М.В. ХАВРОНЕНКО*

ПРОГНОЗИРОВАНИЕ РЕЗУЛЬТАТОВ РАССМОТРЕНИЯ ЗАКОНОПРОЕКТОВ ГОСУДАРСТВЕННОЙ ДУМОЙ РФ: МОДЕЛЬ НЕЙРОННОЙ СЕТИ[1]

Аннотация. В данной статье на основе собранного массива данных с сайта Государственной думы РФ за период с 24 октября 1994 г. по 1 декабря 2022 г. настроены модели машинного обучения и нейронная сеть для прогнозирования итогов рассмотрения законопроектов нижней палатой парламента. Для предварительной обработки данных использовалась модель *rubert-tiny*, для прогнозирования – классификатор случайного леса, логистическая регрессия и модель нейронной сети из трех линейных слоев.

Модели продемонстрировали следующие результаты: 94% точности (метрика *F1* взвешенная) при прогнозировании на основе текстов прилагаемых к законопроекту документов и 87% точности при обучении на параметрах паспорта законопроекта. Обученные только на текстах законопроекта модели демонстрировали точность в 75,6%. Наиболее важным фактором, оказывающим влияние на результат прогноза, оказался текст заключения. Вторым по важности признаком стал «Субъект права законодательной инициативы» с 31,5% значимости в прогнозировании.

На основе объединенных текстовых данных и параметров паспорта законопроекта лучше всего проявил себя алгоритм случайного леса. Среди обученных только на текстовых параметрах алгоритмов на первое место вышла логистическая регрессия. На вероятность принятия законопроекта не оказали существенного влияния текст финансового обоснования, текст пояснительной записки или

* **Хавроненко Максим Викторович,** аспирант факультета политологии, МГУ им. М.В. Ломоносова (Москва, Россия), e-mail: mxavronenko@mail.ru

[1] Статья подготовлена при поддержке грантовой программы Российского общества политологов.

 DOI: 10.31249/poln/2024.03.09

тематика законопроекта. Автором сделаны выводы о направлениях практического применения обученных моделей, а также определены дальнейшие научные проблемы в сфере математического анализа и прогнозирования законотворчества.

Ключевые слова: поименное голосование; Государственная дума; прогнозирование законотворчества; законодательные исследования; нейронные сети; ru-BERT; машинное обучение.

Для цитирования: Хавроненко М.В. Прогнозирование результатов рассмотрения законопроектов Государственной думой РФ: модель нейронной сети // Политическая наука. – 2024. – № 3. – С. 211–240. – DOI: http://www.doi.org/10.31249/poln/2024.03.09

Введение

Государственная дума, состоящая из 450 депутатов и являющаяся наравне с Советом Федерации представительным и законодательным органом власти в Российской Федерации, каждый день принимает решения, от которых зависят жизни миллионов граждан. По данным «Официального интернет-портала правовой информации»[1], на апрель 2024 г. в Российской Федерации действует около 10,7 тыс. федеральных законов, во многие из которых ежегодно вносятся поправки. Несмотря на стремление законодательных органов сделать законотворчество более прозрачным, реальный процесс принятия законопроектов депутатами по-прежнему остается закрытым не только для обычных граждан, но и для многих специалистов.

Какие факторы оказывают влияние на процедуру принятия законопроекта и итоговое решение законодателя в России? Возможно ли прогнозировать вероятность принятия закона, и какие факторы могут являться определяющими?

Ответы на поставленные выше вопросы позволят не только исследовать процесс принятия решений и основные мотивации в голосовании законодателей, но также дадут возможность проследить изменения предпочтений политической элиты, влияющие на развитие страны.

Подобная научная и практическая проблема делает актуальной разработку математических моделей на основе архивных данных для прогнозирования будущих итогов рассмотрения законо-

[1] Официальный интернет-портал правовой информации // Результаты поиска по запросу: действующие Федеральные Законы. – Режим доступа: http://pravo.gov.ru/ips/ (дата посещения: 07.04.2024).

проектов. Построенные модели не только станут прикладным инструментом анализа работы отдельных депутатов, фракций и Государственной думы в целом, но и позволят количественно проследить закономерности и изменения российского парламентаризма, а также эффективность работы отдельных комитетов, что невозможно было бы достичь иными способами.

Обзор литературы

Направление прогнозирования результатов голосования парламента зародилось в мире совсем недавно. Только в конце 1960-х годов в США начали появляться работы, пытающиеся статистически описывать и прогнозировать деятельность законодательных органов власти. Работы таких авторов, как [Cherryholmes, Shapiro, 1969], а также [Matthews, Stimson, 1975] были первыми статьями, анализирующими с помощью методов симуляций и интервью результаты голосований депутатов в парламенте (или *roll-call* данные), определяя основные факторы (как личные, так и институциональные), оказывающие влияние на голосование американских парламентариев по законопроектам. Будущий комплекс теорий и направление по анализу голосований парламентариев получит название «анализ поименного голосования» (*roll call analysis*).

В скором времени появился более эффективный метод, математически описывающий процесс решения депутата – им стал основанный на концепции случайной полезности [McFadden, 1981] метод идеальной точки (*Ideal point model*). *Метод идеальной точки* утверждал, что у каждого парламентария имеется четкая политическая позиция, которую можно представить в виде точки в пространстве [Poole, Rosenthal, 1985] (см. также: [Clinton et al., 2004]. Эта позиция может отражать как предпочтения избирателей и самого депутата, так и другие факторы, стабильно оказывающие влияние на его позицию. В то время как одни исследователи считали, что эти точки отражают публично заявляемые позиции законодателей [Ansolabehere et al., 2001], другие ученые предполагали, что идеальные точки характеризуют истинные предпочтения депутатов [Chiou, Rothenberg, 2003].

В процессе рассмотрения законопроекта каждый парламентарий решает, поддерживать ли ему законопроект, исходя из того, ка-

кое решение ему «ближе» – т.е. в мысленном пространстве сравнивая расстояние от точки, обозначающей его политическую позицию до точек в пространстве, обозначающих голоса за и против. Расстояния в таком случае представляются в виде детерминанты при наличии случайной переменной, обозначающей ошибку.

Эта модель столкнулась с критикой ряда исследователей. Так, она не только не учитывала внешние факторы, способные существенно изменить позицию законодателя с течением времени [Clinton et al., 2004], но также игнорировала возможность единоразового лоббизма [Snyder, Groseclose, 2000], воздержания законодателей от голосования по законопроекту и нехватки данных. Описанные недостатки привели к последующему дополнению модели и разработке новых методов анализа политических предпочтений законодателей.

Дальнейшим развитием теории стали (1) *метод пространственного голосования* (Spatial Voting Model) [Ladha, 1994] (см. также: [Clinton et al., 2004]), представляющий каждого законодателя в виде двух точек в пространстве: одну точку для его голосов за, а другую – для голосов против и вычисляющий на координатном поле дистанции до законопроекта; (2) *тематически настроенный метод идеальной точки* (*Issue-adjusted model Ideal point model*) [Gerrish, Blei, 2012], где для каждого законодателя линейным дискриминантным анализом уточняется его позиция в пространстве, учитывая уже анализ текстов законопроектов; и (3) *разреженный факторный анализ* (sparse factor analysis)[Kim et al., 2018], определяющий позицию парламентариев и законопроектов в пространстве, исходя из текстов документа и голосов соответствующего партийного руководства.

Последующее развитие метода идеальной точки строилось на формировании многомерного пространства признаков и введении дополнительных параметров в определении позиции депутата (*multidimensional ideal vectors model*) [Kraft et al., 2016], а также внедрения Байесовских методов при оценке позиции депутата по законопроекту (*Bayesian ideal point model*) [Gerrish, Blei, 2011].

С развитием технологий, а также увеличением количества доступной информации о заседаниях парламентов, исследователи стали использовать методы обработки больших данных и машинного обучения для прогнозирования результатов голосования депутатов. В качестве данных чаще всего использовали 106–111 сессии

конгресса США, которые включали в себя 4447 законопроектов, 1269 законодателей и 1 837 033 голоса за или против. Основными методами, использованными для прогнозирования результатов голосования депутатов, стали:

- метод байесовской симуляции (*Bayesian simulation*) для уточнения положения законопроектов в многомерном пространстве законодателей метода идеальной точки [Jackman, 2001] (см. также: [Gerrish, Blei, 2011; Wang et al., 2013]) и анализа изменяющихся бинарных матриц и текстов законопроектов [Wang et al., 2010];
- метод опорных векторов (*Support Vector Machines*) для уменьшения размерности параметров документов с последующим прогнозированием на основе алгоритма обратного распространения ошибки (*back propagation learning algorithm*) [Khashman, Khashman, 2016] и алгоритма случайного леса (*Random forest sentiment analysis*) итогов голосования депутатов [Budhwar et al., 2018];
- сверточные нейронные сети (Convolutional Neural Networks) [Korn, Newman, 2020] и рекурентные нейронные сети (*recurrent neural networks*) [Yang et al., 2021].

После 2012 г. исследователи решили обратить внимание не только на описание и прогнозирование голосов парламентариев, но и на общую вероятность принятия законопроекта. Этому способствовало то, что ранее использованные параметры для обучения моделей могли быть использованы повторно в процессе решения смежной задачи. На данный момент существует три такие научные работы. Поскольку наша работа непосредственно относится к этому направлению исследований, разберем каждую из них.

«Текстовые предикторы выживания законопроекта в комитетах Конгресса» [Yano et al., 2012]

Целью работы является определение основных факторов, способствующих «выживанию» законопроекта в профилированном комитете Конгресса США. В данной работе авторами рассматриваются почти 52 тыс. законопроектов, из которых 6828 получили одобрение комитетов и были рассмотрены законодателями.

Логистическая регрессия была использована для классификации законопроектов на основе таких параметров как: характеристики закона, название комитета, субъект законодательной инициати-

вы, количество членов в комитете, их партийная принадлежность и текст законопроекта.

Авторами разработана модель, прогнозирующая результаты рассмотрения комитетами законопроектов с 83% точности на Десятикратном перекрестно-проверенном эксперименте (*ten-fold cross-validated experiment*). Наиболее важными признаками в прогнозировании являлись: членство субъекта законодательной инициативы в партии большинства (0,525), членство субъекта в партии большинства и комитете одновременно (0,233) и партийная принадлежность субъекта законодательной инициативы к демократам (0,135). Данная работа является единственным исследованием, фокусирующим внимание на прохождении законопроекта в комитете.

«Прогнозирование и анализ законотворчества с помощью векторных представлений слов и ансамблевой модели» [Nay, 2017]

Целью работы является разработка модели прогнозирования результата принятия законопроектов Конгрессом США на основе 70 тыс. законопроектов, внесенных за период 2001–2015 гг., среди которых только 2513 были приняты.

В качестве данных использовались параметры и преобразованные в векторное представление с помощью модели нейронной сети тексты законопроектов. Из всех версий выдвигаемых законопроектов автором использовалась только первая. Параметрами законопроектов являлись: регион субъекта законодательной инициативы, партийная принадлежность автора, время нахождения законодателя в парламенте и другие характеристики.

Для прогнозирования итогов принятия законопроектов автором использовались модель случайного леса и ансамбли других моделей машинного обучения. Качество моделей оценивалось с помощью логарифмической оценки (*log score*), Бриеровской оценки (*brier score*) и метрики площади под кривой ошибок *Roc Auc*, позволяющих работать с вероятностными прогнозами.

Выводом этой статьи стало то, что модель, обученная на данных последних конгрессов США и использующая только текст законопроекта, превосходит модель, использующую только параметры законопроекта. В то же время модель, обученная на более

старых заседаниях и использующая только контекст законопроекта, превосходит модель, использующую только тексты законопроектов. Но во всех случаях использование текста как дополнительного параметра добавляет модели прогностическую силу. Текст законопроекта и доля членов палаты от партии большинства среди субъектов законодательной инициативы оказали наибольшее влияние на прогностический потенциал модели. При этом полный текст законопроекта является самым важным предиктором в модели.

«Использование искусственного интеллекта для прогнозирования голосования в законодательных органах Конгресса США» [Bari et al., 2021]

Целью работы также являлось проверка прогностического потенциала моделей машинного обучения на данных о результатах рассмотрения законопроектов в Сенате США. Для этого были использована информация за 113–115 конгрессы США (2013–2019 гг.), включавшие в себя 32 тыс. законопроектов, из которых 1031 документ прошел обе палаты Конгресса, в то время как 31578 не были приняты ни одной из палат. Основными параметрами законопроектов являлись номер законопроекта, тип, название, субъект законодательной инициативы, дата внесения в парламент, комитет, короткая справка о законопроекте и другие.

L2-логистическая регрессия, линейный метод опорных векторов, лес решений, многослойный перцептрон, метод К-ближайших соседей и ансамблевые методы были использованы для прогнозирования итогов принятия законопроектов.

Наиболее эффективные результаты показали ансамблевые методы (80% доля правильных ответов (*accuracy*), 94% точность (*precision*) и полнота (*recall*)), многослойный перцептрон (79% доля правильных ответов, 94% точность и полнота) и дерево решений (75% доля правильных ответов, 92% точность и полнота). Определяющими признаками стали субъект законодательной инициативы, время года, в которое законопроект был предложен, и извлеченные из краткого описания текста законопроекта признаки.

Подводя итог: среди всех работ рассматриваемого направления, наилучшее значение качества продемонстрировано в модели [Yano et al., 2012] – 83%. Их собранная база данных насчитывала

51,762 законопроекта. Основным выводом статей, использовавших описанную выше методологию, стало то, что помимо общих параметров законопроектов необходимо также использовать их текст, поскольку он позволяет улучшить качество предлагаемых моделей и обнаруживает закономерности, недоступные при поверхностном рассмотрении условий внесения законопроектов. Более подробно литература обоих направлений исследований представлена в таблицах 1 и 2 в приложении.

Как законопроекты становятся законами в Российской Федерации

Процесс прохождения всех необходимых стадий рассмотрения законопроекта и его принятие в Российской Федерации регулируется Конституцией Российской Федерации, регламентами обеих палат Федерального собрания, Федеральными законами «О статусе сенатора Российской Федерации и статусе депутата Государственной Думы Федерального Собрания Российской Федерации» и «О парламентском контроле», а также другими нормативными актами[1].

Согласно Конституции РФ, правом законодательной инициативы обладают следующие субъекты: Президент РФ, Совет Федерации РФ, члены Совета Федерации РФ, депутаты Государственной думы РФ, Правительство РФ, законодательные (представительные) органы субъектов РФ, а также Конституционный и Верховный суды РФ по вопросам их ведения.

Законопроект рассматривается в трех чтениях в Государственной Думе РФ, затем направляется на рассмотрение в Совет Федерации, после этого Президенту РФ и в случае поддержки всеми законодательными органами и президентом вступает в силу. Рассмотрим процесс прохождения подробнее, остановившись на деталях, которые могут быть полезны при построении модели.

При первоначальном внесении в Государственную думу законопроекта субъект законодательной инициативы также должен предста-

[1] Официальный интернет-портал Совета Федерации // Страница энциклопедического справочника: Законодательный процесс. – Режим доступа: http://council.gov.ru/services/reference/9373/ (дата посещения: 31.03.24).

вить (1) текст законопроекта, (2) пояснительную записку с изложенной концепцией предлагаемого законопроекта и обоснованием необходимости его принятия, (3) финансово-экономическое обоснование и, при необходимости, согласно 104 статье (части 3) Конституции РФ, отзыв правительства, список актов федерального законодательства, подлежащих признанию утратившими силу, приостановлению, изменению или принятию в связи с принятием федерального закона, а также другие документы[1]. Помимо этого, свои заключения на законопроект дают комитеты-исполнители, Счетная палата, а также правовое управление Государственной думы. Мы обращаем на это внимание, поскольку тексты именно этих документов будут являться одними из параметров, используемыми для прогнозирования результатов принятия законопроектов в Государственной Думе.

Во время пленарного заседания депутаты заслушивают выступления и вопросы относительно вносимого законопроекта и принимают решения голосовать за, против или воздержаться от голосования. В отдельных случаях воздерживающаяся позиция депутата может препятствовать прохождению законопроекта в парламенте, поскольку для принятия законопроекта не будет набрано минимальное число голосов [Карягин, 2023].

После обсуждения законопроекта в первом чтении, кроме определенных случаев, устанавливается срок представления поправок к законопроекту. Поправки рассматриваются во втором чтении, и законопроект со всеми внесенными изменениями принимается или отклоняется. При этом третье чтение служит для коррекционных изменений правового и лингвистического характера, и внесение правок концептуального характера в законопроект не допускается.

Одобренные в трех чтениях законопроекты направляются в течение пяти дней на рассмотрение в Совет Федерации, где за 14-дневный срок должны быть рассмотрены. Подобная же процедура действует и в отношении Президента РФ перед финальным опубликованием закона. Президент также вправе обратиться в

[1] Постановление Государственной Думы Федерального Собрания Российской Федерации от 22 января 1998 года № 2134-II ГД // СПС КонсультантПлюс.

Конституционный суд РФ для проверки конституционности представленного закона[1].

Совет Федерации и Президент могут отклонить направленный им закон. В таком случае, кроме варианта снятия законопроекта, также предполагается создание согласительной комиссии между Государственной думой и Советом Федерации, которая, доработав спорные части законопроекта, повторно направляет законопроект согласно процедуре его принятия.

Понимание процедуры принятия законопроектов чрезвычайно важно, поскольку позволяет сформулировать следующие **гипотезы**, далее проверяемые с помощью математической модели.

Гипотеза 1: Логистическая регрессия и алгоритм случайного леса являются эффективными методами бинарной классификации законопроектов на принятые и не принятые.

Гипотеза 2: Субъект законодательной инициативы является одним из важнейших параметров, оказывающих влияние на принятие законопроекта.

Гипотеза 3: Текст законопроекта является одним из важнейших параметров, оказывающих влияние на принятие законопроекта.

Гипотеза 4: Текст пояснительной записки к законопроекту является одним из важнейших параметров, оказывающих влияние на принятие законопроекта.

Гипотеза 5: Текст финансового обоснования к законопроекту является одним из важнейших параметров, оказывающих влияние на принятие законопроекта.

Гипотеза 6: Текст заключений к законопроекту является одним из важнейших параметров, оказывающих влияние на принятие законопроекта.

Гипотеза 7: Модель рассмотрения законопроектов в различных созывах Государственной думы отличается, поэтому качество прогнозов вырастет при использовании данных только последних созывов.

Гипотеза 8: Использование параметров паспорта законопроекта совместно с текстовыми данными приложенных документов приведет к улучшению качества прогнозов в сравнении с опорой только на одно направление.

---

[1] Постановление Государственной Думы Федерального Собрания Российской Федерации от 22 января 1998 года № 2134-II ГД // СПС КонсультантПлюс.

Данные и их анализ

Получение данных

С помощью программы автоматического сбора данных на языке программирования *Python* с использованием библиотеки *Beautiful Soup* нами получены данные по всем законопроектам, вносимым в Государственную думу РФ и размещенных на интернет портале[1] за период 24.10.1994–01.12.2022. Общее количество законопроектов составило 27 176.

На странице каждого законопроекта нами была собрана информация о номере закона, его названии, комментарии к названию, паспортных данных законопроекта, в том числе о (1) субъекте права законодательной инициативы, (2) предмете ведения, (3) форме законопроекта, (4) ответственном комитете, (5) отрасли законодательства, (6) тематическом блоке законопроекта, (7) профильном комитете. Помимо этого, также была собрана информация о датах рассмотрения законопроекта в каждом чтении, а также документах, которые прикладываются к каждому рассматриваемому законопроекту (тексте законопроекта, финансового обоснования, пояснительной записки и всех заключений разных организаций).

Далее вся информация была занесена в единую базу для последующего анализа с помощью библиотек машинного обучения. База данных насчитывает 75 параметров, примерное количество единиц информации составило 1 млн 588 тыс.[2]

Статистический анализ

Проведем общий статистический анализ процесса принятия законопроектов Государственной думой РФ за период с 17 декабря 1995 г. до 1 декабря 2022 г.

---

[1] Система обеспечения законодательной деятельности. – Режим доступа: sozd.duma.gov.ru (дата посещения: 08.05.2024).

[2] Собранные данные и код размещены по ссылке: Duma_law_prediction_model // GitHub. – Mode of access: github.com/Onroof/Duma_law_prediction_model (accessed: 24.05.2024).

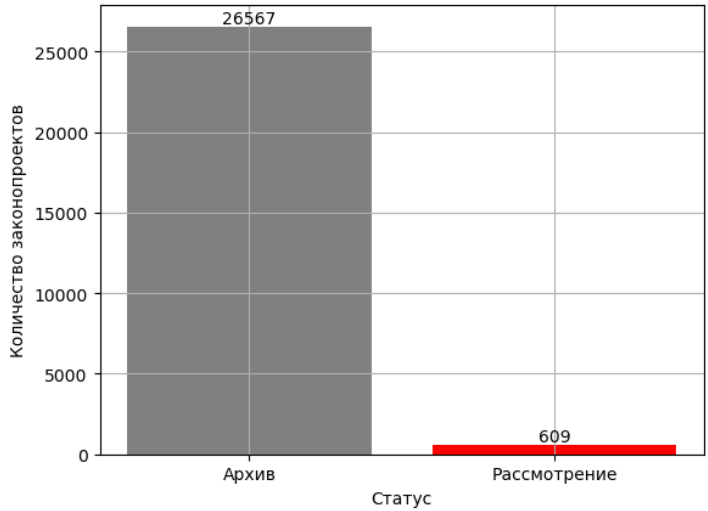

Рис. 1
Распределение законопроектов: архив и рассмотрение (все созывы)

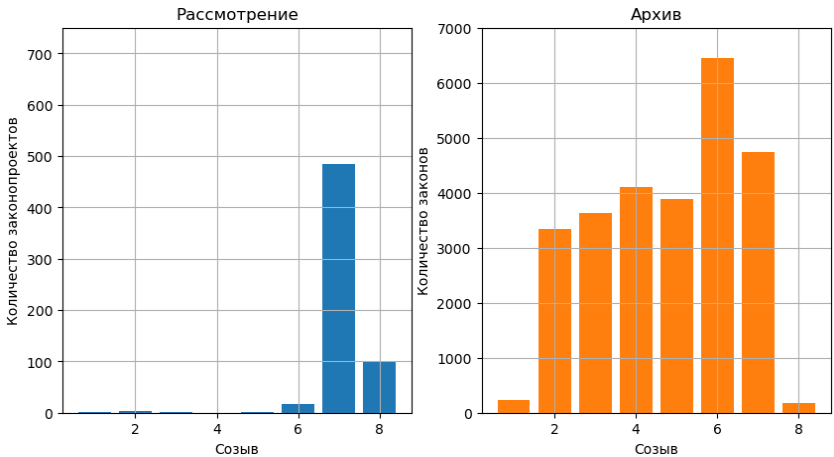

Рис. 2
Распределение рассматриваемых и архивных законопроектов (все созывы)

Таблица 1

Распределение рассматриваемых и архивных законопроектов (все созывы)

	1 созыв	2 созыв	3 созыв	4 созыв	5 созыв	6 созыв	7 созыв	8 созыв
Рассм.	0	2	3	1	2	17	484	100
Архив	239	3342	3628	4101	3877	6454	4745	181

Общее количество рассмотренных за все созывы законопроектов – 27 176 (см. рис. 1). При этом количество законопроектов, имеющих статус «в рассмотрении», насчитывает 609, где 484 законопроекта остались в работе с VII созыва Государственной думы РФ (см. табл. 3 и рис. 2).

Если не учитывать первый созыв, информации по которому недостаточно на сайте sozd.duma.gov, и последний созыв, который до сих пор идет, среднее количество законопроектов, рассматриваемых в каждом созыве, – 4357.

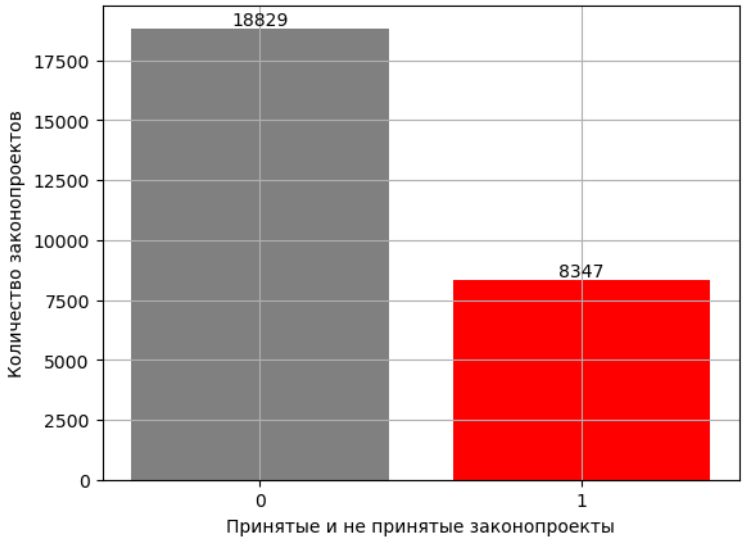

Рис. 3

Распределение законопроектов: принятые и не принятые (все созывы)

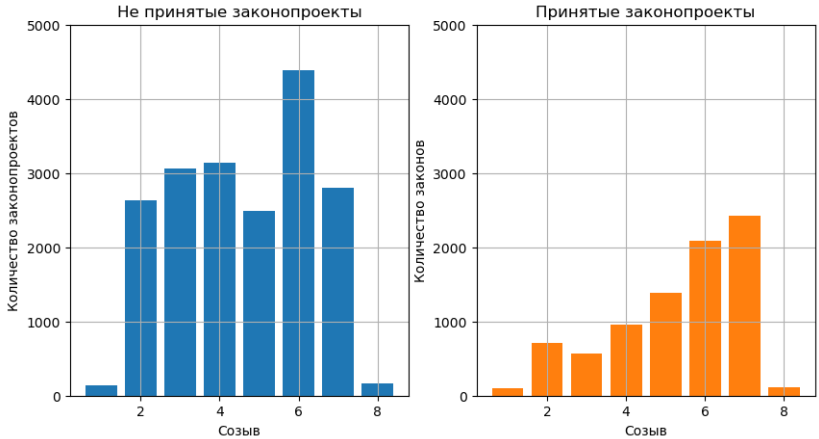

Рис. 4.
Распределение принятых и не принятых законопроектов: 8 созывов

Таблица 2
Распределение принятых и не принятых законопроектов: 8 созывов

	1 созыв	2 созыв	3 созыв	4 созыв	5 созыв	6 созыв	7 созыв	8 созыв
Не принят	140	2636	3059	3141	2495	4388	2806	164
Принят	101	709	570	960	1384	2083	2423	1117

В свою очередь, распределение по принятию законопроектов следующее: общее количество принятых законопроектов за восемь созывов насчитывает 8347, а не принятых законопроектов – 18 829 (см. рис 3). Средний же процент принятия законопроектов 31% (см. рис. 5 и табл. 2).

При этом заметен тренд на увеличение количества принятых законопроектов в Государственной думе за один созыв (см. рис. 4, прав.) Это может являться как следствием того, что все больше мест в парламенте занимает партия власти, что увеличивает вероятность принятия выдвигаемых ей законопроектов, так и повышением общей эффективности рассмотрения законопроектов в парламенте или стремлением разобрать накопившиеся завалы. При этом пик непринятых законопроектов в VI созыве связан с воз-

росшим числом выдвигаемых законопроектов в этот созыв (см. рис. 2). Вопрос дисконтинуитета как средства активизации рассмотрения законопроектов и блокирования политических решений рассмотрен в работе [Помигуев, Алексеев, 2021].

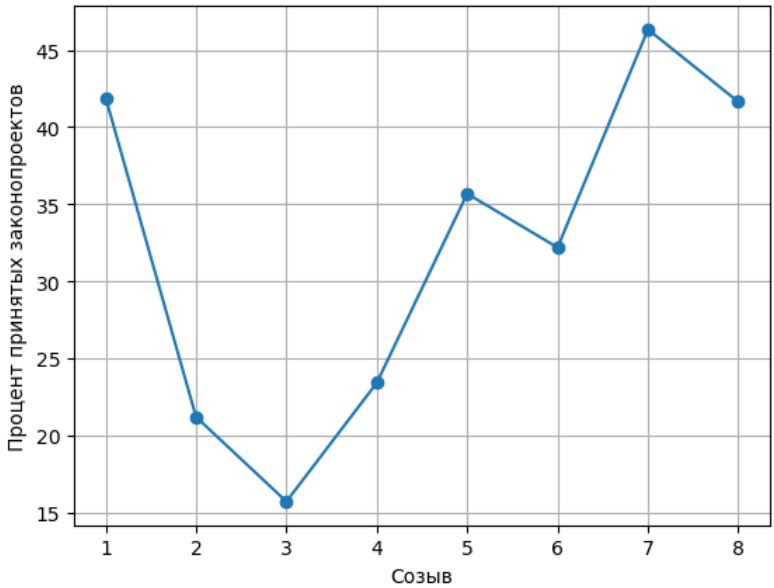

Рис. 4
Процент принятых в трех чтениях законопроектов по созывам

Таблица 3
Процент принятых в трех чтениях законопроектов по созывам

	2	**3**	**4**	**5**	**6**	**7**	**8**
Процент	21.19	15.7	23.4	35.68	32.19	46.34	41.64

Подготовка данных

Прежде чем передавать собранные данные модели машинного обучения, необходимо их предварительно обработать. Процесс

преобразования данных для нашей модели можно разделить на работу (1) с параметрами законопроектов и (2) с текстами вносимых документов.

Работа с параметрами законопроектов. Поскольку большинство моделей показывает лучшие результаты при работе с численными данными, мы трансформировали классы таких параметров как «Субъект права законодательной инициативы», «Предмет ведения», «Тематический блок законопроектов», «Отрасль законодательства» и других в численный вид с помощью функции *CountEncoder* библиотеки *scikit-learn Python* (Для категориального признака она заменяет названия групп на их количество). В то же самое время мы разделили все данные, связанные с датами прохождения каждого чтения, на год-месяц-день и выделили их в отдельные столбцы для того, чтобы иметь единый тип данных.

Работа с текстами вносимых документов. После скачивания всех прилагаемых к законопроектам документов и обновления *doc* форматов в *docx*, с помощью алгоритмов автоматического сбора данных были получены и занесены в базу тексты документов, формируя четыре столбца значений: «Текст законопроекта», «Пояснительная записка», «Финансовое обоснование» и «Текст заключений». Если в большинстве столбцов информация ячейки формировалась из текста одного прилагаемого документа (в случае дублирования нами была использована последняя версия документа), то ячейка «Текст заключения» формировалась путем сложения текстов всех заключений на законопроект. Помимо текстов документов нами также были введены параметры, отвечающие за длину текста в соответствующих ячейках.

Методология

Языковая модель RuBERT-tiny

Поскольку многие модели машинного обучения не могут использовать необработанные текстовые данные для прогнозирования и классификации, мы использовали модель *rubert-tiny* (на основе модели *BERT*), чтобы преобразовать тексты документов в одномерный массив, который, совместно с параметрами законо-

проекта, задали в качестве независимой переменной для определения результатов принятия законопроекта.

BERT (*Bidirectional Encoder Representations from Transformers*) – это нейронная сеть на основе трансформеров, которая была обучена на огромном корпусе текстовых данных для представления слов в контексте. В отличие от ранних моделей, *BERT* использует двунаправленную модель языка, что позволяет учесть контекст как слева, так и справа от рассматриваемого слова. Это делает BERT очень эффективным инструментом для широкого спектра задач, таких как создание вопросно-ответной системы, классификации текстов, извлечения именованных сущностей и многих других [Devlin et al., 2018].

Модель *BERT* использовалась нами для преобразования текстов прилагаемых к законопроекту документов в векторы. Она была выбрана из-за способности модели учитывать контекст и захватывать более сложные взаимосвязи между словами, а также из-за отсутствия необходимости предварительной обработки данных, такой как создание словарей или учет частеречной разметки.

Таким образом, с помощью предварительно обученной модели *ru-BERT* мы выполнили деление текстов прилагаемых к законопроектам документов на отдельные слова, и сгенерировали на их основе векторное пространство признаков с сохранением контекстуального смысла предложений. Общим итогом работы стало получение 1872 признаков.

Для последующего запуска нейронной сети и моделей машинного обучения мы преобразовали все полученные признаки (1872 из текстов документов и 30 из параметров законопроекта) в единую одномерную матрицу и провели разделение данных на обучающую и тестовую выборки в соотношении 75:25. Подготовив данные, мы стали тестировать их на разных алгоритмах.

Метрики качества

Для оценки качества прогнозирования и классификации моделей могут использоваться различные метрики в зависимости от характеристик используемых данных. Наша база данных являлась несбалансированной не только по количеству законов – в первом и последнем созывах их на порядок меньше, но и по факту принятия –

на 8347 принятых законопроектов приходится 18 829 не принятых. Поэтому в качестве метрики оценки качества моделей был использован **F1 макробалл (*F1 macro score*)**, который является балансом между полнотой (*recall*) и точностью (*precision*) и наиболее актуален для оценки качества моделей на несбалансированных данных. Напишем его формулу:

$$F1_{macro} = \Sigma \text{ (F1 scores)} / \text{количество классов.}$$

Поскольку *micro F1 score* придает одинаковое значение всем наблюдениям вне зависимости от их количества (что может повлиять на конечный результат), *macro F1* приравнивает все классы между собой, тем самым позволяя уменьшить влияние несбалансированности на результаты модели. В рамках эксперимента кроме *F1 macro score* нами также была применена метрика F1 взвешенная (*F1 weighted score*), которая также учитывает роль каждого класса в распределении данных (при этом не отдавая большого предпочтения только одному). F1 метрика принимает значения от 0 до 1, где значения близкие к 0 означают плохой прогностический потенциал модели, а значения, близкие к 1 – очень хороший.

Кроме метрики макро F1, мы также использовали такие метрики как сбалансированная доля правильных ответов (*balanced accuracy*), *roc auc* и метрика перекрестной проверки. Такие метрики как чувствительность (*sensitivity*) и специфичность (*specificity*) не использовались из-за несбалансированности данных, о чем мы написали выше.

Нейронная сеть

Помимо известных моделей машинного обучения, таких как логистическая регрессия и алгоритм случайного леса, нами также была использована модель нейронной сети, имеющей следующую структуру:

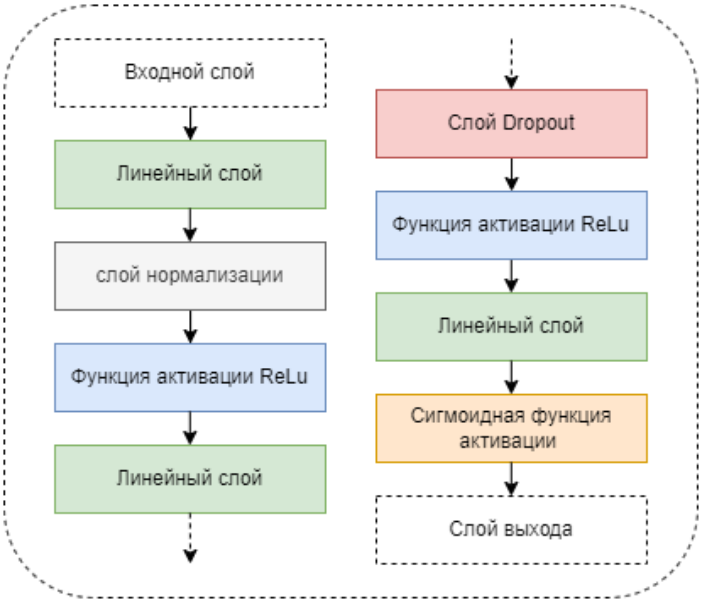

Рис. 6
Схема структуры многослойной нейронной сети

Линейные слои являются основным строительным блоком нашей нейронной сети. Они представляют собой матричные операции, в которых каждый нейрон в слое получает входные значения от предыдущего слоя, умножает их на свои веса и передает результат следующему слою. Линейные слои обычно состоят из нескольких нейронов и каждый нейрон имеет свои собственные веса, которые нужно подобрать во время обучения сети. Эти слои позволяют моделировать нелинейные зависимости в данных. Поскольку каждый нейрон в слое может вычислять линейную комбинацию входных значений, а затем передавать результат через нелинейную функцию активации, такую как *ReLU* или сигмоид, модель способна обучиться распознавать сложные нелинейные зависимости между входными данными и выходными метками. В нашей модели мы использовали всего три линейных слоя, чтобы избежать усложнения модели (и ее переобучения соответственно) [Montufar et al., 2014].

Следующим составным элементом нашей нейронной сети являются *слои нормализации*. Они использованы для того, чтобы улучшить производительность модели, обеспечить ее стабильность и скорости сходимости, что, соответственно, улучшает ее точность. Существует несколько вариантов слоев нормализации, одним из которых является пакетная нормализация (Batch Normalization), которая работает с «пакетами» данных. Слои нормализации позволяют уменьшить переобучение, что улучшает качество предсказаний модели на новых данных [Ba et al., 2016].

Слои *функции активации* также являются необходимыми компонентами нейронной сети, так как они вводят нелинейность в вычисления и позволяют модели аппроксимировать сложные нелинейные отображения между входом и выходом. Без функции активации все линейные слои в нейронной сети могут быть объединены в один линейный слой, что значительно снизит способность модели к обучению сложных отображений. Нами в процессе построения модели были выбраны две функции активации *ReLu* и одна сигмоидная функция активации. Кратко опишем каждую из них:

- Слои функции активации *ReLu* (*Rectified Linear Unit*) позволяют вводить нелинейность в модель и захватывать сложные зависимости в данных. Эти слои отбрасывают все отрицательные значения входа, а положительные значения передают без изменения. Это означает, что слои *ReLu* могут улучшать способность модели выделять важные признаки и снижать влияние шума в данных [Sharma et al., 2017].
- Сигмоидный (*Sigmoid*) слой функции активации используется в конце нейронной сети для задач бинарной классификации, чтобы ограничить выход модели в диапазоне от 0 до 1. Это позволяет интерпретировать выход модели как вероятность отнесения объекта к нужному классу [Sharma et al., 2017].

Исключающие слои (*Dropout layers*) используются для борьбы с переобучением в нейронных сетях. Во время обучения *Dropout* случайным образом отключает некоторые нейроны в сети, что позволяет избежать сильной корреляции между ними и уменьшить эффект переобучения. Во время прямого прохода через слой *Dropout* каждый нейрон имеет определенную вероятность p быть отключенным (в нашем случае стоял $p = 0.2$). Таким образом, для каждой эпохи обучения разные подмножества нейронов будут

отключены, что уменьшает эффект переобучения и способствует обобщению нейронной сети [Srivastava, 2013].

Функция потерь. В качестве функции потерь нами была использована бинарная потеря перекрестной энтропии (*BCELoss binary cross-entropy loss*) – функция, используемая в задачах бинарной классификации. Функции потерь необходимы нейронным сетям для того, чтобы определять, насколько хорошо модель выполняет поставленную задачу. Она вычисляет разницу между предсказаниями модели и правильными ответами (так называемыми метками или метками классов) на обучающих данных [Ruby, Yendapalli, 2020].

В случае же *BCELoss* вычисление потери идет на каждом примере, сравнивая прогнозы модели с соответствующими метками классов и усреднением их по всей выборке. Эта функция потерь имеет преимущество перед другими в том, что она хорошо работает с вероятностными выходами, которые можно интерпретировать как вероятности принадлежности к каждому классу, что очень полезно для нашей задачи.

Оптимизатор. Оптимизаторы в нейронных сетях используются для обновления параметров модели (например, весов и смещений) в процессе обучения таким образом, чтобы минимизировать функцию потерь и повысить точность предсказания модели на новых данных.

В качестве оптимизатора нами был использована Адаптивная оценка моментов *Adam* (*Adaptive Moment Estimation*) – один из наиболее популярных оптимизаторов в нейронных сетях. Он автоматически настраивает скорость обучения в соответствии с градиентами каждого параметра и приводит к быстрой сходимости при обучении на больших данных. Этот оптимизатор эффективно сочетается с бинарной потерей перекрестной энтропии (*Binary Cross Entropy Loss*) для задач бинарной классификации.

Результаты

На основе собранных и предварительно обработанных данных с помощью описанных в прошлом параграфе моделей мы прогнозировали итоговое значение принятия законопроекта: будет он принят или нет. Таким образом, значения «x» принимает одномер-

ный массив признаков, среди которых (1) текстовые признаки, преобразованные в векторы с помощью *rubert_tiny* и (2) признаки на основе характеристик законопроекта: его тематики, внесшего субъекта и так далее, а значения «y» принимают значения 0 или 1, характеризующие то, был ли законопроект принят или нет. Для определения основных параметров модели мы также варьируем набор признаков, попеременно исключая характеристики законопроектов, тексты приложенных документов, а также варьируя блок данных созывов для обучения. Рассмотрим подробно результаты классификации моделей, представленные в табл. 4.

Первоначально мы запустили модель на основе данных всех созывов, но вскоре обнаружили, что качество модели растет, если анализировать данные только пятого созыва и позднее. Это подтверждается двумя моментами: (1) неравномерностью распределений данных с пробелами между первыми и последними четырьмя созывами и (2) характерными чертами последних созывов: возросшую роль партии «Единая Россия» в парламенте, большее количество законов в шестом созыве, а также председательство В.В. Володина – неформальные процедуры принятия законопроектов при нем претерпели изменения. Таким образом, для обучения моделей мы решили использовать только данные 5–8 созывов, что увеличило качество модели минимум на 3% (см. табл. 10).

Качество прогнозов. Лучшие результаты были получены при обучении логистической регрессии на текстах всех прилагаемых к законопроекту документов (93,7% точности) и обучении только на текстах заключений (92% точности). Алгоритм случайного леса показал лучшие результаты на параметрах законопроектов (87,3% точности), а также на текстах заключений (87,7% точности). Нейронная сеть показала лучшие результаты на текстах заключений и всех текстах, прилагаемых к законопроектам (92% точности). Наиболее точные предсказания по обоим направлениям: как анализу текстов прилагаемых документов, так и анализу параметров законопроектов дают модели, которые первоначально вводились нами как дополнительные: логистическая регрессия и случайный лес.

Таблица 4

Оценка качества классификации различных алгоритмов в зависимости от стартовых параметров модели

	Алгоритмы классификации		
Входящие параметры	**Случайный лес**	**Логистическая регрессия**	**Нейронная сеть (1000 эпох)**
Тексты документов и параметры законопроектов всех созывов	F1 macro: 0.825 F1 weight: 0.853 Bal. acc: 0.8 roc_auc: 0.99 cross_val_score: [0.72, 0.88, 0.84, 0.77, 0.47]	F1 macro: 0.824 F1 weight: 0.836 Bal. acc: 0.785 roc_auc: 0.89 cross_val_score: [0.81, 0.84 0.835. 0.775, 0.635]	F1 Macro: 0.737 F1 Weighted: 0.784 Bal. acc.: 0.715 roc_auc: 0.824
Тексты документов и параметры законопроектов только 5–8 созывов	F1 macro: 0.864 F1 weight: 0.872 Bal. acc: 0.857 roc_auc: 0.99 cross_val_score: [0.83, 0.85, 0.89, 0.84, 0.85]	F1 macro: 0.842 F1 weight: 0.853 Bal. acc: 0.83 roc_auc: 0.92 cross_val_score: [0.837, 0.825, 0.835, 0.846, 0.82]	F1 macro: 0.821 F1 weight: 0.830 Bal. acc.: 0.815 roc_auc: 0.909
Только тексты всех документов 5–8 созыва	F1 macro: 0.869 F1 weight: 0.876 Bal. acc: 0.86 roc_auc: 0.99 cross_val_score: [0.83, 0.84, 0.87, 0.8, 0.86]	F1 macro: 0.934 **F1 weight: 0.937** Bal. acc: 0.93 roc_auc: 0.986 cross_val_score: [0.927, 0.91, 0.94, 0.91, 0.92]	F1 macro: 0.917 **F1 weight: 0.921** Bal. acc.: 0.915 roc_auc: 0.976
Только параметры законопроектов 5–8 созыва	F1 macro: 0.865 **F1 weight: 0.873** Bal. acc: 0.86 roc_auc: 0.99 cross_val_score: [0.56, 0.81, 0.8, 0.8, 0.68]	F1 macro: 0.84 F1 weight: 0.85 Bal. acc: 0.83 roc_auc: 0.92 cross_val_score: [0.84, 0.826, 0.83, 0.84, 0.82]	F1 macro: 0.831 F1 weight: 0.839 Bal. acc.: 0.828 roc_auc: 0.912
Только текст законопроектов 5–8 созыва	F1 macro: 0.657 F1 weight: 0.686 Bal. acc: 0.654 roc_auc: 0.96 cross_val_score: [0.65, 0.64, 0.645, 0.654, 0.66]	F1 macro: 0.735 F1 weight: 0.756 Bal. acc: 0.73 roc_auc: 0.81 cross_val_score: [0.74, 0.72, 0.73, 0.725, 0.71]	F1 macro: 0.722 F1 weight: 0.741 Bal. acc.: 0.714 roc_auc: 0.810
Только текст финансового обоснования 5–8 созыв	F1 macro: 0.65 F1 weight: 0.68 Bal. acc: 0.647 roc_auc: 0.95 cross_val_score: [0.6, 0.62, 0.615, 0.62, 0.61]	F1 macro: 0.7 F1 weight: 0.72 Bal. acc: 0.687 roc_auc: 0.792 cross_val_score: [0.685, 0.69, 0.69, 0.7, 0.68]	F1 macro: 0.698 F1 weight: 0.721 Bal. acc.: 0.692 roc_auc: 0.780
Только текст пояснительной записки 5–8 созыва	F1 macro: 0.68 F1 weight: 0.7 Bal. acc: 0.67 roc_auc: 0.96 cross_val_score: [0.69, 0.67, 0.67, 0.64, 0.67]	F1 macro: 0.753 F1 weight: 0.77 Bal. acc: 0.746 roc_auc: 0.84 cross_val_score: [0.75, 0.76, 0.76, 0.74, 0.74]	F1 macro: 0.746 F1 weight: 0.763 Bal. acc.: 0.738 roc_auc: 0.833
Только текст заключений 5–8 созыва	F1 macro: 0.869 **F1 weight: 0.877** Bal. acc: 0.862 roc_auc: 0.991 cross_val_score: [0.856, 0.855, 0.89, 0.827, 0.864]	F1 macro: 0.916 **F1 weight: 0.92** Bal. acc: 0.915 roc_auc: 0.97 cross_val_score: [0.9, 0.885, 0.917, 0.885, 0.917]	F1 macro: 0.914 **F1 weight: 0.917** **Bal. acc.:** **0.921** roc_auc: 0.975

На основе *всех параметров* модели прогнозируют результаты принятия законопроектов с 87,2% точности (случайный лес); *только на основе текстов законопроектов* – с 75,6% точности (логистическая регрессия); *финансового обоснования и пояснительной записки* – 72 и 77% точности соответственно (логистическая регрессия). Лучшие результаты выделены в табл. 10 жирным шрифтом.

В случае работы как с текстовыми данными, так и параметрами законопроектов лучше всего проявил себя алгоритм случайного леса. В случае работы с текстами на первое место вышла логистическая регрессия, а за ней следовала нейронная сеть и случайный лес соответственно.

Одним из главных аспектов работы с моделями является определение признаков, оказывающих наибольшее влияние на результаты прогнозирования. Поскольку текстовые данные документов были преобразованы в одномерный набор признаков, выделить наиболее важные слова и предложения на данном этапе не представляется возможным, однако мы можем рассмотреть влияние «глобальных» факторов на результаты прогнозирования.

Так, важнейшим глобальным признаком модели является текст всех заключений на законопроект. Этот признак включал в себя как заключения комитетов-исполнителей и правового управления, так и правительства и Центрального банка. Этот параметр является определяющим для модели, поскольку только на его основе возможно прогнозирование с 92%-ной точностью. Можно с уверенностью сказать, что *законопроекты в Государственной думе принимаются, во многом опираясь на позицию органов исполнительной власти и других компетентных институтов*. Более подробно выводы о влиянии каждого типа заключений будут сделаны в следующих работах.

Вторым по важности признаком на основе параметров законопроектов является «Субъект права законодательной инициативы». В моделях классификации по карте характеристик законопроектов этот параметр имеет 31,5% значимости в прогнозировании. Это объясняется наличием парламентского большинства у партии власти и общим консенсусом, выработанным между партиями с начала нулевых годов [Панов, Сулимов, 2023] (см. также: [Помигуев, 2016]). Следующим по важности является признак «Длина текста заключений», имеющий 20% значимости (что коррелирует с

результатами, озвученными ранее). Остальные параметры играют не столь важную роль в прогнозировании: так, «Год первого чтения» имеет 6% значимости, длина текста законопроекта – 4, «ответственный комитет» – 3,7, длина финансового обоснования – 3,3% и так далее.

Таким образом, ожидаемое наибольшее влияние на вероятность принятия законопроекта оказывает такой параметр, как текст заключения и субъект законодательной инициативы. Интересным результатом является слабая роль длины финансового обоснования, поскольку предполагалось, что в случае наличия короткого комментария про отсутствие необходимости выделения дополнительных средств после принятия законопроекта шанс принятия законопроекта значительно увеличится.

Таким образом, **подтверждаются** следующие заявленные нами гипотезы.

- **Гипотеза 1** об эффективности методов Логистической регрессии и алгоритма случайного леса в решении задач бинарной классификации законопроектов на принятые и не принятые.
- **Гипотеза 2** о существенном влиянии субъекта законодательной инициативы на вероятность его принятия.
- **Гипотеза 6** о существенном влиянии текста заключений к законопроекту на вероятность его принятия.
- **Гипотеза 7** о повышении качества прогностического потенциала модели при использовании данных только последних созывов.
- **Гипотеза 8** о повышении качества прогностического потенциала модели при одновременном использовании как параметров паспорта законопроекта, так и текстовых параметров.

Следующие гипотезы на основании полученной информации **нами отклоняются**.

- **Гипотеза 3** о существенном влиянии текста законопроекта на вероятность его принятия.
- **Гипотеза 4** о существенном влиянии пояснительной записки законопроекта на вероятность его принятия.
- **Гипотеза 5** о существенном влиянии финансового обоснования законопроекта на вероятность его принятия.

Обсуждение

В рамках развития направления математического анализа и прогнозирования законотворческой деятельности Федерального собрания Российской Федерации также возможна разработка следующих научных направлений.

1. **Прогнозирование голосования депутатов Государственной думы**. Используя имеющиеся данные по результатам голосований конкретных депутатов по законопроектам и характеристик законопроектов, возможно создание реляционной базы с последующим прогнозированием результата голосования отдельного парламентария на основе прошлой истории его голосов и характеристик законопроекта.

2. **Прогнозирование времени принятия законопроектов.** На основе статистики прохождения каждого чтения возможна разработка модели машинного обучения, прогнозирующей время принятия законопроектов. Сложности данного направления заключаются в необходимости предварительной обработки данных, поскольку они будут иметь распределение с длинным хвостом из-за имеющихся выбросов – «долгих» законопроектов.

3. **Сетевой анализ законопроектов.** Используя тексты принятых федеральных законов и информации о принадлежности их к конкретному направлению, возможно построение сети внутритекстовых ссылок для последующего сетевого анализа влиятельности каждого законопроекта для всей сети законотворчества и кластеризации законопроектов по их тематическим блокам.

4. **Автоматическая генерация текстов законопроектов**. На основании базы собранных текстов законопроектов и прилагаемых к ним документам возможна разработка нейронной модели, генерирующей по запросу пользователя макет законопроекта, текст заключения, финансовое обоснование или пояснительную записку. Данное направление столкнется с рядом сложностей технического характера, таких как оформление ссылок на законопроекты или формирование разметки документов, однако данные проблемы возможно решить за счет генерации сети ссылок законопроектов и обращения к ресурсам *html* разметки.

5. **Определение эффективности комитетов**. Используя данные о количестве и текстах законопроектов, а также общем числе комитетов, возможно разработать модель машинного обуче-

ния, уменьшающую пространство размерностей законопроектов и рассчитывающую наиболее эффективное количество и тематические направления комитетов Государственной думы РФ.

M.V. Khavronenko[*]

Predicting the outcomes of consideration of bills in the State Duma using a neural network model[1]

Abstract. In this paper, we trained our machine learning and neural network models to predict the outcome of the bills' consideration in the Russian State Duma. We used data collected from October 24, 1994 to December 1, 2022. A rubert-tiny model was used for data preprocessing, a random forest classifier, logistic regression and a neural network model of 3 linear layers were used for prediction.

The models demonstrated qualitative results on real-life data: 94% accuracy was achieved by using attached documents' texts as the models' parameters and 87% accuracy by training on the data from the bill's passport. Based on the text of the draft alone, the model's accuracy accounted for 75.6%. The most important factor influencing the prediction result was the text of the Governmental conclusion. The second most important parameter influencing the results was the "Subject of the right of legislative initiative" with 31.5% of significance in the models' prediction.

Random forest algorithm performed best when working with combined text data and bill passport parameters while logistic regression and neural network showed promising results based on textual parameters alone. The probability of bill's adoption was not significantly influenced by the financial justification text, the explanatory note text or the subject matter of the bill. The author draws conclusions about the practical applications of the trained models, as well as identifies further scientific problems in the field of mathematical analysis and prediction of lawmaking.

Keywords: roll-call voting; Russian Federation State Duma; forecasting of lawmaking; legislative research; neural networks; Ru-bert; machine learning; legal tech.

For citation: Khavronenko M.V. Predicting the outcomes of consideration of bills in the State Duma using neural network model. *Political science* (RU). 2024, N 3, P. 211–240. DOI: http://www.doi.org/10.31249/poln/2024.03.09

[*] **Khavronenko Maxim,** Lomonosov Moscow State University (Moscow, Russia), e-mail: KhavronenkoMax@gmail.com

[1]This article was supported by the Russian Society of Political Scientists grant program.

References

Ansolabehere S., Snyder Jr J.M., Stewart III C. Candidate positioning in US House elections. *American journal of political science*. 2001, Vol. 45, N 1, P. 136–159. DOI: https://doi.org/10.2307/2669364

Ba J.L., Kiros J.R., Hinton G.E. Layer normalization. *arXiv preprint* arXiv:1607.06450.2016.

Bari A., Brower W., Davidson C. Using artificial intelligence to predict legislative votes in the United States congress. *2021 IEEE 6 th International Conference on Big Data Analytics (ICBDA)*. 2021, P. 56–60.

Budhwar A., Kuboi T., Dekhtyar A., Khosmood F. Predicting the vote using legislative speech. *Proceedings of the 19 th annual international conference on digital government research: governance in the data age*. 2018, P. 1–10.

Cherryholmes C.H., Shapiro M.J. Representatives and roll calls: a computer simulation of voting in the eighty-eighth congress. *Bobbs-Merrill*, Indianapolis, 1969, 196 p.

Chiou F.Y., Rothenberg L.S. When pivotal politics meets partisan politics. *American journal of political science*. 2003, Vol. 47, N 3, P. 503–522. DOI: https://doi.org/10.2307/3186112

Clinton J., Jackman S., Rivers D. The statistical analysis of roll call data. *American political science review*. 2004, Vol. 98, N 2, P. 355–370. DOI: https://doi.org/10.1017/s0003055404001194

Devlin, J., Chang, M.W., Lee, K., Toutanova, K Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018.

Gerrish S.M., Blei D.M. Predicting legislative roll calls from text. *Proceedings of the 28 th International Conference on Machine Learning*. ICML 2011, 2011.

Gerrish S., Blei D. How they vote: Issue-adjusted models of legislative behavior. *Advances in neural information processing systems*. 2012, Vol. 25, P. 2753–2761.

Jackman S. Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking. *Political analysis*. 2001, Vol. 9, N 3, P. 227–241. DOI: https://doi.org/10.1093/polana/9.3.227

Karyagin M.E. The deputies voting strategies: what open data says about the work of the State Duma? *Political science (RU)*. 2023, N 1, P. 303–321. DOI: http://www.doi.org/10.31249/poln/2023.01.13 (In Russ.)

Khashman Z., Khashman A. Anticipation of political party voting using artificial intelligence. *Procedia computer science*. 2016, Vol. 102, P. 611–616. DOI: https://doi.org/10.1016/j.procs.2016.09.450

Kim I.S., Londregan J., Ratkovic M. Estimating spatial preferences from votes and text. *Political analysis*. 2018, Vol. 26, N 2, P. 210–229. DOI: https://doi.org/10.1017/pan.2018.7

Korn J.W., Newman M.A. A deep learning model to predict congressional roll call votes from legislative texts. *Machine learning and applications: an international journal (MLAIJ)* 2020, Vol. 7, N 3–4.

Kraft P., Jain H., Rush A.M. An embedding model for predicting roll-call votes. *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016, P. 2066–2070.

Ladha K.K. Coalitions in congressional voting. *Public choice*. 1994, Vol. 78, P. 43–63. DOI: https://doi.org/10.1007/bf01053365

Matthews D.R., Stimson J.A. Yeas and nays: Normal decision-making in the US House of Representatives. New York: Wiley-Interscience, 1975, 190 p.

McFadden D. Econometric models of probabilistic choice. In: Manski C., McFadden D. (eds). *Structural analysis of discrete data with econometric applications.* Cambridge: MIT Press, 1981, P. 198–272.

Montufar, G.F., Pascanu, R., Cho, K., & Bengio, Y. On the number of linear regions of deep neural networks. *Advances in neural information processing systems*. 2014, Vol. 4, P. 2924–2932.

Nay J.J. Predicting and understanding law-making with word vectors and an ensemble model. *PloS one*. 2017, Vol. 12, N 5, P. e0176999. DOI: https://doi.org/10.1371/journal.pone.0176999

Panov P.V., Sulimov K.A. Ideological «harmony» in the State Duma: emergence, content, boundaries. *Political science (RU)*. 2023, N 1, P. 61–91. DOI: http://www.doi.org/10.31249/poln/2023.01.03 (In Russ.)

Pomiguev I.A. The Council of the State Duma: real veto player or a technical executive? Polis. *Political studies*. 2016, N 2, P. 171–183. DOI: https://doi.org/10.17976/jpps/2016.02.12 (In Russ.)

Pomiguev I.A., Alekseev D.V. Resetting bills: discontinuity as a political technology for blocking policy decision. *Polis. Political studies*. 2021, N 4, P. 176–191. DOI: https://doi.org/10.17976/jpps/2021.04.13 (In Russ.)

Poole K.T., Rosenthal H. A spatial model for legislative roll call analysis. *American journal of political science*. 1985, P. 357–384. DOI: https://doi.org/10.2307/2111172

Ruby, U., Yendapalli, V. (Binary cross entropy with deep learning technique for image classification. *International journal of advanced trends in computer science and engineering*. 2020, Vol. 9, N 10. DOI: https://doi.org/10.30534/ijatcse/2020/175942020

Sharma S., Sharma S., Athaiya A. Activation functions in neural networks. *International journal of engineering applied sciences and technology*. 2017, Vol. 6, N 12, P. 310–316. DOI: https://doi.org/10.33564/ijeast.2020.v04i12.054

Snyder Jr J.M., Groseclose T. Estimating party influence in congressional roll-call voting. *American journal of political science*. 2000, Vol. 44, N 2, P. 193–211. DOI: https://doi.org/10.2307/2669305

Srivastava N. *Improving neural networks with dropout*. Toronto, 2013, 26 p. Mode of access: https://www.cs.toronto.edu/~nitish/msc_thesis.pdf (accessed: 28.05.2024)

Wang, E., Liu, D., Silva, J., Carin, L., Dunson, D. Joint analysis of time-evolving binary matrices and associated documents. *Advances in neural information processing systems*. 2010, Vol. 23.

Wang, E., Salazar, E., Dunson, D., Carin, L. Spatio-temporal modeling of legislation and votes. *Bayesian analysis*. 2013, Vol. 8, N 1, P. 233–268. DOI: https://doi.org/10.1214/13-ba810

Yang, Y., Lin, X., Lin, G., Huang, Z., Jiang, C., Wei, Z Joint representation learning of legislator and legislation for roll call prediction. *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*. 2021, P. 1424–1430. DOI: https://doi.org/10.24963/ijcai.2020/198

Yano T., Smith N.A., Wilkerson J. Textual predictors of bill survival in congressional committees. *Proceedings of the 2012 conference of the North American chapter of*

the Association for computational linguistics: human language technologies. 2012, P. 793–802.

Литература на русском языке

Карягин М.Е. Индивидуальные стратегии голосования парламентариев: что открытые данные говорят о работе депутатов Государственной думы? // Политическая наука. – 2023. – № 1. – С. 303–321. – DOI: http://www.doi.org/10.31249/poln/2023.01.13

Панов П.В., Сулимов К.А. Идеологическая «гармония» в Государственной думе: возникновение, содержание, границы // Политическая наука. – 2023. – № 1. – С. 61–91. – DOI: http://www.doi.org/10.31249/poln/2023.01.03

Помигуев И.А. Совет Государственной думы: реальный вето-игрок или технический исполнитель? // Полис. Политические исследования. – 2016. – № 2. – С. 171–183. – DOI: https://doi.org/10.17976/jpps/2016.02.12

Помигуев И.А., Алексеев Д.В. Обнуление законопроектов: дисконтинуитет как технология блокирования политических решений // Полис. Политические исследования. – 2021. – № 4. – С. 176–191. – DOI: https://doi.org/10.17976/jpps/2021.04.13