



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Московский институт электроники и математики им. А.Н. Тихонова

Проект #2101, Изучение временных рядов рейтингов компаний, занимающихся экономической деятельностью

Исполнители:

Липатов Данила Вячеславович,
МСМТ243

Руководитель, инициатор:
Зотов Леонид Валентинович

Консультант:
Игнатовская Валерия
Анатольевна

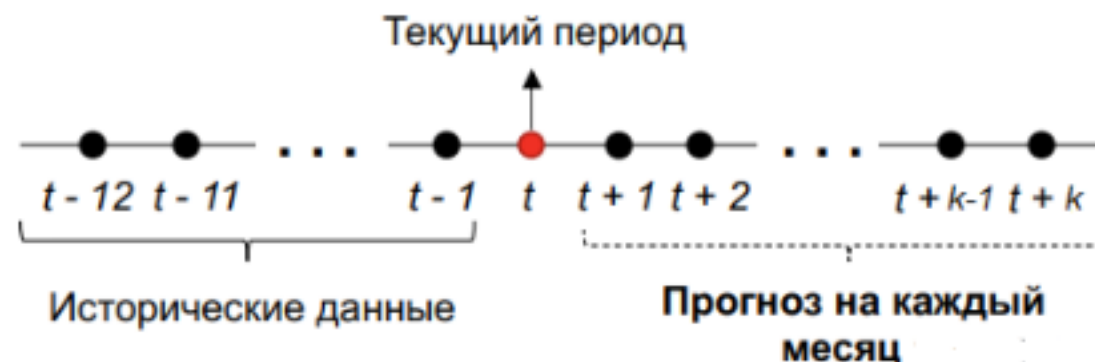
Москва, 2025

Суть проекта

Мотивация:

- Существует классическая процедура оценки финансовой устойчивости компаний, именуемая кредитным рейтингованием.
- Кредитные рейтинги широко используются инвесторами, кредиторами и другими участниками рынка для оценки вероятности исполнения обязательств компаний. Однако, учитывая изменчивость экономической среды, задача прогнозирования изменений рейтингов становится всё более актуальной
- Разработка подходов к предсказанию кредитных рейтингов компаний на будущее позволяет не только повысить точность оценки их финансовой устойчивости, но и принять своевременные меры для минимизации рисков и оптимизации инвестиционных стратегий.

Основная идея: Важно уметь предсказывать дальнейшие состояния кредитного рейтинга компании (объекта) на основе данных о его поведении.





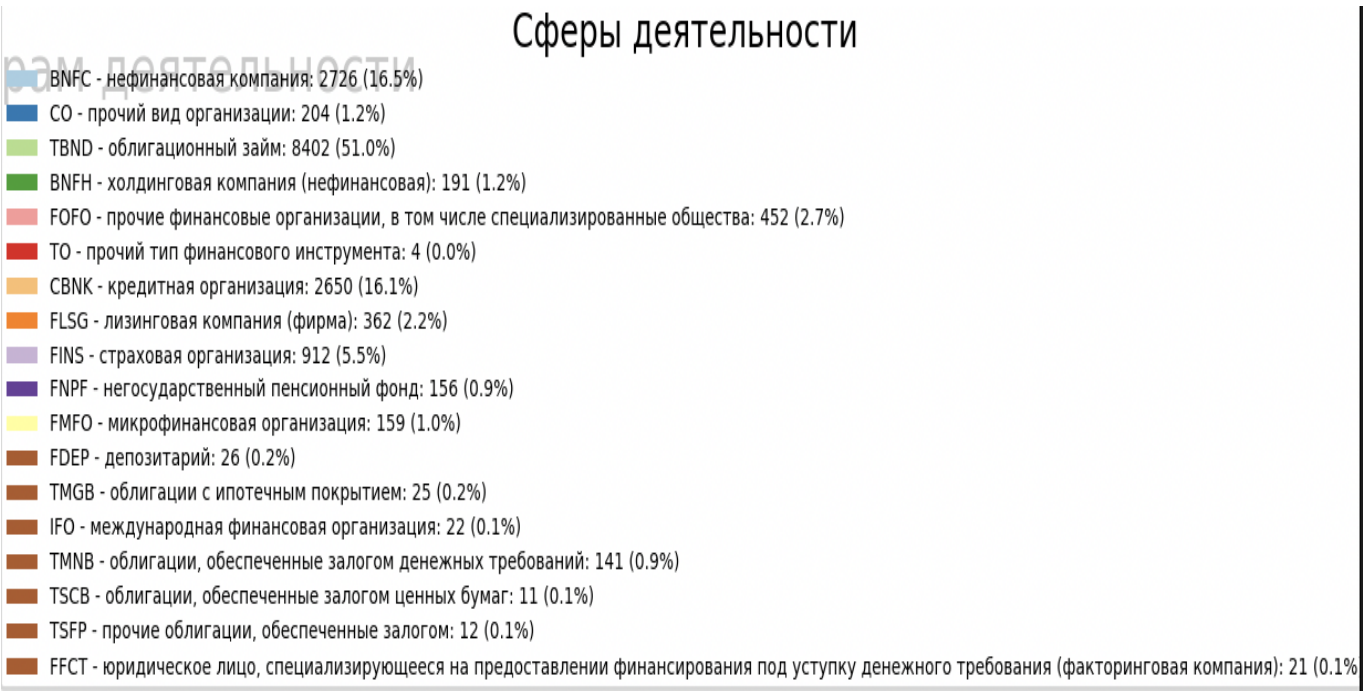
Цель и задачи проекта

Цель проекта: предложить алгоритм прогнозирования кредитного рейтинга объекта на несколько будущих периодов

Задачи проекта:

- изучить возможные состояния кредитных рейтингов и переходы между ними
- провести сбор данных из официальных источников
- получить качественное описание собранных данных

Данные открытого репозитория



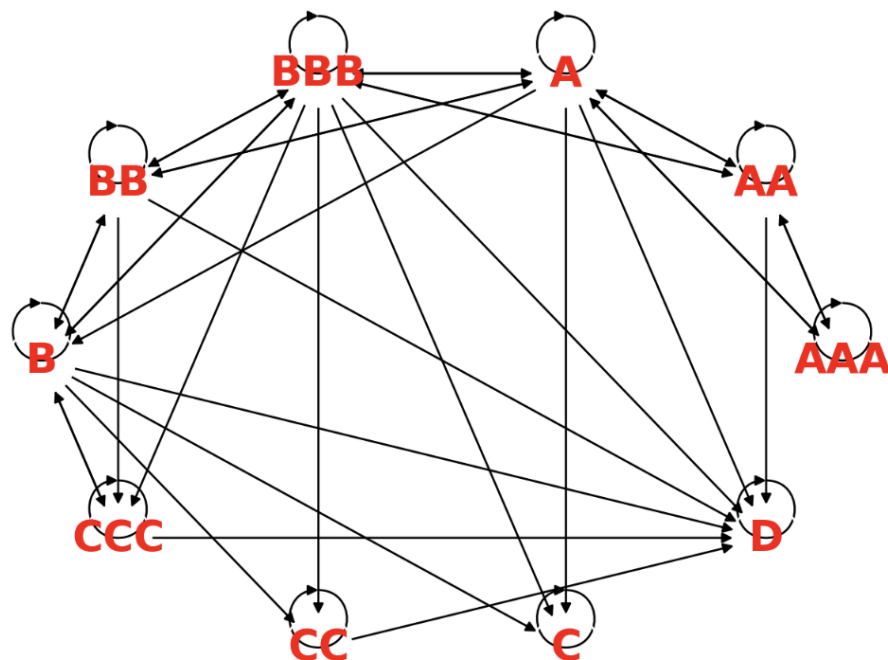
Все рейтинги, присвоенные объекту

Значение кредитного рейтинга	Наименование КРА	Дата публикации кредитного рейтинга
ruAA-	АО "Эксперт РА"	17.12.2024
AA-(RU)	АКРА (АО)	19.04.2024
AA-(RU)	АКРА (АО)	18.05.2023
AA-(RU)	АКРА (АО)	14.03.2023
AA-(RU)	АКРА (АО)	27.05.2022
AA-(RU)	АКРА (АО)	07.06.2021
A+(RU)	АКРА (АО)	16.06.2020
A+(RU)	АКРА (АО)	24.06.2019
A+(RU)	АКРА (АО)	28.06.2018
A+(RU)	АКРА (АО)	28.06.2017

Цель и задачи проекта

- изучить и применить марковские модели и модели классификации с помощью моделирования в Python, на доступных данных
- провести тестирование, оценить их качество
- сравнить результаты
- выявить сезонность (при существовании)

Пример переходов между состояниями



Основные уровни категорий (AAA — D):

1.AAA:

Наивысший рейтинг. Указывает на исключительную надежность и стабильность эмитента. Риски дефолта практически отсутствуют.

2.AA:

Очень высокая кредитоспособность. Риски минимальны, но чуть выше по сравнению с AAA.

3.A:

Высокая способность выполнять финансовые обязательства. Возможны незначительные риски при изменении экономических условий.

4.BBB:

Умеренная кредитоспособность. Эмитент достаточно стабилен, но более чувствителен к неблагоприятным условиям.

5.BB:

Спекулятивный уровень. Существует вероятность финансовых трудностей в случае ухудшения внешней среды.

6.B:

Повышенный риск дефолта. Эмитент зависит от благоприятной экономической конъюнктуры.

7.CCC:

Высокий риск дефолта. Финансовое состояние нестабильно, возможна реструктуризация долгов.

8.CC:

Очень высокая вероятность дефолта. Эмитент практически неспособен выполнять свои обязательства.

9.C:

На грани дефолта. Финансовые обязательства могут выполняться лишь частично.

10.D:

Дефолт. Эмитент не выполняет свои финансовые обязательства.

Математическая модель

$\mathbb{Y} = \{AAA, AA+, AA, \dots, D\}$ или $\{1, 2, 3, \dots, 20\}$ – множество состояний

$\mathbb{X} = \{X_1, X_2, \dots, X_m\}$ – множество уникальных объектов для которых определены цепочки рейтингов (их изменения во времени)

$X_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ – цепочка состояний рейтинга во времени для i -го объекта

где $x_{it} \in \mathbb{Y}$ – рейтинг i -ого объекта в момент времени t

(для упрощения все показатели были отсортированы от максимального к минимальному и переведены в ранг)

Алгоритм прогнозирования следующего состояния цепи

$$h_l: \mathbb{X} \rightarrow \mathbb{Y}, \quad h_l(x) = y,$$

где $l = \{1, 2, 3, \dots\}$ соответствует сфере деятельности объекта (см. Приложение)

Итоговая задача

Построить алгоритм $a = h_l$, который минимизирует некоторый функционал

$$Q(a, X) \rightarrow \min,$$

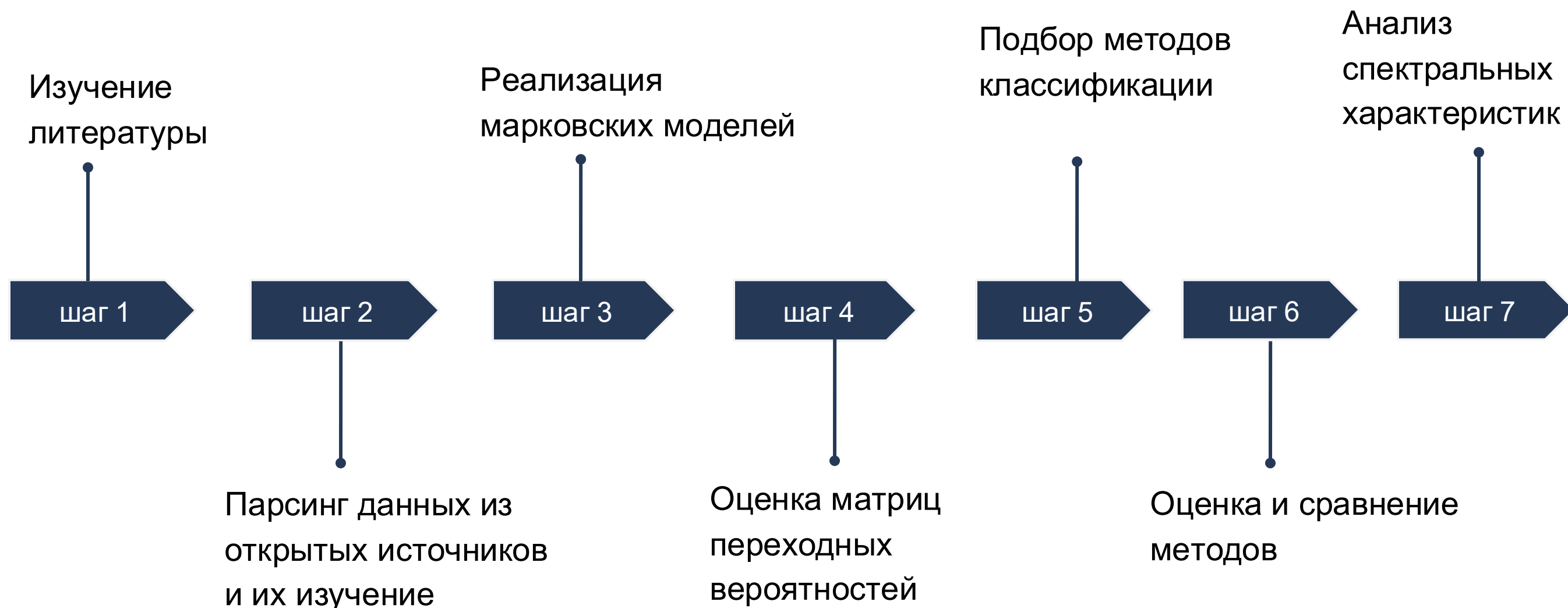
например, минимизирует ошибку прогноза по всем компаниям.

Ожидаемые результаты и этапы реализации

Ожидаемый результат:

Методология (алгоритм) для прогнозирования состояний кредитного рейтинга на несколько периодов для любого набора данных.

Основные этапы:



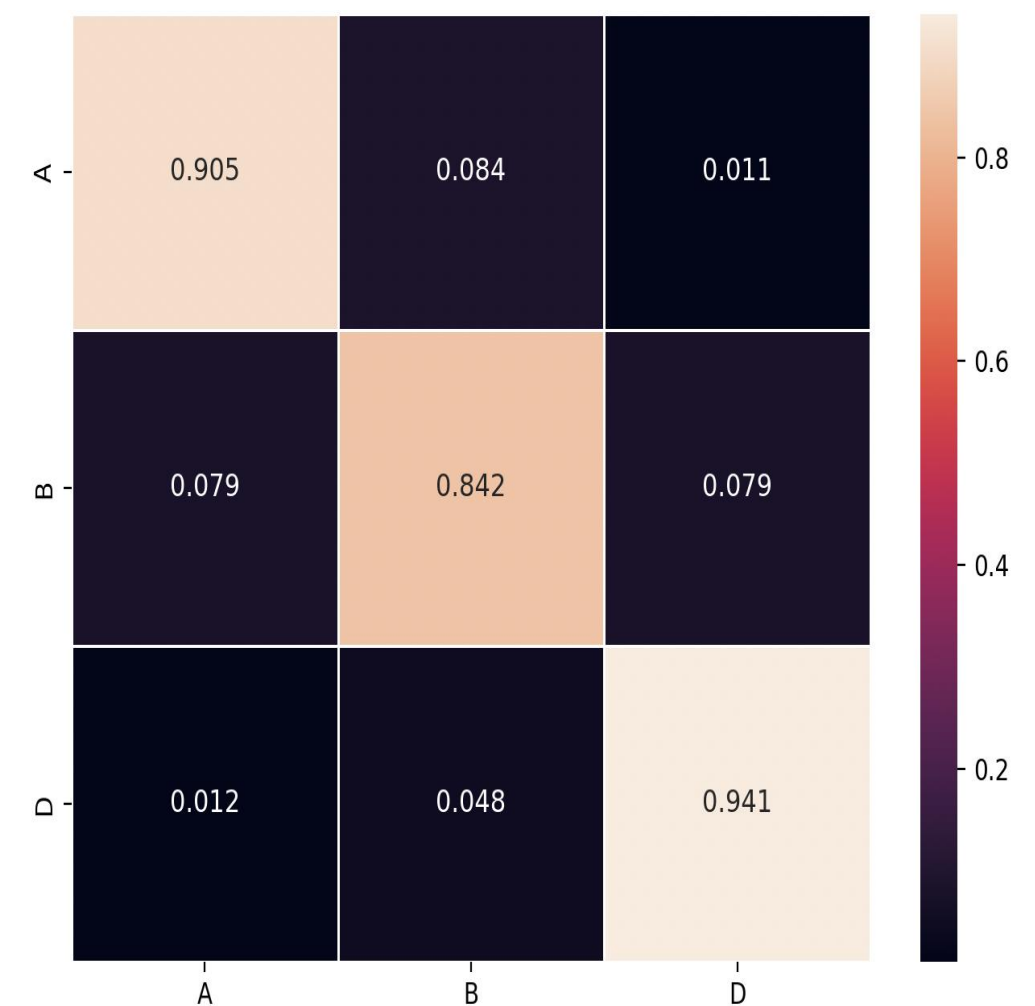
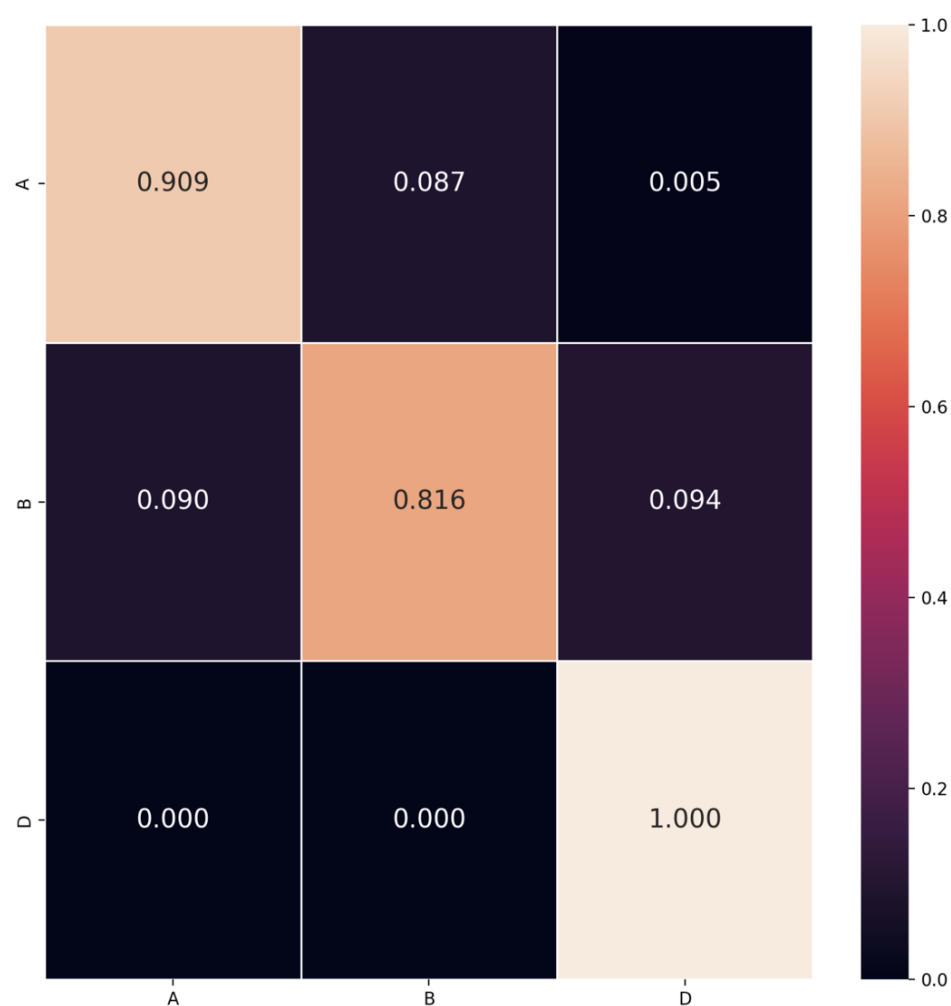
Марковская цепь 1-ого порядка

Для марковской цепи с непрерывным временем (или метод дюрации) используется очень удобная формула для оценки элементов матрицы Q методом максимального правдоподобия (MLE)

$$\hat{\lambda}_{i,j} = \frac{N_{i,j}(T)}{\int_0^T Y_i(s) ds}$$

Использование байесовского подхода с априорным распределением Дирихле позволяет гибко моделировать переходы между рейтингами с учетом наблюдаемых данных.

$$\hat{p}_{i,j} = E[p_{i,j}|N_{i,j}] = \frac{N_{i,j} + a_{i,j}}{\sum_{k=1}^K (N_{i,k} + a_{i,k})}$$



Марковская цепь 2-ого порядка

В некоторых случаях также полезно рассматривать процесс как на цепь Маркова 2-ого порядка — процесс, удовлетворяющий условию

$$P(X_t = j | X_s = i, X_{s-1} = i_{s-1}, \dots, X_1 = i_1) = P(X_t = j | X_s = i, X_{s-1} = i_{s-1})$$

Чтобы реализовать это, в качестве пространства состояний S нужно взять упорядоченные 2-кортежи значений S и рассмотреть процесс на этом пространстве как цепь Маркова 1-ого порядка. При этом множество состояний цепи Маркова сильно возрастает, для случая с 3 состояниями:

$$S = \{AA, AB, AD, BA, BB, BD, DA, DB, DD\}$$

Однако есть ряд недостатков — переходы $DB \rightarrow BD$ из экспертных оценок не могут быть настолько высокими, как в данном примере. Поэтому либо эти состояния надо убирать, либо делать самопереход = 1



Модель классификации (случайный лес)

Алгоритм строит ансамбль из T решающих деревьев:

$$\mathcal{F} = \{h_t(x)\}_{t=1}^T, h_t: \mathbb{R}^d \rightarrow Y, t \in \{1, 2, \dots, T\}$$

Каждое дерево h_t обучается на случайной подвыборке $\mathcal{D}_t \subseteq \mathcal{D}$ полученной методом бутстрапа (случайной выборкой с возвращением).

При построении каждого дерева в каждом узле случайным образом выбирается подмножество признаков

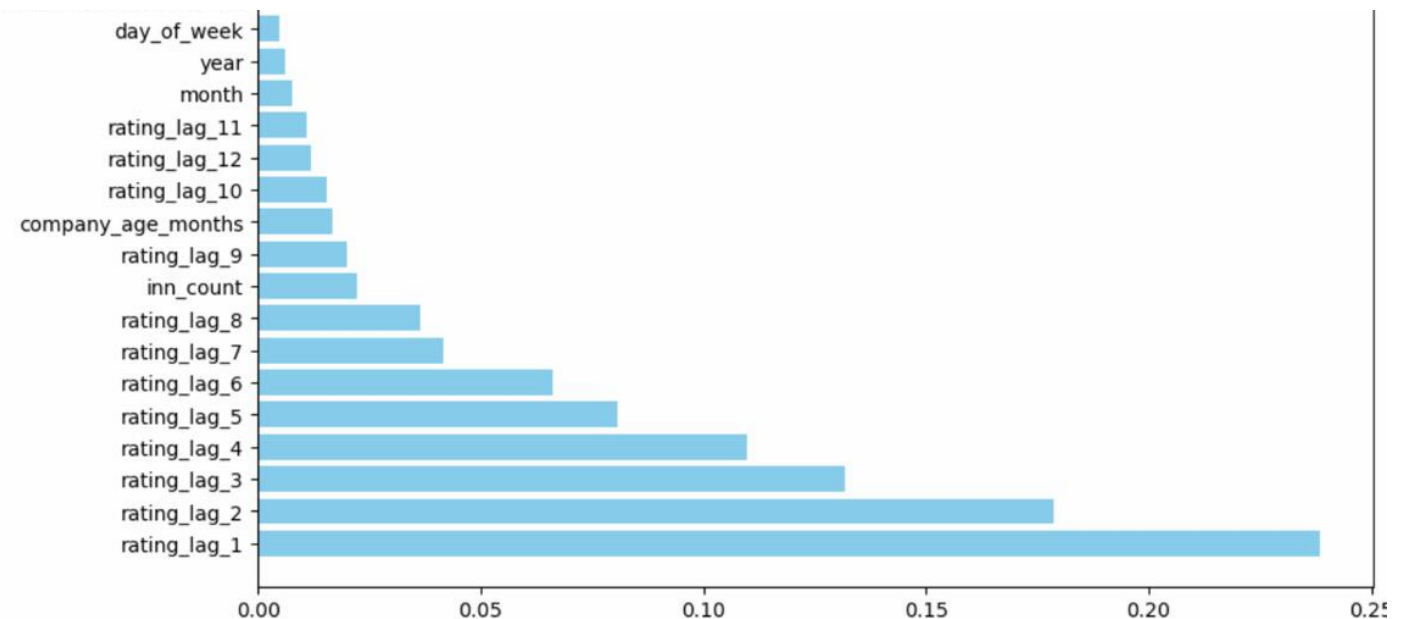
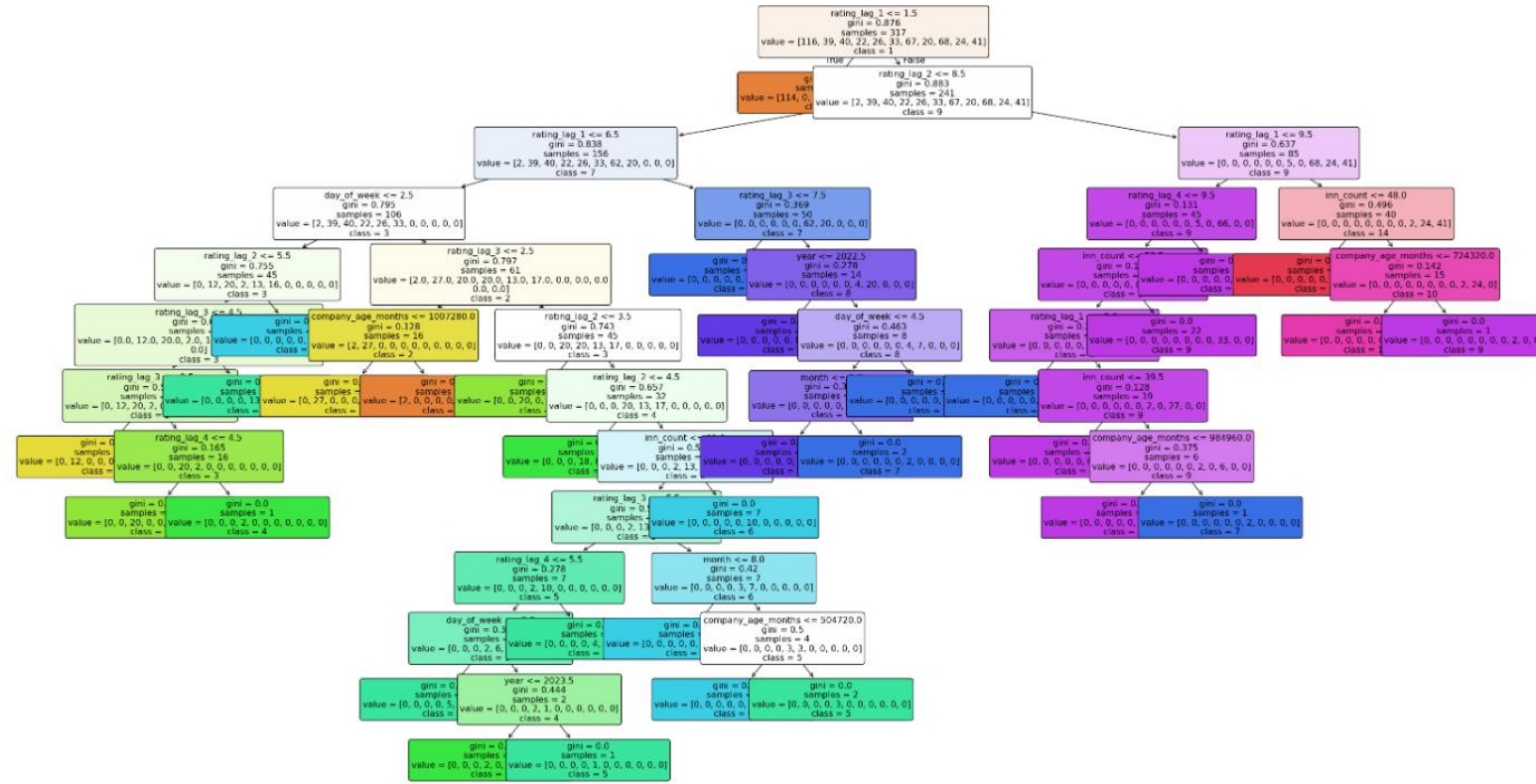
$$\mathcal{F}_m \subset \{1, 2, \dots, d\}, |\mathcal{F}_m| = m \ll d$$

Выбирается оптимальное разбиение пространства:

$$(j^*, r^*) = \arg \min_{j \in \mathcal{F}_m, r \in \mathbb{R}} \mathcal{I}(S_{left}, S_{right})$$

$$S_{left} = \{(x_i, y_i) \in S | x_i^j \leq r\}$$

$$S_{right} = \{(x_i, y_i) \in S | x_i^j > r\}$$



Спектральные характеристики / матрица кросс-корреляции

Параметр Херста показывает:

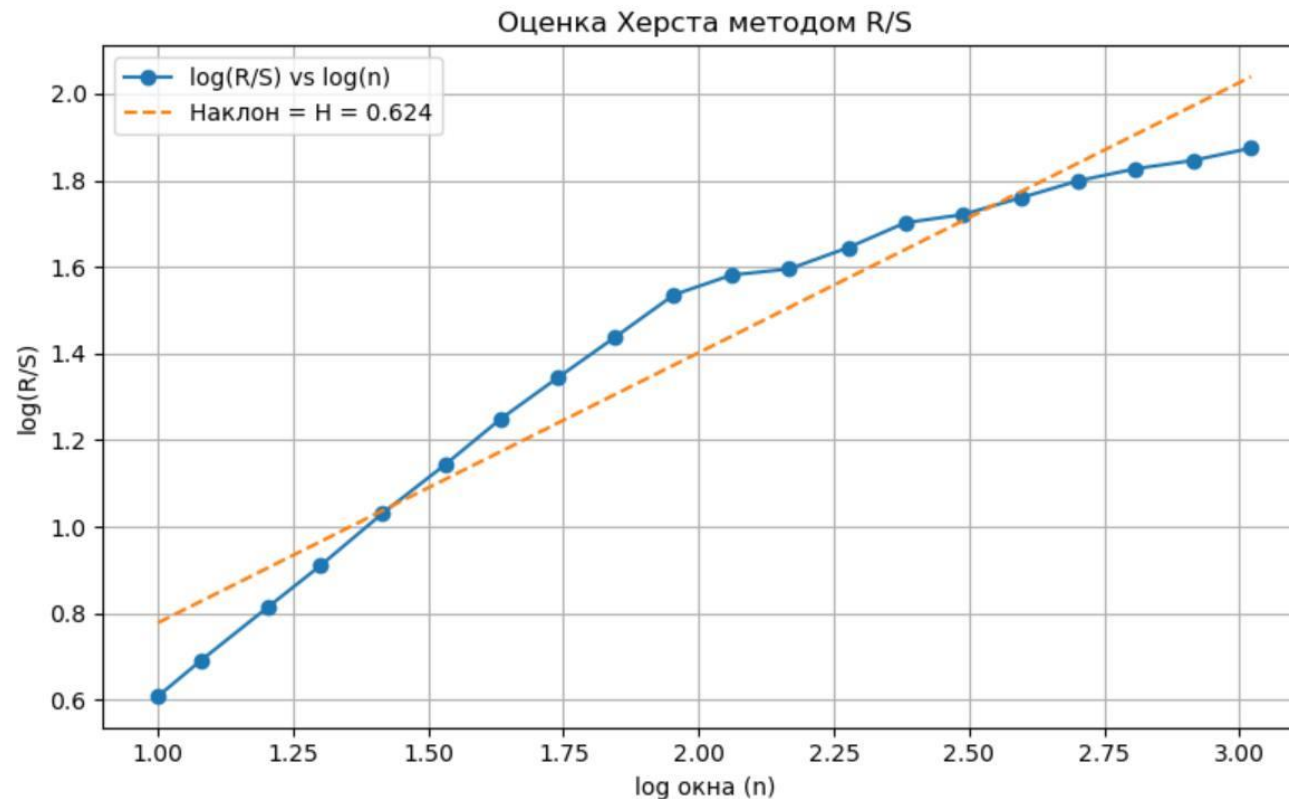
- $H = 0.5$: поведение похоже на случайное блуждание;
- $H > 0.5$: персистентное поведение — рост (или падение) рейтинга, скорее всего, продолжится;
- $H < 0.5$: антиперсистентное поведение — рост сменяется падением и наоборот.

Временной ряд разбивается на $\left[\frac{T}{n}\right]$ непересекающихся сегментов длины n . Для каждого сегмента вычисляется:

1. Среднее значение: $\bar{r}_n = \frac{\sum_{t=1}^n r_t}{n}$
2. Кумулятивные отклонения: $X_t = \sum_{k=1}^t r_k - \bar{r}_n$
3. Размах: $R(n) = \max_{1 \leq t \leq n} X_t - \min_{1 \leq t \leq n} X_t$
4. Нормализованный размах: $\frac{R(n)}{S(n)}$

Далее из уравнения находится параметр Херста:

$$\log\left(\frac{R(n)}{S(n)}\right) = H \cdot \log(n) + \log(c)$$



Спектральные характеристики / матрица кросс-корреляции

Параметр Херста показывает:

- $H = 0.5$: поведение похоже на случайное блуждание;
- $H > 0.5$: персистентное поведение — рост (или падение) рейтинга, скорее всего, продолжится;
- $H < 0.5$: антиперсистентное поведение — рост сменяется падением и наоборот.

Предполагается, что спектр $S(f)$ подчиняется степенному закону:

$$S(f) \sim \frac{1}{f^\beta}, f \rightarrow 0$$

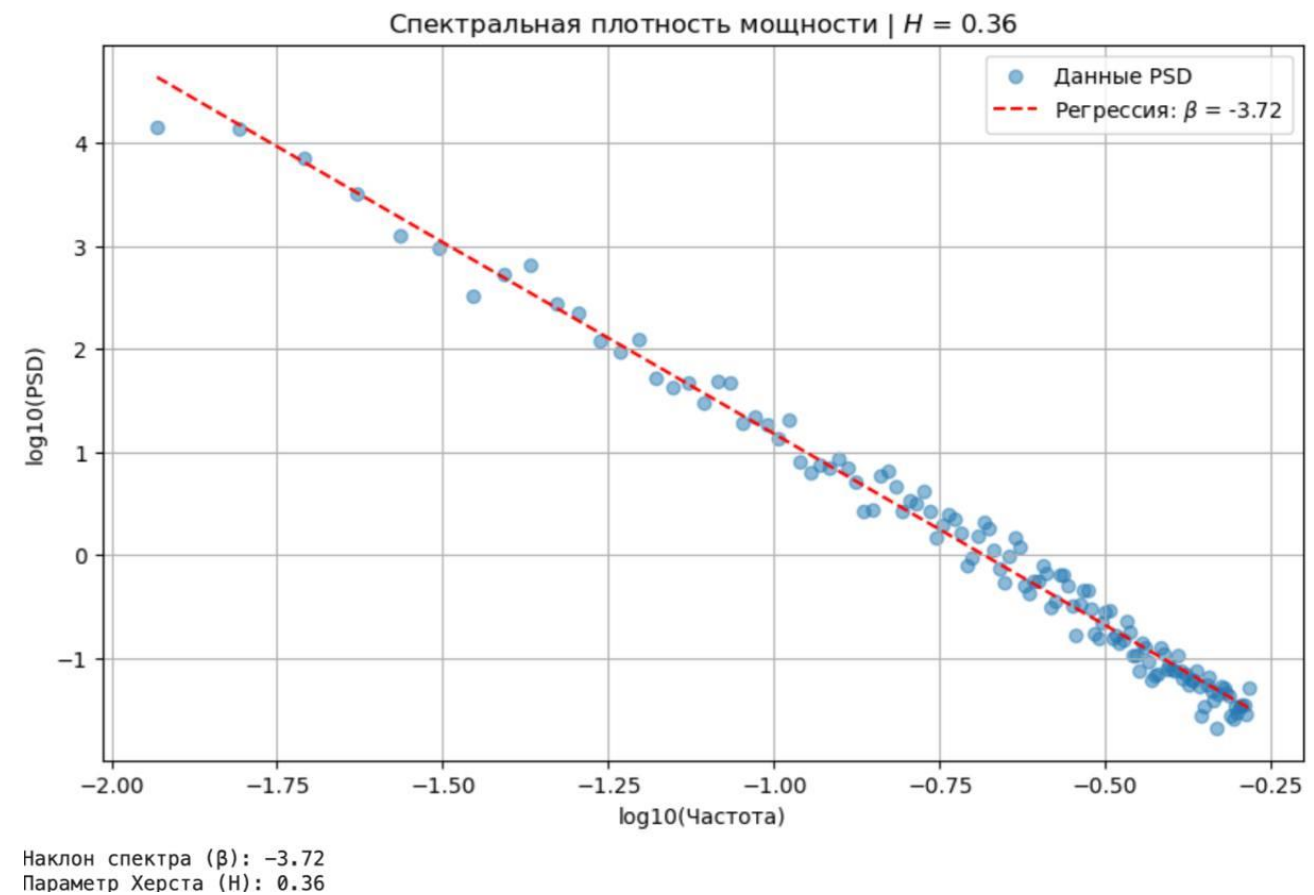
где

$S(f)$ - мощность сигнала на частоте f ,

β - наклон в логарифмической шкале.

Параметр Херста при этом выражается через наклон β следующим образом:

$$H = \frac{1 + \beta}{2}$$



Спектральные характеристики / матрица кросс-корреляции

Для анализа линейных связей между этими рядами строится матрица кросс-корреляции

$$C \in \mathbb{R}^{N \times N}$$

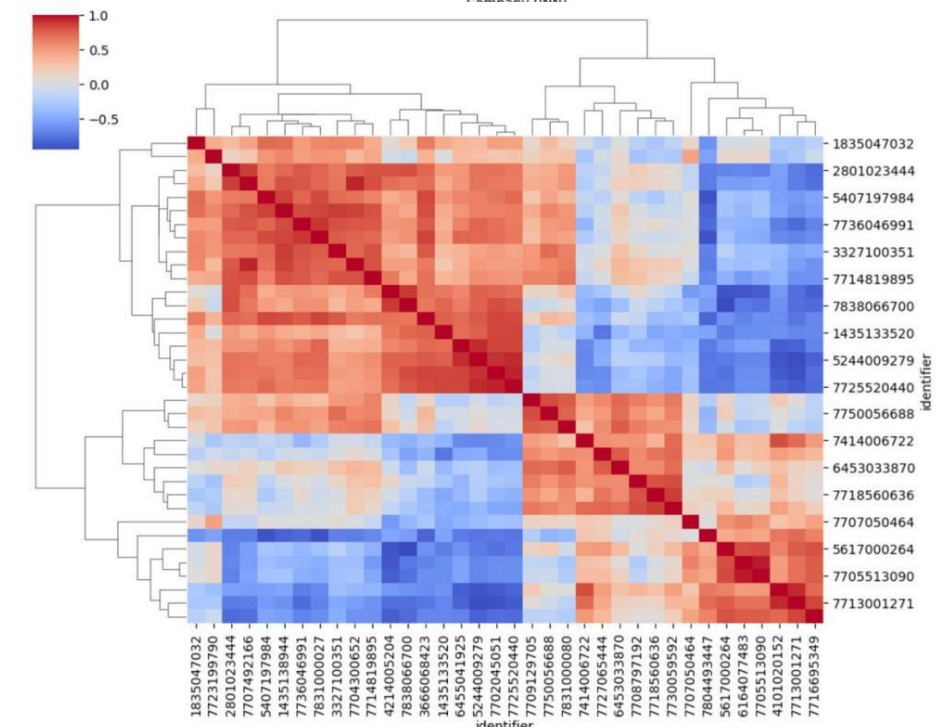
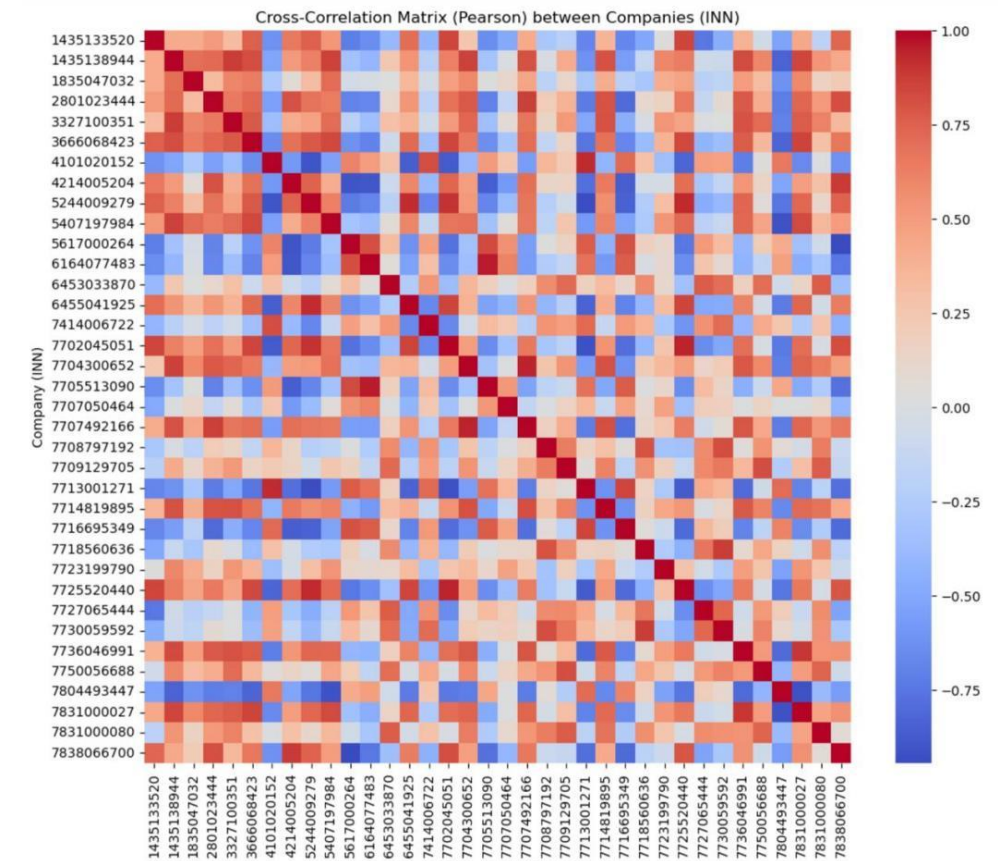
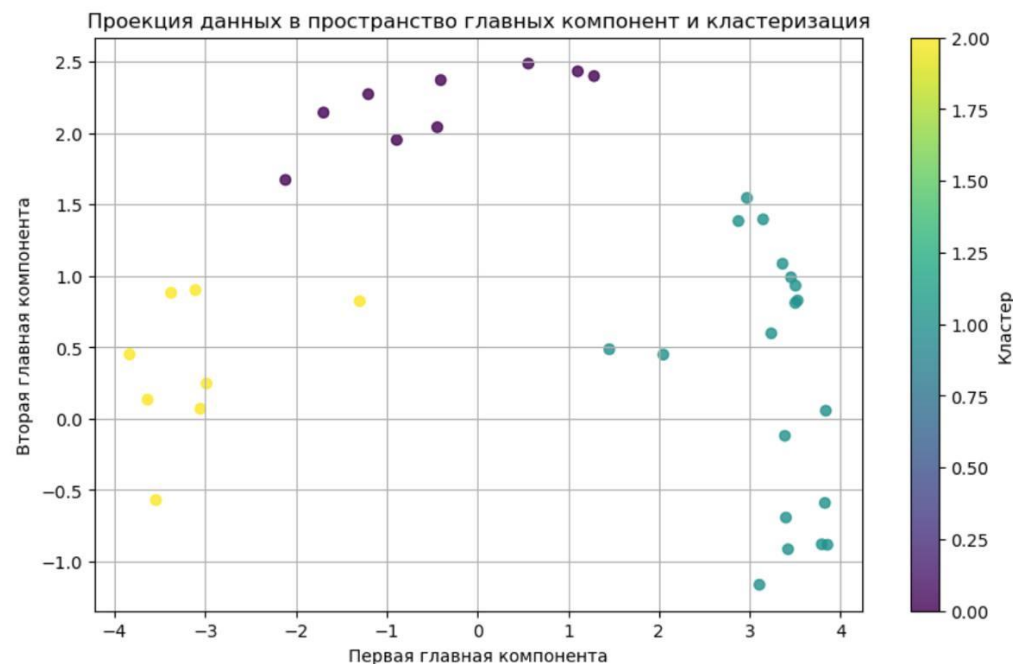
Где каждый элемент $C_{i,j}$ определяется как нормированное ковариационное выражение:

$$C_{i,j} = \frac{1}{T} \sum_{t=1}^T \frac{(r_t^i - \bar{r}^i)(r_t^j - \bar{r}^j)}{\sigma^i \sigma^j}$$

Данная матрица C будет либо ту же размерность, сколько N цепочек мы имеем, либо меньше, в зависимости от разнообразности данных и кол-ва переходов внутри цепочек.

Сингулярное разложение (SVD) для матрицы C имеет вид:

$$C = U \Sigma V^T$$



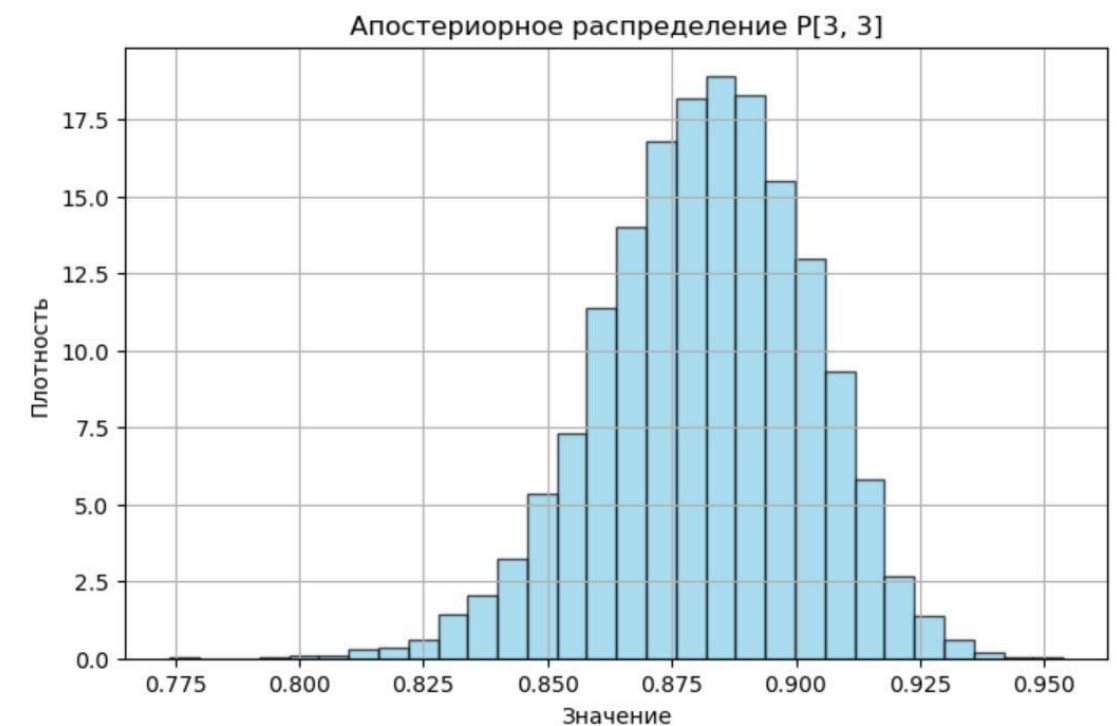
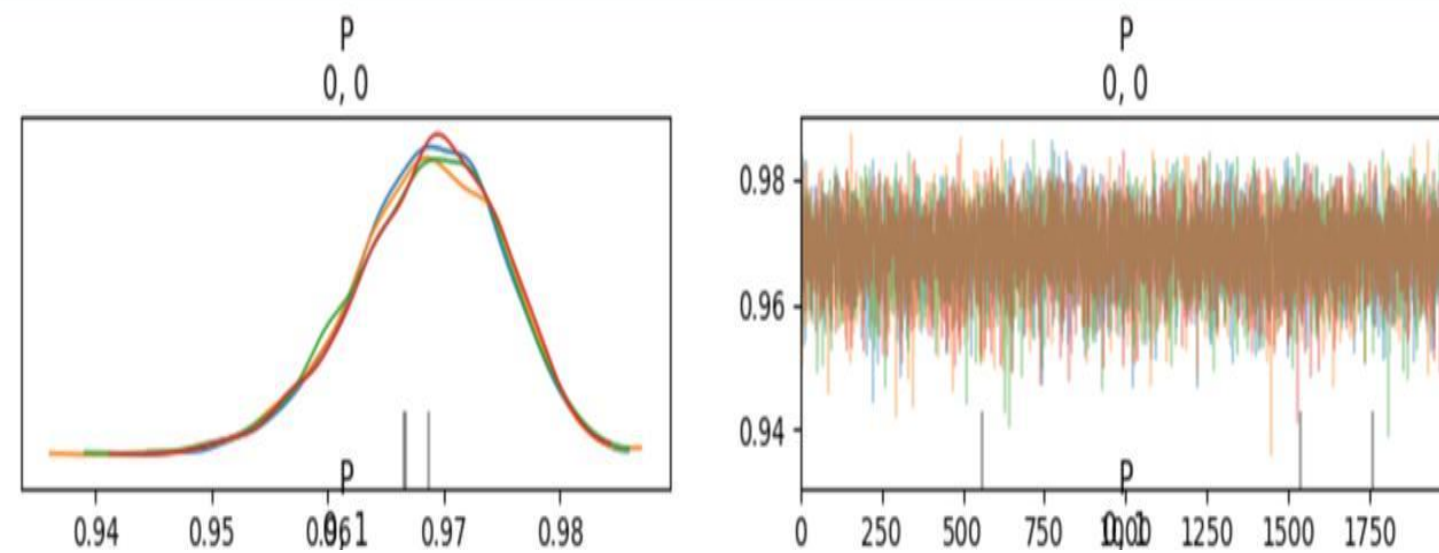
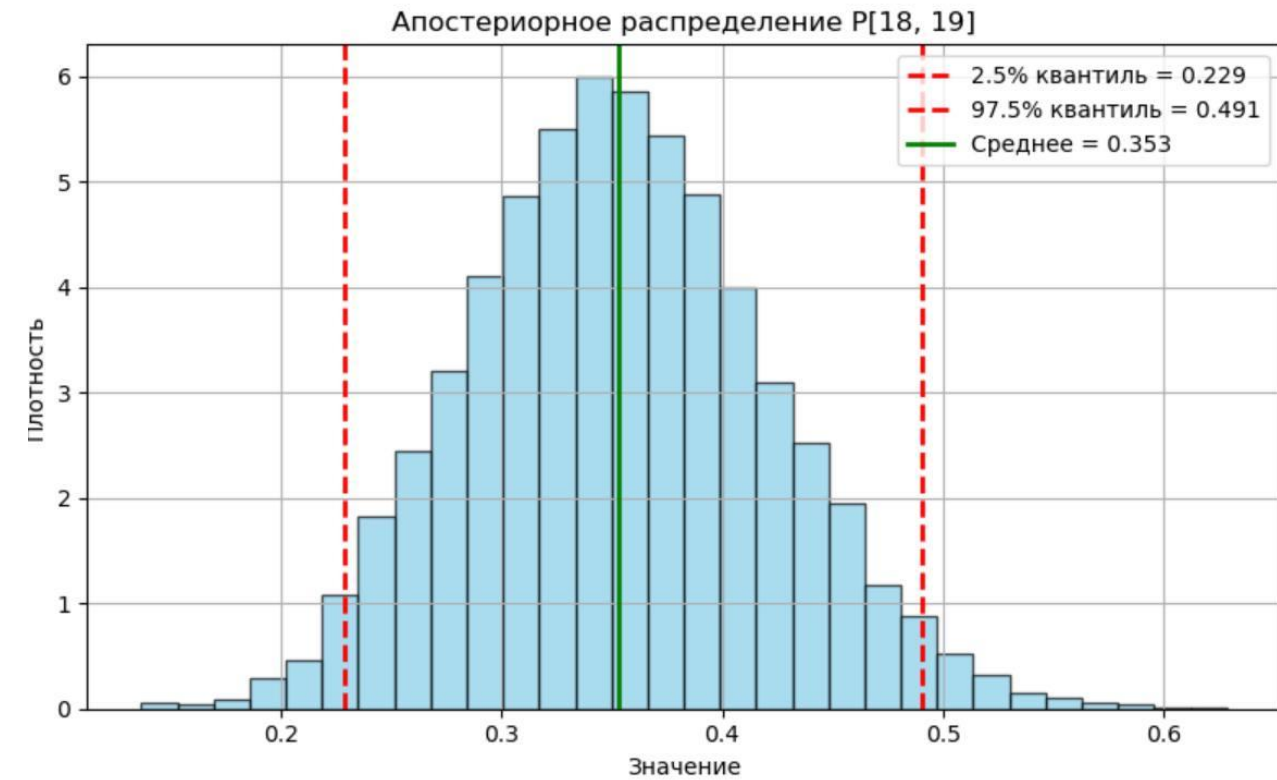
Марковская цепь Монте-Карло (МСМС)

Задача: сгенерировать выборку из апостериорного распределения:

$$p(\theta|r_{1:t}) = \frac{p(r_{1:T}|\theta)p(\theta)}{p(r_{1:T})}$$

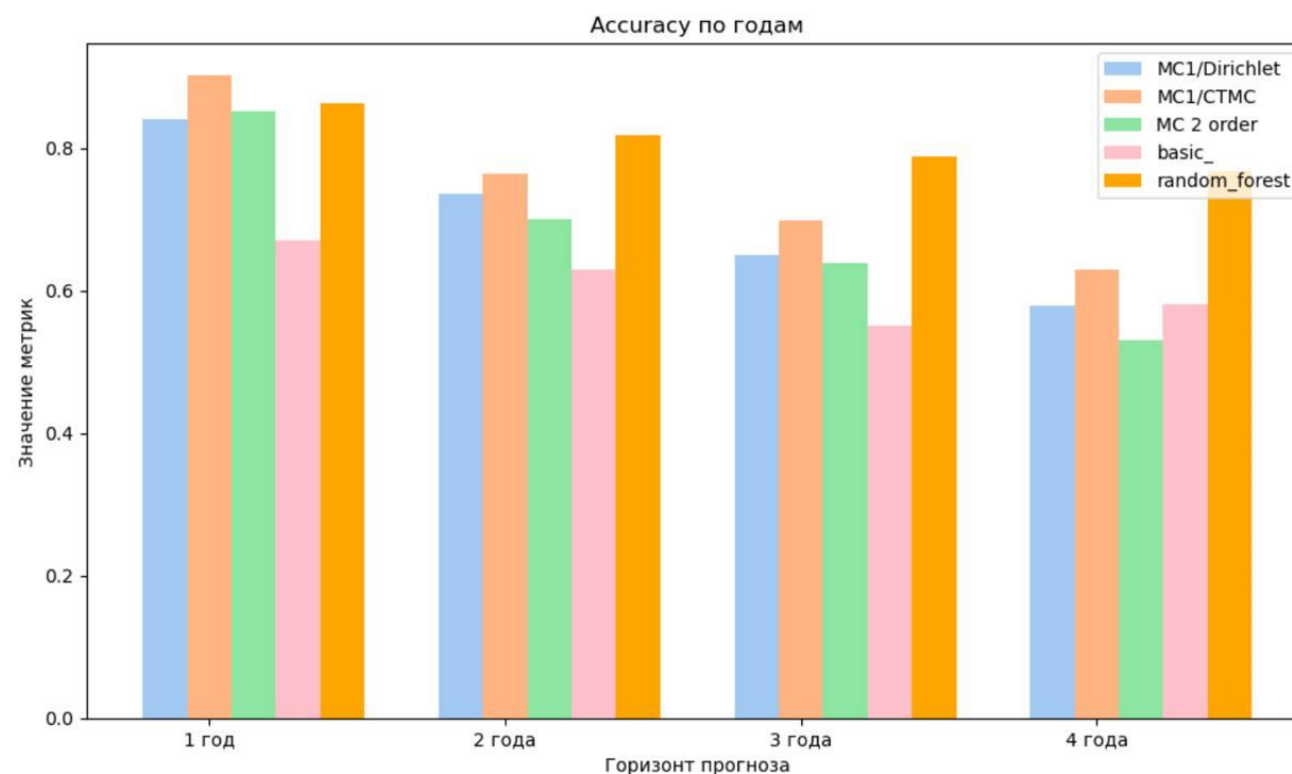
Вместо точного вычисления распределения $p(\theta|r_{1:T})$ метод МСМС строит марковскую цепь $\{\theta^n\}_{n=1}^N$, стационарное распределение которой совпадает с нужным $p(\theta|r_{1:T})$.

После «разогрева» (burn-in) цепь начинает производить зависимые, но корректно распределенные выборки из апостериорного распределения.

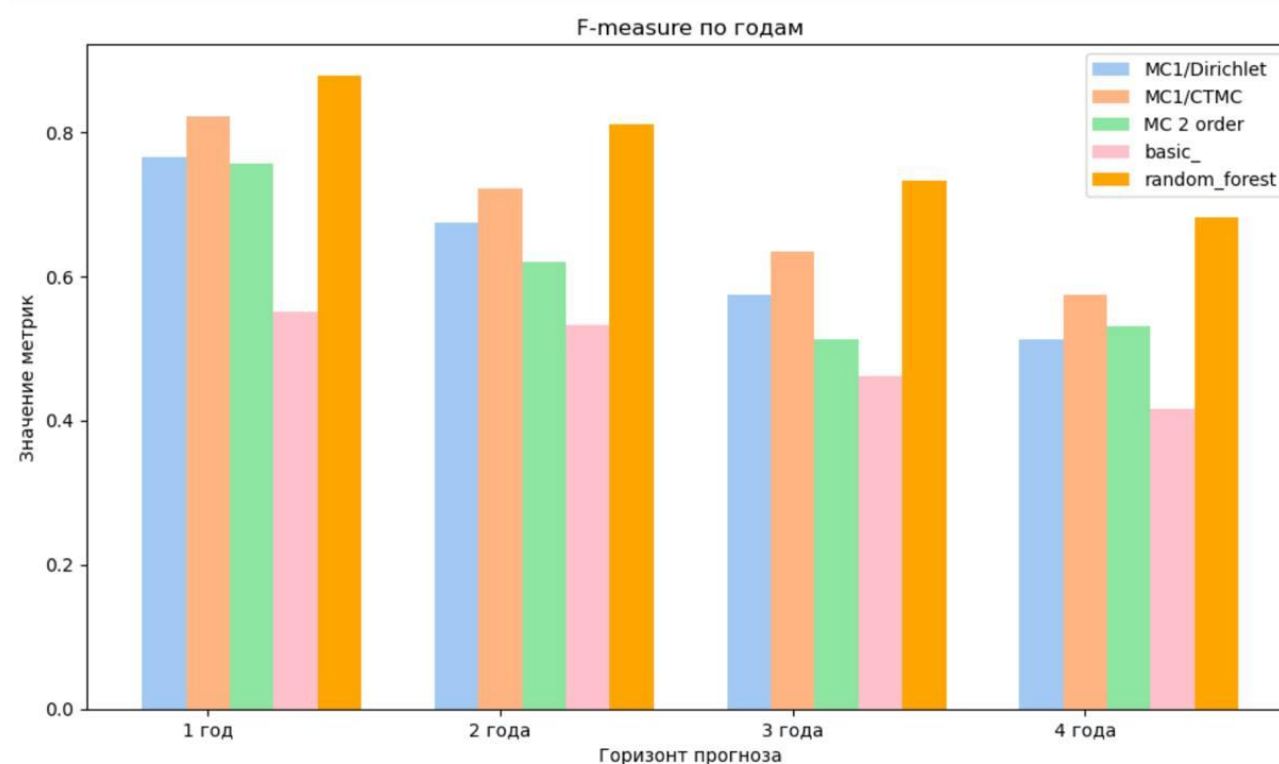




Результаты на метриках классификации



Метод	1 год	4 год	Δ_abs	Δ_rel (%)
Dirichlet	0.8406	0.5795	0.2611	31.06%
CTMC	0.9015	0.6296	0.2719	30.17%
2-й порядок	0.8520	0.5304	0.3216	37.73%
Basic	0.6706	0.5817	0.0889	13.26%
Random Forest	0.8628	0.7676	0.0952	11.04%

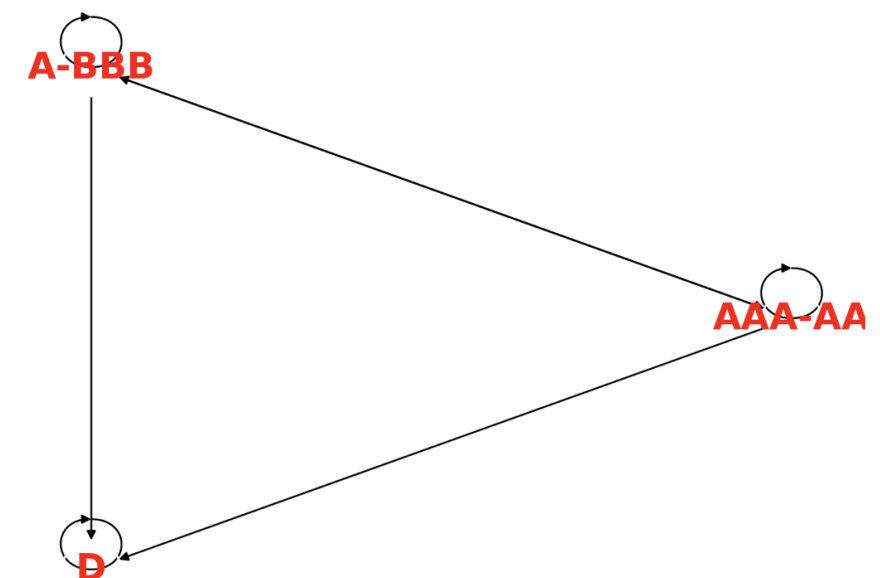
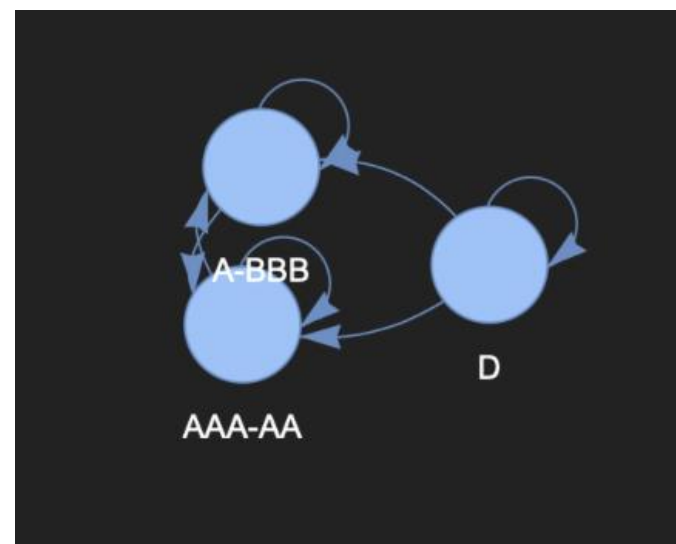
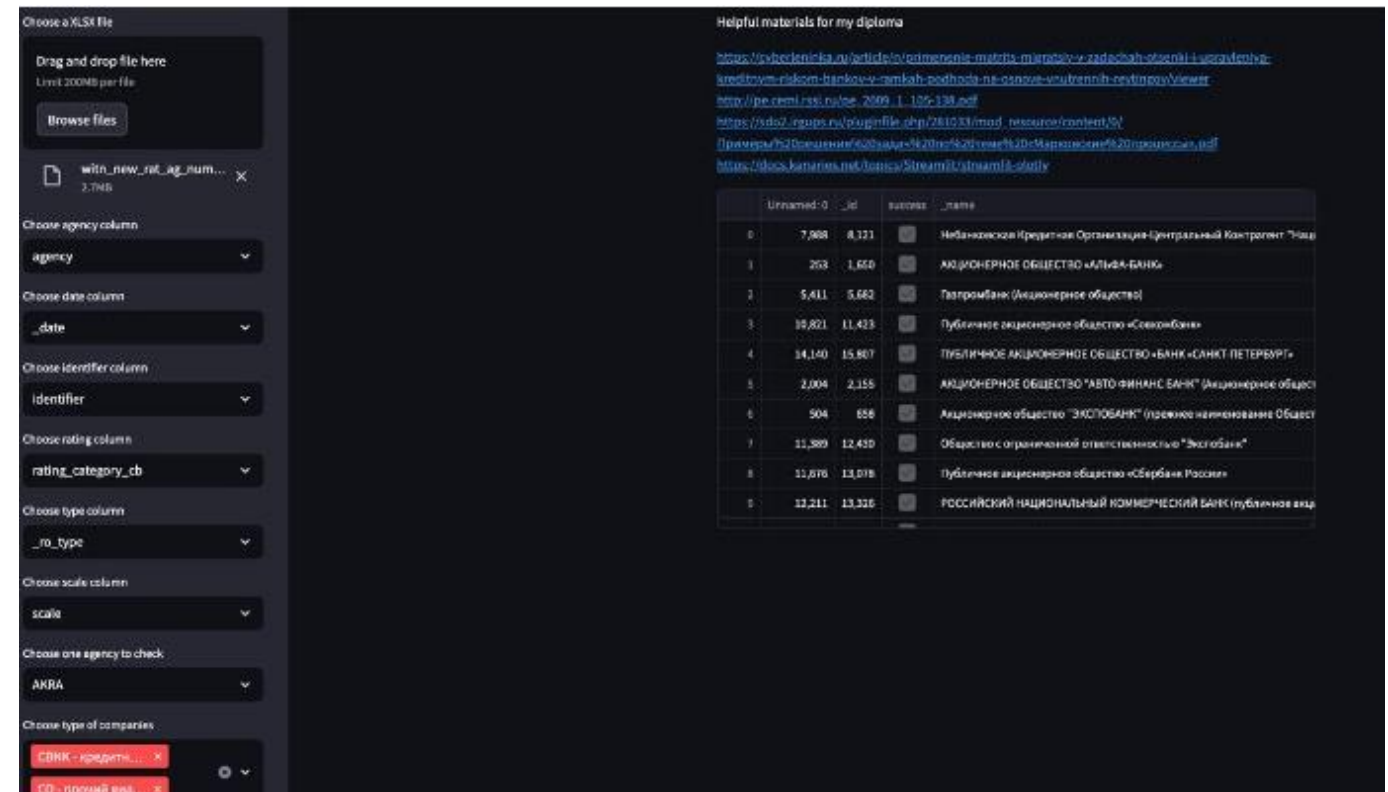


Метод	1 год	4 год	Δ_abs	Δ_rel (%)
Dirichlet	0.7655	0.5124	0.2531	33.07%
CTMC	0.8228	0.5753	0.2475	30.09%
2-й порядок	0.7566	0.5304	0.2262	29.89%
Basic	0.5506	0.4165	0.1341	24.35%
Random Forest	0.8779	0.6814	0.1965	22.38%

Веб-интерфейс

В рамках данной работы также был реализован веб-интерфейс для удобного взаимодействия с данными и быстрого расчета реализованных методов. Было использовано достаточно много разнообразных библиотек, позволяющих визуализировать данные, графы и матрицы переходных вероятностей.

На базе веб-фреймворка строился сам сервис и все дополнительные функции. При помощи `fitter` – строились распределения, при помощи `seaborn` – тепловые карты матриц переходных вероятностей, при помощи `NetGraph` – графы и прочие вспомогательные библиотеки



Заключение

Основное достижение данного проекта состоит в разработке алгоритмов предсказания типа карточного клиента на несколько будущих периодов вперёд. Алгоритмы были основаны на случайном лесе и марковской модели. Ключевой особенностью проекта является изучение и предсказание всех состояний кредитных рейтингов, в то время как на основе приведённого анализа литературы видно, что работы в мире посвящены предсказанию лишь отдельных состояний.

В рамках работы были изучены и реализованы новые для российского рынка подходы, которые несут в себе большой потенциал. Также в рамках работы с марковской моделью была построена цепь Маркова 2-го порядка, которая местами демонстрирует улучшение результатов по сравнению с цепью 1 порядка.

Также были применены относительно новые методы для потенциального прогнозирования состояний – МСМС и проанализированы спектральные характеристики наиболее представительного временного ряда





Роли в команде

Липатов
Данила

- Изучение литературы
- Парсинг данных
- Обработка и изучение данных
- Изучение и применение марковских моделей (построение графов переходов, матрицы переходных вероятностей)
- Изучение и реализация на Python алгоритмов классификации данных (случайный лес)
- Изучение и реализация на Python методов оценки матриц переходных вероятностей (бутстрап, MCMC)
- Анализ спектральных характеристик данных кредитных рейтингов
- Реализация веб-интерфейса для расчета на базе Streamlit (опционально)



Список литературы

- Афанасьева Л.Г., Булинская Е.В. Случайные процессы в теории массового обслуживания и управления запасами. - М.: изд-во МГУ, 1980 - Гл. 3, [ссылка](#)
- Régis, Daniel Evangelista; Artes, Rinaldo. Using multi-state markov models to identify credit card risk // 2016, [link](#)
- David Lando Analyzing rating transitions and rating drift with continuous observations // Journal of Banking & Finance. – 2002. – No26. – С. 423 – 444.
- Rating transitions forecasting: a filtering approach. - hal.science, [link](#)
- Bayesian Inference for Issuer Heterogeneity in Credit Ratings Migration. – fdic, [link](#)
- Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. - MIT Press, 2012, [link](#)
- Христофоров А.В. ТЕОРИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ В ГИДРОЛОГИИ. - М.: издательство МГУ, 1994 - 139 с.



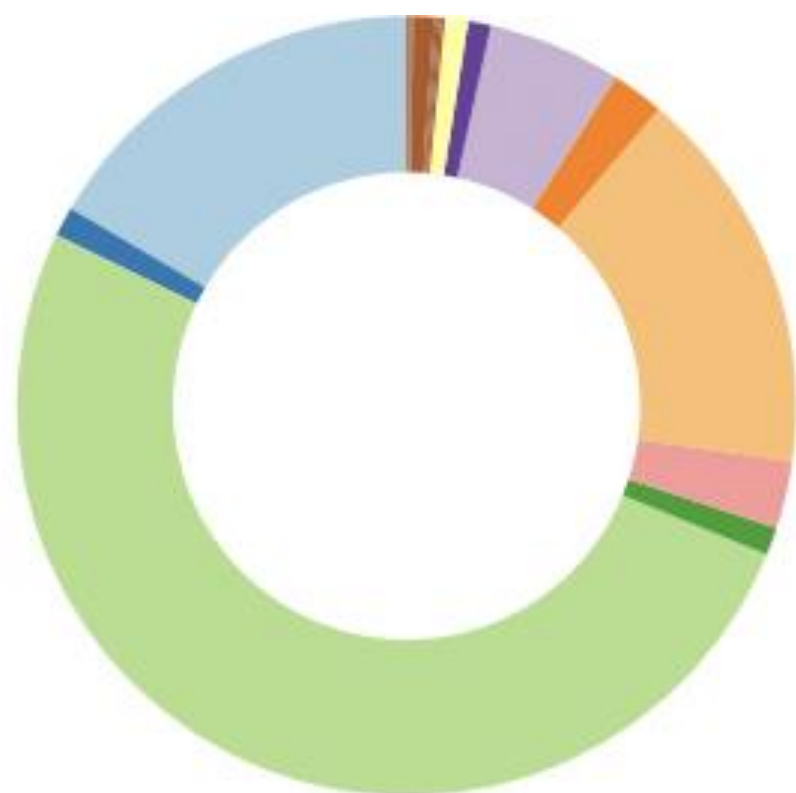
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Руководитель: Зотов Леонид Валентинович,
Izotov@hse.ru

Консультант: Игнатовская Валерия Анатольевна,
vignatovskaya@hse.ru

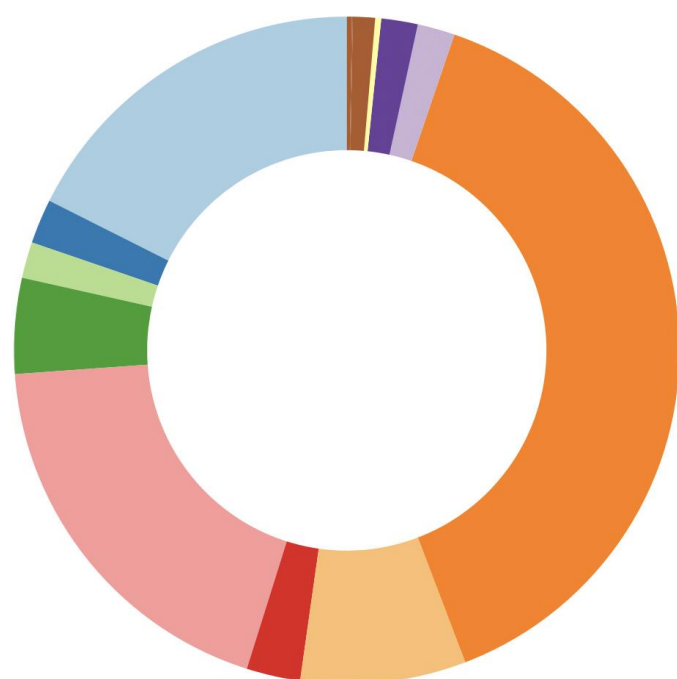
Исполнители: Липатов Данила Вячеславович,
dvlipatov@edu.hse.ru

Приложение

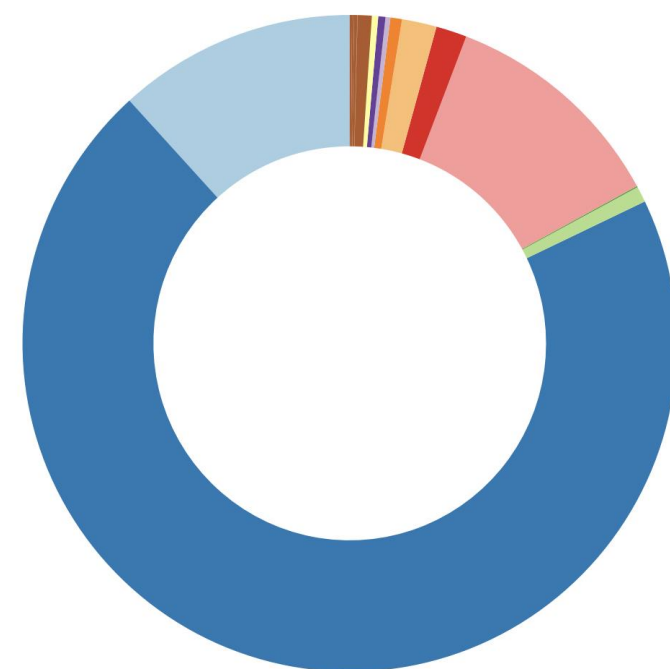


BNFC (нефинансовая компания)	— 2726 (16.5%)
CO (прочий вид организации)	— 204 (1.2%)
TBND (облигационный заем)	— 8402 (51.0%)
BNFH (холдинговая компания, нефинансовая)	— 191 (1.2%)
FOFO (прочие финансовые организации, включая специализированные общества)	— 452 (2.7%)
ТО (прочий тип финансового инструмента)	— 4 (0.0%)
CBNK (кредитная организация)	— 2650 (16.1%)
FLSG (лизинговая компания/фирма)	— 362 (2.2%)
FINS (страховая организация)	— 912 (5.5%)
FNPF (негосударственный пенсионный фонд)	— 156 (0.9%)
FMFO (микрофинансовая организация)	— 159 (1.0%)
FDEP (депозитарий)	— 26 (0.2%)
TMGB (облигации с ипотечным покрытием)	— 25 (0.2%)
IFO (международная финансовая организация)	— 22 (0.1%)
TMNB (облигации, обеспеченные залогом денежных требований)	— 141 (0.9%)
TSCB (облигации, обеспеченные залогом ценных бумаг)	— 11 (0.1%)
TSFP (прочие облигации, обеспеченные залогом)	— 12 (0.1%)
FFCT (юридическое лицо, специализирующееся на финансировании под уступку денежных требований)	— 21 (0.1%)

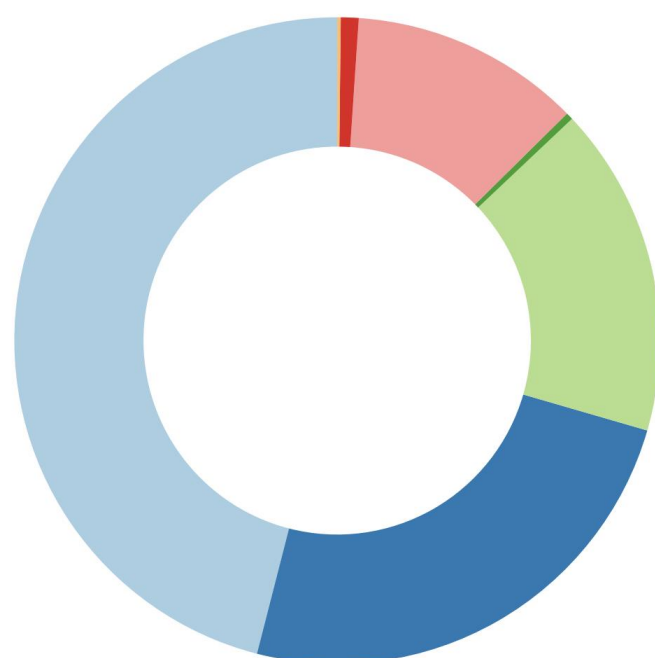
Приложение



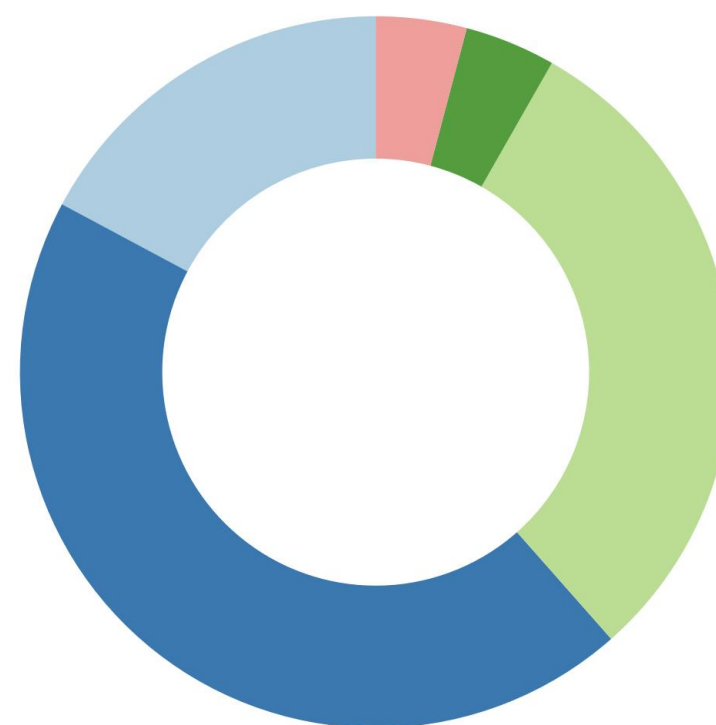
KPA – 1



KPA – 2



KPA – 3



KPA – 4



Приложение

Навыки, которые планируется приобрести во время реализации проекта

- Поиск, сбор, обработка и анализ информации из научных источников, приобретение навыка работы с англоязычными и отечественными статьями и учебными пособиями
- Приобретение навыка формализации прикладных задач, построение математических моделей задач
- Формирование и описание алгоритмов для решения практических задач
- Углубление знаний теории случайных процессов, машинного обучения
- Углубление навыков программирования в Python и R