

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра информатики и программирования

**РАЗРАБОТКА СИСТЕМЫ ПРЕДСКАЗАНИЯ ИСХОДА ТЕННИСНЫХ
МАТЧЕЙ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 441 группы

направления 02.03.03 Математическое обеспечение и администрирование
информационных систем

факультета компьютерных наук и информационных технологий

Шкодина Максима Сергеевича

Научный руководитель:

Старший преподаватель кафедры ИиП _____ А.А. Казачкова
подпись, дата

Зав. кафедрой:

к. ф.-м. н. _____ М.В. Огнева
подпись, дата

Саратов 2022

ВВЕДЕНИЕ

Актуальность темы. Прогнозирование – это оценка возможных путей развития какого-либо события. Оценка предполагаемого исхода может не только помочь в планировании, но и при правильном использовании информации может принести прибыль. Для предсказания результата спортивного события, как правило, используются информация предыдущих матчей, и на ее основе можно сделать предположение об исходе предстоящего события. На результат может влиять множество непредсказуемых факторов, например: психоэмоциональное состояние игроков, скрытые травмы, полученные до поединка, а также погодные условия на момент проведения матча. Но на основе большой выборки, эти факторы нивелируются, а это позволяет предсказать вероятность возникновения того или иного события, как во время, так и после завершения матча.

В наше время во всемирной сети можно найти множество результатов исходов матчей почти по любому виду спорта, однако не все найденные данные могут подойти для построения модели. Информация о матчах должна содержать все ключевые параметры, которая учитывает выработанная модель.

Теннис – это вид спорта, в котором соперничают либо два игрока, либо две команды, состоящие из двух игроков. Задачей соперников — теннисиста или теннисистки — является при помощи ракеток отправлять мяч на сторону соперника так, чтобы тот не смог его отразить, не более чем после первого падения мяча на игровом поле на половине соперника.

Прогнозирование результатов теннисных матчей является непростой задачей, но стоит учитывать, что в поединке участвует меньшее количество игроков, чем в других видах спорта, следовательно на исход спортивного события будет влиять меньше количество параметров.

Тема прогнозирования результатов теннисных матчей является достаточно актуальной, ведь каждый человек, кто делает ставки на исход событий рано или поздно задумывается, о том, какие параметры поединка больше всего влияют на результат матча.

Существуют сервисы, которые предсказывают исход поединка в теннисе, но большинство из них определяет вероятность победы игрока до матча, и выбирают матчи только по мужскому теннису. Например, система сервиса OhMyBet предлагает коэффициент доходности ROI в 12% и использует машинное обучение для выдачи прогноза на матч. В данной работе рассматривается другой подход к прогнозированию, т.е система, которая отбирает матч и делает прогноз по определенному алгоритму.

Цель бакалаврской работы: разработать систему предсказания исхода теннисных матчей в режиме реального времени. Для достижения поставленной цели были определены **следующие задачи:**

1. Проанализировать предметную области;
2. Выбрать подходящий сайт для сбора информации о теннисных матчах;
3. Изучить структуру сайта;
4. Изучить необходимые материалы по библиотекам для анализа данных;
5. Создать парсер для сбора информации;
6. Проанализировать информацию, собранную в ходе парсинга;
7. Выработать стратегию для предсказания выигрыша;
8. Ознакомиться с работой Telegram API (Application Program Interface);
9. Создать систему мониторинга матчей по стратегии;
10. Протестировать работу системы.

Методологические основы анализа теннисных матчей представлены в работах Тристана Барнетта, Стеффана Кларка, Вильяма Кнотенбельта, а также Магнет Мудуруски.

Теоретическая значимость бакалаврской работы заключается в создании способа предсказания исхода теннисного матча без использования цепей Маркова, а также машинного обучения.

Практическая значимость бакалаврской работы заключается в разработке способа сбора информации в режиме реального времени

Структура и объём работы. Бакалаврская работа состоит из введения, 4 разделов, заключения, списка использованных источников и 4 приложений. Общий объем работы – 107 страниц, из них 46 страниц – основное содержание, включая 6 рисунков, цифровой носитель в качестве приложения, список использованных источников информации – 20 наименований.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Системы предсказаний исхода теннисных матчей»

посвящен обзору способов анализа теннисных матчей.

Существуют различные способы предсказания теннисных матчей. Большинство современных моделей для прогнозирования тенниса используют иерархические стохастические выражения на основе цепей Маркова. Также существуют подходы к анализу, основанные на машинном обучении.

Цепь Маркова – это система перехода между различными состояниями в определенном пространстве состояний. Важным свойством данной системы является то, что следующее состояние системы зависит только от текущего. Если представить, что счёт в гейме – это пространство состояний, а переходы между состояниями – вероятности того, что первый игрок выиграет или проиграет очко, то получается цепь Маркова, которая отражает стохастическую прогрессию счета в гейме.

Машинным обучением называется раздел искусственного интеллекта, изучающий алгоритмы, способные обучаться или адаптировать свою структуру на основании обработанной выборки данных. Машинное обучение с учителем решает задачу построения функции из набора помеченных обучающих примеров, где помеченный пример — это пара, состоящая из вектора на входе и желаемого значения на выходе.

В контексте тенниса историческая статистика матчей может использоваться для формирования выборки обучающих примеров. Для отдельного матча входной вектор может содержать различные признаки матча и игроков, а выходным значением будет исход матча. Отбор релевантных признаков – это одна из основных проблем построения эффективного алгоритма машинного обучения.

Второй раздел «Методы получения и анализа информации»

посвящен способам получения информации из интернета, а также содержит

информацию о библиотеках на языке программирования Python, которые способны помочь в анализе полученной информации.

Программа, дающая возможность компьютеру «читать», то есть анализировать текст, сравнивая предложенные слова с имеющимися во Всемирной сети, называется парсером. Сфера применения таких программ действительно широка, но они практически не отличаются и работают по одному алгоритму. Меняются лишь цели и способы сбора информации.

В последнее время появилось множество библиотек на языке Python, которые облегчают процесс парсинга информации. Основными 3 являются: Scrapy, BeautifulSoup, а также Selenium web driver, но у каждой библиотеки есть свои минусы и плюсы.

Третий раздел «Описание предметной области» посвящен описанию правил в теннисе.

Четвертый раздел «Разработка системы предсказания исхода теннисных матчей» посвящен разработке системы реального времени, которая позволяет отслеживать матчи по определенной стратегии, а также давать сигнал пользователю на совершение ставки. Для сбора статистики был реализован парсер, который собирает информацию с сайта <https://www.livesport.com> по следующему алгоритму:

1. Переход по ссылке игрока в рейтинге WTA;
2. Для каждого игрока определен список сыгранных матчей за всю карьеру. Осуществляем переход по каждой такой ссылке;
3. Открывается окно с информацией о матче с 3 разделами: обзор матча, статистика, ход матча;
4. Собираем информацию о счете сетов в матче, и далее кликаем на вкладку статистика;
5. Сохраняем для каждого следующие параметры: процент первой подачи, очки выигранные на подаче, очки выигранные на приеме и переходим в раздел ход матча;

6. Для каждого сета сохраняем информацию о ходе сета, а также последовательность очков в каждом гейме.

Далее полученный данные проанализированы и выведена следующая стратегия:

1. Если счет матча 3-1, 3-2, 1-3, 2-3, 3-3, то он попадает в первую корзину, в которой проверяется наличие брейка.
2. Если брейк случился, то далее матч попадет во вторую корзину, в которой проверяется второе условие стратегии. Пользователю отправляется оповещение о возможной ставке
3. Если матч подошел по 2 критерию, то сразу проверяется и 3 условие. В случае прохождения этих двух условия матч попадет в 3 корзину, которая необходима для проверки результатов сета.

Анализ данных по этой стратегии показал:

1. Процент выигрыша по стратегии составляет 64%, что является очень хорошим результатом;
2. Процент выигрыша по стратегии не учитывая условие подачи составляет всего 52%. Отсюда можно сделать вывод, что последнее условие стратегии оказывает большое значение.
3. Около 14 % матчей подходит для ставки по стратегии, а это каждый 7 матч, поэтому человеку очень сложно отследить все матчи.

Для проверки корректности результатов исходный датафрейм был случайным образом разбит на 10, равных по количеству матчей, наборов данных. В среднем по каждому отдельному датафрейму получились аналогичные результаты с отклонением в 3 процента.

В ходе работы была разработана система, которая оповещает пользователей о подходящей по стратегии ставке. Программа была выложена на хостинговый сервис Heroku, который предоставляет бесплатный доступ для одной программы. Данная система дала сигнал на ставку для 567 матчей. По результатам собранной статистики по матчам получились следующие результаты: ставка выиграла для 361 матчей, проиграла для 206. Процент

выигрышных ставок составил 63,6%. Для подсчета прибыльности стратегии был взят коэффициент возврата инвестиций ROI, который выражается следующей формулой:

$$ROI = \frac{\text{сумма выплат} - \text{сумма ставок}}{\text{изначальный размер банка}} * 100\%$$

Для каждого матча делалась ставка в размере 100 условных единиц. В итоге общая сумма выплат составила 58428 у.е, сумма ставок составила 56700 у.е, изначальный размер банка – 20000 у.е. В результате расчетов коэффициент ROI составил 8.64%.

ЗАКЛЮЧЕНИЕ

Целью данной работы было создание системы реального времени, которая находит матчи по стратегии и отправляет оповещения о них в Telegram канал. В ходе данной работы были изучены все необходимые библиотеки для анализа данных и сбора информации. Также был реализован парсер, который собирает информацию о сыгранных матчах в женской лиге. После чего была проанализирована статистика выигрышных ставок и подсчитана примерная прибыль. Для подбора матчей была разработана система реального времени, которая анализирует матчи и отправляет оповещения пользователям. Данная работа имеет дальнейшее развитие: можно добавить стратегии для других видов спорта, также добавить бота, который продавать различные типы подписок пользователям и следить за сроком их истечения. Тем самым получится полностью независимая система, которая будет нуждаться только в оплате хостинга. Также можно добавить другие методы отправки оповещений, например сделать поддержку других социальных сетей таких, как VK, WhatsApp, Viber, Facebook, а также сделать мобильное приложение, где будет появляться всплывающее окно.

Основные источники информации:

1. T. Barnett and S. R. Clarke. Combining player statistics to predict outcomes of tennis matches. IMA Journal of Management Mathematics, 16:113–120, 2005.
2. W. J. Knottenbelt, D. Spanias, and . A common-opponent stochastic model for predicting the outcome of professional tennis matches. Computers and Mathematics with Applications, 64:3820–3827, 2012.
3. A. M. Madurska. A Set-By-Set Analysis Method for Predicting the Outcome of Professional Singles Tennis Matches. Technical report, Imperial College London, London, 2012.
4. Хабр: «Парсинг сайтов: как с точки зрения закона выглядит один из самых полезных ИТ — инструментов по миру (и в России)?» [Электронный

ресурс]. URL: <https://habr.com/ru/post/340302/> (дата обращения: 10.01.2022).

5. Хабр: «Python-пакеты для Data Science» [Электронный ресурс]. URL: <https://habr.com/ru/company/ruvds/blog/525834/> (дата обращения: 15.01.2022).
6. DevGang: «Python: Scrapy, Selenium, BeautifulSoup что лучше для парсинга веб сайтов» [Электронный ресурс]. URL: <https://dev-gang.ru/article/python-scrapy-selenium-beautiful-soup-czto-luczshe-dlja-parsinga-veb-saitov-rmv3n1q3us> (дата обращения 10.01.2022)