikit-learn.org/stable/>

29. Линейная регрессия. URL: <https://neurohive.io/ru/osnovy-data-science/linejnaja-regressija/>

30. Случайный лес.

URL: [https://dya-](https://dya-konov.org/2016/11/14/%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9-%D0%BB%D0%B5%D1%81-random-forest/)

[konov.org/2016/11/14/%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9-%D0%BB%D0%B5%D1%81-random-forest/](https://dya-konov.org/2016/11/14/%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9-%D0%BB%D0%B5%D1%81-random-forest/)

31. XGBoost. URL: <https://neerc.ifmo.ru/wiki/index.php?title=XGBoost>

32. Регрессия Lasso.

URL: <https://ranalytics.github.io/data-mining/042-Regularization.html>

33. Ридж-регрессия.

URL: <https://www.mygreatlearning.com/blog/what-is-ridge-regression/>

34. MAPE.

URL: <https://www.statisticshowto.com/mean-absolute-percentage-error-mape/>

## **RECOMMENDATION SYSTEM FOR SELECTION OF PLAYERS IN TEAM SPORTS BUILT ON THE BASIS OF MACHINE LEARNING**

**R. R. Shigapov**<sup>1[0000-0002-3163-0959]</sup>, **A. A. Ferenetz**<sup>2[0000-0002-7859-9901]</sup>

*Institute of Information Technology and Intelligent Systems, Kazan Federal University, 35 Kremlevskaya str., Kazan, 420008*

<sup>1</sup> shigapov.rinat.2000@gmail.com, <sup>2</sup> ist.kazan@gmail.com

### **Abstract**

This article describes the development of a recommender system for selecting players based on machine learning. The system introduced the example of hockey with the possibility of expanding its use in various team sports. For each sport different roles and characteristics of the players were considered. The article analyzes information about hockey, football, basketball and volleyball. The characteristics of the players are structured and divided into general groups. For each parameter coefficients are displayed that show the impact on the result of the match. Various machine learning algorithms were used to build the model. The web interface of the application has been created.

**Keywords:** *sports, hockey, selection of players, recommender system, machine learning*

### **REFERENCES**

1. *Gorelikov V.A.* Transfer of a player in professional sports as a marketing product of a sports club // *Science and sport: current trends.* 2021. V. 9, No. 3. S. 115–124. <https://doi.org/10.36028/2308-8826-2021-9-3-115-124>
2. Why did you switch? KHL summer transfers that have not brought any benefit yet. URL: <https://www.championat.com/hockey/article-4497349-igroki-neudachno-vystupayuschie-za-novye-kluby-v-khl-lindholm-tarasov-bartoshak-bryukvin-shalunov-dansk-rejdeborn.html>
3. The level of maturity of Russian sports technology. URL: <https://rb.ru/analytics/russia-sporttech/>
4. Top 10 most popular sports in the world in 2021. URL: <https://themoney.co/ru/top-10-des-sports-les-plus-populaires-au-monde-en-2021/>
5. Russians named the most popular sports. URL: [https://www.magram.ru/news/rossiyane\\_nazvali\\_samye\\_populyarnye\\_vidy\\_sporta.html](https://www.magram.ru/news/rossiyane_nazvali_samye_populyarnye_vidy_sporta.html)

6. InStat official website. URL: <https://instatsport.com/>
  7. IceBerg official website. URL: <https://www.icebergsports.com/>
  8. DataVolley4 official website.  
URL: <https://www.dataproject.com/Products/GLOBAL/en/Volleyball/DataVolley4>
  9. Hockey statistics. URL: <https://www.eliteprospects.com/>
  10. Football stats. URL: <https://en.whoscored.com/Statistics>
  11. Basketball statistics. Where to look?  
URL: <https://bukmekerov.net/stati/cool/statistika-na-basketbol-gde-iskat/>
  12. Volleyball statistics. URL: <http://www.volleyservice.ru/>
  13. International Football Association Board (IFAB).  
URL: <https://www.theifab.com/>
  14. Volleyball Rules.  
URL: <https://www.sportzone.ru/sport/rules.html?sport=volleyball>
  15. Official Basketball Rules 2020. URL: <https://russiabasket.ru/Files/Documents/%d0%9e%d1%84%d0%b8%d1%86%d0%b8%d0%b0%d0%bb%d1%8c%d0%bd%d1%8b%d0%b5%20%d0%9f%d1%80%d0%b0%d0%b2%d0%b8%d0%bb%d0%b0%20%d0%91%d0%b0%d1%81%d0%ba%d0%b5%d1%82%d0%b1%d0%be%d0%bb%d0%b0%202020%20v.1.1.pdf>
  16. Official Hockey Rules 2021–2022.  
URL: <https://fhr.ru/upload/iblock/0a2/Ofitsialnaya-kniga-Pravil-IHKHF.pdf>
  17. The fastest hockey players. URL: <https://www.sport-express.ru/hockey/nhl/reviews/bure-ili-makdevid-samy-e-bystrye-hokkeisty-v-istorii-nhl-1344222/>
  18. Positions of players in hockey and their functions.  
URL: <https://dcntrsport.com/poziczii-igrokov-v-hokkee-i-ih-funkczii/>
  19. Official website of the Kontinental Hockey League (KHL).  
URL: <https://www.khl.ru/>
  20. Official website of the National Hockey League. URL: <https://www.nhl.com/>
  21. Reserve goalkeepers kill time in different ways.  
URL: <https://www.nhl.com/ru/news/rezervnye-vratari-po-raznomu-ubivaiut-vremia/c-874329>
  22. Hockey players on viewing contracts in the KHL.
-

URL: <https://www.championat.com/hockey/article-3808815-hokkeisty-na-prosmotrovyyh-kontraktah-v-khl-v-sezone-201920.html>

23. The official website of the Python programming language.

URL: <https://www.python.org/>

24. Requests. URL library documentation: <https://requests.readthedocs.io/en/latest/>

25. Beautiful Soup Library Documentation.

URL: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

26. HC Dynamo and HC MVD ceased to exist.

URL: <https://sportrbc.ru/news/5755069a9a79474acad120d9>

27. The craziest massacre in the history of Russian hockey

28. Official website of the scikit-learn library. URL: <https://scikit-learn.org/stable/>

29. Linear Regression.

URL: <https://neurohive.io/en/osnovy-data-science/linejnaja-regressija/>

30. Random Forest.

URL: <https://dya->

[konov.org/2016/11/14/%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9-%D0%BB%D0%B5%D1%81-random-forest/](https://dya-konov.org/2016/11/14/%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D1%8B%D0%B9-%D0%BB%D0%B5%D1%81-random-forest/)

31. XGBoost. URL: <https://neerc.ifmo.ru/wiki/index.php?title=XGBoost>

32. Regression Lasso.

URL: <https://ranalytics.github.io/data-mining/042-Regularization.html>

33. Ridge Regression.

URL: <https://www.mygreatlearning.com/blog/what-is-ridge-regression/>

34. MAPE.

URL: <https://www.statisticshowto.com/mean-absolute-percentage-error-map/>

## СВЕДЕНИЯ ОБ АВТОРАХ



**ШИГАПОВ Ринат Рустемович** – выпускник бакалавриата Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, г. Казань.

**Rinat SHIGAPOV** – graduate student of the Institute of Information Technologies and Intelligent Systems, Kazan (Volga region) Federal University, Kazan.

e-mail: shigapov.rinat.2000@gmail.com

ORCID: 0000-0002-3163-0959



**ФЕРЕНЕЦ Александр Андреевич** – старший преподаватель кафедры программной инженерии Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета, г. Казань.

**Alexander FERENETS** – senior lecturer of Software Engineering of Institute of Information Technologies and Intelligent Systems KFU

e-mail: ist.kazan@gmail.com

ORCID: 0000-0002-7859-9901

*Материал поступил в редакцию 20 августа 2022 года*