آموزش پروژه‌محور پردازش صوت و گفتار با پایتون

فصل دوم: ماشین لرنینگ روی صدا

درس 6

ورود به دنیای احساسات

شروع پروژه ۱

طراح و مدرس: مصطفی کرمانی‌نیا

تحت نظارت علمی: دکتر مهدی سیفی‌پور

Made with GAMMA

ماشینِ تشخیص احساسات

خواندن ذهن از روی صدا

تا اینجا به کامپیوتر یاد دادیم صدا را «بشنود». حالا می‌خواهیم یاد بدهیم آن را «حس کند»! آیا ماشین می‌تواند بفهمد یک نفر عصبانی است یا فقط دارد بلند حرف می‌زند؟

هدفِ پروژه اول ما

می‌خواهیم یک ارتش از داده‌های صوتی را پردازش کنیم و اولین هوش مصنوعی «تشخیص احساسات» (Emotion Recognition) را بسازیم.

Made with GAMMA

نقشه راه: ساخت کارخانه داده

ایستگاه اول: RAVDESS
آشنایی با دیتابیس بازیگران احساساتی

ایستگاه دوم: پردازش گروهی
استخراج جادوی MFCC از صدها فایل با حلقه For

ایستگاه سوم: X و Y
تفکیک ویژگی‌ها (صدا) و برچسب‌ها (احساس)

ایستگاه چهارم: آماده‌سازی نهایی
نرمال‌سازی و تقسیم داده‌ها برای جلوگیری از تقلب ماشین

ایستگاه پنجم: کدتایم!
شروع برنامه‌نویسی پروژه

Made with GAMMA

دیتاست RAVDESS

صداهای استاندارد شده

۲۴ بازیگر حرفه‌ای، جملات کاملاً یکسانی را با احساسات مختلف (شادی، خشم، ترس، غم و...) بیان کرده‌اند.

چرا این دیتاست فوق‌العاده است؟

چون کلمات در همه فایل‌ها یکسان است! هوش مصنوعی مجبور می‌شود فقط روی «لحن» و «اثر انگشت صوتی» تمرکز کند، نه معنی کلمات.

کدهای پنهان در نام فایل‌ها

هر فایل یک اسم رمزی دارد (مثلاً 03-01-05). احساسِ آن صدا در همین اعداد پنهان شده که در پایتون استخراج می‌کنیم.

دانلود با اینترنت ملی

github.com/ZenvilleErasmus/RAVDESS-emotions-speech-audio-only

Made with GAMMA

اتوماسیون استخراج

پردازش دستی ممنوع!

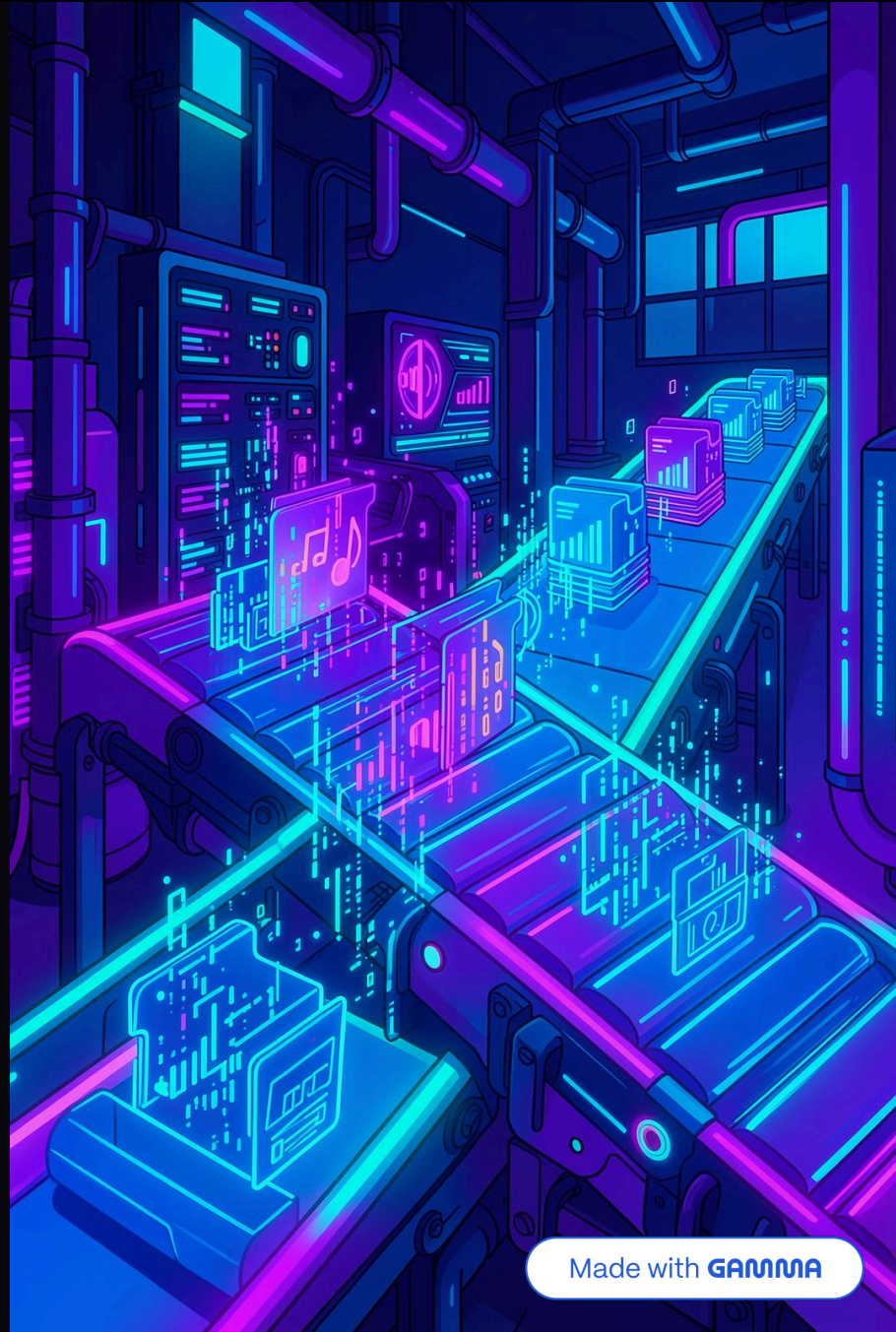
نمی‌توانیم صدها فایل صوتی را یکی‌یکی به Librosa بدهیم. برای تغذیه ماشین، نیازمند اتوماسیون هستیم.

حلقه استخراج ویژگی‌ها

یک حلقه For می‌نویسیم که با سرعت بالا وارد پوشه‌ها شود، فایل‌ها را بخواند، MFCC و ZCR و Spectral Centroid را استخراج کند و در یک لیست بزرگ ذخیره کند.

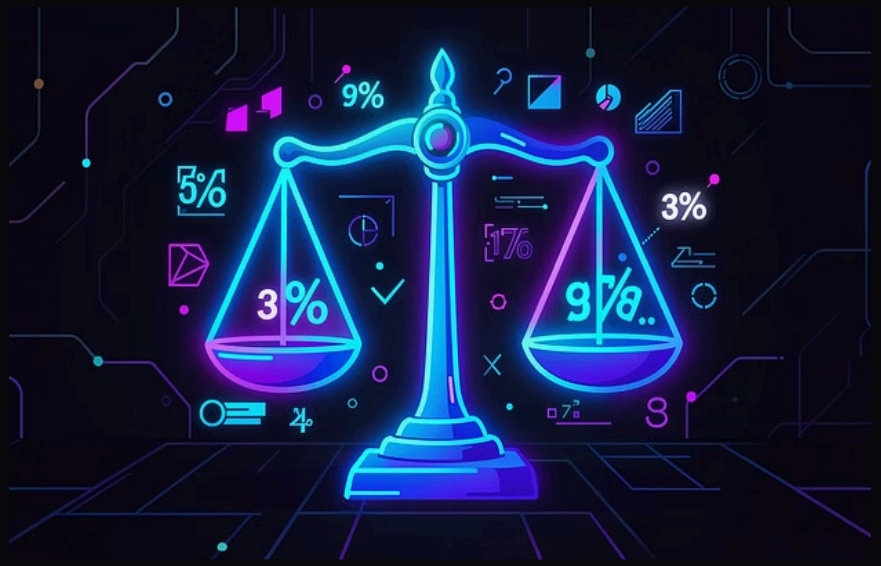
تبدیل صوت به ماتریس

کل دیتاستِ صوتی تبدیل به یک جدول غول‌پیکر از اعداد (آرایه NumPy) می‌شود که آماده تزریق به الگوریتم است.



Made with GAMMA

تعادل در دنیای اعداد: نرمال‌سازی



نقش ویژگی‌ها (X) و برچسب‌ها (Y)

ماتریس X همان اعداد استخراج‌شده (MFCC و ZCR و Spectral Centroid) هستند و بردار Y همان احساسات (شادی، خشم).

مشکل مقیاس‌ها در هوش مصنوعی

بعضی اعداد ریاضی خیلی بزرگ و بعضی خیلی کوچکند. مدل‌های یادگیری ماشین به اعداد بزرگ‌تر، ناخودآگاه وزن و اهمیت بیشتری می‌دهند!

راه‌حل: ابزار StandardScaler

همه ویژگی‌ها را هم‌قد و هم‌مقیاس می‌کنیم تا ماشین با عدالت کامل تصمیم بگیرد و در یادگیری الگوها دچار خطا نشود.

Made with GAMMA

روز امتحان: Train-Test Split

جلوگیری از تقلب ماشین!

اگر تمام ۱۰۰٪ فایل‌ها را به ماشین بدهیم، او جواب‌ها را «حفظ» می‌کند (Overfitting) و قدرت تحلیلش را از دست می‌دهد.

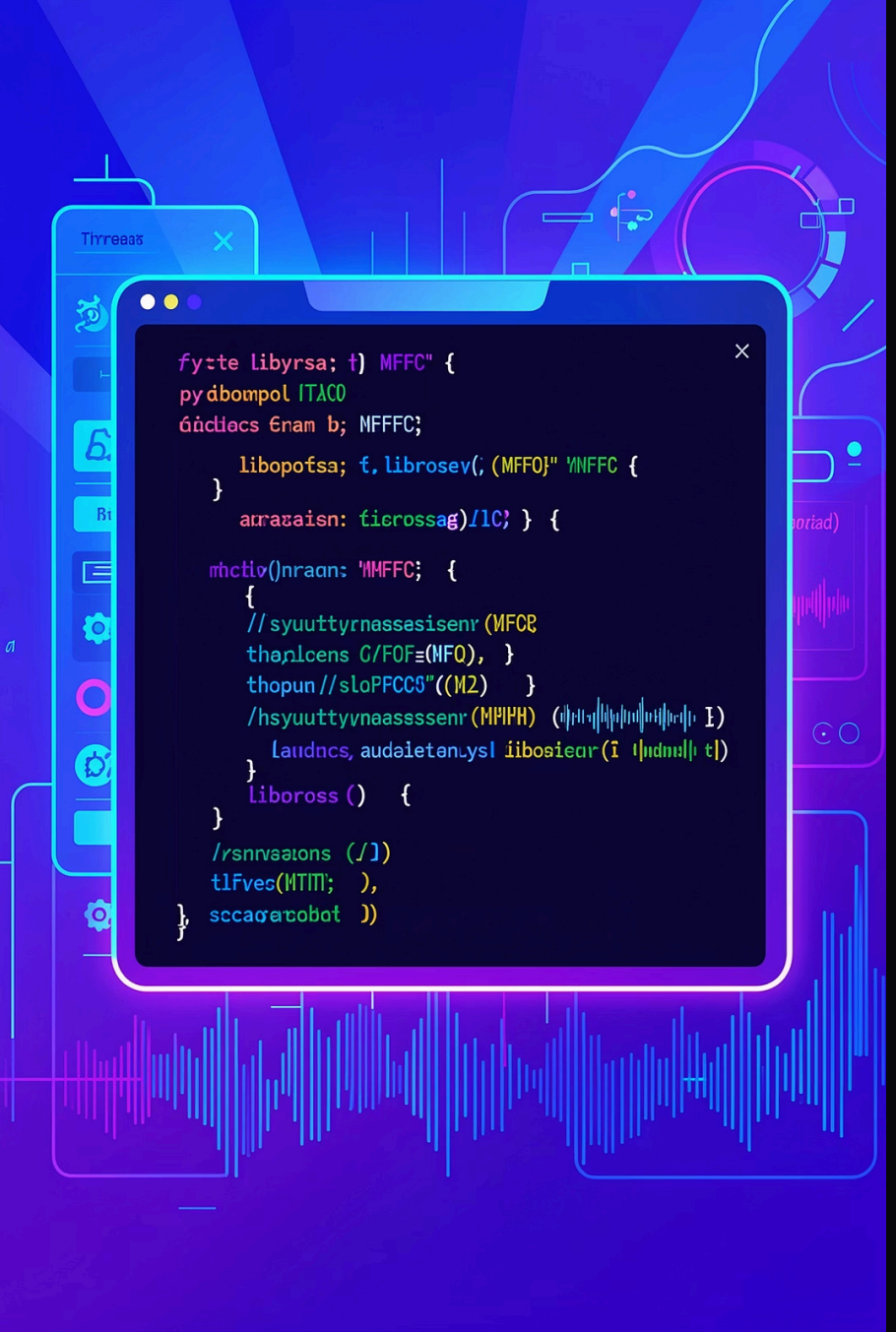
آموزش (Train Data) — ۸۰٪

معمولاً ۸۰٪ از فایل‌های صوتی را به ماشین می‌دهیم تا الگوهای خشم، شادی و غم را در آن‌ها کشف کند.

آزمون (Test Data) — ۲۰٪

۲۰٪ از داده‌ها را قایم می‌کنیم. بعد از آموزش، این فایل‌های کاملاً جدید را به ماشین می‌دهیم تا در دنیای واقعی از او امتحان بگیریم!

Made with GAMMA



کدتایم!

شروع پروژه: پردازش گروهی داده‌ها

لود کردن دیتاست RAVDESS

استخراج برچسب احساسات از نام فایل‌ها با پایتون

حلقه For و MFCC

نوشتن اسکریپت اتوماسیون برای پردازش صدها فایل

تقسیم و نرمال‌سازی

استفاده از Scikit-Learn برای StandardScaler و Train-Test Split

Made with GAMMA