Применимость ИИ для прогноза популярности книги по начальным логам чтения

Исследовательский отчёт

Исполнитель: Боровинский Арсен Исаевич
ИТ-университет
Пермь, 2019 г.

Мотивация

- В образовании хайп по использованию ИИ для индивидуализации образовательной траектории
- Одна из подзадач: как узнать что созданный ресурс станет популярным и будет полезен при обучении

Задача

- Протестировать перспективность применения нейронных сетей для задачи предсказания популярности учебного ресурса в условиях ограниченных знаний об взаимодействии пользователя с ресурсом

Методы

- Обучение нейронной сети для задачи предсказания популярности книг в формате PDF, размещённых в электронной библиотеке вуза по логам чтения книг.

формулу для определения ускорения силы тяжести

$$g = 4\pi^2 \frac{l_{np}}{T^2} \qquad (2)$$

Период колебаний T физического маятника можно измерить непосредственно с помощью секундомера, а приведенную длину l_{np}

Кафедра общей физики ПГНИУ
Лаборатория механики и молекулярной физики
Лабораторная работа № 109

непосредственно измерить нельзя. Поэтому необходимо выразить приведенную длину через величины, доступные прямому измерению. По теореме Гюйгенса-Штейнера момент инерции I маятника относительно любой оси подвеса равен

$$I = I_0 + ml^2 \qquad (3)$$

где I_0 – момент инерции маятника относительно оси, параллельной оси подвеса и проходящей через центр инерции маятника, l – расстояние от оси подвеса до центра инерции. Подставив (3) в формулу (1), получим

Методика исследования

Есть 500 тыс. логов чтения:

- h – час открытия книги;
- p – число прочитанных страниц;
- t – общее время пока документ открыт;
- nid – идентификатор книги;
- s – статус доступа (1 – доступен, 0 – не доступен)

1	6	ebook	510735	История зарубежной литературы. Современная английская литература	29
1	2	ebook	380407	№110 Изучение собственных колебаний на примере пружинного маятника	29
1	101	ebook	93398	Список населенных мест Пермской губернии. Верхотурский уезд.	1261
1	1	ebook	76163	Очерк состояния кустарной промышленности в Пермской губернии.	27554
1	4	ebook	513758	Поиски и разведка месторождений полезных ископаемых. Т. 1 Прогнозирование и поиски месторождений	60

Выходной нейрон

Популярность книги описывается рейтингом R, посчитанному как сумма сессий для конкретной книги за всё время существования книги и являющейся суммой с весовыми коэффициентами логов по книге

$R(nid) = \Sigma(1 + 0.1*p)$, где сумма берётся по всем логам для книги с идентификатором nid.

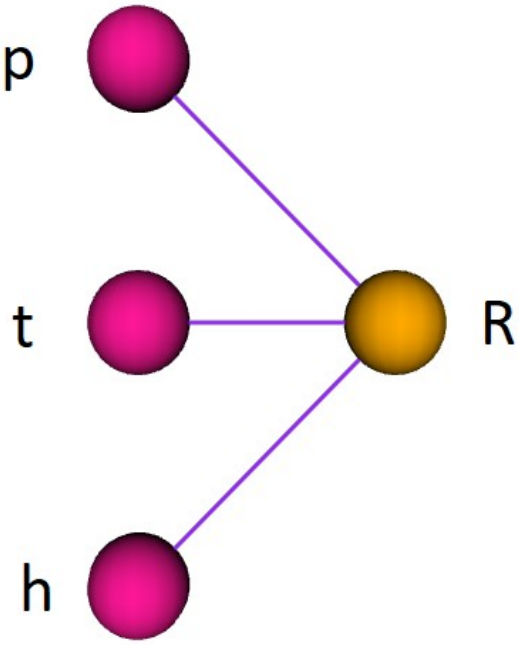
Модель 1

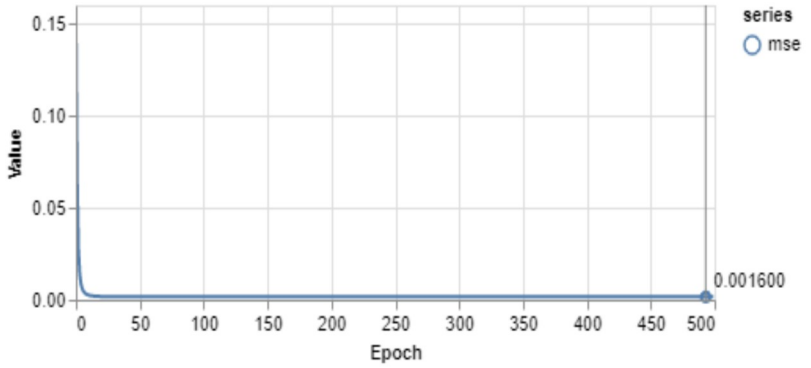
Входные нейроны:

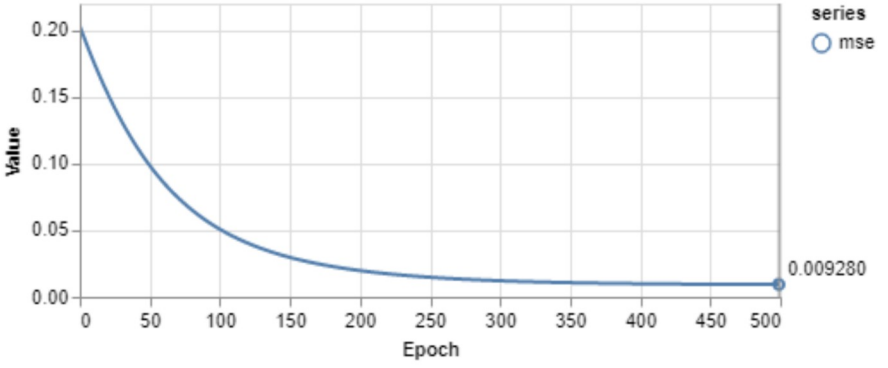
- p – число прочитанных страниц;
- t – общее время пока документ открыт;
- h – час открытия книги.

Выходной нейрон:

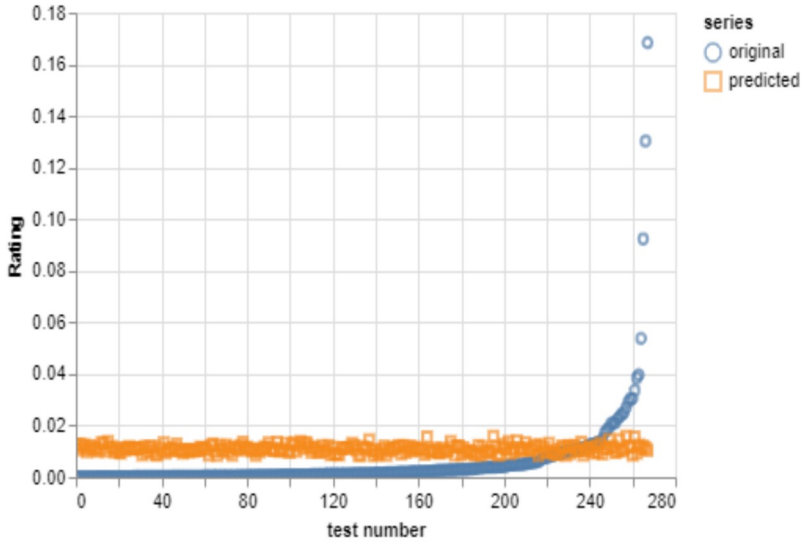
- R – рейтинг (популярность) книги.



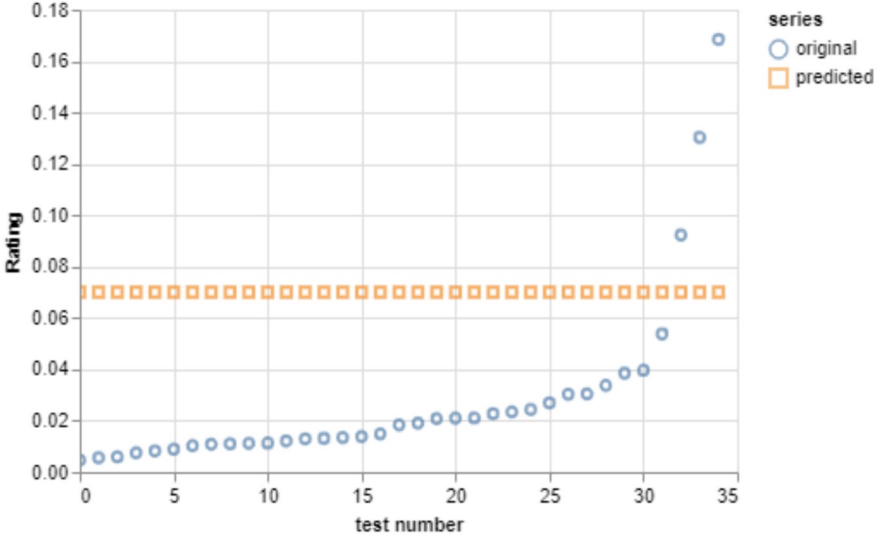




Model Predictions vs Original Data



Model Predictions vs Original Data



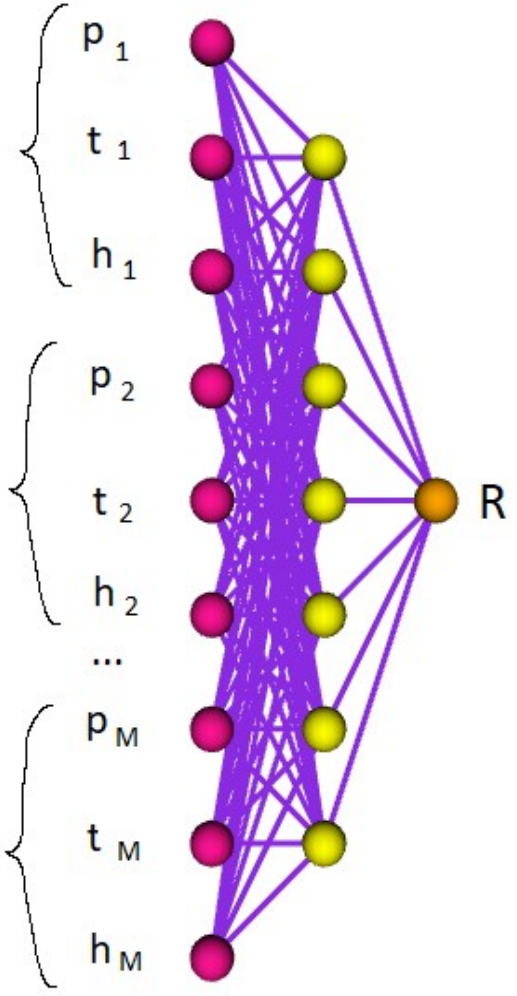
Модель 2

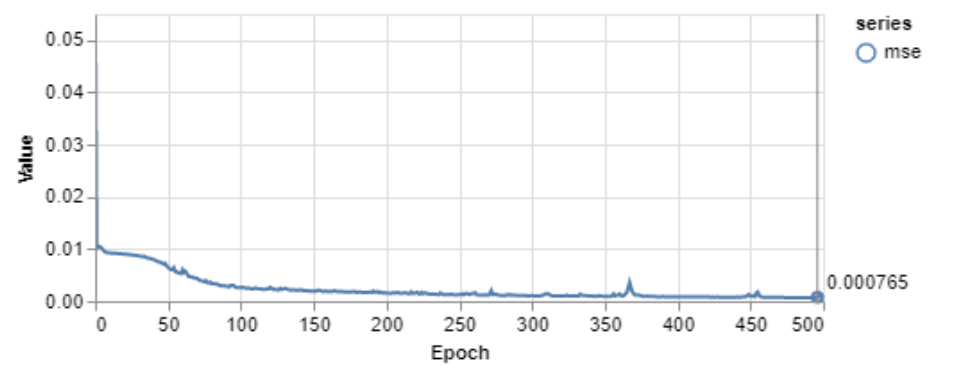
Сгруппированы логи по книгам (nid).

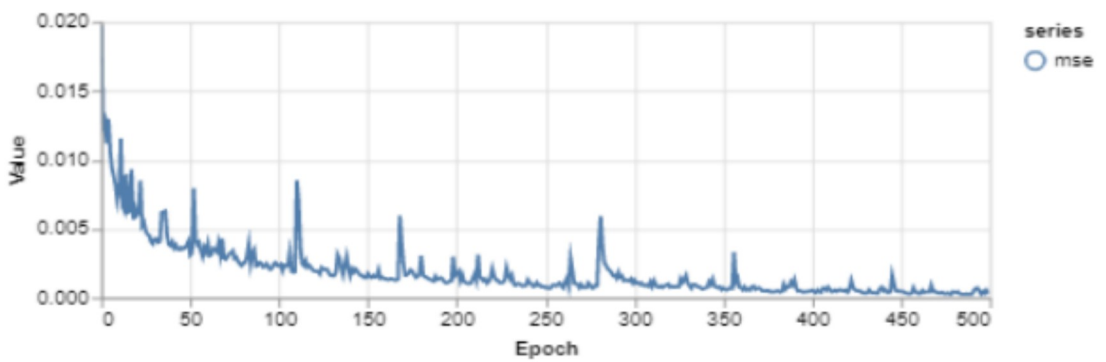
Обучение по M первых логов для книги с идентификатором nid.

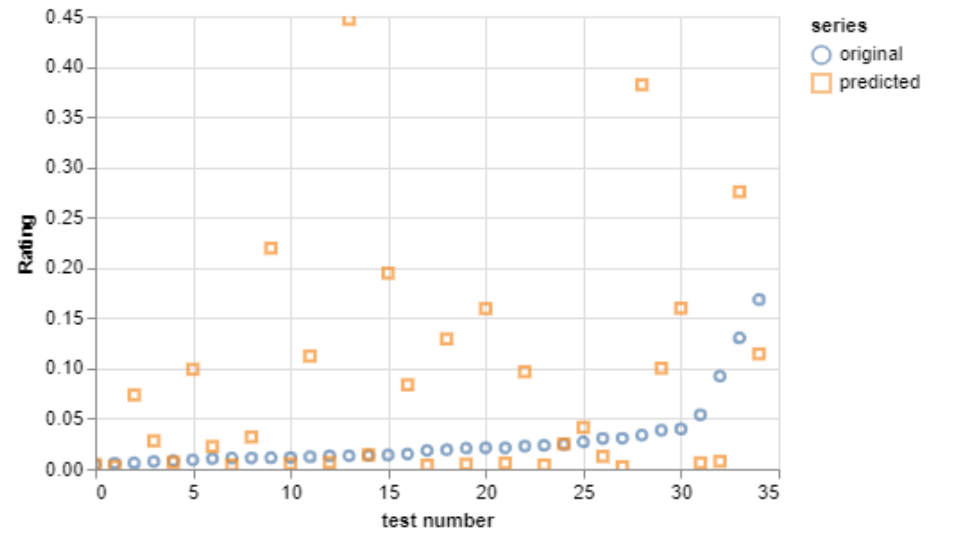
Выходной нейрон:

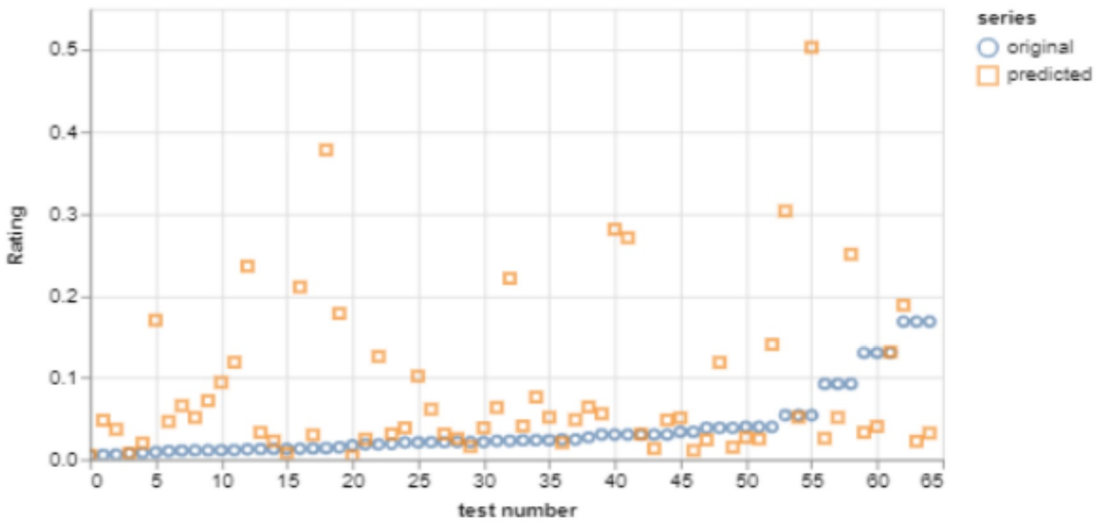
- R – рейтинг (популярность) книги.











6 скрытых нейронов

10 скрытых нейронов

Модель 3

Сгруппированы логи по книгам и берутся только такие логи, когда точно было взаимодействие с книгой

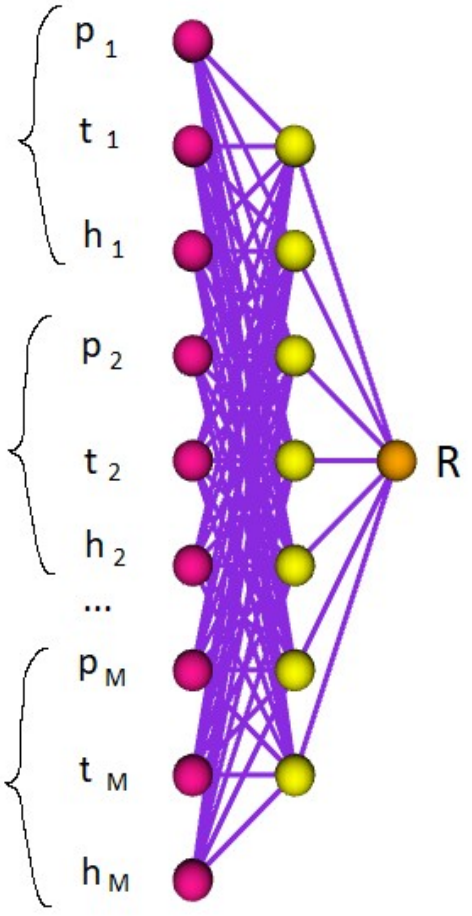
- 1 < p < 20;
- t > 15 сек;
- логи берутся только в течении 2018 года для книг, загруженных раньше 2018 года;
- R перерасчитывается на логах 2018 года;
- Логи с R > 0.2 * Max (R) откидываются.

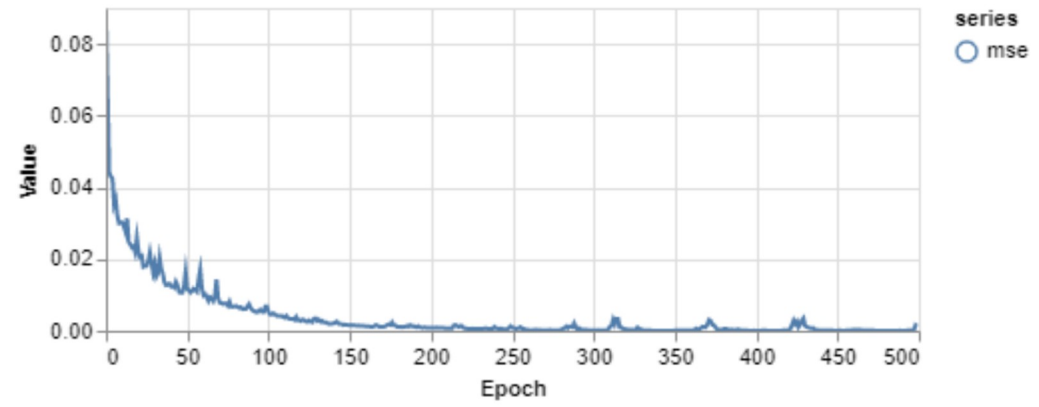
10 < M < 100.

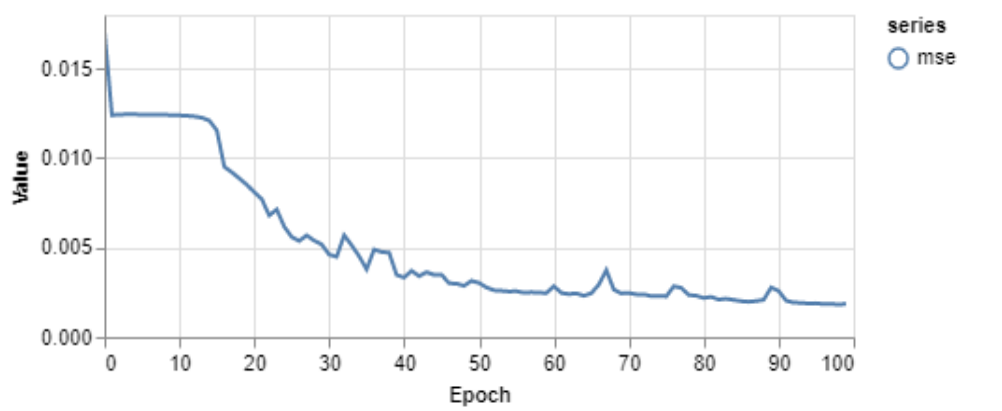
Один скрытый слой: 6-10 нейронов.

Выходной нейрон:

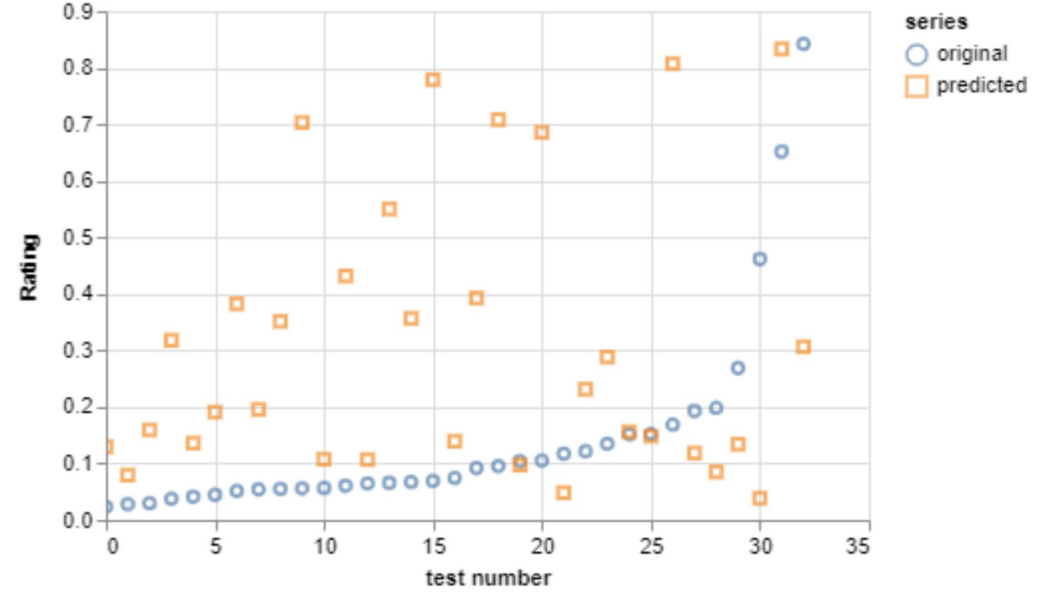
- R – рейтинг (популярность) книги на основе логов 2018 года.



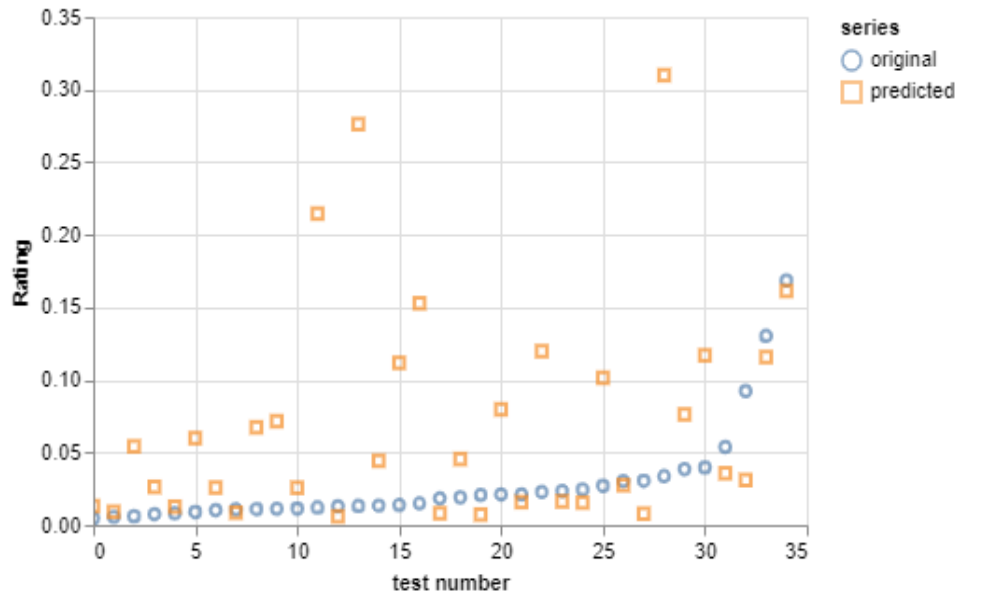




Model Predictions vs Original Data



Model Predictions vs Original Data



Выводы

- Гипотезу о возможности предсказания популярности (рейтинга) электронной книги по М актам взаимодействия с книгой подетвердить не удалось.

Возможные причины

- Отсутствет связь популярности с подробностью чтения;
- Данные не кластеризованы;
- Необходим более детальный трекинг.