

User cases при использовании моделей

Для разработки моделей была выбрана “категория” акций: компании, которые вкладываются в технологии ИИ и/или генетические. Был разбит набор данных на train/test субъективно.

1. Применение модели дало “открытие”

Пользователь не обратил внимание, но модель нашла (в данном случае компанию CIAN — неочевидно, но компания занимается разработками в ИИ области, скриншот из Интернет прилагается).



Циан запустил умного помощника с применением...

[cian.ru](#) › Новости › Новости компании 

ЦИАН - новость о рынке недвижимости от 2024-04-16 - **Циан** запустил умного помощника с применением технологий Искусственного интеллекта...

2. Расширение активов

За счет высокого recall модель удачно включила в результаты акции подходящих компаний, которых не было в тренировке (компании без описаний, только с категорией от брокера):

- GEMC ЕвроМедЦентр (скриншот прилагается: “ДНК” — генетика);
- HHRU HeadHunter: очевидно, применяет ML/AI в работе.

Европейский Медицинский Центр EMC - частная платная...

emcmos.ru

Диагностика ДНК и индивидуальное лечение: почему онкология — не приговор.

Грубо, доля верных ответов

Приблизительно оценим вручную долю верных ответов в inference модели:

в результатах inference выдал 23 объекта:

4 явно неподходящих (авиакомпания, Русолово, алюминий, Мостотрест).

Получается грубо accuracy $\approx 0,83$.

Обновление модели (Модель 1Б)

После изучения полученных результатов набор помеченных данных train/test был дополнен (компанией CIAN). Поскольку размеченный набор данных очень маленький (особенно по положительному классу), то каждый объект имеет значение.

Грубо, доля верных ответов

Приблизительно оценим вручную долю верных ответов в inference модели Модель 1Б:

в результатах inference выдал 18 объектов:

1 явно неподходящая (Мостотрест).

Получается грубо ассигасу $\approx 0,94$.

При этом из результатов не пропала ни одна компания, в которой есть высокая уверенность принадлежности к категории, исчезли некоторые компании “под вопросом” (мало информации, чтобы понять их принадлежность).

Результаты

Результаты: “открытий” больше не стало, но субъективно улучшилась доля верных ответов, модель стала более прицельной.

Практическое использование модели в течение нек. времени показало, что 100 % качество не достигается (и пользователь может отметить единицы объектов), но модель полезна. Можно сравнить с моделями распознавания пальцев или лиц: работает не 100 %, но в работе помогает.

Использование модели позволяет

- расширить число акций в активе, потому что модель фильтрует акции беспристрастно и внимательно с высоким покрытием, исключается человеческий фактор;
- пользователь автоматически получит информацию о новых появляющихся на бирже компаниях, которые будут захвачены классификатором;
- исследовать рынок по выбранному категориальному критерию, потому что модель может «заметить» акции, на которые не обратил внимание человек (например, в случаях, когда неочевидно, что акция на самом деле подходит под критерии, а модель такие случаи может обнаружить) и привести к некоторым «открытиям»;
- также, модель экономит время при многократном/периодическом использовании inference, потому что человеку пришлось бы неоднократно просматривать и анализировать информацию по всем акциям, на такую работу ушло бы больше времени.