

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

Хорошо



Михаил Шардин ★

личный блог



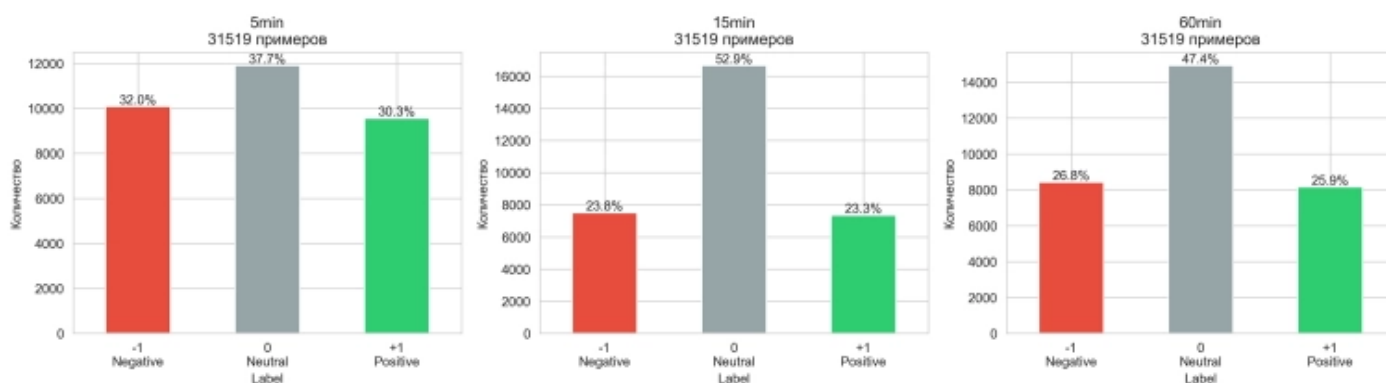
17 октября 2025, 06:35

+ Подписаться

## Как я погружаюсь в ML и понимаю, что всё начинается с данных

Экспериментирую с ML. Несколько недель ковыряюсь в данных и всё больше понимаю — алгоритмы это не главное. Главная боль — подготовка данных.

Баланс классов по временным периодам



Уже несколько недель разбираюсь с машинным обучением. Не ради статьи на Смартлабе — захотелось понять, как это устроено изнутри. И чем глубже погружаюсь, тем больше понимаю: вся «магия» моделей начинается задолго до самого обучения.

Собрать данные — не проблема. Подготовить их — проблема. И особенно сложно — сбалансировать классы.

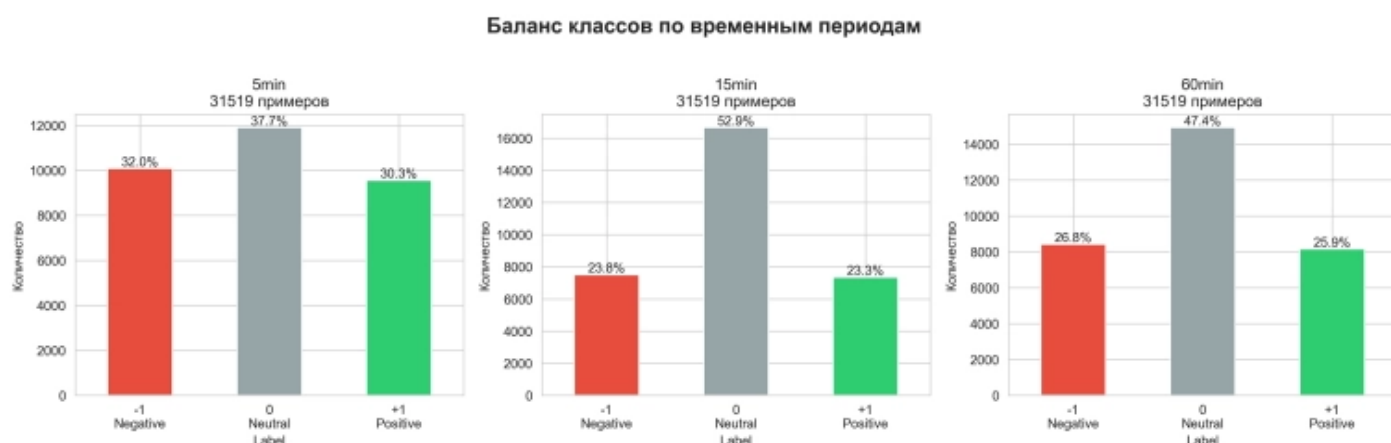
Когда модель учится отличать категории, она должна видеть их примерно в равных долях. Если один класс встречается гораздо чаще других, модель быстро «разучивается думать» и начинает просто угадывать самый частый вариант. В итоге вроде бы всё обучилось, но результат — в мусорку.

На практике это выглядит так:

Откройте счёт в ВТБ Мои Инвестиции

Введите текст комментария

После долгих экспериментов и отбраковки однотипных записей я получил такую картину:



Выглядит вроде бы прилично, но за этими цифрами — десятки вариантов фильтрации, пересемплирования и чистки.

Главная проблема — нейтральные примеры. Их очень много, и именно они чаще всего «смазывают» сигналы в данных.

Удалить слишком много — модель потеряет способность видеть естественные переходные состояния.

Удалить слишком мало — она начнёт всё считать нейтральным.

Это [в продолжение темы](#).

Поэтому подбор баланса — это не механическая операция, а прямо сложно.

Ведь машинное обучение — это не про «нажал кнопку и получил результат» как в ChatGPT. Это скорее про методичное, почти исследовательское выравнивание весов, классов и смыслов.

**Если кто-то тоже бьётся с дисбалансом классов — делитесь опытом в комментариях.**

Особенно интересны реальные кейсы, где удалось найти золотую середину между балансом и сохранением репрезентативности данных.

**Автор:**

Михаил Шардин

[Моя онлайн-визитка](#)

[Telegram «Умный Дом Инвестора»](#)

17 октября 2025

Откройте счёт в ВТБ Мои Инвестиции



## Михаил Шардин

📍 Пермь

👤 314 📊 2 949

📅 с 23 января 2019

🔗 +HreHDn1F5CZjN...

+ Подписаться

### 24 КОММЕНТАРИЯ

Сначала старые ▾



Михаил

17 октября 2025, 07:00



Никак, потому что не имею дело с классами, а предсказываю распределение доходности

🔗 👍 +1 💬



Ho\_Chu

17 октября 2025, 07:20



рискнул бы не согласиться с тезисом о неважности алгоритмов

взять хотя бы вчерашний день

из-за произошедших событий придется заново обучать некоторые алгоритмы, хотя другие демонстрируют поразительную стойкость к такого рода явлениям, при этом все используют почти один и тот же входной набор данных

но, кажется мне, что даже новое переобучение не поможет... а вот почему не поможет, — это надо сильно подумать

🔗 👍 +1 💬



Максим Павлов

17 октября 2025, 09:54



Смотря какой алгоритм используешь. Но почти в любом алгоритме машинного обучения есть параметр по типу «scale\_pos\_weight», который делает больше веса в обучении на примеры минорного класса. Но лучший вариант — это добрать выборки)

🔗 👍 +1 💬

Откройте счёт в ВТБ Мои Инвестиции

Ещё 7 комментариев

Напишите комментарий...



ОТПРАВИТЬ

## Читайте на SMART-LAB:

Рейтинг кредитоспособности	BB-.ru со стабильным прогнозом от <a href="#">RUS</a>
Объем выпуска	250 000 000 рублей
Ставка купона	- 25% на первый год - 24% на второй год
УТМ	26,92%
Срок обращения / duration в оферте PUT	4,5 года / 1,64 года
Оферта PUT	Через 2 года обращения
Оферта CALL	Через 2,25 года обращения
Купонный период	91 день
Номинал серии облигаций	1 000 рублей
Ориентир даты размещения	23 октября 2025
Организатор / Андеррайтер	ООО ИК «Иволга Капитал»
Основные финансовые показатели эмитента	
Источник: отчетность РСБУ, млн рублей	
Выручка LTM	3 323
Чистая прибыль LTM	456

### Обновление параметров выпуска РДВ Технолоджи (BB-.ru, YTM 26,92%)

Обновленные предварительные параметры : — Кредитный рейтинг: BB-.ru со стабильным прогнозом — 250 млн р. — 4,5 года...



Иволга Капитал

21.10.2025



### Золото и серебро — пересмотр целей

В стратегии на IV квартал 2025 года аналитики «Финама» указывали цель для цены на золото на уровне \$4200 за тройску...



Финам Брокер

21.10.2025

### Автоконвертация расписок "Циана"

Компания «Циан» объявила о старте автоматической конвертации депозитарных расписок CIAN PLC в акции российского МКАО «ЦН». Разбираемся, как это повлияет на инвесторов и на что...



НДФЛка

21.10.2025

Откройте счёт в ВТБ Мои Инвестиции

[О смартлабе](#)[Реклама](#)[Полная версия](#)

Московская Биржа является спонсором ресурса smart-lab.ru  
Источник: ПАО Московская Биржа

Откройте счёт в ВТБ Мои Инвестиции