Модель логистической регрессии предиктивной аналитики в распределении дел государственных защитников

Новиков Всеволод Александрович
к.т.н., доцент, доцент кафедры информационных систем и технологий, Кубанский институт профессионального образования, vsenovik@bk.ru,

Новикова Анастасия Александровна
к.т.н., доцент, доцент кафедры информационной безопасности, Краснодарский университет МВД России, gingary@yandex.ru,

Барковская Дарья Леонидовна
студентка, специальность «Юриспруденция», Кубанский институт профессионального образования, 930copper@tiffincrane.com

Статья посвящена применению предиктивной аналитики для управления нагрузкой государственных защитников. Если можно заранее понять, насколько рискованным будет дело, это даёт основу для более рационального распределения. Сложные дела должны попадать к более опытным защитникам. Раньше такие методы использовались в основном для прогнозирования судебных решений. Перенос этого подхода на задачу распределения дел пока изучен слабо.
Эмпирическую базу составили открытые данные Верховного суда Индии. На их основе в платформе Loginom построена модель логистической регрессии. Она оценивает вероятность отклонения заявленных требований и относит дело к одному из уровней риска. Прогноз становится формальным критерием при назначении защитника. Чем выше риск, тем опытнее специалист. Такой подход не только снижает вероятность неблагоприятного исхода. Он делает распределение дел более прозрачным и обоснованным.
Индийский контекст специфичен, прямой перенос модели в Россию невозможен. Но логика подхода адаптируема. По мере цифровизации судопроизводства и накопления массива отечественных судебных актов подобные инструменты могут найти применение и здесь.
Ключевые слова: государственный защитник, предиктивная аналитика, логистическая регрессия, судебные решения, распределение дел, риск-ориентированный подход.

Введение

Проблема, которая остро стоит сегодня – это неравенство в качестве юридической помощи между частными и назначаемыми защитниками. Особенно это касается социально незащищённых граждан. Исследования показывают, что исход дела зависит не только от фактических обстоятельств и применимого права. Он зависит также от того, как устроена система государственной защиты, кто конкретно ведёт дело, сколько у человека опыта, какая его загрузка и в чём специализация.

Механизмы назначения защитников чаще всего формально случайны, основаны на ручном распределении нагрузки. Такие механизмы не учитывают прогнозируемую сложность дела, а также вероятность неблагоприятного исхода для подзащитного. Отсюда и возникает задача построения формализованной, прозрачной и проверяемой системы, которая опирается на данные судебной практики и позволяет более рационально распределять дела.

Научная новизна предлагаемого подхода состоит в переносе методов предиктивной аналитики на задачу управления ресурсами системы государственной защиты. Такой подход уже широко применяемых для прогнозирования исходов судебных споров. Предполагается использовать открытые массивы данных о судебных делах. На их основе строится модель, которая по формализованным признакам дела оценивает риск неблагоприятного исхода (осуждения, строгого приговора, отказа в удовлетворении требований).

Новизна работы состоит в том, что прогноз становится входным параметром алгоритма распределения дел. Дела высокого риска закрепляются за более опытными и статистически результативными адвокатами. Дела низкого риска будут закрепляться за защитниками с меньшим стажем или стажёрами под контролем наставника. В итоге формируется управленческая модель, в которой качество защиты по тяжёлым делам перестаёт зависеть от случайного назначения или субъективного выбора.

Практическая значимость состоит в том, что подобная система позволяет формализовать и обосновать распределение дел между государственными защитниками. Снижается риск скрытого «ручного управления» и неравномерной загрузки отдельных адвокатов. Учёт прогнозируемой сложности и рисков по делу даёт возможность гарантировать минимальный стандарт качества защиты по тем категориям дел, где цена ошибки особенно высока (тяжкие составы, уязвимые группы, риск длительной изоляции). Интеграция такой модели в цифровую инфраструктуру судов и адвокатуры вписывается в общую тенденцию цифровизации правосудия и внедрения решений на основе данных. Это уже декларируется в стратегиях развития судебной системы и правосудия в РФ. При соблюдении правовых и этических ограничений подобный инструмент может стать одним из средств выравнивания доступа к квалифицированной юридической помощи.

1. Анализ литературы

1.1. Эмпирика по эффективности государственной защиты

Ряд обзоров систематизирует исследования о том, как разные модели предоставления юридической помощи малоимущим влияют на исход дел [1, 3]. Обобщение ключевых работ показывает, что более современные исследования, учитывающие демографию, криминальный бэкграунд и другие факторы, фиксируют заметные различия в результатах в зависимости от типа и качества защиты. В некоторых юрисдикциях публично финансируемая защита демонстрирует даже несколько лучшие результаты, чем частные адвокаты [2]. Однако в среднем устойчиво подтверждается другое. Структура системы защиты малоимущих и качество конкретного защитника статистически значимо влияют на вероятность осуждения и длину срока [7]. В сумме это создаёт серьёзную фактологическую базу для тезиса о том, что кто именно защищает подсудимого и как организована система защиты, важно не меньше, чем формальные нормы права. И это прямое обоснование для попыток формализовать и оптимизировать распределение дел.

1.2. Рандомное распределение дел и влияние личности адвоката

Ключевой класс работ использует случайное распределение дел как естественный эксперимент, чтобы измерить «способности адвоката» без сильной систематической ошибки отбора. В исследовании «Using Random Case Assignment to Investigate Attorney Ability» показано, что при случайном назначении адвокатов на уголовные

дела различия в результативности между адвокатами остаются существенными [7]. Авторы фиксируют, что подсудимые, случайно получившие адвоката в 90-м перцентиле по эффективности, имеют на 14 процентных пунктов меньшую вероятность лишения свободы по сравнению с теми, кому достался адвокат в 10-м перцентиле. Ожидаемый срок снижается в среднем на 1,2 месяца [7]. Эти данные подтверждают две вещи:

1. Рандомизация действительно даёт более «честную» картину влияния адвоката.

2. Даже при формальном равенстве процедур исходы заметно зависят от индивидуальных характеристик защитника — опыта, специализации, нагрузки. На этом фоне идея системно учитывать прогнозируемый риск дела и распределять более сложные кейсы к более сильным адвокатам получает эмпирическую основу [7].

2. ИИ и предиктивные модели в правосудии и для защитников

В последние годы активно появляются работы о применении ИИ для поддержки общественных защитников и расширения доступа к правосудию [4]. Исследования показывают, что защитники видят в ИИ прежде всего инструмент ускорения рутинных задач: поиск релевантной практики, подготовка черновиков ходатайств, быстрый вход в новую тему [4, 6]. В другом направлении развивается информационный поиск и поиск юридической информации (LCR) для общественных защитников. Модели подбирают релевантные прецеденты и нормы под конкретные фактические обстоятельства дела, повышая качество аргументации без прямого воздействия на процессуальные решения о назначении адвоката.

Параллельно существует обширный блок работ по предсказанию исходов судебных дел на базе открытых датасетов высших судов [5]. Созданы наборы данных и модели, демонстрирующие приемлемую точность прогнозирования исхода на основе текстовых и структурированных признаков дела. На уровне практики появляются сервисы, предлагающие ИИ-прогноз шансов выиграть суд для граждан. Это свидетельствует о технологической зрелости самой идеи предиктивной аналитики в правосудии.

В российском контексте ИИ уже внедряется в суды, но пока преимущественно во вспомогательных задачах (распознавание речи и автоматическое составление протоколов, поиск нормативных актов, подготовка проектов судебных решений и анализ документов). Это создаёт институциональную базу для последующего перехода от чисто вспомогательных инструментов к более содержательным системам поддержки решений, в том числе в сфере организации государственной защиты [1].

3. Место предлагаемого подхода на фоне литературы

Обзор показывает несколько вещей. Эффективность систем госзащиты и отдельных защитников эмпирически подтверждённо влияет на исход дел [1, 2, 7]. Случайное распределение дел используется как метод исследования, но не как способ оптимизации качества защиты [7]. ИИ и предиктивные модели уже успешно применяются для анализа судебных данных и поддержки работы защитников. Однако не найдено работ, где предсказание риска по делу служит формальным основанием для распределения дел между государственными защитниками по уровню опыта [4, 5, 6]. Таким образом, предлагаемый в статье подход занимает нишу на стыке трёх направлений и обладает научной новизной.

4. Источник и структура данных

Для построения предиктивной модели использовались открытые данные Верховного суда Индии, полученные с официального сервиса публикации судебных решений. Индия выбрана в качестве объекта исследования, поскольку национальное законодательство допускает массовую публикацию текстов судебных актов высшей судебной инстанции в машиночитаемом виде. Это облегчает построение статистических моделей на реальной судебной практике [3].

В выборку включены 42342 дела Верховного суда Индии. Каждое дело описано следующими полями: уникальный идентификатор дела; категория дела (гражданское, уголовное, неопределённое); вид производства (апелляция, конституционная жалоба, специальное ходатайство и т.п.); служебная информация (номер, состав суда, даты); полный текст судебного решения; итоговый исход (принято, отклонено, иное) [3]. Для целей прогнозирования исхода дела и последующего распределения дел по уровню риска использовались только дела с однозначным результатом (принято или отклонено). Наблюдения с неопределённым исходом и неопределённой категорией дела исключались на этапе предобработки. Это позволило получить очищенную выборку с бинарной целевой переменной «исход дела».

5. Предобработка и формирование признаков

Текстовые данные судебных решений имеют значительный объём и сложную структуру. В связи с этим на первом этапе выполнялась их структуризация и агрегирование в набор признаков, пригодных для интеллектуального анализа. Предобработка включала несколько шагов.

Категориальные поля. Категория дела, вид производства, состав суда нормализовались и кодировались в виде фиктивных переменных. По тексту решения выделялись простые количественные признаки (общая длина, наличие ключевых лексем, характеризующих исход, количество «доказательных» предложений). Там, где это было необходимо, текст агрегировался в интегральные индикаторы, например индекс сложности дела по длине и разнообразию правовых норм. Сформированный набор данных импортировался в аналитическую платформу Loginom, где выполнялись последующие этапы моделирования.

6. Постановка задачи и выбор модели

Целью моделирования является оценка вероятности неблагоприятного исхода дела для стороны заявителя, то есть отклонения требований. Задача бинарная (принято или отклонено). Поэтому логистическая регрессия стала естественным выбором базового метода. Она даёт непосредственные оценки вероятности и позволяет интерпретировать вклад отдельных факторов.

В Loginom задача формулировалась так. Зависимая переменная — бинарный исход дела (1 — требования удовлетворены, 0 — отклонены). Независимые переменные — категория дела (уголовное или гражданское), вид производства, длина решения, индикаторы наличия ключевых лексем, характеристики состава суда и другие. Итоговый результат модели — оценочная вероятность удовлетворения требований по конкретному делу. Для защиты от переобучения и повышения устойчивости результатов использовались встроенные механизмы регуляризации (L2-регуляризация, Ridge) и автоматический отбор значимых факторов, доступные в модуле логистической регрессии Loginom.

7. Использование предсказаний для распределения дел по уровню риска

Полученная модель возвращает для каждого дела числовое значение — оценочную вероятность неблагоприятного исхода. На основе этой вероятности предлагается вводить градацию уровней риска. Низкий — при вероятности менее 0,4; средний — в диапазоне 0,4–0,7; высокий — свыше 0,7. Шкала задаётся административно и может калиброваться в зависимости от приоритетов системы государственной защиты.

Для связки прогноза с назначением защитника «опытность» предлагается задавать несколькими формальными индикаторами (стаж по соответствующей категории дел, числом таких дел за последний период и долей дел без неблагоприятного исхода при сопоставимой тяжести). На основе этих показателей каждому защитнику присваивается уровень квалификации по типам дел (например, базовый или повышенный). Алгоритм назначения формулируется как правило. Дела высокого риска направляются защитникам с более высоким уровнем по данной категории с учётом текущей нагрузки. Дела низкого риска допускают назначение менее опытных защитников или стажёров под наставничеством. Это позволяет напрямую связать предиктивную модель, реализованную в Loginom, с практической задачей управления ресурсами государственной защиты.

8. Ограничения исследования

Использование предиктивной аналитики для распределения дел между защитниками ставит вопросы о прозрачности алгоритма, рисках скрытой дискриминации и недопустимости подмены профессионального усмотрения адвоката или суда автоматизированными рекомендациями. Эти аспекты требуют дальнейшего этико-правового анализа и нормативного регулирования. Кроме того, модель не претендует на точечный прогноз каждого дела. Она рассматрива-

ется именно как инструмент раннего риск-ориентированного распределения нагрузки, который требует последующей адаптации и перекалибровки при переносе в другие юрисдикции. Следует также отметить, что использованные данные включают как уголовные, так и гражданские дела. Разработанный метод риск-стратификации универсален и может быть применён к любой категории дел при наличии бинарного исхода, интересующего сторону защиты. Однако при адаптации для системы государственных защитников в уголовном процессе потребуется дополнительная настройка признакового пространства.

9. Экспериментальная часть

Для построения предиктивной модели распределения дел по уровню риска использовалась очищенная выборка судебных дел Верховного суда Индии. Эксперимент проводился в аналитической платформе Loginom и включал полный цикл обработки данных. Нами использовались импорт исходного файла, проверка и настройка типов полей, фильтрация пропусков и аномальных наблюдений, обучение модели логистической регрессии и оценка её качества.

На первом этапе в Loginom был создан отдельный пакет. Через узел импорта загружался файл в формате CSV, содержащий сведения о 42342 судебных дел. При загрузке проверялась корректность распознавания разделителя столбцов и кодировки. Затем выполнялась настройка типов шкал для каждого поля. Идентификаторы дел, тип производства и категория дела задавались как номинальные (категориальные) признаки. Числовые показатели, как количественные. Текстовые поля с полным текстом решения и обоснованием исхода отмечались как строковые и на данном этапе использовались только для формирования агрегированных индикаторов.

Для оценки качества исходных данных в Loginom применялся визуализатор «Качество данных». Это позволило получить сводную информацию о количестве пропусков, распределении значений и наличии потенциально проблемных полей. На основе диагностики были идентифицированы столбцы с единичными пропусками и поля, полностью дублирующие информацию других признаков. Их исключили из дальнейшего анализа, чтобы не создавать искусственную коллинеарность.

Следующим шагом стала фильтрация записей с некорректно заполненными ключевыми полями. Из выборки исключались дела с неопределённой категорией (Undetermined), а также наблюдения без однозначного исхода (Accept или Reject). Это обеспечило бинарный характер целевой переменной для последующей логистической регрессии. Дополнительно с помощью узла фильтрации удалялись строки с пропусками в критически важных полях (категория дела, тип производства). Их восстановление по косвенным признакам в рамках эксперимента не предполагалось.

Для борьбы с аномальными значениями и выбросами использовались базовые статистические критерии. По количественным полям рассчитывались медиана и межквартильный размах. Наблюдения с экстремальными значениями за пределами разумного диапазона (условно, выше 99-го перцентиля) рассматривались как аномальные. Часть таких записей исключалась, часть оставлялась при условии, что они соответствуют реально сложным делам и не являются артефактом загрузки. Это позволило снизить влияние технических выбросов на оценку коэффициентов модели.

После очистки данных был сформирован набор признаков, используемых в качестве входных переменных модели. В него вошли категориальные признаки (категория дела, вид производства, некоторые характеристики состава суда) и агрегированные количественные признаки (длина судебного решения в словах, количество предложений, отнесённых к разряду «доказательных», индикаторы наличия ключевых лексем, связанных с удовлетворением или отклонением требований). Категориальные признаки преобразовывались в бинарные индикаторы методом one-hot кодирования средствами Loginom. Для числовых переменных при необходимости выполнялась нормализация (масштабирование). Это повышает устойчивость алгоритма оптимизации при обучении логистической регрессии.

На очищенной и подготовленной выборке была построена модель логистической регрессии с использованием стандартного узла «Логистическая регрессия» в Loginom. В качестве зависимой переменной выступал бинарный исход дела (1 — требования заявителя удовлетворены, 0 — отклонены). В качестве независимых переменных задавались все отобранные категориальные и количественные признаки.



Рис. 1. Готовая модель назначения общественных защитников

Данные разделялись на обучающую и тестовую подвыборки в пропорции 80/20. В обучающую попали первые 80% дел по хронологическому признаку, в тестовую — оставшиеся 20%. Это имитирует реальную ситуацию прогнозирования результатов по более поздним делам на основе предыдущей практики. В настройках узла логистической регрессии включалась L2-регуляризация для снижения риска переобучения и использовался встроенный пошаговый отбор переменных (stepwise), автоматически исключающий статистически незначимые факторы.

После обучения модель применялась к тестовой выборке. Для каждого дела рассчитывалась оценочная вероятность благоприятного исхода для заявителя. На основе фактического исхода и прогнозов вычислялись стандартные метрики качества: доля правильно классифицированных наблюдений (accuracy), точность и полнота для класса «неблагоприятный исход», а также площадь под ROC-кривой (AUC-ROC). Это позволило оценить дискриминирующую способность модели.

Полученная модель продемонстрировала удовлетворительную предсказательную способность и интерпретируемую структуру коэффициентов. Это позволило использовать её прогнозы как основу для построения шкалы уровней риска дела и последующей концепции алгоритмического распределения дел между государственными защитниками.

Вывод

По результатам эмпирического исследования показано, что логистическая регрессия обеспечивает удовлетворительное качество прогнозирования исхода дела для заявителя на основе ограниченного набора формализованных признаков. Это подтверждает возможность использования сравнительно простых моделей предиктивной аналитики для предварительной оценки риска неблагоприятного исхода по конкретному делу.

Предложенная шкала стратификации дел по уровням риска на основе оценочной вероятности неблагоприятного исхода создаёт формальный инструмент для выделения групп дел, требующих привлечения наиболее опытных государственных защитников. Использование таких уровней риска в алгоритме назначения защитников позволяет перейти от фактически случайного или ручного распределения к риск-ориентированной модели управления ресурсами государственной защиты.

Сформулированная схема сопоставления уровней риска дел и формализованных индикаторов «опытности» защитников (стаж, количество дел соответствующей категории, доля дел без неблагоприятного исхода) демонстрирует, что предиктивная модель может быть интегрирована в практику распределения дел без вмешательства в собственно судебное решение. Важным результатом является то, что предлагаемый инструмент ориентирован на повышение прозрачности и воспроизводимости административных решений, а не на автоматизацию правосудия как такового.

Анализ ограничений и рисков показал, что применение предиктивной аналитики в сфере государственной защиты требует учёта вопросов алгоритмической справедливости, недопустимости скрытой дискриминации и сохранения профессиональной автономии адвокатов. Модель должна рассматриваться исключительно как вспомогательный механизм поддержки управленческих решений и подлежит регулярной переоценке и перекалибровке при переносе в другие юрисдикции и при изменении структуры данных.

Полученные результаты подтверждают, что использование предиктивной аналитики для риск-ориентированного распределения дел между государственными защитниками обладает потенциалом повышения равномерности нагрузки и выравнивания доступа к квалифицированной юридической помощи для социально уязвимых категорий граждан.

Заключение

Проведённое исследование продемонстрировало принципиальную реализуемость интеграции предиктивной аналитики в процессы организационного управления системой государственной защиты без подмены судебного усмотрения автоматизированными решениями. Использование логистической регрессии в качестве базовой модели показало, что даже агрегированные количественные и категориальные признаки судебных дел позволяют конструировать практико-ориентированные инструменты риск-стратификации.

С учётом текущего уровня цифровизации судебной системы Российской Федерации и постепенного расширения доступа к массивам судебных актов предложенный подход может быть адаптирован к национальному контексту. Условия для этого — формирование репрезентативных датасетов и разработка процедур внешней валидации моделей. Особое значение будет иметь институциональное закрепление требований к прозрачности алгоритмов, порядку их аудита и распределению ответственности между разработчиками, органами судебной власти и профессиональным сообществом адвокатов.

В дальнейших исследованиях целесообразно расширить набор признаков за счёт более глубокого анализа текстов судебных актов. Также имеет смысл исследовать альтернативные классы моделей (градиентный бустинг, нейронные сети) с учётом требований интерпретируемости и провести пилотные эксперименты на данных российских судов и региональных систем юридической помощи. Это позволит оценить устойчивость полученных результатов в других юрисдикциях и выработать практические рекомендации по внедрению риск-ориентированных подходов в деятельность органов государственной защиты.

Литература

1. Агафонова, А. Г. Цифровизация судебной власти как фактор повышения эффективности правосудия / А. Г. Агафонова, М. В. Радаева. — 2025.

2. Куликова, Н. В. Искусственный интеллект в правовой предиктивной аналитике: совершенствование судебных стратегий / Н. В. Куликова // Вестник ВГУЭС. — 2022

3. A/HRC/53/31. Report of the Special Rapporteur on the independence of judges and lawyers / UN Human Rights Council. — 2023.

4. Aletras, N. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective / N. Aletras [et al.] // PeerJ Computer Science. — 2016.

5. Anderson, J. M. How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes / J. M. Anderson, P. Heaton // Yale Law Journal. — 2012. — Vol. 122. — P. 154–217.

6. Corbett-Davies, S. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning / S. Corbett-Davies, S. Goel // Annual Review of Statistics and Its Application. — 2018.

7. Human rights lawyers at risk / Human Rights House Foundation. — 2021.

A Logistic Regression Model of Predictive Analytics in Public Defender Case Allocation

Novikov V.A., Novikova A.A., Barkovskaya D.L.

Kuban Institute of Professional Education, Krasnodar University of the Ministry of Internal Affairs of Russia, Kuban Institute of Professional Education

This article examines the use of predictive analytics to manage public defender caseloads. Knowing the risk of a case in advance provides the basis for more rational allocation. Complex cases should be assigned to more experienced defenders. Previously, such methods were used primarily to predict court decisions. The application of this approach to case allocation has been poorly studied.

The empirical base was provided by open data from the Supreme Court of India. A logistic regression model was built on this data in the Loginom platform. It assesses the probability of rejection of the stated claims and assigns the case to one of the risk levels. This prediction becomes a formal criterion for assigning a defender. The higher the risk, the more experienced the defender. This approach not only reduces the likelihood of an unfavorable outcome but also makes case allocation more transparent and justified. The Indian context is unique, and a direct transfer of the model to Russia is impossible. However, the logic of the approach is adaptable. As legal proceedings become digitalized and a body of domestic judicial decisions accumulates, similar tools may find application here as well.

Keywords: public defender, predictive analytics, logistic regression, court decisions, case assignment, risk-based approach.

References

1. Agafonova, A. G. Digitalization of the judiciary as a factor in increasing the efficiency of justice / A. G. Agafonova, M. V. Radaeva. — 2025.
2. Kulikova, N.V. Artificial intelligence in legal predictive analytics: improving judicial strategies / N.V. Kulikova // Vestnik VGUES. — 2022
3. A/HRC/53/31. Report of the Special Rapporteur on the independence of judges and lawyers / UN Human Rights Council. — 2023.
4. Aletras, N. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective / N. Aletras [et al.] // PeerJ Computer Science. — 2016.
5. Anderson, J. M. How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes / J. M. Anderson, P. Heaton // Yale Law Journal. - 2012. - Vol. 122. - P. 154–217.
6. Corbett-Davies, S. The Measure and Mismeasure of Fairness: A Critical Review of Fair Machine Learning / S. Corbett-Davies, S. Goel // Annual Review of Statistics and Its Application. — 2018.
7. Human rights lawyers at risk / Human Rights House Foundation. — 2021.