

Основные принципы:

Рекомендательный алгоритм по подбору адвокатов не имеет прямых аналогов в мире. Наша методика сходна с новейшими способами измерения качества работы врачей [1] и оценкой результативности спортсменов [2].

Основной принцип следующий – мы используем продвинутое машинное обучение и ИИ для того, чтобы на основе доступных для измерения обстоятельств уголовного дела предсказать его исход. Наши алгоритмы позволяют предсказать бинарные исходы (заккрытие или продолжение дела, условных срок или реальный и пр.) с высокой точностью.

При этом алгоритмы не учитывают те обстоятельства, на которые может прямо повлиять адвокат. Мы предполагаем, что значительная часть необъясненных алгоритмом результатов приходится именно на качество работы адвоката.

Далее мы сравниваем прогнозы с реальными исходами. Принцип оценки прост – *если адвокат чаще своих коллег добивается для доверителей результатов лучше, чем прогнозировала модель, то он работает более эффективно*. Рекомендательная система будет предлагать именно таких адвокатов, причем применительно к вашей категории дела.

Источник наших данных – тексты судебных решений, которые размещаются в открытом доступе на сайтах судов общей юрисдикции согласно №262-ФЗ от 2008 г. Суммарно мы анализируем более 5 млн. судебных решений за период с 2010 по 2024 гг. ФИО адвокатов берутся из Единого государственного реестра адвокатов, которые размещаются в открытом доступе на официальном сайте Министерства Юстиции РФ: <https://minjust.gov.ru/ru/pages/advokaty-rossijskoj-federacii/>

[1] Hall B.L., Huffman K.M., Hamilton B.H., Paruch J.L., Zhou L., Richards K.E., Cohen M.E., Ko C.Y. Profiling Individual Surgeon Performance Using Information from a High-Quality Clinical Registry: Opportunities and Limitations // Journal of the American College of Surgeons. – 2015. – Vol. 221. – № 5. – P. 901-913. DOI: 10.1016/j.jamcollsurg.2015.07.454

[2] Grasseti L., Bellio R., Di Gaspero L., Fonseca G., Vidoni P. An extended regularized adjusted plus-minus analysis for lineup management in basketball using play-by-play data // IMA Journal of Management Mathematics. – 2021. – Vol. 32. – № 4. – P. 385-409. DOI: 10.1093/imaman/dpaa022.

Учет различных факторов в рейтинге:

-Прогноз исходов судебных решений учитывает такие обстоятельства как: состав преступления (статья УК РФ, часть статьи), рецидивизм, наличие смягчающих и отягчающих обстоятельств, суд, регион, (включая социально-демографический состав населения, образование), год и месяц принятия решения. Отдельно мы учитываем при прогнозе исходов дел социально-демографические характеристики населенного пункта, что позволяет более точно соблюдать принцип сравнения “при прочих равных”.

- К факторам, на которые может повлиять адвокат, относятся сделка с правосудием («особый порядок»), избранная мера пресечения, а также исходы судебных решений (от закрытия дела до назначения реального срока заключения).

- Наш алгоритм также учитывает множество нюансов:

а) *количество адвокатов в деле*: если результат достигнут в команде, то каждый адвокат получает несколько меньшую прибавку к личному рейтингу, чем если бы он работал один;

б) *последовательность этапов рассмотрения дела*: если алгоритм предсказывал условный срок, а дело закончилось реальным, то адвокат не получит большого бонуса за то, что добился 3 лет лишения свободы, а не 4-ех.

с) *количество дел, которые ведет адвокат* – если оно небольшое, то алгоритм показывает более скромную оценку адвоката (рейтинг приближается к 5-6) поскольку небольшое количество дел не позволяет оценить вклад адвоката достаточно точно.

д) *сделки с правосудием* – дела, которые идут в «особом порядке» представляют большую часть

Ограничения данных:

Наша рекомендательная система имеет ряд ограничений, которые важно принимать во внимание. Ниже описаны ограничения в порядке их значимости для пользователей нашей рекомендательной системы.

А) Мы оцениваем только судебную стадию работы. Важной частью работы адвоката является досудебная стадия, более того, многие хорошие адвокаты работают только на стадии до суда. Если дело не поступило в суд – то это хорошо для доверителя, но это также означает, что об этом деле нет почти никакой публичной информации, потому мы не можем применить к ним наши алгоритмы.

Б) Неполнота данных. Некоторые судебные решения по тем или иным причинам не размещаются в открытом доступе или же публикуются с задержкой доходящей до 1-2 лет. К тому же в некоторых текстах судебных решений изымаются ФИО участников процессов, включая данные адвоката. Это означает, что профиль адвоката может быть неполным. Система поиска может недоучитывать самые новые дела, которые выиграл или проиграл адвокат.

Г) Дела с двумя и более подзащитными. На данный момент мы не учитываем в рейтинге дела, в которых одновременно представлены несколько подзащитных и/или одному подзащитному предъявлено несколько обвинений. Это связано исключительно с техническими сложностями при автоматизированном извлечении из текстов приговоров информации отдельно для каждого из обвиняемых (например, кто-то мог признать вину, а кто-то нет, набор смягчающих и отягчающих факторов может различаться и т.д.). Мы работаем над решением данной задачи.

Д) Небольшое количество адвокатов остаются неидентифицированными, поскольку имеют полных тезок по фамилии, имени и отчеству. К сожалению, в некоторых случаях мы не можем отличить дела, которые вели разные адвокаты, потому не включаем эту информацию в нашу рекомендательную систему. В дополнение к этому в реестре адвокатов иногда встречаются ошибки, которые также могут мешать корректно связывать дела с адвокатами.

Выводы: таким образом, рекомендательная система предоставляет важную дополнительную информацию при поиске адвоката, но не может и не должна использоваться в качестве единственного критерия при выборе защитника.

Система может подсказать, что какой-то из защитников является наиболее эффективным в определенном типе дел. Однако она не может и не должна использоваться как свидетельство, что адвокат работает недостаточно эффективно – не исключено, что адвокат не попадает в список рекомендаций именно потому, что нам доступна неполная информация.

Этапы прогнозирования судебного решения

В рамках разработки методологии оценки эффективности адвокатов предусмотрено последовательное моделирование ключевых характеристик итоговых судебных решений по уголовным делам. Прогнозирование осуществляется поэтапно, что обусловлено логикой движения дела и наличием нескольких важных развилок, каждая из которых

формирует самостоятельный объект анализа. Такая поэтапная структура позволяет учитывать специфические факторы, влияющие на вероятность наступления того или иного исхода, и обеспечивает методологическую корректность сопоставления прогнозируемых и фактических результатов.

На **первом этапе** проводится оценка вероятности выбора судом формы итогового решения: постановления или приговора. Эти два исхода кардинально различаются между собой, так как в случае завершения рассмотрения дела постановлением подсудимый не получает судимость. Решения оформляются постановлением в случае прекращения дела, возвращения дела прокурору, примирения сторон или применения судебного штрафа. Приговор, в свою очередь, также может привести к отсутствию судимости в случае оправдания подсудимого, но это намного более редкий случай. Таким образом, постановление суда по итогам рассмотрения дела считается намного более благоприятным исходом, нежели чем провозглашение приговора, следовательно способность адвоката систематически добиваться такого исхода будет свидетельствовать о его высоком профессионализме и эффективности. Модели, применяемые на данном этапе, обучаются на полной совокупности уголовных дел, что позволяет обеспечить всестороннее отражение факторов, влияющих на исход первой процессуальной развилки.

На **втором этапе** прогнозируются параметры исхода внутри категории дел, завершившихся приговором. На данном уровне анализируется вероятность назначения наказания в виде реального лишения свободы по сравнению со всеми альтернативными исходами, возможными при вынесении приговора. Указанный этап представляет собой ключевой момент оценки тяжести последствий для подсудимого. Модели обучаются на подвыборке дел, по которым фактически был вынесен приговор, что обеспечивает релевантность используемой информации и исключает влияние структурных особенностей иных процессуальных решений.

На **третьем этапе** проводится прогнозирование продолжительности реального лишения свободы для тех дел, по которым такая мера наказания была назначена. Основная цель указанного этапа заключается в выявлении возможностей влияния адвоката на снижение тяжести правовых последствий для подсудимого в тех случаях, когда назначение тюремного наказания является неизбежным. Модели обучаются исключительно на подвыборке дел, содержащих сведения о фактически назначенном реальном сроке, что обеспечивает корректность оценки и учет специфики данной категории наблюдений.

Совокупное применение трех указанных этапов прогнозирования обеспечивает построение детализированной структуры ожидаемых исходов, позволяющей сопоставлять фактические результаты судебного разбирательства с предсказанными значениями на каждом этапе рассмотрения дела в суде. Такой подход дает возможность формировать комплексную и количественно обоснованную оценку профессиональной деятельности адвокатов, основанную на отклонениях фактических исходов от статистически ожидаемых значений.

Применяемые методы машинного обучения

В рамках реализации методологии прогнозирования судебных решений были использованы современные методы машинного обучения, обеспечивающие получение статистически обоснованных и воспроизводимых предсказаний по различным категориям уголовных дел. Основными моделями, примененными в процессе работы, стали ансамблевые методы **Random Forest** и **Extreme Gradient Boosting**, а также обобщенные

линейные модели (GLM). Ансамблевые методы были выбраны в качестве приоритетных инструментов в связи с их способностью формировать устойчивые прогнозы за счет агрегирования большого числа деревьев решений, функционирующих как в независимом, так и во взаимосвязанном режиме. Random Forest применяется для построения совокупности деревьев на основе случайных подвыборок предикторов и наблюдений, что позволяет минимизировать риск переобучения и обеспечить высокую стабильность. Extreme Gradient Boosting основан на процедуре последовательного улучшения качества решений путем минимизации функции потерь на каждом шаге и демонстрирует высокую точность при наличии сложных взаимодействий между переменными.

Обобщенные линейные модели рассматривались в качестве базового метода для сопоставления результатов и формирования контрольного набора предсказаний. Однако в целях оптимизации вычислительных ресурсов обучение данного типа моделей было исключено из итоговой процедуры, поскольку предварительные испытания показали, что показатели точности GLM систематически уступали ансамблевым методам на большинстве подвыборок данных.

Для всех моделей применялась система критериев оценки качества предсказаний. В задачах бинарной классификации, характерных для первого и второго этапов прогнозирования, использовался показатель AUC, позволяющий оценить способность модели различать неблагоприятные и благоприятные исходы. В задачах количественного прогнозирования, относящихся к третьему этапу, использовался показатель RMSE, направленный на измерение среднеквадратичного отклонения предсказаний от фактических значений. Оценка качества проводилась посредством процедуры кросс-валидации с разбиением исходных данных на пять блоков, что обеспечило объективность и статистическую надежность полученных результатов.

По итогам обучения для каждой группы наблюдений на всех этапах выбиралась модель, продемонстрировавшая наилучшие показатели качества. Такой подход позволяет обеспечить применение наиболее релевантных и оптимальных алгоритмов в каждом конкретном контексте и является необходимым условием формирования корректной и воспроизводимой системы прогнозирования судебных решений.

Методика расчета рейтинга эффективности адвокатов

Для получения оценки эффективности адвоката на каждом из трех этапов рассчитывается разница между предсказанным и наблюдаемым исходами. Делается это в несколько шагов. Так, на первом шаге рассчитывается «сырая» дельта по следующим формулам:

$$\delta_{k,s} = \gamma_{k,s} - \rho_{k,s} \quad (1a)$$

$$\delta_{k,s} = \hat{x}_{k,s} - x_{k,s} \quad (1b)$$

Формула (1a) используется при предсказании бинарных исходов, в ней $\gamma_{k,s} \in \{0, 1\}$ – фактический результат; $\rho_{k,s}$ – предсказание наилучшей моделью вероятности исхода. В свою очередь формул (1b) используется для прогнозирования значения непрерывной переменной (количества месяцев реального заключения), в ней $\hat{x}_{k,s}$ – наилучшее предсказание модели; $x_{k,s}$ –

фактическое значение предсказываемой переменной. В обеих формулах k – дело, s – стадия.

Если на стадии $s - 1$ была предсказана вероятность успеха $\rho_{k,s-1}$, но фактически успех не был достигнут, тогда каждому адвокату в деле s на стадии k назначается штраф $\varphi_{k,s-1}$ по следующей формуле:

$$\varphi_{k,s-1} = \begin{cases} 0, & \text{if } \gamma_{k,s-1} = 1 \text{ or } \rho_{k,s-1} \leq \rho \\ \lambda \cdot \frac{1}{1-\rho}(\rho_{k,s-1} - \rho), & \text{else} \end{cases} \quad (2)$$

где λ – максимально возможный штраф (в текущем варианте реализации алгоритма $\lambda = 0.2$), а ρ – порог вероятности хорошего исхода, при котором мы даем штраф (в текущем варианте $\rho = 0.6$).

Затем мы применяем штрафы со всех предыдущих стадий в соответствии со следующей формулой:

$$\delta_{k,s}^{\cdot} = \delta_{k,s} - |\delta_{k,s}| \cdot \sum_{t < s} \varphi_{k,t} \quad (3)$$

где $\delta_{k,s}^{\cdot}$ – итоговое значение дельты после применения штрафов.

На следующем шаге мы штрафует адвокатов за командную работу в соответствии с формулой (4):

$$\dot{\partial}_{k,s} = \delta_{k,s}^{\cdot} - |\delta_{k,s}^{\cdot}| \cdot \min\{\mu, n_k - 1\} \cdot \vartheta \quad (4)$$

где ϑ – размер наказания за одного дополнительного адвоката, участвующего в деле (в текущем варианте $\vartheta = 0.1$); μ – максимальное количество дополнительных адвокатов, за которых накладывается штраф (в текущем варианте $\mu = 5$), n_k – общее количество адвокатов в деле k . В итоге $\dot{\partial}_{k,s}$ – итоговая дельта для адвоката, который работал над делом k на стадии s в команде из n_k адвокатов.

Теперь пусть $\Omega_{i,s}$ – множество дел, в которых адвокат i участвовал на стадии s , тогда:

$$\bar{\delta}_{i,s} = \frac{1}{N_{i,s}} \sum_{k \in \Omega_{i,s}} \dot{\partial}_{k,s} \quad (5)$$

где $N_{i,s} = |\Omega_{i,s}|$. Таким образом $\bar{\delta}_{i,s}$ – средняя дельта адвоката i на стадии s . Однако если у адвоката по какой-то причине мало дел на стадии s , то полученное значение может оказаться нестабильным в силу небольшого количества наблюдений. В связи с этим мы рассчитываем сглаженную оценку для адвокатов, у которых $N_{i,s} < \alpha_s$, где α_s – минимальное количество дел, достаточное для получения устойчивой $\bar{\delta}_{i,s}$ (в текущем

варианте $\forall s \alpha_s = 10$). Итоговое значение дельты рассчитывается по следующей формуле:

$$\bar{\Delta}_{i,s} = \begin{cases} \bar{\delta}_{i,s}, & \text{if } N_{i,s} \geq \alpha_s \\ \frac{N_{i,s}}{\alpha_s} \cdot \bar{\delta}_{i,s} + \frac{\alpha_s - N_{i,s}}{\alpha_s} \cdot \hat{\Delta}_s, & \text{else} \end{cases} \quad (6)$$

где $\bar{\Delta}_{i,s}$ – итоговая средняя дельта адвоката i на стадии s , участвующая в дальнейших расчетах; $\hat{\Delta}_s$ – среднее значение дельты на стадии s среди всех адвокатов.

В результате мы получаем некоторое распределение множества $D_s = \{\bar{\Delta}_{i,s}\}$. Пусть $m_s = |D_s|$, теперь для каждого адвоката i мы получаем $\tau_{i,s}$ – значение эмпирической функции распределения (ECDF). Для этого мы пользуемся следующей формулой:

$$\tau_{i,s} = \frac{1}{m_s} \sum_j I(\bar{\Delta}_{j,s} \leq \bar{\Delta}_{i,s}) \quad (7)$$

где $I(\cdot)$ – индикаторная функция, принимающая значение 1, когда условие в скобках выполняется, и 0 в ином случае. Получив $\tau_{i,s}$ для каждой стадии s мы теперь можем получить $\bar{\tau}_i$ – среднее значение эмпирической функции распределения адвоката i по всем стадиям, взвешенное на количество кейсов в каждой стадии. Пусть $p_i = \{s | N_{i,s} > 0\}$ – множество стадий, в которых адвокат i участвовал хотя бы единожды. Тогда получим $\bar{\tau}_i$ по следующей формуле:

$$\bar{\tau}_i = \frac{1}{\sum_{p_i} N_{i,p_i}} \sum_{p_i} N_{i,p_i} \cdot \tau_{i,s} \quad (8)$$

Вновь получаем некоторое распределение множества $D = \{\bar{\tau}_i\}$, немного изменив формулу (7) получим, что:

$$r_i = \frac{1}{|D|} \sum_j I(\bar{\tau}_j \leq \bar{\tau}_i) \quad (9)$$

где r_i – значение ECDF для i -го адвоката. Используя его, мы теперь можем получить дециль i -го адвоката воспользовавшись следующей формулой:

$$R_i = q \text{ если } \frac{q-1}{10} < r_i \leq \frac{q}{10}, q = 1, \dots, 10 \quad (10)$$

R_i – это то число, которое видит пользователь сайта проекта в качестве общей оценки эффективности работы адвоката i в уголовном судопроизводстве по некоторой статье.