

محمد جواد رنجبر

810101173

پروژه امتیازی درس پردازش گفتار:

تشخیص احساسات زبان فارسی با استفاده از یادگیری عمیق

دکتر ویسی

بهار 1403

Contents

[گزارش فایل‌ کد ساده 4](#_Toc171279878)

[لود کردن کتابخانه‌ها و دانلود داده‌ها 4](#_Toc171279879)

[حالت oversample کردن 17](#_Toc171279880)

[آموزش سه کلاسه 17](#_Toc171279881)

[Undersample کردن 18](#_Toc171279882)

[آموزش با دیگر مجموعه داده‌ها 18](#_Toc171279883)

[تصویر 1 نصب کتابخانه و مجموعه داده 4](#_Toc171279891)

[تصویر 2 به دست آوردن طول داده‌ها 4](#_Toc171279892)

[تصویر 3 تعداد از داده‌ها 5](#_Toc171279893)

[تصویر 4 ویژگی‌های اماری دیتاست 5](#_Toc171279894)

[تصویر 5 توزیع جنسیت 5](#_Toc171279895)

[تصویر 6 توزیع کلاس‌ها 5](#_Toc171279896)

[تصویر 7 هیستوگرام داده‌ها 6](#_Toc171279897)

[تصویر 8 هیستوگرام برای هر کلاس‌ 6](#_Toc171279898)

[تصویر 9 هیستوگرام طول صوت‌ها 7](#_Toc171279899)

[تصویر 10 باکس پلات داده‌ها 7](#_Toc171279900)

[تصویر 11 باکس پلات جنسیت‌ها 8](#_Toc171279901)

[تصویر 12 داده‌های آموزش، ارزیابی و آزمون 8](#_Toc171279902)

[تصویر 13 بارگزاری داده‌ها 9](#_Toc171279903)

[تصویر 14 پیش‌پردازش داده‌ها 9](#_Toc171279904)

[تصویر 15 تبدیل داده‌ها به فرمت torch 10](#_Toc171279905)

[تصویر 16 لود کردن مدل 10](#_Toc171279906)

[تصویر 17 تعریف call back 11](#_Toc171279907)

[تصویر 18 تعریف دقت 11](#_Toc171279908)

[تصویر 19 تعریف آرگمان‌های آموزش 12](#_Toc171279909)

[تصویر 20 ساخت trainer 12](#_Toc171279910)

[تصویر 21 نمایش عملکرد مدل در حین آموزش 13](#_Toc171279911)

[تصویر 22 پیش‌بینی داده‌های آزمون 13](#_Toc171279912)

[تصویر 23 گزارش طبقه‌بندی 13](#_Toc171279913)

[تصویر 24 ماتریس درهمریختگی 14](#_Toc171279914)

[تصویر 25 تابع استخراج ویژگی 15](#_Toc171279915)

[تصویر 26 استخراج ویژگی از داده‌ها 15](#_Toc171279916)

[تصویر 27 آموزش مدل SVM 16](#_Toc171279917)

[تصویر 28 پیش‌بینی مدل SVM 16](#_Toc171279918)

[تصویر 29 نمایش ماتریس درهمریختگی 16](#_Toc171279919)

[تصویر 30 مدل MLP 17](#_Toc171279920)

[تصویر 31 oversample کردن داده‌ها 17](#_Toc171279921)

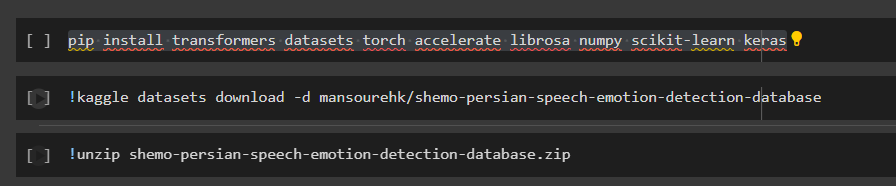
[تصویر 32 حذف کلاس‌های اقلیت 18](#_Toc171279922)

[تصویر 33 under sample کردن داده‌ها 18](#_Toc171279923)

# گزارش فایل‌ کد ساده

لود کردن کتابخانه‌ها و دانلود داده‌ها

ابتدا کتابخانه‌های مورد نیاز را نصب کرده و داده‌ها را از منبع مورد نظر دانلود می‌کنیم:



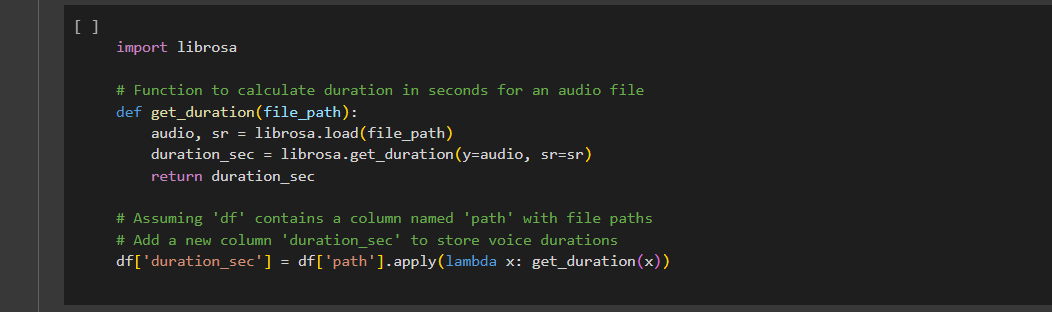
تصویر 1 نصب کتابخانه و مجموعه داده

همچنین داده‌ها را از حالت زیپ خارج می‌کنیم.

تمیز سازی و نمایش داده‌ها:

حال داده‌ها به دو دسته زن و مرد تقسیم شده‌اند در این بخش ما به جنسیت‌ آن‌ها کار ندارم و فقط قصد داریم یک مدل عمومی برای تشخیص احساسات بسازیم، بنابراین داده‌ها را با یکدیگر ترکیب می‌کنیم و یک دیتافریم از کل داده‌ها با احساسات مربوط به آن‌ها نگه می‌داریم.

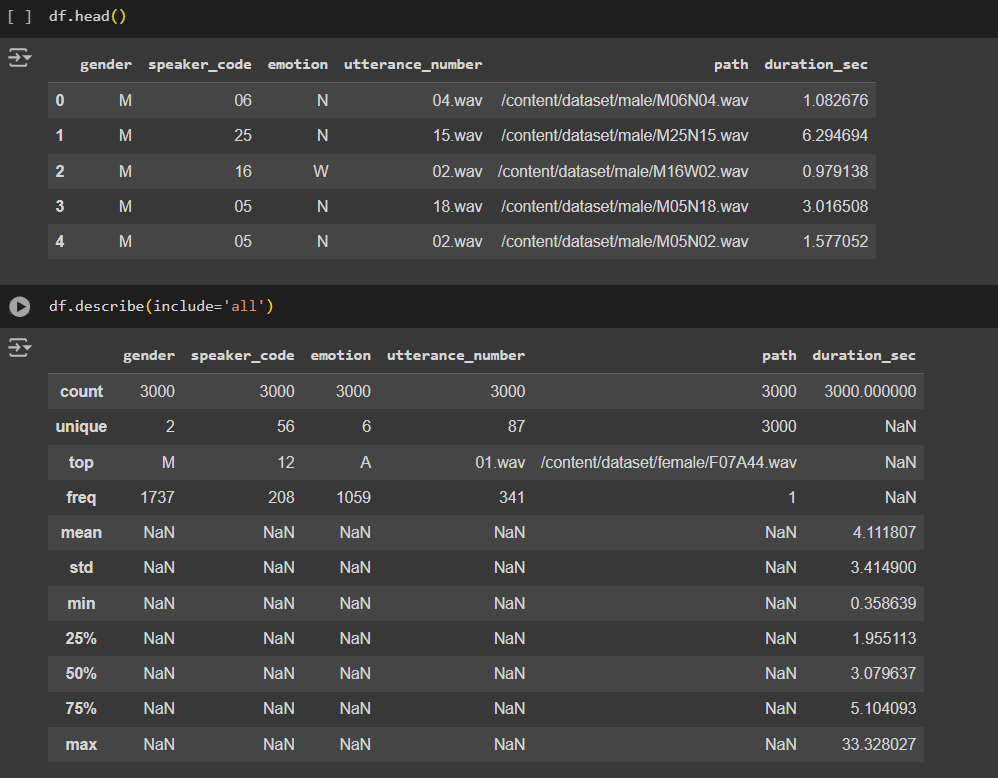
همچنین مدت زمان هر داده را نیز استخراج می‌کنیم و دیتافریم می‌گذاریم.



تصویر 2 به دست آوردن طول داده‌ها

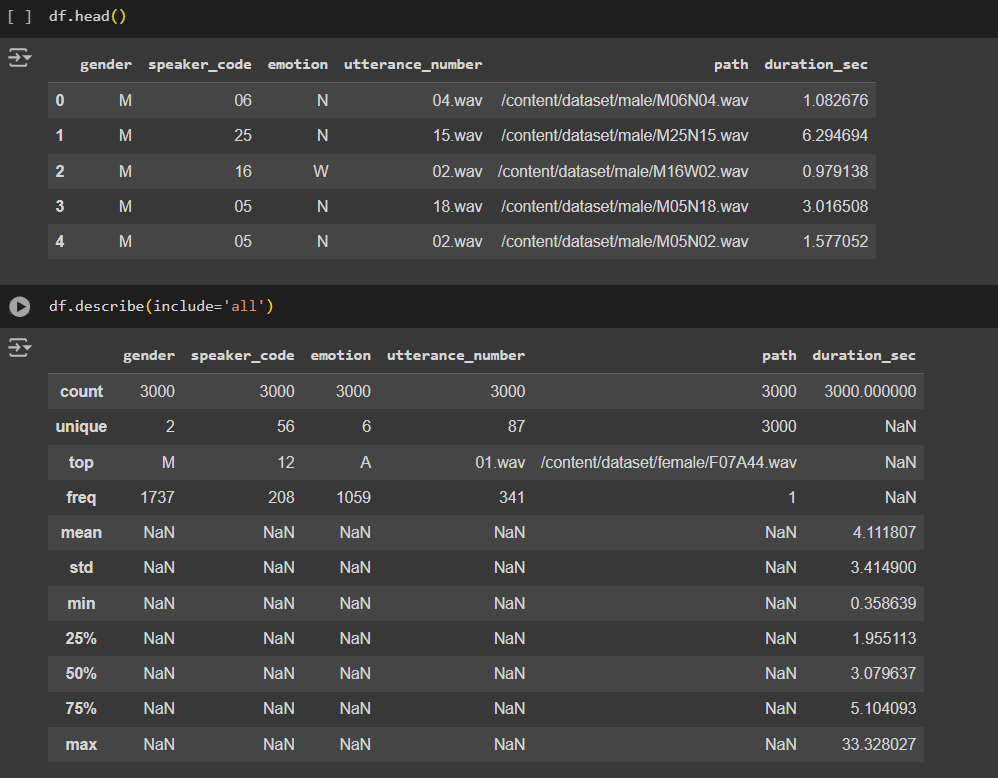
نمایش داده‌ها:

ابتدا دیتا فریم به شکل زیر می‌باشد:

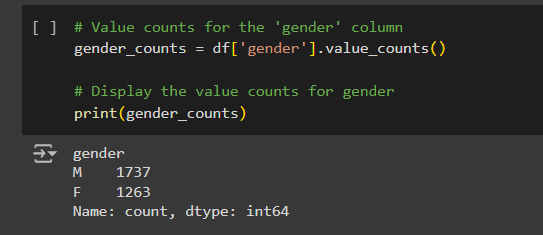


تصویر 3 تعداد از داده‌ها

همچنین داده‌ها به شکل زیر می‌باشد:

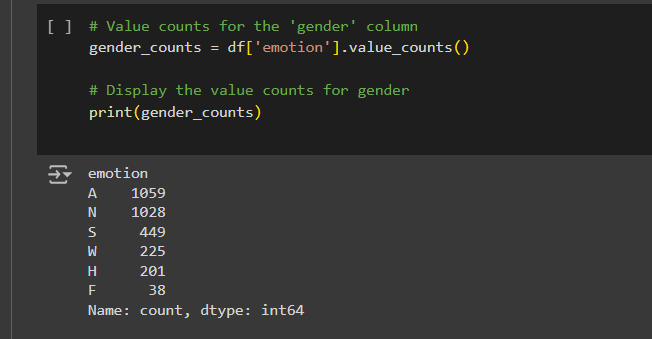


تصویر 4 ویژگی‌های اماری دیتاست

دارای 3000 داده هستیم که به صورت میانگین 4 ثانیه هستند. همچنین توزیع جنسیت زن و مرد به شکل زیر می‌باشد:  


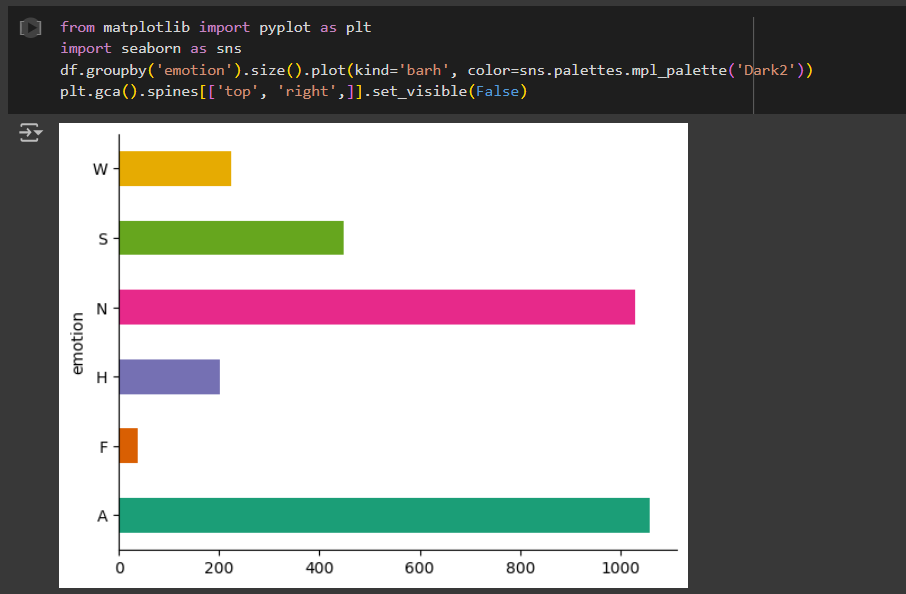
تصویر 5 توزیع جنسیت

همچنین توزیع احساسات مختلف به شکل زیر می‌باشد:



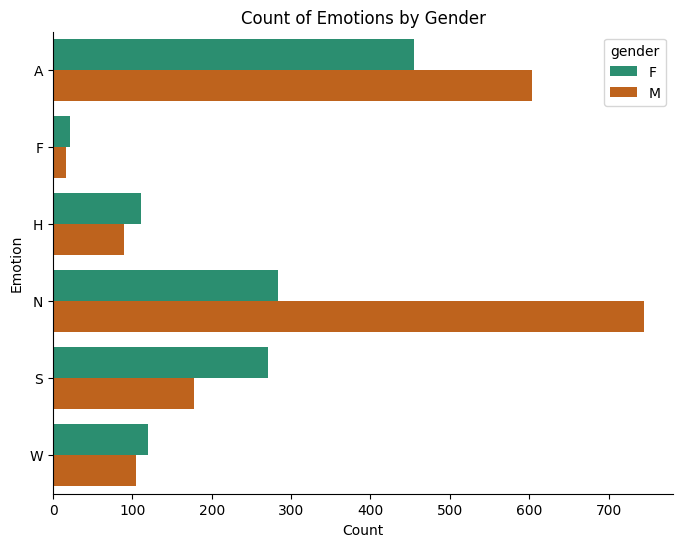
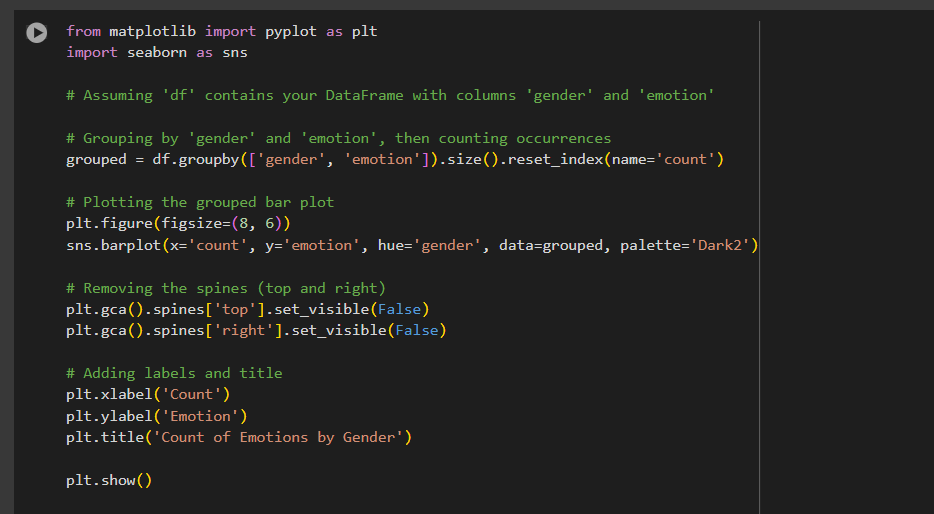
تصویر 6 توزیع کلاس‌ها

همچنین هیستگورام توزیع داده‌ها به شرح زیر می‌باشد:



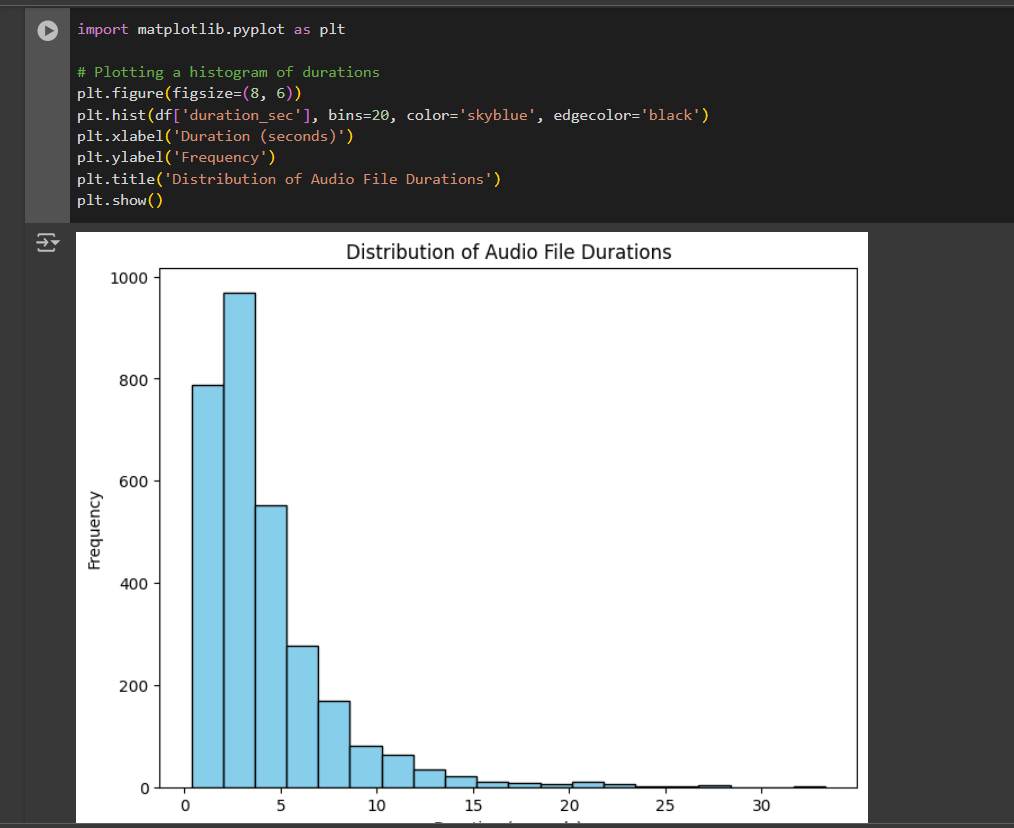
تصویر 7 هیستوگرام داده‌ها

همچنین توزیع احساسات برای جنسیت‌های مختلف به شکل زیر به دست می‌آید:



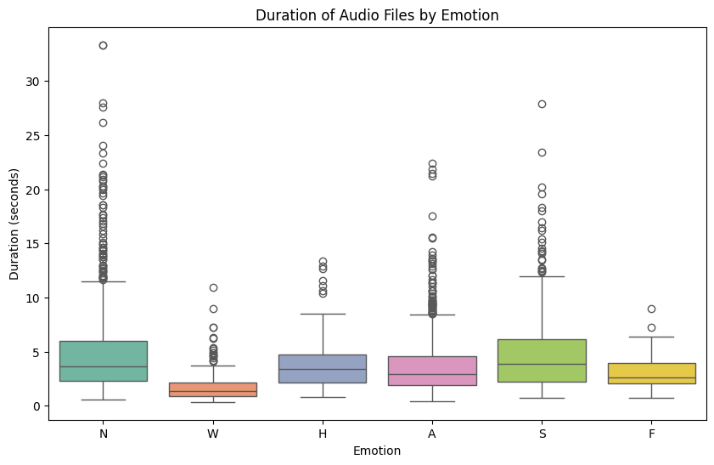
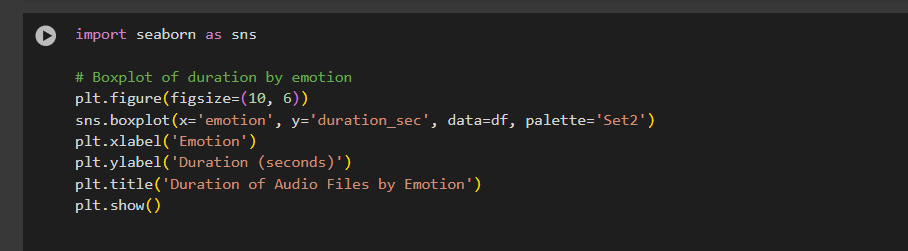
تصویر 8 هیستوگرام برای هر کلاس‌

همچنین هیستوگرام مدت زمان صوت‌ها به شکل زیر می‌باشد:



تصویر 9 هیستوگرام طول صوت‌ها

همچین باکس پلات مربوط به احساسات و مدت زمان‌ آن‌ها به شکل زیر می‌باشد:



تصویر 10 باکس پلات داده‌ها

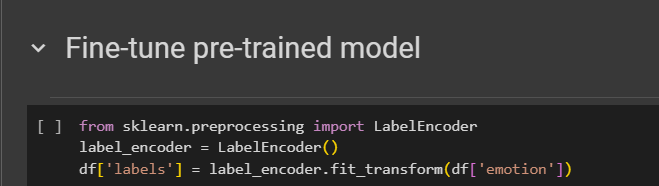
همچنین برای جنسیت‌های مختلف به شکل زیر می‌باشد:



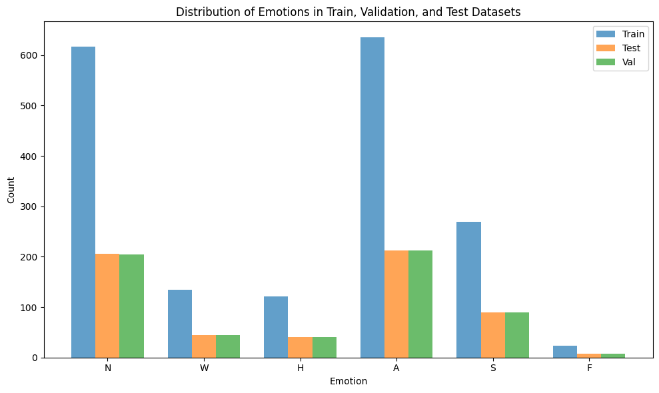
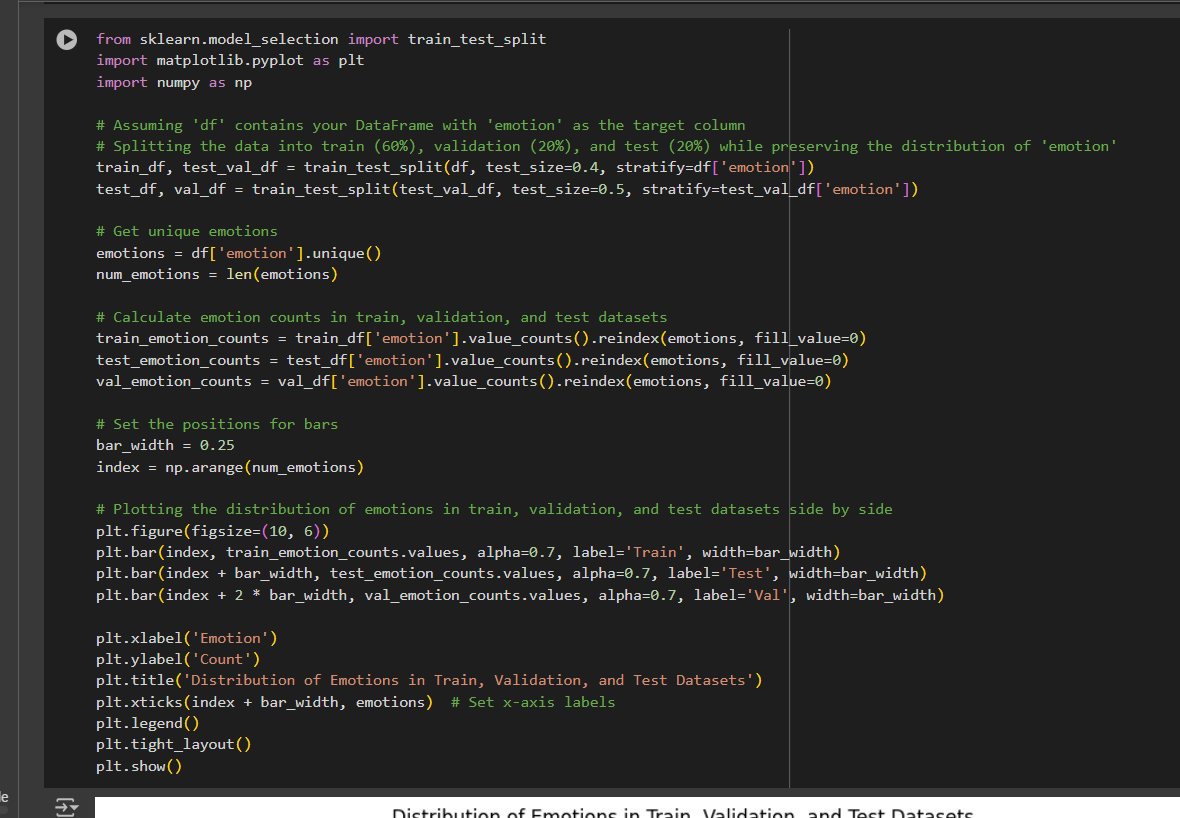
تصویر 11 باکس پلات جنسیت‌ها

آماده‌سازی داده‌ها برای آموزش:

ابتدا داده‌ها را از حروف انگلیسی به اعداد تبدیل می‌کنیم تا مدل بتواند آن‌ها را پیش‌بینی کند:



حال داده‌ها را به سه دسته آموزش، ارزیابی و آزمون تقسیم می‌کنیم، در این تقسیم توزیع داده‌ها در کلاس‌های مختلف حفظ می‌شود:



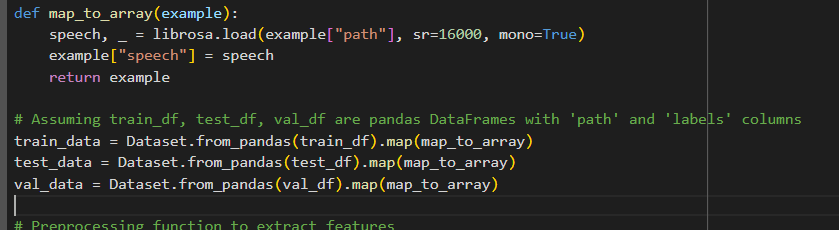
تصویر 12 داده‌های آموزش، ارزیابی و آزمون

**تعریف تابع map\_to\_array**

ابتدا تابع map\_to\_array تعریف شده که وظیفه‌اش بارگذاری فایل‌های صوتی و تبدیل آن‌ها به آرایه‌های numpy است. از کتابخانه‌ی librosa برای این منظور استفاده شده است.

**تبدیل داده‌ها به فرمت Dataset**

داده‌های آموزشی (train\_df)، آزمایشی (test\_df) و اعتبارسنجی (val\_df) که در قالب DataFrameهای pandas هستند، به فرمت Dataset تبدیل شده و سپس تابع map\_to\_array روی آن‌ها اعمال می‌شود.



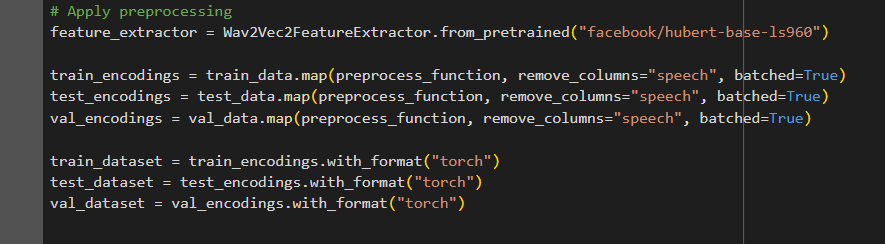
تصویر 13 بارگزاری داده‌ها

**تعریف تابع پیش‌پردازش**

یک تابع پیش‌پردازش تعریف شده که ویژگی‌های صوتی را استخراج کرده و برچسب‌ها را اضافه می‌کند. از feature\_extractor که با استفاده از مدل پیش‌آماده‌ی facebook/hubert-base-ls960 ایجاد شده، استفاده می‌شود.

**اعمال تابع پیش‌پردازش**

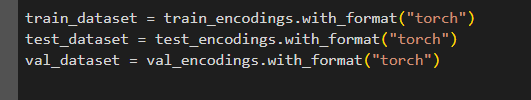
با استفاده از feature\_extractor، تابع پیش‌پردازش روی داده‌های آموزشی، آزمایشی و اعتبارسنجی اعمال می‌شود.



تصویر 14 پیش‌پردازش داده‌ها

**تبدیل داده‌ها به فرمت PyTorch**

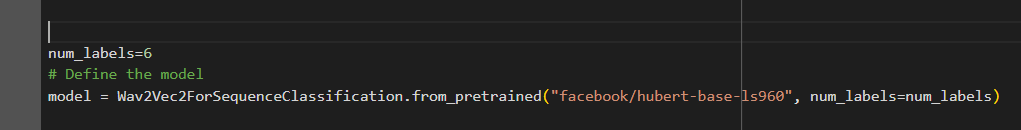
در نهایت، داده‌های کدگذاری شده به فرمت PyTorch تبدیل می‌شوند تا برای استفاده در مدل‌های یادگیری عمیق آماده شوند.



تصویر 15 تبدیل داده‌ها به فرمت torch

**تعریف مدل و آماده‌سازی محیط**

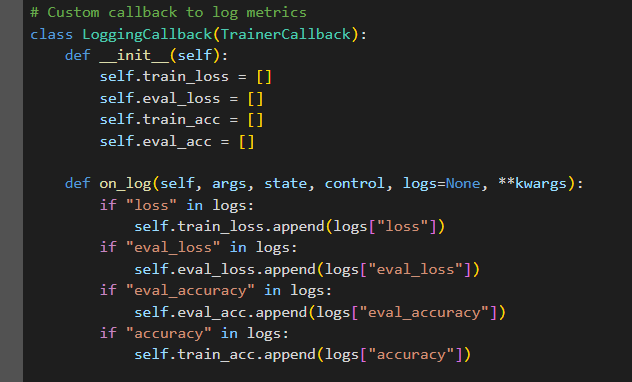
در این کد، مدل Wav2Vec2ForSequenceClassification از کتابخانه‌ی transformers برای انجام وظیفه‌ی طبقه‌بندی توالی‌های صوتی استفاده شده است. مدل بر پایه‌ی facebook/hubert-base-ls960 است و برای طبقه‌بندی ۶ دسته تنظیم شده است.



تصویر 16 لود کردن مدل

**تعریف Callback سفارشی برای ثبت معیارها**

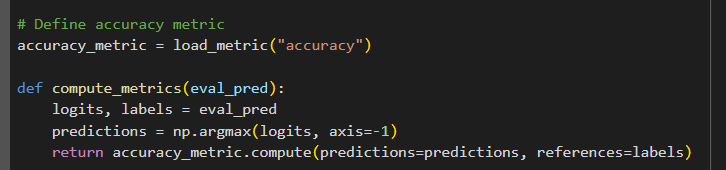
برای ثبت معیارهای آموزشی و ارزیابی، یک Callback سفارشی به نام LoggingCallback تعریف شده است. این کلاس معیارهای مانند train\_loss، eval\_loss، train\_acc و eval\_acc را ذخیره می‌کند.



تصویر 17 تعریف call back

**تعریف معیار دقت**

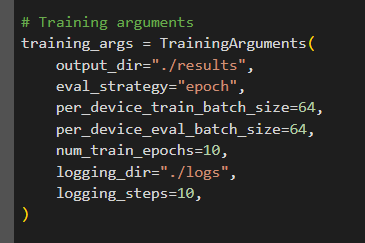
برای محاسبه‌ی معیار دقت از کتابخانه‌ی datasets استفاده شده است.



تصویر 18 تعریف دقت

**تنظیمات آموزشی**

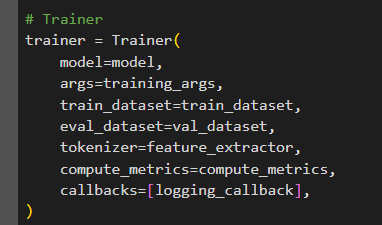
پارامترهای آموزشی با استفاده از کلاس TrainingArguments تنظیم شده‌اند. این پارامترها شامل تنظیمات مربوط به تعداد اپوک‌ها، اندازه‌ی بچ‌ها و مکان ذخیره‌سازی نتایج و لاگ‌ها هستند.



تصویر 19 تعریف آرگمان‌های آموزش

**تعریف Trainer**

کلاس Trainer برای مدیریت آموزش و ارزیابی مدل استفاده شده است. این کلاس مدل، داده‌های آموزشی و اعتبارسنجی، توکنایزر، معیارهای محاسبه‌شده و Callback سفارشی را دریافت می‌کند.



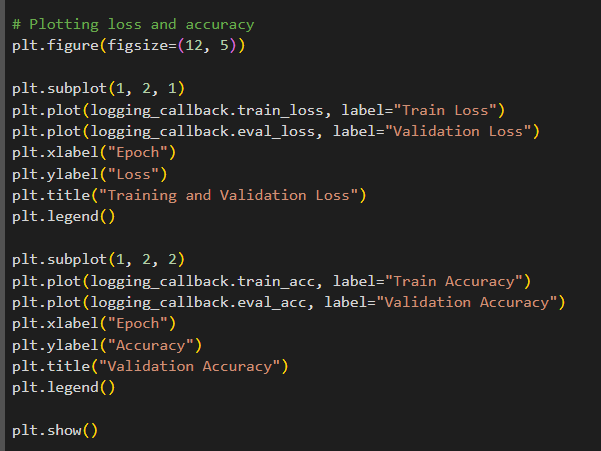
تصویر 20 ساخت trainer

**آموزش مدل**

مدل با استفاده از trainer.train() آموزش داده می‌شود و سپس با trainer.evaluate(test\_dataset) ارزیابی می‌شود.

**رسم نمودارهای از دست دادن و دقت**

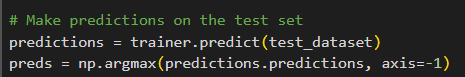
در نهایت، نمودارهای از دست دادن و دقت برای دوره‌های آموزشی و ارزیابی رسم می‌شوند.



تصویر 21 نمایش عملکرد مدل در حین آموزش

**پیش‌بینی بر روی مجموعه‌ی تست**

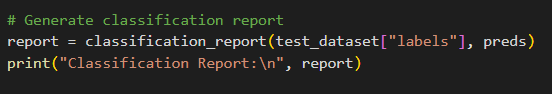
مدل آموزش‌دیده با استفاده از trainer.predict بر روی مجموعه‌ی تست اعمال می‌شود و پیش‌بینی‌ها استخراج می‌شوند.



تصویر 22 پیش‌بینی داده‌های آزمون

**تولید گزارش طبقه‌بندی**

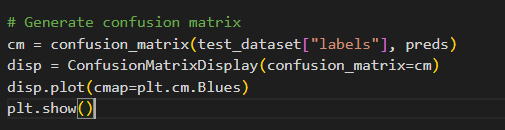
گزارش طبقه‌بندی که شامل معیارهایی مثل دقت، بازخوانی و امتیاز F1 برای هر کلاس است، با استفاده از classification\_report از کتابخانه‌ی sklearn تولید می‌شود.



تصویر 23 گزارش طبقه‌بندی

**تولید ماتریس درهمریختگی**

ماتریس درهمریختگی برای نمایش تعداد نمونه‌های درست و نادرست طبقه‌بندی شده برای هر کلاس تولید می‌شود و با استفاده از ConfusionMatrixDisplay از کتابخانه‌ی sklearn رسم می‌شود.



تصویر 24 ماتریس درهمریختگی

تمام بخش‌های بالا برای مدل‌های دیگر تکرار شده است، تنها فرق این است که نام مدل‌های دیگر برای دانلود داده شده‌اند.

مدل‌های کلاسیک:

برای این مدل‌ها باید از روش‌های استخراج ویژگی مربوط به صوت استفاده کنیم بنابراین داریم:

**تعریف تابع extract\_features**

تابع extract\_features که سه پارامتر mfcc، chroma و mel دارد که تعیین می‌کنند کدام ویژگی‌ها استخراج شوند. ورودی این تابع مسیر فایل صوتی (file\_path) است.

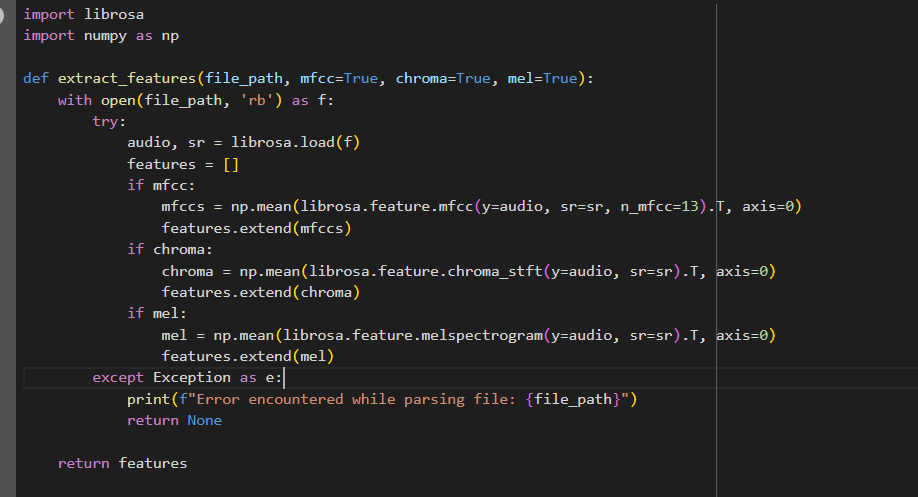
 **بارگذاری فایل صوتی**: فایل صوتی با استفاده از librosa.load بارگذاری می‌شود که به طور خودکار نرخ نمونه‌برداری (sampling rate) را نیز برمی‌گرداند.

**استخراج MFCC**: اگر پارامتر mfcc برابر با True باشد، ویژگی‌های MFCC با استفاده از librosa.feature.mfcc استخراج شده و میانگین آن‌ها محاسبه و به لیست ویژگی‌ها اضافه می‌شود.

**استخراج کرومای استروف**: اگر پارامتر chroma برابر با True باشد، ویژگی‌های کرومای استروف با استفاده از librosa.feature.chroma\_stft استخراج شده و میانگین آن‌ها محاسبه و به لیست ویژگی‌ها اضافه می‌شود.

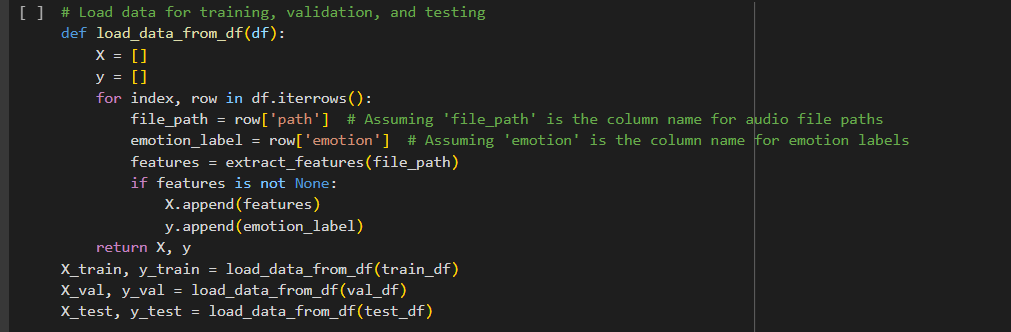
**استخراج مل‌سپکتروگرام**: اگر پارامتر mel برابر با True باشد، ویژگی‌های مل‌سپکتروگرام با استفاده از librosa.feature.melspectrogram استخراج شده و میانگین آن‌ها محاسبه و به لیست ویژگی‌ها اضافه می‌شود.

.



تصویر 25 تابع استخراج ویژگی

حال با استفاده از دیتافریم ویژگی‌های مربوط به همه‌ی صوت‌ها را استخراج می‌کنیم:



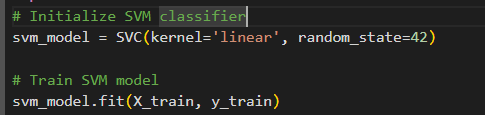
تصویر 26 استخراج ویژگی از داده‌ها

حال مدل‌های مختلفی را برای این کار آموزش می‌دهیم:

مدل SVM

**تعریف و آموزش مدل SVM**

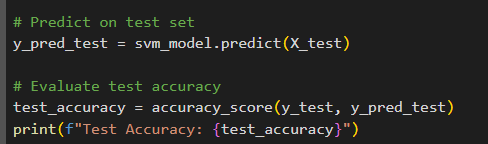
ابتدا مدل SVM با هسته‌ی خطی (linear kernel) و حالت تصادفی ۴۲ تعریف می‌شود. سپس مدل با استفاده از داده‌های آموزشی آموزش داده می‌شود.



تصویر 27 آموزش مدل SVM

**پیش‌بینی و ارزیابی مدل**

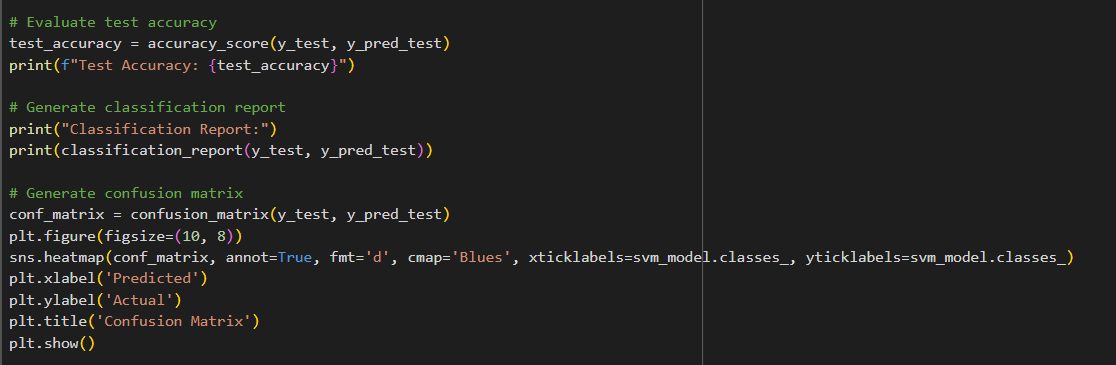
مدل آموزش‌دیده بر روی مجموعه‌ی تست اعمال شده و پیش‌بینی‌ها محاسبه می‌شوند. سپس دقت تست و گزارش طبقه‌بندی محاسبه و نمایش داده می‌شوند.



تصویر 28 پیش‌بینی مدل SVM

**تولید و نمایش ماتریس درهمریختگی**

ماتریس درهمریختگی گمی برای نمایش تعداد نمونه‌های درست و نادرست طبقه‌بندی شده برای هر کلاس تولید و با استفاده از seaborn رسم می‌شود.

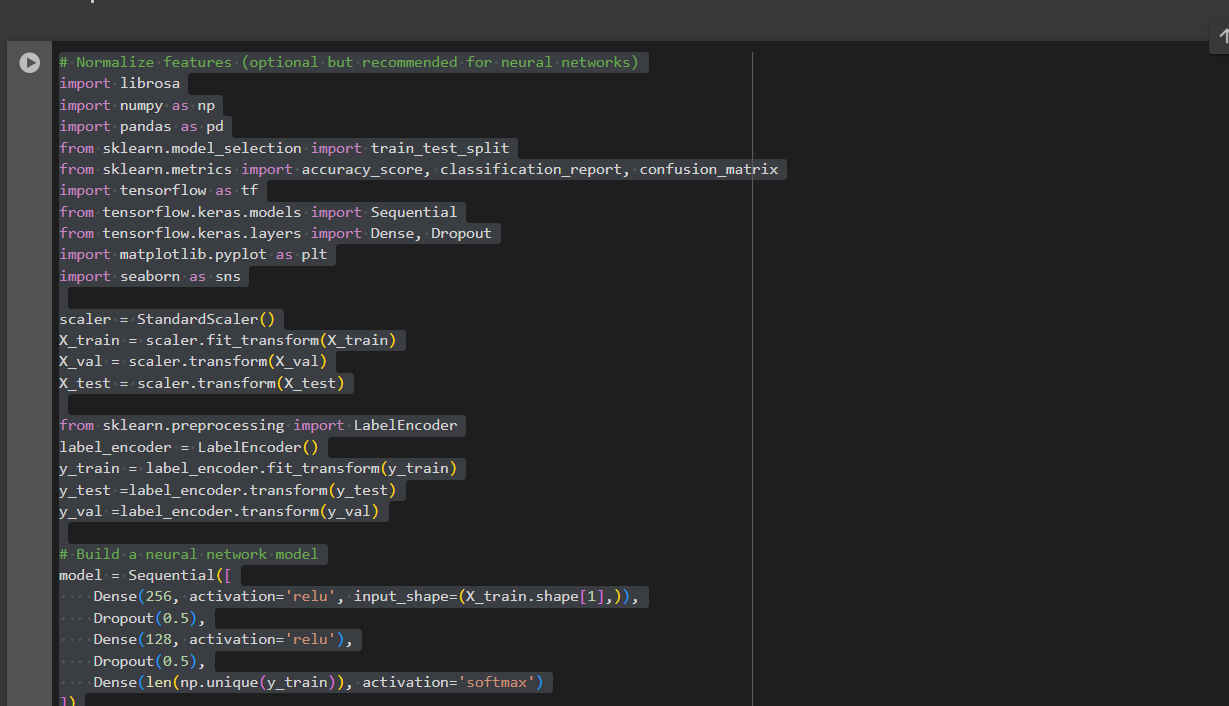


تصویر 29 نمایش ماتریس درهمریختگی

برای مدل‌های کلاسیک دیگر نیز همین فرایند کد تکرار می‌شود.

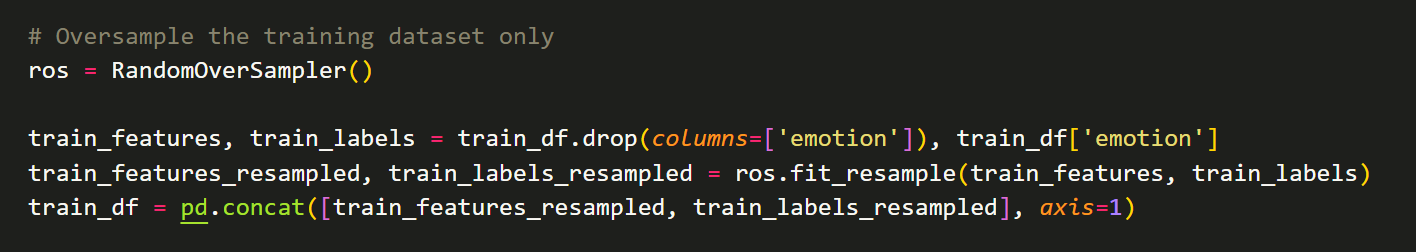
مدل عمیق:

برای عمیق ابتدا، یک شبکه‌ی عصبی ساده چند لایه می‌سازیم و سپس مانند مدل‌های بالا آن را آموزش داده و عملکرد آن را روی داده‌های مختلف بررسی می‌کنیم.



تصویر 30 مدل MLP

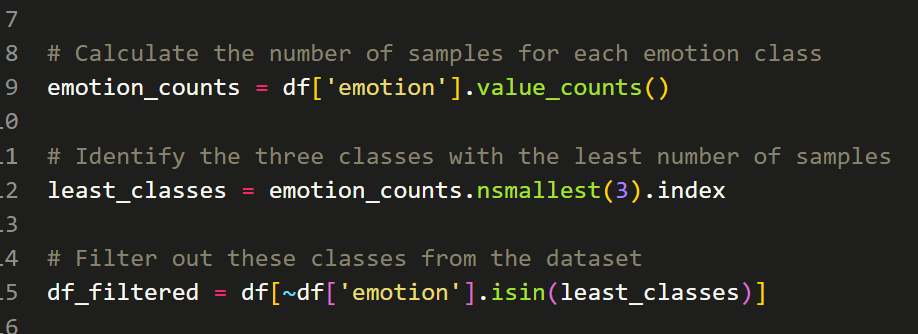
## حالت oversample کردن



تصویر 31 oversample کردن داده‌ها

تفاوت این کد فقط در بخش oversample کردن داده‌های آموزش هست که با کد بالا انجام شده است.

## آموزش سه کلاسه

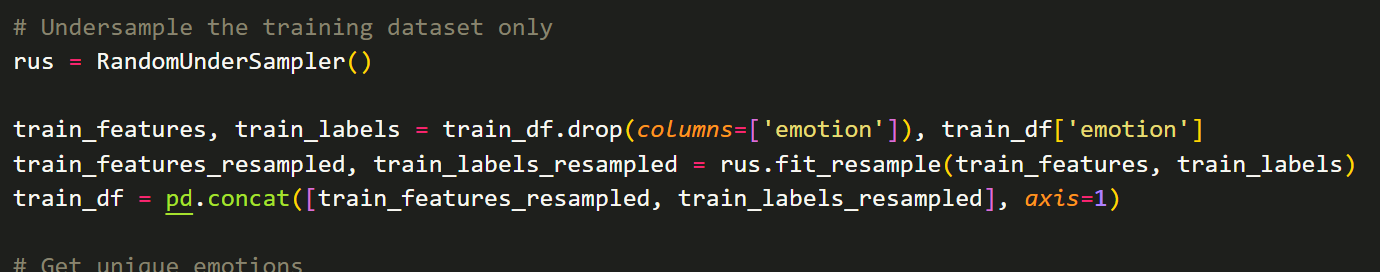


تصویر 32 حذف کلاس‌های اقلیت

کلاسهایی که کمترین تعداد داده را داشتند حذف کردیم.

## Undersample کردن

داده‌های ترین را undersample می‌کنیم:



تصویر 33 under sample کردن داده‌ها

## آموزش با دیگر مجموعه داده‌ها

فرمت هرکدام از این مجموعه داده‌ها متفاوت است بنابراین ابتدا هرکدام را باز می‌کنیم و با فرمت یکسان برای آموزش آماده می‌کنیم، دقت شود داده‌های تست، همچنان مجموعه داده اصلی هستند.