

Министерство образования и науки Российской Федерации
Московский физико-технический институт
(государственный университет)

Физтех-школа прикладной математики и информатики
Кафедра финансовых технологий

Выпускная квалификационная работа бакалавра

**Исследование влияния операционных
метрик на метрики удовлетворенности
клиента**

Автор:

Студент группы Б05-104
Карпов Никита Сергеевич

Научный руководитель:

научная степень
Ишмеев Марат Рашидович

Научный консультант:

научная степень

АННОТАЦИЯ

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	5
1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СТАТИСТИЧЕСКОГО АНАЛИЗА И МАШИННОГО ОБУЧЕНИЯ	6
1.1 Коэффициенты корреляции	6
1.1.1 Корреляция Кендалла-тау	6
1.1.2 Корреляция Пирсона	6
1.1.3 Корреляция Спирмена	7
1.2 Примеры вычисления коэффициентов корреляции	7
1.3 Статистические критерии проверки значимости коэффициен- тов корреляции	8
1.3.1 t-критерий для коэффициента корреляции Пирсона . .	8
1.3.2 z-преобразование Фишера для корреляций	8
1.3.3 Критерии значимости для корреляций Кендалла и Спирмена	9
1.3.4 Примеры применения критериев для проверки значи- мости коэффициентов корреляции	9
1.4 Доверительные интервалы для доли (бернуллиевой величины)	10
1.4.1 Классические подходы построения интервалов	10
1.4.2 Сравнение и обсуждение доверительных интервалов . .	11
1.4.3 Примеры вычисления доверительных интервалов	13
1.5 Методы и модели машинного обучения	13
1.5.1 Логистическая регрессия: теоретические аспекты и интерпретация результатов	13
1.5.2 Модель градиентного бустинга	15
1.5.3 Интерпретация моделей машинного обучения с помо- щью SHAP	16
2 ПРАКТИЧЕСКИЙ АНАЛИЗ ВЛИЯНИЯ ОПЕРАЦИОННЫХ МЕТ- РИК НА УРОВЕНЬ CSAT	19

2.1	Подготовка и описание данных	19
2.1.1	Описание используемой витрины данных	19
2.1.2	Операционные и неоперационные признаки в исследовании	19
2.2	Исследование данных с использованием статистических критериев	19
2.2.1	Анализ зависимости CSAT от времени жизни таска	19
2.2.2	Оценка необходимости проведения А/В теста	19
2.3	Построение и анализ моделей машинного обучения	19
2.3.1	Построение логистической регрессии и анализ ее коэффициентов	19
2.3.2	Применение модели CatBoost и анализ важности признаков с помощью SHAP	19
2.3.3	Анализ влияния различных факторов на оценки пользователей	19
3	РЕЗУЛЬТАТЫ И ВЫВОДЫ ИССЛЕДОВАНИЯ	20
3.1	Оценка эффективности сокращения времени выполнения задач	20
3.2	Интерпретация и практические рекомендации на основе моделей	20
	БИБЛИОГРАФИЧЕСКИЙ СПИСОК	22
	ПРИЛОЖЕНИЕ А	24

ВВЕДЕНИЕ

1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ СТАТИСТИЧЕСКОГО АНАЛИЗА И МАШИННОГО ОБУЧЕНИЯ

1.1 Коэффициенты корреляции

Коэффициенты корреляции используются для измерения силы и направления связи между двумя переменными. В зависимости от типа данных и целей исследования применяются различные коэффициенты: коэффициент Пирсона — для линейной связи между количественными переменными, коэффициенты Спирмена и Кендалла — для ранговой корреляции, устойчивой к выбросам и нелинейностям.

1.1.1 Корреляция Кендалла-тау

Коэффициент τ Кендалла измеряет степень согласованности пар наблюдений. Он основан на подсчёте числа согласованных (C) и несогласованных (D) пар:

$$\tau = \frac{C - D}{\frac{1}{2}n(n - 1)}$$

где n — общее число наблюдений. Метод обладает высокой устойчивостью к выбросам и предпочтителен при наличии нечисловых шкал или неполной информации о распределении данных.

Коэффициент Кендалла был предложен в работе М. Кендалла в 1938 году [1].

1.1.2 Корреляция Пирсона

Коэффициент корреляции Пирсона r — наиболее распространённая мера линейной зависимости между двумя количественными переменными:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

где \bar{x} и \bar{y} — средние значения переменных x и y . Значение $r \in [-1, 1]$, где ± 1 указывает на идеальную линейную зависимость. При этом $r = 0$ не означает отсутствие связи, а лишь отсутствие линейной зависимости.

Метод предложен К. Пирсоном в 1896 году [2].

1.1.3 Корреляция Спирмена

Коэффициент Спирмена ρ оценивает монотонную зависимость между переменными и определяется как коэффициент Пирсона между ранжированными значениями:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

где d_i — разность между рангами x_i и y_i в выборке. Подходит для оценки силы связи в случае нелинейных, но монотонных зависимостей.

Впервые описан Ч. Спирменом в 1904 году [3].

1.2 Примеры вычисления коэффициентов корреляции

Рассмотрим выборку из пяти наблюдений:

$$X = \{2, 4, 6, 8, 10\}, \quad Y = \{1, 3, 4, 7, 9\}$$

Пирсон: $\bar{X} = 6, \bar{Y} = 4.8$

$$r = \frac{\sum (X_i - 6)(Y_i - 4.8)}{\sqrt{\sum (X_i - 6)^2 \cdot \sum (Y_i - 4.8)^2}} = \frac{40}{\sqrt{40 \cdot 40.8}} \approx 0.99$$

Спирмен: Ранги $X = \{1, 2, 3, 4, 5\}, Y = \{1, 2, 3, 4, 5\}, \rho = 1$

Кендалл: Все пары согласованные $\Rightarrow \tau = 1$

Таким образом, все коэффициенты показывают высокую положительную корреляцию.

1.3 Статистические критерии проверки значимости коэффициентов корреляции

Для оценки статистической значимости коэффициентов корреляции применяются специализированные критерии, зависящие от типа корреляционной меры и распределения данных. Рассмотрим основные подходы, применяемые на практике.

1.3.1 t-критерий для коэффициента корреляции Пирсона

Для проверки гипотезы $H_0 : \rho = 0$ (отсутствие линейной связи) используется t-распределение:

$$t = r \cdot \sqrt{\frac{n-2}{1-r^2}}$$

где r — выборочный коэффициент корреляции Пирсона, n — размер выборки. Статистика t подчиняется t-распределению с $n - 2$ степенями свободы. Полученное значение сравнивается с критическим, соответствующим заданному уровню значимости α .

Метод используется при соблюдении условий нормальности и независимости наблюдений [4].

1.3.2 z-преобразование Фишера для корреляций

Для оценки значимости разности между двумя независимыми коэффициентами корреляции или построения доверительного интервала для одного r применяется преобразование Фишера:

$$z = \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$$

где z асимптотически нормально распределён с дисперсией $1/(n-3)$. Это позволяет использовать стандартные нормальные границы при проверке гипотез:

$$Z = \frac{z - z_0}{\sqrt{1/(n-3)}}$$

Метод применяется как при сравнении корреляций между группами, так и при проверке отличия r от заданного значения [5].

1.3.3 Критерии значимости для корреляций Кендалла и Спирмена

Коэффициенты Спирмена ρ и Кендалла τ требуют иных подходов. При достаточно большой выборке (обычно $n > 10$) их значения стандартизируются, и используется аппроксимация нормальным распределением:

Для Спирмена:

$$Z = \frac{\rho \cdot \sqrt{n-1}}{\sqrt{1-\rho^2}}$$

Для Кендалла:

$$Z = \frac{3 \cdot \tau \cdot \sqrt{n(n-1)}}{\sqrt{2(2n+5)}}$$

Значения Z сравниваются с критическим значением нормального распределения. Альтернативно можно использовать точные p-value при малых n (например, через перестановочные тесты) [6].

1.3.4 Примеры применения критериев для проверки значимости коэффициентов корреляции

Рассмотрим пример применения t-критерия для коэффициента Пирсона. Для выборки из $n = 20$ пар наблюдений получено $r = 0.57$. Тогда

$$t = 0.57 \cdot \sqrt{\frac{20-2}{1-0.57^2}} \approx 3.03$$

Критическое значение t_{crit} для $\alpha = 0.05$, $df = 18$ составляет 2.101. Так как $3.03 > 2.101$, нулевая гипотеза отклоняется — корреляция статистически значима.

Для сравнения двух коэффициентов корреляции с использованием z -преобразования Фишера:

$$\begin{aligned} r_1 &= 0.65, \quad r_2 = 0.30, \quad n_1 = n_2 = 50 \\ z_1 &= 0.5 \ln \left(\frac{1 + 0.65}{1 - 0.65} \right) \approx 0.775, \quad z_2 \approx 0.309 \\ Z &= \frac{0.775 - 0.309}{\sqrt{1/(50 - 3) + 1/(50 - 3)}} \approx 3.72 \end{aligned}$$

При $Z > 1.96$, разница между коэффициентами статистически значима на уровне 0.05.

1.4 Доверительные интервалы для доли (бернуллиевой величины)

Доверительный интервал (ДИ) для доли — это интервал, в котором с заданной вероятностью содержится истинное значение параметра p распределения Бернулли. Он широко применяется в задачах бинарной классификации, в анализе успехов/неудач, а также в А/В тестировании.

1.4.1 Классические подходы построения интервалов

WILSON SCORE INTERVAL

Wilson-интервал улучшает симметричность классического нормального приближения. Он основан на перестроенной статистике и формулируется следующим образом:

$$\begin{aligned} \hat{p}_W &= \frac{\hat{p} + \frac{z^2}{2n}}{1 + \frac{z^2}{n}}, \quad ME = \frac{z}{1 + \frac{z^2}{n}} \cdot \sqrt{\frac{\hat{p}(1 - \hat{p})}{n} + \frac{z^2}{4n^2}} \\ CI &= \hat{p}_W \pm ME \end{aligned}$$

где z — квантиль стандартного нормального распределения. Этот метод особенно точен при малых n и при \hat{p} , близком к 0 или 1 [7].

JEFFREYS INTERVAL

Интервал Джеффриса основан на байесовском подходе с априорным распределением Бета(0.5, 0.5). Доверительный интервал строится как:

$$CI = \text{Beta}_{\alpha/2}^{-1}(x + 0.5, n - x + 0.5)$$

где x — число успехов. Интервал обладает хорошими частотными свойствами и симметрией [7].

CLOPPER–PEARSON INTERVAL

Точный Clopper–Pearson интервал строится на основе биномиального распределения:

$$CI = [B(\alpha/2; x, n - x + 1), B(1 - \alpha/2; x + 1, n - x)]$$

где B — обратная функция неполной бета-функции. Несмотря на точность, интервал часто критикуется за избыточную ширину [8].

AGRESTI–COULL INTERVAL

Agresti–Coull-интервал — это аппроксимация Wilson-интервала, предлагающая удобное приближение:

$$\tilde{n} = n + z^2, \quad \tilde{x} = x + \frac{z^2}{2}, \quad \tilde{p} = \frac{\tilde{x}}{\tilde{n}}, \quad CI = \tilde{p} \pm z \cdot \sqrt{\frac{\tilde{p}(1 - \tilde{p})}{\tilde{n}}}$$

Он используется благодаря простоте вычислений и приемлемой точности [9].

1.4.2 Сравнение и обсуждение доверительных интервалов

Сравнение различных подходов показывает, что нормальное приближение (так называемый Wald-интервал) часто имеет заниженную фактическую надёжность, особенно при малых объёмах выборки ($n < 30$) или при

оценках вероятности, близких к 0 или 1. Поэтому он не рекомендуется для практического применения.

Wilson и Jeffreys интервалы демонстрируют лучшие частотные свойства и надёжность. Они особенно полезны в ситуациях, когда выборка небольшая или доля успехов близка к краевым значениям. Wilson-интервал рекомендуется при необходимости симметричной аппроксимации и строгом контроле ширины интервала, в том числе в A/B тестировании и биостатистике. Jeffreys-интервал актуален в байесовском контексте и показывает стабильные частотные свойства даже при очень малых n .

Clopper–Pearson интервал, будучи «точным», часто даёт чрезмерно широкие границы, особенно при небольшом числе наблюдений. Он подходит, когда важна консервативность, например, в фармацевтических испытаниях или регулируемой среде, где требуется минимизировать вероятность ложноположительных решений.

Agresti–Coull-интервал — хорошая альтернатива для быстрой оценки: он не требует сложных вычислений и обеспечивает удовлетворительную точность. Рекомендуется для практических приложений, когда важен баланс между вычислительной простотой и приемлемой статистической точностью.

Таким образом, выбор метода зависит от контекста:

- **Wilson:** рекомендуется как надёжный метод общего назначения;
- **Jeffreys:** подходит для байесовских и частотных оценок при малых n ;
- **Clopper–Pearson:** применяется при необходимости строгости и регуляторных ограничений;
- **Agresti–Coull:** используется в прикладных задачах при больших n ;
- **Wald:** не рекомендуется (только в случае $n > 100$ и $p \approx 0.5$).

Сравнительный анализ этих подходов можно найти в обобщающем исследовании Brown et al. (2001) [7].

1.4.3 Примеры вычисления доверительных интервалов

Пусть в эксперименте из $n = 100$ наблюдений получено $x = 20$ успехов ($\hat{p} = 0.2$), при уровне доверия 95%:

- **Wilson:** $CI \approx [0.129, 0.296]$
- **Jeffreys:** $CI \approx [0.129, 0.298]$
- **Clopper–Pearson:** $CI \approx [0.127, 0.299]$
- **Agresti–Coull:** $CI \approx [0.133, 0.300]$

Эти результаты демонстрируют, что при небольших выборках все четыре метода дают схожие оценки, но Wilson и Jeffreys обеспечивают лучшую симметрию и сдержанность по ширине интервала.

1.5 Методы и модели машинного обучения

1.5.1 Логистическая регрессия: теоретические аспекты и интерпретация результатов

Логистическая регрессия является одной из базовых моделей бинарной классификации. Её цель — оценить вероятность принадлежности объекта к одному из двух классов в зависимости от набора признаков. В отличие от линейной регрессии, логистическая модель использует логистическую функцию для получения значений в интервале $(0, 1)$, что позволяет интерпретировать результат как вероятность.

Математически модель имеет следующий вид:

$$P(y = 1 | x) = \frac{1}{1 + \exp(-z)}, \quad \text{где } z = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

ЛОГАРИФМ ШАНСОВ И ИНТЕРПРЕТАЦИЯ КОЭФФИЦИЕНТОВ.

Для удобства анализа логистическая регрессия представляется через логарифм отношения шансов:

$$\log \left(\frac{P(y = 1 | x)}{1 - P(y = 1 | x)} \right) = \beta_0 + \sum_{j=1}^k \beta_j x_j$$

Каждый коэффициент β_j указывает, насколько изменяется логарифм шанса положительного исхода при увеличении x_j на единицу. При экспоненцировании получаем отношение шансов e^{β_j} — во сколько раз изменяется шанс при увеличении соответствующего признака.

ПРОВЕРКА ЗНАЧИМОСТИ КОЭФФИЦИЕНТОВ.

Для проверки гипотезы $H_0 : \beta_j = 0$ применяются статистические тесты:

- **Тест Вальда (Wald test)**: используется асимптотическая нормальность оценок $\hat{\beta}_j$. Статистика:

$$z = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim \mathcal{N}(0, 1)$$

Здесь $SE(\hat{\beta}_j)$ — стандартная ошибка оценки коэффициента. При больших выборках z сравнивается с квантилем стандартного нормального распределения. Если $|z| > z_{\alpha/2}$, нулевая гипотеза отвергается, и коэффициент считается статистически значимым.

- **Тест отношения правдоподобий (likelihood ratio test)**: сравниваются две модели — с и без x_j . Статистика:

$$G = -2 \cdot (\log L_0 - \log L_1)$$

где L_0 — логарифм функции правдоподобия модели без признака x_j , L_1 — с ним. При верной H_0 статистика G имеет распределение χ^2 с 1 степенью свободы.

Оба подхода используются в современных реализациях логистической регрессии (например, в `statsmodels` и `R glm()`), позволяя судить о статистической значимости каждого признака.

ОБУЧЕНИЕ МОДЕЛИ.

Оценка коэффициентов β_j проводится с помощью метода максимального правдоподобия:

$$L(\beta) = \prod_{i=1}^n P(y_i | x_i; \beta) = \prod_{i=1}^n \sigma(z_i)^{y_i} (1 - \sigma(z_i))^{1-y_i}$$

Вычисление оптимальных коэффициентов обычно осуществляется численными методами (например, градиентным спуском, L-BFGS и др.).

ДОПОЛНИТЕЛЬНЫЕ АСПЕКТЫ.

Логистическая регрессия поддерживает:

- регуляризацию (L1, L2) — для борьбы с переобучением и отбора признаков;
- обобщение на многоклассовый случай — через softmax-регрессию;
- возможность проведения статистических тестов для каждого признака.

Дополнительную информацию о вероятностной интерпретации и тестах значимости см. в материалах К. В. Воронцова на MachineLearning.ru [10], а также в фундаментальных статистических источниках [11].

1.5.2 Модель градиентного бустинга

Градиентный бустинг — это ансамблевый метод построения прогностических моделей, который объединяет множество слабых моделей (обычно решающих деревьев), формируя сильный предсказатель. Алгоритм работает итеративно: каждая последующая модель обучается на ошибках предыдущих, минимизируя заданную функцию потерь.

Основная идея градиентного бустинга заключается в построении модели $F(x)$, которая представляется в виде суммы базовых моделей:

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x)$$

где $h_m(x)$ — очередное решающее дерево, а γ_m — его вес. На каждом шаге алгоритм минимизирует градиент функции потерь \mathcal{L} по направлению ошибки предыдущей итерации:

$$\tilde{y}_i^{(m)} = - \left. \frac{\partial \mathcal{L}(y_i, F(x_i))}{\partial F(x_i)} \right|_{F(x_i)=F_{m-1}(x_i)}$$

ПРЕИМУЩЕСТВА ГРАДИЕНТНОГО БУСТИНГА.

Градиентный бустинг зарекомендовал себя как один из самых мощных алгоритмов машинного обучения, особенно при работе с табличными данными. Среди его преимуществ:

- **Высокая точность:** показывает конкурентные результаты на большинстве задач классификации и регрессии, особенно на Kaggle-соревнованиях [12].
- **Гибкость:** позволяет использовать различные функции потерь, включая логистическую, квадрат ошибки, гингубель и др.
- **Устойчивость к выбросам:** слабые модели (например, деревья с ограниченной глубиной) минимизируют переобучение.
- **Поддержка категориальных признаков и регуляризации (в XGBoost, CatBoost).**
- **Обработка пропусков и автоматическая настройка важности признаков.**

Популярные реализации включают XGBoost [12], LightGBM [13] и CatBoost, каждая из которых оптимизирована для высокой производительности и больших объёмов данных.

1.5.3 Интерпретация моделей машинного обучения с помощью SHAP

Современные модели градиентного бустинга обеспечивают высокую точность предсказаний, но страдают от недостатка интерпретируемости. Для преодоления этой проблемы применяется метод SHAP (SHapley Additive exPlanations), разработанный на основе теории кооперативных игр.

SHAP позволяет количественно оценить вклад каждого признака в конкретное предсказание модели. Метод опирается на концепцию значений Шепли — справедливого распределения выигрыша между участниками коалиционной игры.

ФОРМУЛА SHAP-ЗНАЧЕНИЙ.

Для заданной модели f и объекта с признаками $x = (x_1, x_2, \dots, x_M)$ предсказание модели можно разложить в виде суммы:

$$f(x) = \phi_0 + \sum_{j=1}^M \phi_j$$

где:

- ϕ_0 — базовое значение (обычно это среднее значение предсказания по обучающей выборке),
- ϕ_j — вклад признака x_j в отличие текущего предсказания от базового.

SHAP-значение ϕ_j для признака x_j рассчитывается как:

$$\phi_j = \sum_{S \subseteq \{1, \dots, M\} \setminus \{j\}} \frac{|S|! \cdot (M - |S| - 1)!}{M!} \cdot [f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)]$$

Здесь:

- S — подмножество признаков, не включающее j ,
- $f_S(x_S)$ — предсказание модели, обученной только на признаках из S ,
- $f_{S \cup \{j\}}(x_{S \cup \{j\}})$ — предсказание модели с добавленным x_j .

Вычисление формулы в лоб требует экспоненциального числа вызовов модели. Однако для деревьев реализованы эффективные алгоритмы расчёта SHAP-значений за полиномиальное время [15].

ОСОБЕННОСТИ SHAP В CATBOOST.

В библиотеке CatBoost реализован оптимизированный алгоритм расчёта SHAP-значений на основе внутренней структуры модели. Он учитывает

разбиения, категориальные признаки и симметрию деревьев. Базовое значение ϕ_0 задаётся как среднее предсказание модели:

$$\phi_0 = \frac{1}{n} \sum_{i=1}^n f(x_i)$$

Вклад каждого признака ϕ_j можно интерпретировать как изменение в предсказании относительно ϕ_0 при учёте значения x_j . Если $\phi_j > 0$, признак увеличивает предсказание модели, и наоборот.

ПРЕИМУЩЕСТВА SHAP:

- **Аддитивность:** предсказание модели точно разлагается в сумму вкладов признаков.
- **Универсальность:** SHAP применим ко всем моделям, включая ансамбли, нейросети, линейные модели.
- **Глобальная и локальная интерпретация:** возможен анализ как по всей выборке, так и для отдельного объекта.
- **Справедливость:** SHAP удовлетворяет аксиомам симметрии, эффективности и нулевого вклада.

SHAP активно используется в задачах, где требуется объяснимость решений: в здравоохранении, кредитном скоринге, судебной аналитике и рекомендательных системах.

2 ПРАКТИЧЕСКИЙ АНАЛИЗ ВЛИЯНИЯ ОПЕРАЦИОННЫХ МЕТРИК НА УРОВЕНЬ CSAT

2.1 Подготовка и описание данных

2.1.1 Описание используемой витрины данных

2.1.2 Операционные и неоперационные признаки в исследовании

2.2 Исследование данных с использованием статистических критериев

2.2.1 Анализ зависимости CSAT от времени жизни таска

2.2.2 Оценка необходимости проведения A/B теста

2.3 Построение и анализ моделей машинного обучения

2.3.1 Построение логистической регрессии и анализ ее коэффициентов

2.3.2 Применение модели CatBoost и анализ важности признаков с помощью SHAP

2.3.3 Анализ влияния различных факторов на оценки пользователей

3 РЕЗУЛЬТАТЫ И ВЫВОДЫ ИССЛЕДОВАНИЯ

3.1 Оценка эффективности сокращения времени выполнения задач

3.2 Интерпретация и практические рекомендации на основе моделей

3 ЗАКЛЮЧЕНИЕ

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1/2), 81–93. <https://doi.org/10.2307/2332226>
2. Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London A*, 187, 253–318. <https://doi.org/10.1098/rsta.1896.0007>
3. Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72–101. <https://doi.org/10.2307/1412159>
4. Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, 65(1), 23–28. <https://doi.org/10.1007/BF02294183>
5. Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a small sample. *Metron*, 1, 3–32. Reprinted in *Collected Papers of R. A. Fisher*.
6. Zimmerman, D. W., & Zumbo, B. D. (1997). Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *Journal of Experimental Education*, 65(1), 71–86. <https://doi.org/10.1080/00220973.1996.9943798>
7. Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Statistical Science*, 16(2), 101–133. <https://doi.org/10.1214/ss/1009213286>
8. Clopper, C. J., & Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4), 404–413. <https://doi.org/10.1093/biomet/26.4.404>

9. Agresti, A., & Coull, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *The American Statistician*, 52(2), 119–126. <https://doi.org/10.1080/00031305.1998.10480550>
10. Воронцов, К. В. (2010). Логистическая регрессия: вероятностная модель бинарной классификации. *MachineLearning.ru*.
URL: http://www.machinelearning.ru/wiki/index.php?title=Logistic_regression
11. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. John Wiley & Sons. <https://doi.org/10.1002/9781118548387>
12. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). <https://doi.org/10.1145/2939672.2939785>
13. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30. https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html
14. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
15. Lundberg, S. M., Erion, G., & Lee, S.-I. (2018). Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv preprint arXiv:1802.03888*. <https://arxiv.org/abs/1802.03888>
16. CatBoost Documentation — SHAP Values. URL: <https://catboost.ai/docs/en/concepts/shap-values>

ПРИЛОЖЕНИЕ А