

МРНТИ 28.23.25

DOI: <https://doi.org/10.62687/STJ.2.2.2026.24>

РЕКОМЕНДАЦИЯ КУЛЬТУРНО-РАЗВЛЕКАТЕЛЬНЫХ СОБЫТИЙ В УСЛОВИЯХ ЭКСТРЕМАЛЬНОГО ХОЛОДНОГО СТАРТА: ЦЕННОСТЬ АФИШИ, ПОГОДНОГО И КАЛЕНДАРНОГО КОНТЕКСТА (НА ДАННЫХ БИЛЕТНОЙ ПЛАТФОРМЫ КАЗАХСТАНА)

¹Б.Б. Маратов*, ²Ж. Ламашева^{1,2}НАО Евразийский национальный университет имени Л. Н. Гумилёва, г. Астана, Казахстан

*e-mail: 040901550381@enu.kz

Б.Б. Маратов – магистрант 1 курса образовательной программы «Бизнес-аналитика в IT», НАО «Евразийский национальный университет имени Л. Н. Гумилёва», г. Астана, Республика Казахстан, e-mail: maratov.b@enu.kz, <https://orcid.org/0009-0001-9053-3649>

Ж. Ламашева – PhD, ассоциированный профессор, НАО «Евразийский национальный университет имени Л. Н. Гумилёва», г. Астана, Республика Казахстан.

Аннотация. Настоящая статья посвящена задаче персонализированной рекомендации культурно-развлекательных событий в условиях экстремального холодного старта, характерного для билетных онлайн-платформ. Исследование выполнено на реальных обезличенных данных национальной билетной платформы Казахстана, охватывающих 12 715 взаимодействий 10 657 пользователей с 2 074 событиями в четырёх городах страны за период с марта 2023 г. по февраль 2024 г. Особенностью данных являются медианное число взаимодействий на пользователя, равное единице, доля «холодных» пользователей в контрольном месяце 83 % и около 50 % покупок, приходящихся на события-премьеры, не имеющие истории продаж. В рамках единого протокола строгой временной валидации на шести месячных фолдах сопоставлены 13 алгоритмов: от базовых моделей популярности до гибридного градиентного ранжировщика LightGBM (LambdaRank) с погодными, календарными и афишными признаками и смешиванием рангов методом обратной ранговой фузии (RRF). Статистическая значимость различий оценивалась методом парного бутстрепа (5000 репликатов, 95 % доверительные интервалы). Установлено, что знание афиши следующего месяца обеспечивает наибольший и статистически значимый прирост качества ($\Delta\text{MAP}@10 = +0,0084$; $p < 0,0001$), превышающий эффект от любых усложнений собственно алгоритма ранжирования; лучшая конфигурация достигает $\text{MAP}@10 = 0,0578$, что на 17 % выше сильнейшего неперсонализированного базового алгоритма. Погодный контекст демонстрирует прирост на границе статистической значимости ($p = 0,053$), календарные признаки сами по себе значимого эффекта не дают. Полученные результаты обосновывают приоритет интеграции данных афиши в продуктивные рекомендательные контуры событийных платформ.

Ключевые слова: рекомендательные системы, холодный старт, ранжирование, градиентный бустинг, LambdaRank, контекстно-зависимые рекомендации, событийные платформы, $\text{MAP}@10$.

EVENT RECOMMENDATION UNDER EXTREME COLD-START CONDITIONS: THE VALUE OF SCHEDULE, WEATHER AND CALENDAR CONTEXT (BASED ON DATA FROM A KAZAKHSTANI TICKETING PLATFORM)

¹B.B. Maratov*, ²Zh. Lamasheva^{1,2}L. N. Gumilyov Eurasian National University, Astana, Kazakhstan

*e-mail: 040901550381@enu.kz

B.B. Maratov – 1st year Master's student of the educational program «Business Analytics in IT», L. N. Gumilyov Eurasian National University, Astana, Kazakhstan, e-mail: maratov.b@enu.kz, <https://orcid.org/0009-0001-9053-3649>

Zh. Lamasheva – PhD, Associate Professor, L. N. Gumilyov Eurasian National University, Astana, Kazakhstan.

Abstract. This article addresses the problem of personalized recommendation of cultural and entertainment events under the extreme cold-start conditions typical of online ticketing platforms. The study is based on real anonymized data from a national ticketing platform of Kazakhstan, covering 12,715 interactions of 10,657 users with 2,074 events in four cities of the country from March 2023 to February 2024. The dataset is characterized by a median of one interaction per user, an 83 % share of cold-start users in the holdout month, and about 50 % of purchases falling on premiere events with no sales history. Within a unified protocol of strict temporal validation on six monthly folds, 13 algorithms were compared: from basic popularity models to a hybrid gradient-boosted ranker LightGBM (LambdaRank) enriched with weather, calendar and schedule features and rank blending via reciprocal rank fusion (RRF). Statistical significance was assessed using a paired bootstrap procedure (5,000 replicates, 95 % confidence intervals). It was found that knowledge of the next month's event schedule provides the largest and statistically significant quality gain ($\Delta\text{MAP}@10 = +0.0084$; $p < 0.0001$), exceeding the effect of any refinement of the ranking algorithm itself; the best configuration achieves $\text{MAP}@10 = 0.0578$, which is 17 % above the strongest non-personalized baseline. The weather context demonstrates a gain at the boundary of statistical significance ($p = 0.053$), while calendar features alone yield no significant effect. The results substantiate the priority of integrating schedule data into production recommendation pipelines of event platforms.

Keywords: recommender systems, cold start, learning to rank, gradient boosting, LambdaRank, context-aware recommendation, event platforms, $\text{MAP}@10$.

ЭКСТРЕМАЛДЫ «СУЫҚ СТАРТ» ЖАҒДАЙЫНДА МӘДЕНИ-ОЙЫН-САУЫҚ ІС-ШАРАЛАРЫН ҰСЫНУ: АФИШАНЫҢ, АУА РАЙЫ МЕН КҮНТІЗБЕЛІК КОНТЕКСТІҢ ҚҰНДЫЛЫҒЫ (ҚАЗАҚСТАНДЫҚ БИЛЕТ ПЛАТФОРМАСЫНЫҢ ДЕРЕКТЕРІ НЕГІЗІНДЕ)

¹Б.Б. Маратов*, ²Ж. Ламашева

^{1,2}Л. Н. Гумилев атындағы Еуразия ұлттық университеті КеАҚ, Астана қ., Қазақстан

*e-mail: 040901550381@enu.kz

Б.Б. Маратов – «IT-дегі бизнес-аналитика» білім беру бағдарламасының 1-курс магистранты, «Л. Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ, Астана қ., Қазақстан Республикасы, e-mail: maratov.b@enu.kz, <https://orcid.org/0009-0001-9053-3649>

Ж. Ламашева – PhD, қауымдастырылған профессор, «Л. Н. Гумилев атындағы Еуразия ұлттық университеті» КеАҚ, Астана қ., Қазақстан Республикасы.

Аңдатпа. Бұл мақала билет сату онлайн-платформаларына тән экстремалды «суық старт» жағдайында мәдени-ойын-сауық іс-шараларын дербестендірілген түрде ұсыну мәселесіне арналған. Зерттеу Қазақстанның ұлттық билет платформасының нақты иесіздендірілген деректері негізінде орындалды: 2023 жылғы наурыз бен 2024 жылғы ақпан аралығында елдің төрт қаласындағы 2 074 іс-шарамен 10 657 пайдаланушының 12 715 өзара әрекеттесуі қамтылды. Деректердің ерекшелігі - бір пайдаланушыға шаққандағы өзара әрекеттесулердің медианасы бірге тең, бақылау айындағы «суық» пайдаланушылардың үлесі 83 % және сатып алулардың шамамен 50 %-ы сату тарихы жоқ премьер-іс-шараларға тиесілі. Қатаң уақыттық валидацияның бірыңғай хаттамасы аясында алты айлық фолдта 13 алгоритм салыстырылды: базалық танымалдылық модельдерінен бастап ауа райы, күнтізбелік және афишалық белгілермен байытылған әрі рангілерді кері рангілік фузия (RRF) әдісімен араластыратын гибридік градиенттік LightGBM (LambdaRank) ранжирлеушісіне дейін. Айырмашылықтардың статистикалық маңыздылығы жұптық бутстреп әдісімен бағаланды (5000 реплика, 95 % сенімділік интервалдары). Келесі айдың афишасын білу сапаның ең

үлкен әрі статистикалық маңызды өсімін қамтамасыз ететіні анықталды ($\Delta \text{MAP}@10 = +0,0084$; $p < 0,0001$), бұл ранжирлеу алгоритмінің өзін кез келген күрделендіру әсерінен асып түседі; ең үздік конфигурация $\text{MAP}@10 = 0,0578$ мәніне жетеді, бұл ең күшті дербестендірілмеген базалық алгоритмнен 17 %-ға жоғары. Ауа райы контексті статистикалық маңыздылық шекарасындағы өсімді көрсетеді ($p = 0,053$), ал күнтізбелік белгілер өздігінен маңызды әсер бермейді. Алынған нәтижелер афиша деректерін іс-шара платформаларының өндірістік ұсыныс контурларына интеграциялаудың басымдығын негіздейді.

Кілт сөздер: ұсыныс жүйелері, суық старт, ранжирлеу, градиенттік бустинг, LambdaRank, контекстке тәуелді ұсыныстар, іс-шара платформалары, $\text{MAP}@10$.

Введение. Стремительная цифровизация сферы культурного потребления привела к тому, что билетные онлайн-платформы стали основным каналом дистрибуции билетов на кинопоказы, концерты, театральные постановки, спортивные и детские мероприятия. В условиях непрерывно обновляющейся афиши пользователь сталкивается с проблемой информационной перегрузки, а платформа - с необходимостью персонализации выдачи: своевременная и релевантная рекомендация напрямую влияет на конверсию, удержание аудитории и экономику сервиса. Основным инструментом решения этой задачи являются рекомендательные системы, которые за последние два десятилетия стали одним из наиболее востребованных приложений методов машинного обучения (Ricci и др., 2015:1003; Гомзин&Коршунов, 2012:401-417).

Классический арсенал рекомендательных систем включает коллаборативную фильтрацию, основанную на схожести объектов (Sarwar и др., 2001:285-295) или латентных факторных представлениях, получаемых матричной факторизацией (Koren и др., 2009:30-37), в том числе для неявной обратной связи (Hu и др., 2008:263-272); контентные методы, опирающиеся на векторизацию текстовых описаний объектов (Salton&Buckley, 1988:513-523); а также гибридные архитектуры, комбинирующие несколько источников сигнала (Burke, 2002:331-370). Современным стандартом промышленного ранжирования выступают методы обучения ранжированию на основе градиентного бустинга - семейство LambdaRank/LambdaMART (Burgess, 2010:19), эффективно реализованное в библиотеке LightGBM (Ke и др., 2017:3146-3154). Для устойчивого объединения нескольких ранжированных списков широко применяется обратная ранговая фузия (reciprocal rank fusion, RRF), демонстрирующая результаты на уровне и выше обучаемых методов смешивания (Cormack и др., 2009:758-759).

Вместе с тем событийный домен принципиально отличается от традиционных доменов рекомендательных систем - кино- и видеосервисов, электронной коммерции, музыкальных платформ. Событие существует ограниченное время и после проведения безвозвратно покидает каталог; жизненный цикл объекта составляет дни или недели; к моменту формирования рекомендации значительная часть релевантных объектов ещё не имеет ни одного взаимодействия (Macedo и др., 2015:123-130). Тем самым проблема холодного старта (Schein и др., 2002:253-260; Lam и др., 2008:208-211) приобретает в событийном домене экстремальный характер: холодными оказываются одновременно и пользователи, и объекты. Кроме того, посещение офлайн-события, в отличие от потребления цифрового контента, потенциально зависит от внешнего контекста - погодных условий, праздничных и выходных дней (Adomavicius&Tuzhilin, 2005:734-749; Trattner и др., 2018:117-150), что делает событийные рекомендации естественным полигоном для контекстно-зависимых моделей.

Анализ литературы выявляет существенный разрыв между академическими бенчмарками и практическими условиями работы региональных билетных платформ. Абсолютное большинство публикуемых сравнений выполняется на плотных наборах данных типа MovieLens или Netflix, где на пользователя приходится десятки и сотни оценок; при этом показано, что на разреженных данных тщательно настроенные простые базовые алгоритмы нередко превосходят сложные нейросетевые архитектуры (Dacrema и др., 2019:101-109). Систематические исследования на реальных транзакционных данных билетных платформ

Центральной Азии, характеризующихся малым объёмом, экстремальной разреженностью, двуязычным (русским и казахским) описанием контента и выраженной городской локализацией спроса, в открытой печати практически отсутствуют. Не изучен и вопрос о сравнительной ценности различных источников информации - истории продаж, контента, внешнего контекста и афиши предстоящего месяца - в подобных условиях.

Научная новизна настоящего исследования состоит в следующем: (1) впервые на открытых обезличенных данных национальной билетной платформы Республики Казахстан выполнен систематический бенчмарк 13 рекомендательных алгоритмов в рамках единого протокола строгой временной валидации; (2) предложена и количественно оценена концепция «ценности афиши»: показано, что знание каталога событий предстоящего месяца обеспечивает статистически значимый прирост качества, превышающий эффект от усложнения собственно модели ранжирования; (3) впервые для событийного домена Казахстана исследована интеграция погодной климатологии и праздничного календаря в градиентный ранжировщик, построенная без утечек данных из будущего; (4) применена строгая процедура статистической верификации различий на основе парного бутстрепа по наблюдениям «фолд-пользователь», редко используемая в прикладных работах по рекомендательным системам.

Целью работы является количественная оценка сравнительного вклада алгоритмической сложности, контентных сигналов, внешнего контекста и знания афиши в качество рекомендаций культурно-развлекательных событий в условиях экстремального холодного старта. Для достижения цели решались следующие задачи:

1. Сформировать воспроизводимый протокол временной валидации и метрики качества, соответствующие практике событийных платформ.

2. Реализовать и сопоставить базовые, коллаборативные, контентные и гибридные модели рекомендаций с последовательным добавлением погодных, календарных и афишных признаков.

3. Оценить статистическую значимость различий между моделями и выявить признаки, вносящие наибольший вклад в ранжирование.

Материалы и методы исследования. Исследование выполнено на обезличенном наборе транзакционных данных национальной билетной онлайн-платформы Казахстана Ticketon, предоставленном в рамках открытого аналитического конкурса Freedom Data Challenge (Data Science Academy, <https://dsacademy.kz/fdc2025>). Набор охватывает период с марта 2023 г. по февраль 2024 г. и включает 12 715 взаимодействий 10 657 пользователей с 2 074 событиями в четырёх городах: Алматы, Астане, Шымкенте и Павлодаре. Для 4 597 событий доступны метаданные: категория (кино, театр, концерт и др.), текстовое описание на русском и казахском языках, продолжительность, возрастной рейтинг, признак отечественного производства и год выпуска. Каждое взаимодействие снабжено статусом заказа; при построении моделей статусам присваивались веса: оплаченный заказ - 1,0; отменённый или возвращённый - 0,4; созданный, но не оплаченный - 0,2. Текстовые описания очищались от HTML-разметки, гиперссылок и служебных символов; атрибуты пользователя (город, пол, возраст) рассматривались как статические.

Ключевой особенностью набора является экстремальная разреженность: медианное число взаимодействий на пользователя равно единице, лишь 0,1% пар «пользователь-событие» повторяются, 83% пользователей контрольного месяца не имеют истории покупок (холодные пользователи), а около половины всех покупок приходится на события-премьеры, отсутствовавшие в каталоге ранее. Подобная конфигурация практически исключает эффективное применение классической коллаборативной фильтрации в чистом виде и смещает задачу в область гибридных и контекстно-зависимых методов.

Протокол оценивания построен по принципу строгой временной валидации на шести месячных фолдах с датами отсечения с 1 сентября 2023 г. по 1 февраля 2024 г. Для каждого фолда обучающая история включает только взаимодействия, строго предшествующие дате отсечения, а эталоном (ground truth) служат оплаченные покупки в течение соответствующего календарного месяца. Пул кандидатов формировался из событий,

имевших активность в течение 60 дней до даты отсечения. Качество оценивалось метрикой средней точности MAP@10, вычисляемой по официальному скрипту организаторов конкурса, дополнительно рассчитывались Recall@10 и HitRate@10 с раздельным учётом тёплых и холодных пользователей. Итоговая выборка для статистического анализа содержит 6 309 наблюдений вида «фолд-пользователь».

В сравнение включены следующие группы моделей. Базовые модели популярности: GlobalPop и CityPop ранжируют события по популярности с экспоненциальным затуханием во времени (период полураспада 14 суток), во втором случае - с сегментацией по городу пользователя и откатом к глобальному рейтингу. Коллаборативные модели: ItemKNN строит взвешенную по времени разреженную матрицу «пользователь-событие» и оценивает косинусную близость событий; ALS - матричная факторизация для неявной обратной связи (64 латентных фактора, регуляризация 0,05; 30 итераций, коэффициент уверенности 40) (Hu и др., 2008:263-272). Контентная модель Content TF-IDF векторизует двуязычные описания событий (словарь до 60 000 признаков, униграммы и биграммы) и ранжирует кандидатов по косинусной близости к профилю пользователя, агрегированному из векторов просмотренных событий с временным затуханием (Salton&Buckley, 1988:513-523).

Гибридная модель реализована как двухэтапный конвейер «генерация кандидатов - ранжирование». Объединённый список кандидатов формируется из трёх источников: топ-50 по городской популярности, топ-30 по ItemKNN и топ-30 по контентной близости (суммарно 80-120 кандидатов на пользователя). Ранжирование выполняется градиентным бустингом LightGBM с целевой функцией lambdarank и метрикой MAP (400 деревьев, скорость обучения 0,05; 63 листа, минимум 20 наблюдений в листе, субдискретизация строк и признаков 0,9) (Ke и др., 2017:3146-3154; Burges, 2010:19). Вектор описания пары «пользователь-событие» включает 27 базовых признаков шести групп: сигналы популярности с затуханием (в том числе за окна 3, 7, 14 и 30 суток, тренд и ускорение спроса); характеристики жизненного цикла события (возраст с первого появления, давность последней покупки, число уникальных покупателей); контентные и категориальные признаки (категория, продолжительность, возрастной рейтинг, косинусная близость TF-IDF); коллаборативный сигнал ItemKNN; социально-демографические признаки пользователя (город, пол, возраст, совпадение города); поведенческие агрегаты (аффинность к категории и жанру, число взаимодействий, давность последней активности). Для устранения смещения ранжировщика при обучении применялась комбинированная выборка негативных примеров: к позитивам из пула добавлялось до 40 случайных негативов на пользователя. Финальное предсказание опционально смешивалось с городской популярностью методом обратной ранговой фузии RRF со сглаживающей константой 20 (Cormack и др., 2009:758-759), что стабилизирует выдачу в месяцы с аномальной динамикой спроса.

Контекстные расширения строились с исключением утечек информации из будущего. Погодный блок (8 признаков) использует суточные данные открытого метеоархива Open-Meteo (<https://open-meteo.com>) по четырём городам; климатическая норма месяца (средняя температура, доля снежных дней, доля дней с неблагоприятной погодой - минимум ниже -15 °C либо осадки не менее 1 мм) рассчитывалась по опорному периоду 2010-2022 гг., заведомо предшествующему интеракционным данным, а фактическая погода привлекалась только для дат, уже наступивших к моменту отсечения. Календарный блок (4 признака) кодирует число праздничных дней целевого месяца, наличие крупных праздников и долю праздничных покупок события. Афишный блок (3 признака) отражает знание каталога предстоящего месяца: индикатор новизны события, признак отечественного фильма и разницу между годом выпуска и текущим годом; добавление этого блока сопровождается расширением пула кандидатов событиями объявленной афиши, не имеющими истории продаж.

Статистическая значимость различий между моделями оценивалась методом парного бутстрепа (Efron&Tibshirani, 1993:436): значения AP@10 ресемплировались по 6 309 наблюдениям «фолд-пользователь» с одинаковыми индексами для сравниваемых моделей (5 000 репликатов, фиксированное зерно генератора), рассчитывались 95% перцентильные

доверительные интервалы разности и двусторонние р-значения. Все эксперименты воспроизводимы: исходный код и полные результаты доступны в открытом репозитории <https://github.com/m-bikko/article-freedom-datachallenge>. Использованные наборы данных опубликованы там же: журнал взаимодействий (data/train_test.csv), метаданные событий (data/events_description.csv) и суточные погодные данные (data/weather_daily.csv).

Результаты и обсуждение. Сводные характеристики использованного набора данных приведены в таблице 1. Обращает на себя внимание сочетание факторов, каждый из которых в отдельности считается серьёзным осложнением для рекомендательных систем, а в совокупности они формируют режим экстремального холодного старта: единичная медиана взаимодействий исключает построение содержательного профиля для большинства пользователей, а преобладание премьер среди покупок означает, что классический коллаборативный сигнал недоступен для значительной части целевых объектов.

Таблица 1. Основные характеристики набора данных билетной платформы (март 2023 г. - февраль 2024 г.)

Показатель	Значение
Число взаимодействий	12 715
Число пользователей	10 657
Число событий с взаимодействиями	2 074
Число событий с метаданными	4 597
Города	Алматы, Астана, Шымкент, Павлодар
Медиана взаимодействий на пользователя	1
Доля холодных пользователей в контрольном месяце	83 %
Доля покупок на события-премьеры	≈ 50%
Доля повторных пар «пользователь-событие»	0,1 %

Результаты сравнения моделей по метрике MAP@10, усреднённые по шести месячным фолдам, представлены в таблице 2 и на рисунке 1. Верхней границей достижимого качества служит оракульная модель, ранжирующая только фактически купленные события из пула кандидатов: MAP@10 = 0,4653. Разрыв между оракулом и лучшей практической моделью показывает, что доминирующим ограничением является не алгоритм ранжирования, а принципиальная непредсказуемость единичных покупок при столь разреженной истории.

Таблица 2. Качество моделей рекомендаций по метрике MAP@10 (среднее по шести фолдам; в скобках - 95% бутстреп-интервалы)

Модель	MAP@10	95 % ДИ
Oracle (потолок пула кандидатов)	0,4653	[0,453; 0,478]
Hybrid + Context (календарь, без RRF)	0,0405	[0,037; 0,044]
GlobalPop (глобальная популярность)	0,0420	[0,039; 0,045]
Hybrid LGBM (чистый ранжировщик)	0,0424	[0,039; 0,046]
Hybrid + Weather (погода, без RRF)	0,0428	[0,039; 0,047]
ALS (матричная факторизация)	0,0465	[0,043; 0,050]
ItemKNN (коллаборативная фильтрация)	0,0479	[0,044; 0,052]

Модель	MAP@10	95 % ДИ
Hybrid + Context + Schedule (афиша, без RRF)	0,0489	[0,045; 0,053]
CityPop (городская популярность)	0,0492	[0,046; 0,053]
Content TF-IDF (контентная модель)	0,0496	[0,046; 0,053]
Hybrid RRF	0,0544	[0,050; 0,059]
Hybrid + Context RRF (календарь)	0,0558	[0,052; 0,060]
Hybrid + Weather RRF (погода)	0,0572	[0,053; 0,062]
Hybrid + Context + Schedule RRF (афиша)	0,0578	[0,054; 0,062]

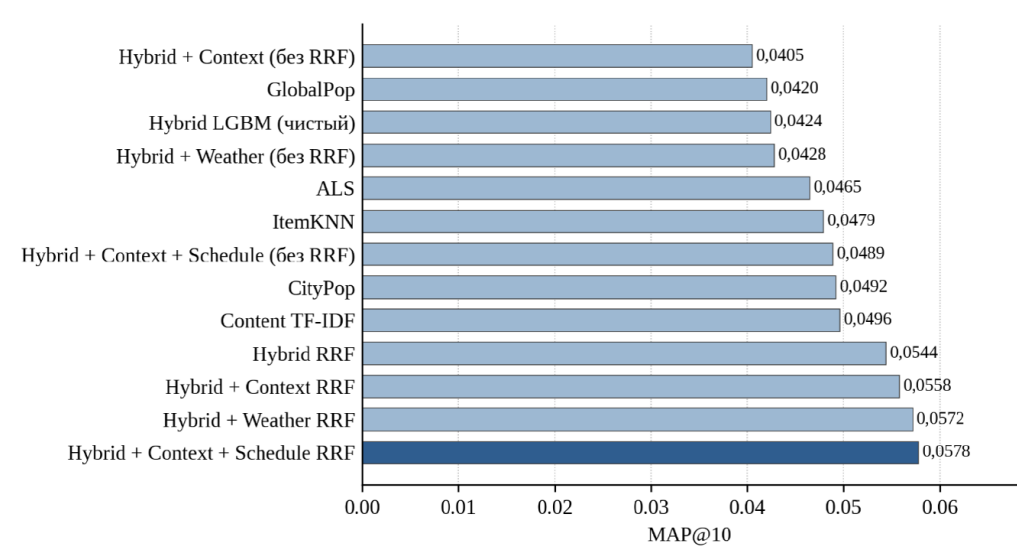


Рисунок 1. Сравнение качества моделей рекомендаций по метрике MAP@10 (среднее по шести месячным фолдам; оракульная модель не показана)

Анализ таблицы 2 позволяет выделить несколько закономерностей. Во-первых, персонализированные коллаборативные модели (ItemKNN, ALS) не превосходят простую городскую популярность: при единичной медиане взаимодействий матрица «пользователь-событие» не содержит достаточного сигнала для обобщения, что согласуется с выводами о силе настроенных базовых алгоритмов на разреженных данных (Dasgupta и др., 2019:101-109). Во-вторых, контентная модель на двуязычных TF-IDF представлениях ($MAP@10 = 0,0496$) оказывается сильнее обоих коллаборативных методов - в условиях дефицита истории текст описания события несёт больше информации, чем следы редких совместных покупок. В-третьих, чистый градиентный ранжировщик уступает базовой городской популярности (0,0424 против 0,0492): при малом числе позитивных примеров помесячная нестабильность обучаемой модели перевешивает её выразительную способность. Смешивание выдачи ранжировщика с популярностным приором методом RRF устраняет этот дефект и выводит гибрид на уровень 0,0544, что статистически значимо выше CityPop ($p = 0,001$). Показательно, что та же закономерность воспроизводится и для контекстных расширений: без RRF конфигурации с погодными (0,0428) и календарными (0,0405) признаками не превосходят базовые алгоритмы, а полная модель с афишей без RRF статистически неотличима от CityPop ($p = 0,896$) - стабилизация популярностным приором является необходимым условием реализации потенциала обучаемого ранжировщика на данных такого объёма.

Результаты проверки статистической значимости ключевых сравнений приведены в таблице 3.

Таблица 3. Парные сравнения моделей (парный бутстреп, 5 000 репликаторов, 6 309 наблюдений «фолд-пользователь»)

Сравнение	$\Delta \text{MAP@10}$	p-значение	Вывод
Hybrid RRF vs CityPop	+0,0052	0,001	значимо: ранжировщик полезен только в смеси с приором
Hybrid LGBM (чистый) vs CityPop	-0,0069	0,002	значимо хуже: нестабильность по месяцам
+ Афиша vs + Календарь	+0,0084	<0,0001	значимо: наибольший единичный прирост
Лучшая модель vs CityPop	+0,0086	<0,0001	значимо: +17 % к сильнейшему базовому алгоритму
+ Погода vs Hybrid RRF	+0,0028	0,053	на границе значимости
+ Календарь vs + Погода	-0,0022	0,088	значимых различий нет
+ Афиша (без RRF) vs CityPop	-0,0003	0,896	без RRF выигрыш афиши не реализуется

Центральный результат исследования - количественная оценка ценности афиши. Добавление всего трёх признаков, кодирующих знание каталога предстоящего месяца, вместе с расширением пула кандидатов объявленными премьерными даёт прирост $\Delta \text{MAP@10} = +0,0084$ ($p < 0,0001$) - больше, чем любая другая модификация модели, включая переход от простой популярности к полному гибриднему конвейеру. Итоговая конфигурация Hybrid + Context + Schedule RRF достигает $\text{MAP@10} = 0,0578$ [0,054; 0,062], превосходя сильнейший неперсонализированный базовый алгоритм на 17% ($p < 0,0001$). Содержательно это объясняется структурой спроса: поскольку около половины покупок приходится на премьеры, ни одна модель, оперирующая только историческими данными, в принципе не способна рекомендовать значительную часть фактически востребованных событий; информация об афише устраняет это структурное ограничение.

Погодный контекст демонстрирует прирост +0,0028 с p-значением 0,053, формально не достигающим конвенционального порога 0,05: эффект следует квалифицировать как индикативный и требующий проверки на большем объёме данных. Календарные признаки в отрыве от афиши значимого эффекта не дают ($p = 0,088$), что интерпретируется следующим образом: праздничный календарь влияет на общий объём спроса, но слабо дифференцирует события между собой внутри месяца. Анализ важностей признаков ранжировщика (по суммарному приросту качества расщеплений) показывает, что наибольший вклад вносят признаки жизненного цикла события (возраст в каталоге, давность последней покупки), городская популярность и возраст пользователя, а контентная близость TF-IDF систематически превосходит по важности классический коллаборативный сигнал - в полном согласии с результатами прямого сравнения одиночных моделей.

Практическое значение полученных результатов состоит в смещении приоритетов разработки: для событийных платформ, работающих в условиях, аналогичных исследованному, интеграция машиночитаемой афиши предстоящего периода в рекомендательный контур (включая генерацию кандидатов из объявленных, но ещё не продаваемых событий) даёт больший и статистически более надёжный эффект, чем усложнение алгоритма ранжирования, и должна выполняться в первую очередь. Дополнительным практическим выводом является необходимость стабилизации обучаемых ранжировщиков популярностным приором (например, методом RRF) при малых объёмах обучающих данных.

Заключение. В работе впервые на открытых обезличенных данных национальной билетной платформы Казахстана выполнено систематическое сравнение 13 рекомендательных алгоритмов в условиях экстремального холодного старта: при медиане в одно взаимодействие на пользователя, 83% холодных пользователей и около 50% покупок, приходящихся на события

без истории продаж. Единый протокол строгой временной валидации на шести месячных фолдах в сочетании с парным бутстрепом по 6 309 наблюдениям обеспечил статистическую корректность выводов.

Главный вывод исследования формулируется как принцип «ценности афиши»: знание каталога событий предстоящего месяца обеспечивает наибольший статистически значимый прирост качества рекомендаций ($\Delta \text{MAP}@10 = +0,0084$; $p < 0,0001$), превышающий эффект любых усложнений собственно модели. Лучшая конфигурация - гибридный градиентный ранжировщик LightGBM с погодными, календарными и афишными признаками, стабилизированный обратной ранговой фузией с городской популярностью, - достигает $\text{MAP}@10 = 0,0578$, что на 17% выше сильнейшего базового алгоритма. Установлено также, что в условиях экстремальной разреженности контентные сигналы двуязычных описаний превосходят классическую коллаборативную фильтрацию, чистый обучаемый ранжировщик без популярностного приора неустойчив, а календарные признаки сами по себе значимого эффекта не дают; влияние погодного контекста находится на границе статистической значимости и требует дальнейшей проверки.

Ограничениями работы являются охват одного годового цикла и четырёх городов, оценка исключительно в офлайн-режиме без онлайн-экспериментов, а также использование фиксированного набора табличных признаков. Перспективными направлениями дальнейших исследований представляются: применение последовательностных и нейросетевых моделей (Quadrana и др., 2018:66) и многоязычных векторных представлений описаний событий; расширение погодного и социокультурного контекста; проведение онлайн-А/В-экспериментов для проверки переносимости офлайн-выводов; а также воспроизведение методики на данных других региональных платформ для оценки обобщаемости принципа «ценности афиши».

Литература

- Гомзин&Коршунов, 2012 - Гомзин А.Г., Коршунов А.В. Системы рекомендаций: обзор современных подходов // Труды Института системного программирования РАН. - 2012. - Т. 22. - С. 401-417. [Rus]
- Adomavicius&Tuzhilin, 2005 - Adomavicius G., Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions // IEEE Transactions on Knowledge and Data Engineering. - 2005. - Vol. 17, № 6. - P. 734-749. DOI: 10.1109/TKDE.2005.99 [Eng]
- Burges, 2010 - Burges C.J.C. From RankNet to LambdaRank to LambdaMART: An overview // Microsoft Research Technical Report MSR-TR-2010-82. - 2010. - 19 p. [Eng]
- Burke, 2002 - Burke R. Hybrid recommender systems: Survey and experiments // User Modeling and User-Adapted Interaction. - 2002. - Vol. 12, № 4. - P. 331-370. DOI: 10.1023/A:1021240730564 [Eng]
- Cormack и др., 2009 - Cormack G.V., Clarke C.L.A., Buettcher S. Reciprocal rank fusion outperforms Condorcet and individual rank learning methods // Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. - 2009. - P. 758-759. DOI: 10.1145/1571941.1572114 [Eng]
- Dacrema и др., 2019 - Dacrema M.F., Cremonesi P., Jannach D. Are we really making much progress? A worrying analysis of recent neural recommendation approaches // Proceedings of the 13th ACM Conference on Recommender Systems. - 2019. - P. 101-109. DOI: 10.1145/3298689.3347058 [Eng]
- Efron&Tibshirani, 1993 - Efron B., Tibshirani R.J. An Introduction to the Bootstrap. - New York: Chapman & Hall, 1993. - 436 p. [Eng]
- Hu и др., 2008 - Hu Y., Koren Y., Volinsky C. Collaborative filtering for implicit feedback datasets // Proceedings of the 8th IEEE International Conference on Data Mining. - 2008. - P. 263-272. DOI: 10.1109/ICDM.2008.22 [Eng]
- Ke и др., 2017 - Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y. LightGBM: A highly efficient gradient boosting decision tree // Advances in Neural Information Processing Systems. - 2017. - Vol. 30. - P. 3146-3154. [Eng]
- Koren и др., 2009 - Koren Y., Bell R., Volinsky C. Matrix factorization techniques for recommender systems // Computer. - 2009. - Vol. 42, № 8. - P. 30-37. DOI: 10.1109/MC.2009.263 [Eng]
- Lam и др., 2008 - Lam X.N., Vu T., Le T.D., Duong A.D. Addressing cold-start problem in recommendation systems // Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication. - 2008. - P. 208-211. DOI: 10.1145/1352793.1352837 [Eng]
- Macedo и др., 2015 - Macedo A.Q., Marinho L.B., Santos R.L.T. Context-aware event recommendation in event-based social networks // Proceedings of the 9th ACM Conference on Recommender Systems. - 2015. - P. 123-130. DOI: 10.1145/2792838.2800187 [Eng]
- Quadrana и др., 2018 - Quadrana M., Cremonesi P., Jannach D. Sequence-aware recommender systems // ACM Computing Surveys. - 2018. - Vol. 51, № 4. - Article 66. - P. 1-36. DOI: 10.1145/3190616 [Eng]
- Ricci и др., 2015 - Ricci F., Rokach L., Shapira B. (eds.) Recommender Systems Handbook. - 2nd ed. - New York: Springer, 2015. - 1003 p. DOI: 10.1007/978-1-4899-7637-6 [Eng]
- Salton&Buckley, 1988 - Salton G., Buckley C. Term-weighting approaches in automatic text retrieval // Information Processing & Management. - 1988. - Vol. 24, № 5. - P. 513-523. DOI: 10.1016/0306-4573(88)90021-0 [Eng]
- Sarwar и др., 2001 - Sarwar B., Karypis G., Konstan J., Riedl J. Item-based collaborative filtering recommendation algorithms // Proceedings of the 10th International Conference on World Wide Web. - 2001. - P. 285-295. DOI: 10.1145/371920.372071 [Eng]
- Schein и др., 2002 - Schein A.I., Popescul A., Ungar L.H., Pennock D.M. Methods and metrics for cold-start recommendations // Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. - 2002. - P. 253-260. DOI: 10.1145/564376.564421 [Eng]
- Trattner и др., 2018 - Trattner C., Oberegger A., Marinho L., Parra D. Investigating the utility of the weather context for point of interest recommendations // Information Technology & Tourism. - 2018. - Vol. 19. - P. 117-150. DOI: 10.1007/s40558-017-0100-9 [Eng]

References

- Gomzin&Korshunov, 2012 - Gomzin, A. G., & Korshunov, A. V. (2012). Sistemy rekomendatsiy: obzor sovremennykh podkhodov [Recommendation systems: a survey of modern approaches]. Trudy Instituta sistemnogo programmirovaniya RAN, 22, 401-417. [In Russ]
- Adomavicius&Tuzhilin, 2005 - Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749. [In Eng]
- Burges, 2010 - Burges, C. J. C. (2010). From RankNet to LambdaRank to LambdaMART: An overview (Microsoft Research Technical Report MSR-TR-2010-82). 19 p. [In Eng]
- Burke, 2002 - Burke, R. (2002). Hybrid recommender systems: Survey and experiments. User Modeling and User-Adapted Interaction, 12(4), 331-370. [In Eng]
- Cormack et al., 2009 - Cormack, G. V., Clarke, C. L. A., & Buettcher, S. (2009). Reciprocal rank fusion outperforms Condorcet and individual rank learning methods. In Proceedings of the 32nd International ACM SIGIR Conference (pp. 758-759). [In Eng]
- Dacrema et al., 2019 - Dacrema, M. F., Cremonesi, P., & Jannach, D. (2019). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In Proceedings of the 13th ACM Conference on Recommender Systems (pp. 101-109). [In Eng]
- Efron&Tibshirani, 1993 - Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. New York: Chapman & Hall. 436 p. [In Eng]
- Hu et al., 2008 - Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. In Proceedings of the 8th IEEE International Conference on Data Mining (pp. 263-272). [In Eng]
- Ke et al., 2017 - Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146-3154. [In Eng]
- Koren et al., 2009 - Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37. [In Eng]
- Lam et al., 2008 - Lam, X. N., Vu, T., Le, T. D., & Duong, A. D. (2008). Addressing cold-start problem in recommendation systems. In Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication (pp. 208-211). [In Eng]
- Macedo et al., 2015 - Macedo, A. Q., Marinho, L. B., & Santos, R. L. T. (2015). Context-aware event recommendation in event-based social networks. In Proceedings of the 9th ACM Conference on Recommender Systems (pp. 123-130). [In Eng]
- Quadrana et al., 2018 - Quadrana, M., Cremonesi, P., & Jannach, D. (2018). Sequence-aware recommender systems. ACM Computing Surveys, 51(4), Article 66, 1-36. [In Eng]
- Ricci et al., 2015 - Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2015). Recommender Systems Handbook (2nd ed.). New York: Springer. 1003 p. [In Eng]
- Salton&Buckley, 1988 - Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513-523. [In Eng]
- Sarwar et al., 2001 - Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web (pp. 285-295). [In Eng]
- Schein et al., 2002 - Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In Proceedings of the 25th Annual International ACM SIGIR Conference (pp. 253-260). [In Eng]
- Trattner et al., 2018 - Trattner, C., Oberegger, A., Marinho, L., & Parra, D. (2018). Investigating the utility of the weather context for point of interest recommendations. Information Technology & Tourism, 19, 117-150. [In Eng]