

## ПРОГНОЗИРОВАНИЕ LTV ПОЛЬЗОВАТЕЛЕЙ ОНЛАЙН-СЕРВИСОВ НА ОСНОВЕ ПОВЕДЕНЧЕСКИХ ДАННЫХ

Астапенкова А.В.<sup>1</sup> Берген Д. С.<sup>1</sup>

Научный руководитель – кандидат экономических наук, доцент Чуракова И. Ю.<sup>1</sup>

<sup>1</sup>Университет ИТМО

367832@niuitmo.ru

### Введение

В условиях высокой конкуренции показатель пожизненной ценности клиента (CLV) является одним из важнейших для оценки долгосрочной финансовой ценности клиентов и в значительной степени влияет на маркетинговые решения и оптимизацию распределения ресурсов [1]. Традиционный метод прогнозирования: модель RFM, фокусируется на анализе трех важных характеристиках взаимодействия с клиентами: давности, частоты и денежной стоимости. Вероятностные модели Pareto/NBD и BG/NBD применяются для прогнозирования будущих показателей покупок и вероятностей активности клиентов, в то время как модель Gamma-Gamma рассчитывает среднюю стоимость транзакции, однако предполагают постоянство поведения клиентов и не учитывают конкурентную динамику. Random Forest, градиентный бустинг, нейронные сети демонстрируют более высокую точность прогнозирования LTV, так как могут выявлять сложные нелинейные зависимости в больших данных [1, 3].

### Основная часть

Для изучения факторов, которые влияют на жизненный цикл пользователя, будет использоваться открытый набор транзакционных данных онлайн-платформы, содержащий информацию о покупательском поведении за многолетний период на протяжении 4 лет. На первом этапе проводится разведочный анализ данных с целью выявления распределений, аномалий и базовых закономерностей поведения пользователей. На втором этапе на основе RFM-анализа и feature engineering формируется расширенный набор поведенческих признаков, включающий динамические характеристики: межпокупочные интервалы, сезонность активности, разнообразие категорий покупок и активность пользователя в первые 90 дней [2]. Затем применяется кластеризация пользователей методом K-Means для сегментации по поведенческим паттернам [1]. На третьем этапе для прогнозирования LTV проводится сравнительный анализ классических вероятностных моделей (BG/NBD, Gamma-Gamma) с алгоритмами машинного обучения — градиентным бустингом (XGBoost, LightGBM) и случайным лесом (Random Forest) [1,3]. Качество моделей оценивается по метрикам MAE, RMSE и R<sup>2</sup>. Для проверки гипотез о статистически значимых различиях между сегментами пользователей применяются t-критерий и критерий Манна-Уитни [2]. На четвертом этапе интерпретация значимости признаков осуществляется методом SHAP (SHapley Additive exPlanations) [3].

### Выводы

Ожидается, что алгоритмы машинного обучения превзойдут классические вероятностные модели по точности прогнозирования LTV, а поведенческие паттерны — разнообразие категорий покупок, частота транзакций и сезонность активности — окажутся более значимыми предикторами ценности пользователя, чем демографические характеристики. Применение SHAP-анализа позволит количественно

оценить вклад каждого поведенческого признака в прогноз и выявить конкретные факторы, в наибольшей мере определяющие LTV пользователей онлайн-сервисов.

### Литература

1. Haddadi A. M., Hamidi H. A hybrid model for improving customer lifetime value prediction using stacking ensemble learning algorithm // *Computers in Human Behavior Reports*. 2025. (18). С. 100616.
2. Sun Y., Liu H., Gao Y. Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model // *Heliyon*. 2023. № 2 (9). С. e13384.
3. Yan Y., Resnick N. A high-performance turnkey system for customer lifetime value prediction in retail brands: Forthcoming in *quantitative marketing and economics* // *Quantitative Marketing and Economics*. 2024. № 2 (22). С. 169–192.