

1. В команде есть два стажера-аналитика. Правильный ответ каждый из стажеров получает в 14 из 17 случаев. Чтобы быть более уверенным в важном решении, менеджер решил дать одну и ту же задачу сразу обоим аналитикам: если оба получают одинаковый ответ, то менеджер его использует, а если ответы разные - выберет один из них наугад. Насколько такой способ повышает шансы менеджера принять верное решение?

Вероятность, что оба ответят верно: $(14/17) * (14/17) = 196/289 \approx 0.6782$

Вероятность, что оба ответят неверно: $(3/17) * (3/17) = 9/289 \approx 0.0311$

Вероятность, что ответы будут разными: $1 - (196/289) - (9/289) = 84/289$ (ДЛЯ 2 ВАРИАНТОВ ГДЕ 1 ответил верно и второй не верно и для варианта, где первый неверно и второй верно)

В этом случае менеджер в половине случаев выберет верный вариант, то есть $84/289 / 2 = 42/289$

Общая вероятность принять верное решение = $196/289 + 42/289 = 0,9688$

И найдем насколько улучшилась вероятность

$0,9688 - 0,8235 = 0,1453$ или $14,53\%$

2. В мешке лежат три кубика: 6-гранный, 12-гранный, 20-гранный. Мы достали один кубик на удачу, подкинули его, и на нем выпало число N. Какова вероятность того, что если мы достанем еще один кубик из оставшихся в мешке и подкинем его, на нем выпадет меньше N?

1) $N = 12$

2) $N = 4$

1.1 Тут задача на условную вероятность, потому что нам нужно найти вероятность что выпало число менее N при условии того, что выпало число 12

P (вероятность что кубик с 6 гранями, при условии что выпало 12) = 0

КУБ С 12 ГРАНЯМИ

$P(H) = 1/3$ (Вероятность того, что кубик с 12 гранями и выпало 12)

$P(E|H) = 1/12$

$P(-H) = 2/3$ (Вероятность что не выпадет куб с 12 гранями)

$P(E|-H) = 1/20$

$P(E) = (1/3 * 1/12) + (1/3 * 1/20) + (1/3 * 0) = 2/45$ (вероятность что вообще выпадет 12, вне зависимости от того будет событие H или нет)

$P(\text{Кубик 12} | \text{Выпало 12}) = (1/3 * 1/12) / 2/45 = 0,625$

Можно сразу найти вероятность того, что выпадет кубик с 20 гранями = $1 - 0,625 = 0,375$

Так как на кубике с 6 гранями не могло выпасть число 12, следовательно эта вероятность равна 0

Чтобы найти вероятность выпадения числа менее 12 воспользуемся декомпозицией сначала

Выпал куб с 12 гранями

$$0,625 * (1/2 * 11/20 + 1/2 * 6/6) = 31/64$$

Выпал куб с 20 гранями

$$0,375 * (1/2 * 11/12 + 1/2 * 6/6) = 23/64$$

Найдем полную вероятность

$$P(x) = 31/64 + 23/64 = 27/32 \text{ или } 0,84375 \text{ или } 84\%$$

1.2 Найдем вероятность что число, выпавшее на новом кубике менее $N = 4$

Механизм тот же, расписывать не буду, нам нужно посчитать изначально вероятность что на кубике с 6 гранями выпало число 4, потом на кубике с 12 гранями выпало 4 и на кубике с 20 гранями выпало 4. Делаем это также по формуле Байеса

Найдем $p(E)$ -вероятность что выпало 4

$$P(E) = (1/3 + 1/6) + (1/3 * 1/12) + (1/3 * 1/20) = 0,1$$

Найдем вероятность для 6

$$(1/3 * 1/6) / 0,1 = 5/9$$

Найдем вероятность для 12

$$(1/3 * 1/12) / 0,1 = 5/18$$

Найдем вероятность для 20

$$(1/3 * 1/20) / 0,1 = 1/6$$

Для кубика с 6 гранями

$$5/9 * (1/2 * 3/12 + 1/2 * 3/20) = 16/144$$

Для кубика с 12 гранями

$$5/18 * (1/2 * 3/6 + 1/2 * 3/20) = 13/144$$

Для кубика с 20 гранями

$$1/6 * (1/2 * 3/6 + 1/2 * 3/12) = 9/144$$

Найдем общую вероятность

$$16/144 + 13/144 + 9/144 = 19/72$$

Или если грубо округлить 0,26 или 26%

3. Чтобы между пользователями Какого-то сайта) было больше доверия, а жизнь мошенников стала сложнее, мы решили попробовать ввести систему отзывов: покупатель может оставить отзыв на продавца. Отзыв может быть просто рейтингом (1-5 звездочек), а может содержать дополнительно какой-то произвольный текст.

а. Предложите метрики и поясните, каким образом их можно использовать для отслеживания прогресса и успешности проекта?

1 Кол-во отзывов о дне, неделю, месяц.

Эта метрика показывает, как сильно юзеры доверяют системе оценок. Чем число выше, тем люди больше доверяют системе

2 Процент того, как часто юзеры оставляют отзыв по сравнению с числом покупок на сайте

Такая же метрика, как и первая, но она более защищена от влияния накрутки, потому что накрутки не будут покупать товар, а потом оставлять плохой отзыв

3 Возможно кол-во слов в отзыве и процент оригинальности для каждого отзыва.

Я имею в виду что отзывы, где просто написано «Плохо», «Не рекомендую» и тд, вполне могут быть накруткой

4 Сколько акков имеют оценку 1, 2, 3 и тд

Данная метрика покажет какие пользователи преобладают на платформе, если этот показатель пойдет вверх, то можно будет сделать о том, что новая система помогла улучшить сервис

5 Время, которое пользователь в среднем тратит на просмотр вакансии или на прочтение отзывов, сделать кнопку посмотреть отзывы и начать смотреть как долго пользователи в среднем / медианно читают отзывы, чем более высок этот рейтинг, тем больше вероятность что проект успешен

6 Кол-во фейковых отзывов

7 Насколько снизился уровень жалоб в Тех поддержку по поводу мошенников, после старта проекта.

Объясняю, теперь пользователь смотрит на отзывы и понимает, что много плохих комментов и не покупает ничего на данном объявлении, следовательно с большей вероятностью проходит мимо мошенника

8 Как быстро продавец реагирует на плохие отзывы, мошенники как правило не будут часто отвечать на них и тем более предложить возврат средств, к примеру и другую помощь пользователю

9 Можно еще сделать опрос, как вам новая система оценивания. Тут все просто, чем больше хороших оценок, тем лучше это изменение)

б. Поскольку Какой сайт) не магазин, а площадка для связи покупателя и продавца, они в большинстве случаев не знаем, произошла ли в действительности сделка и на каких условиях. Для большинства сделок последнее, что нам известно - покупатель нажал кнопку просмотра телефона продавца или связался с ним в чате. Также мы всегда знаем логин (привязанный к email и телефону) продавца, но покупатель может быть незалогиненным. В связи с этим кажется, что есть большой риск накрутки отзывов и недобросовестного использования системы: например, профессиональные участники будут пытаться оставлять негативные отзывы на своих конкурентов и хвалебные на себя. Предположим, что система некоторое время уже работает и у нас есть данные по отзывам и всей активности клиентов: что продавал, что искал, на каких объявлениях смотрел телефоны и т.п. Как оценить масштабы накрутки, т.е. долю фальшивых отзывов?

1 Я бы посмотрел на то сколько пользователь оставляет отзывов за день/неделю/месяц

Если это очень большое число, то можно сделать вывод, что это накрутка

2 Как много времени человек провел на объявлении перед тем, как написать отзыв. Если время короткое, то скорее всего это мошенник

3 Я бы посмотрел, как пользователь пишет отзыв. Если они одинаковые почти что, то это накрутка

4 Если у какого-то аккаунта в среднем было 2–3 отзыва в день, а потом резко пришло 100–200. Это также говорит о накрутке

5 Смотреть на IP, user_agent, cashe пользователя, плюс пользователь может анонимизировать трафик используя сторонние сервисы Okto browser, dolphin и тд

6 Соотношение покупок и кол-ва оценок/отзывов. Если число отзывов растет, а кол-во покупок нет, это говорит о накрутке

7 Можно посмотреть на то, через какое время после регистрации пользователь начинает писать отзывы, ставить оценки, если сразу же, то это накрутка

8 Я бы дополнительно посмотрел на время в которое пишет пользователь, основываясь на его IP, если пользователь пишет в 2-3 ночи много отзывов, это вызывает подозрения

9 Кол-во жалоб от продавцов на фейковый отзыв