**山东大学威海机电学院**

实验

**课 程 名 称： 深度学习概论**

**题 目： 基于Transformer的语音分类**

**专 业 班 级： 21电子1班**

**学 生 姓 名： 冯浩然**

**学 生 学 号：**  **202100800378**

**日 期： 2023.12.25**

**指 导 教 师： 薛鹏**

**机电学院教务部印制**

目录

[一、 实验目的 3](#_Toc154428023)

[二、 实验设备 3](#_Toc154428024)

[三、 实验原理 3](#_Toc154428025)

[四、 实验步骤 4](#_Toc154428026)

[五、 实验结果 6](#_Toc154428027)

[六、 实验心得体会： 7](#_Toc154428028)

# 实验目的

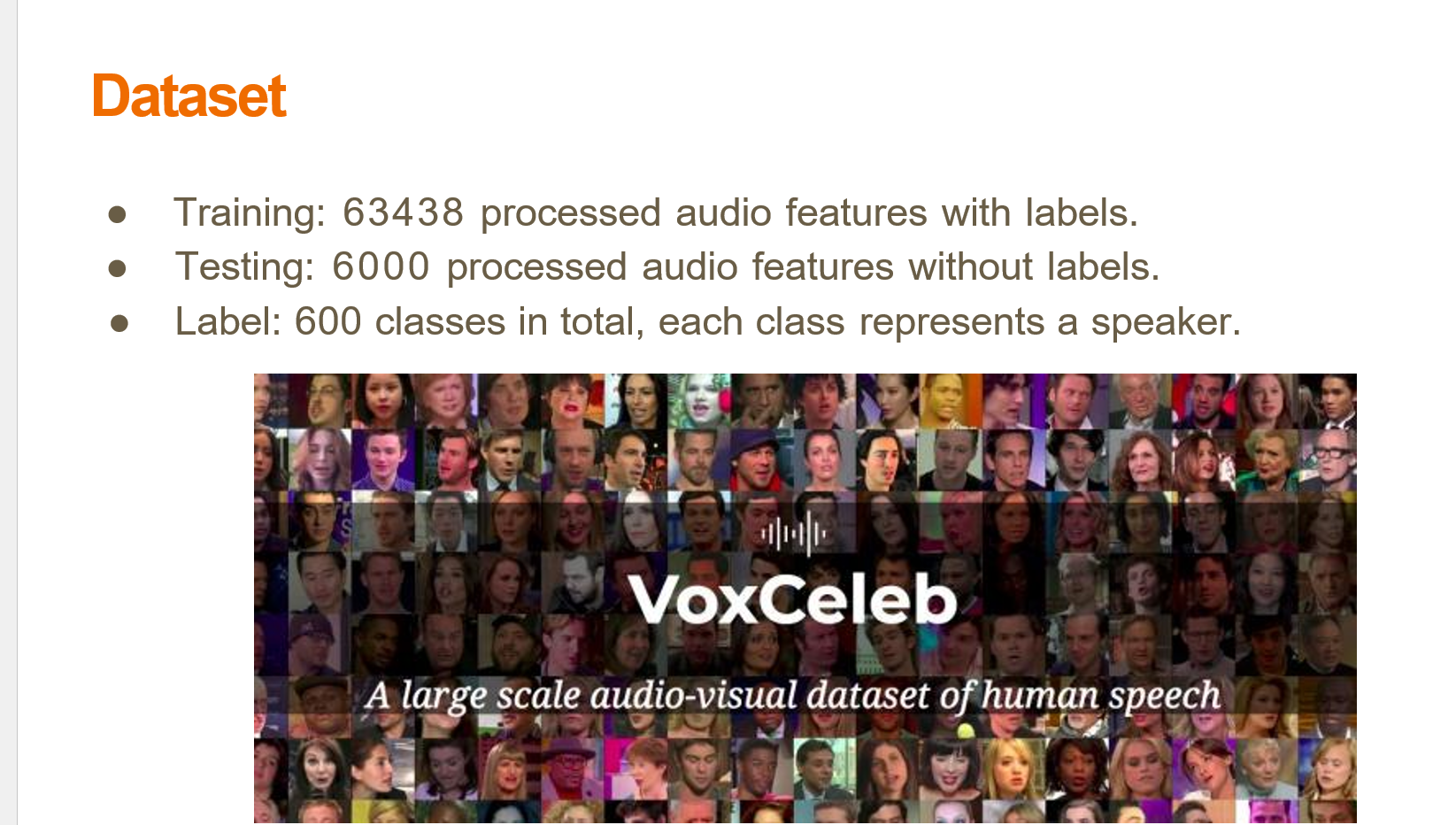
1. 理解和应用深度神经网络来解决多类分类问题。
2. 预测给定语音的说话人类别。
3. 学习使用自注意力和变换器（Transformer）架构。
4. 熟悉PyTorch在处理音频数据的应用。

# **实验设备**

1. 计算机
2. Python编程环境
3. PyTorch深度学习库
4. 相关数据集

# **实验原理**

数据的形式：

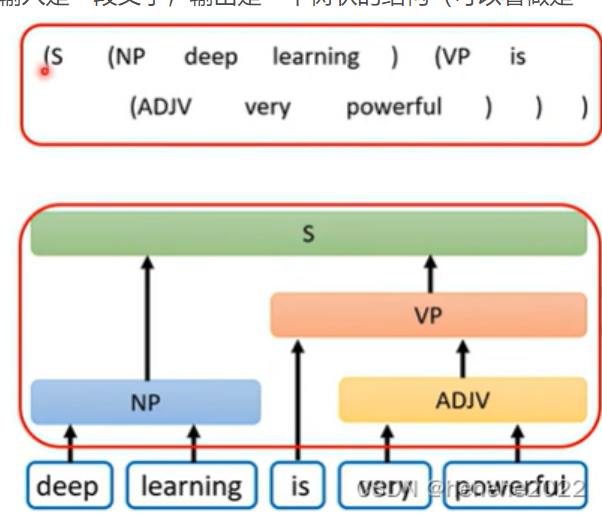
****



transformer就是一个sequence-to-sequence（seq2seq）的模型，并且用到了self-attention；输入是一个序列，输出也是一个序列，但是输出的长度取决于机器。比如语音识别，聊天机器。

seq2seq for syntactic parsing

输入是一段文字，输出是一个树状的结构（可以看做是一个序列）

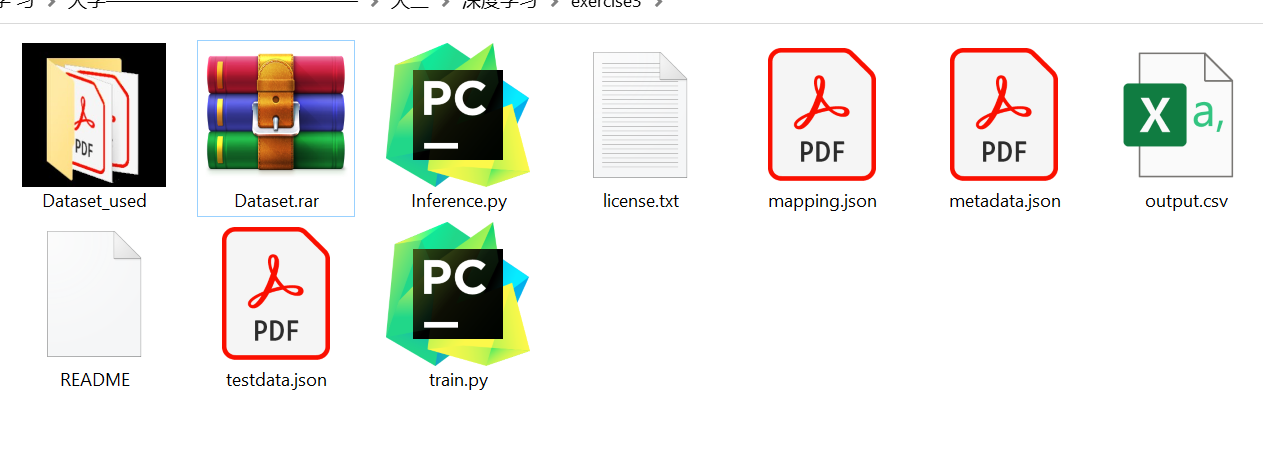


seq2seq for multi-label classification

multi-class classification 是从数个类别中，选择某一个类出来；multi-label classification同一个东西可以属于多个类（class）；机器决定每篇文章属于多少个class。

# **实验步骤**

下载所需要的数据，解压到一个文件夹里（这一步省去配置输入输出路径）

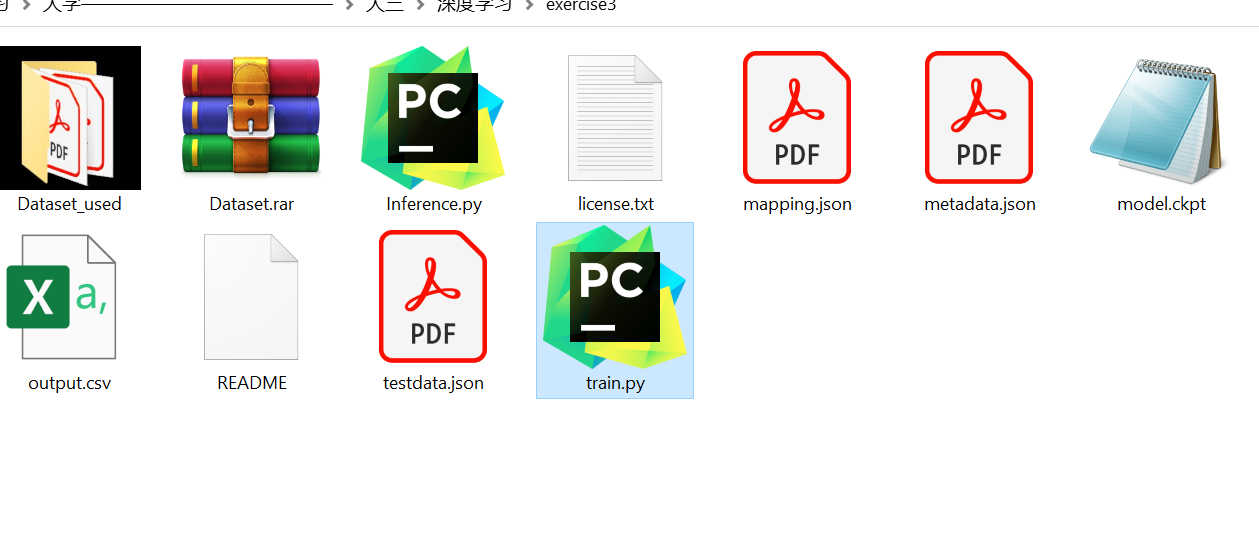


然后打开train.py文件

增加一些对数据的选择，优化

点击运行，等待结果

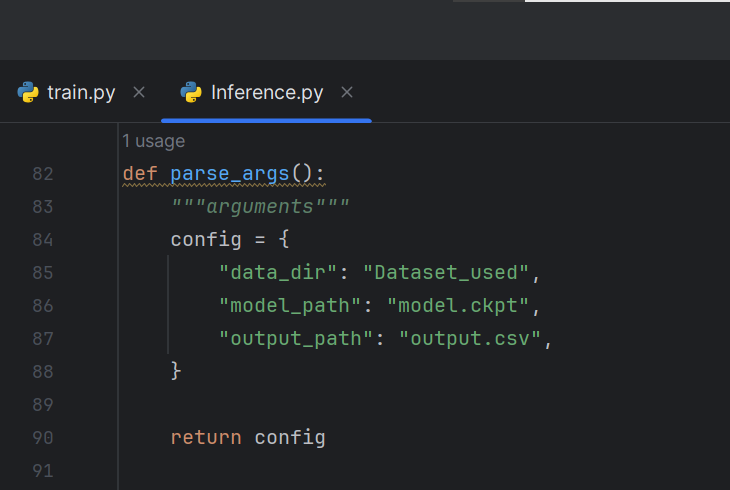
文件夹中出现一个model.ckpt





接着打开Inference.py

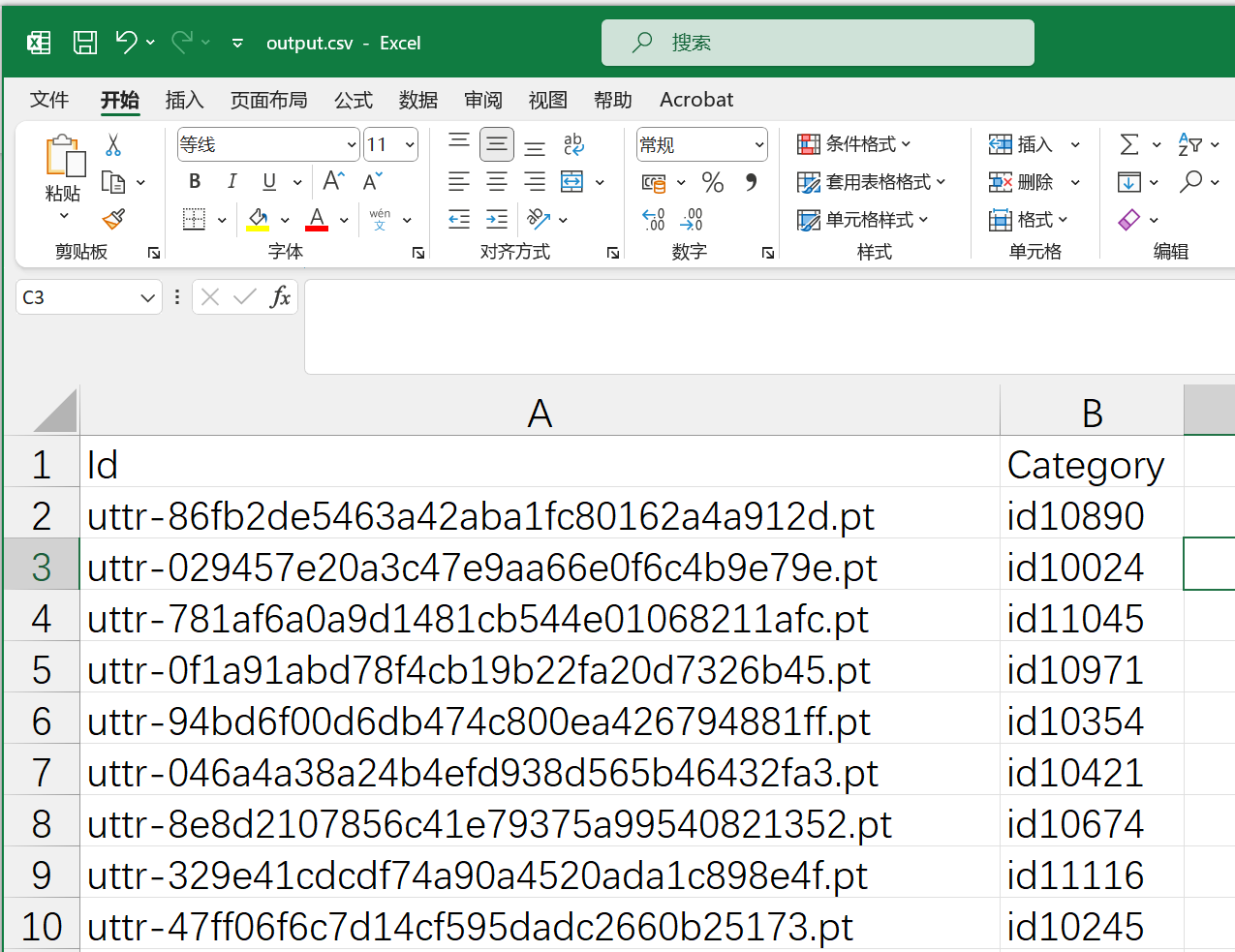
确保文件存在，路径正确



运行

# **实验结果**

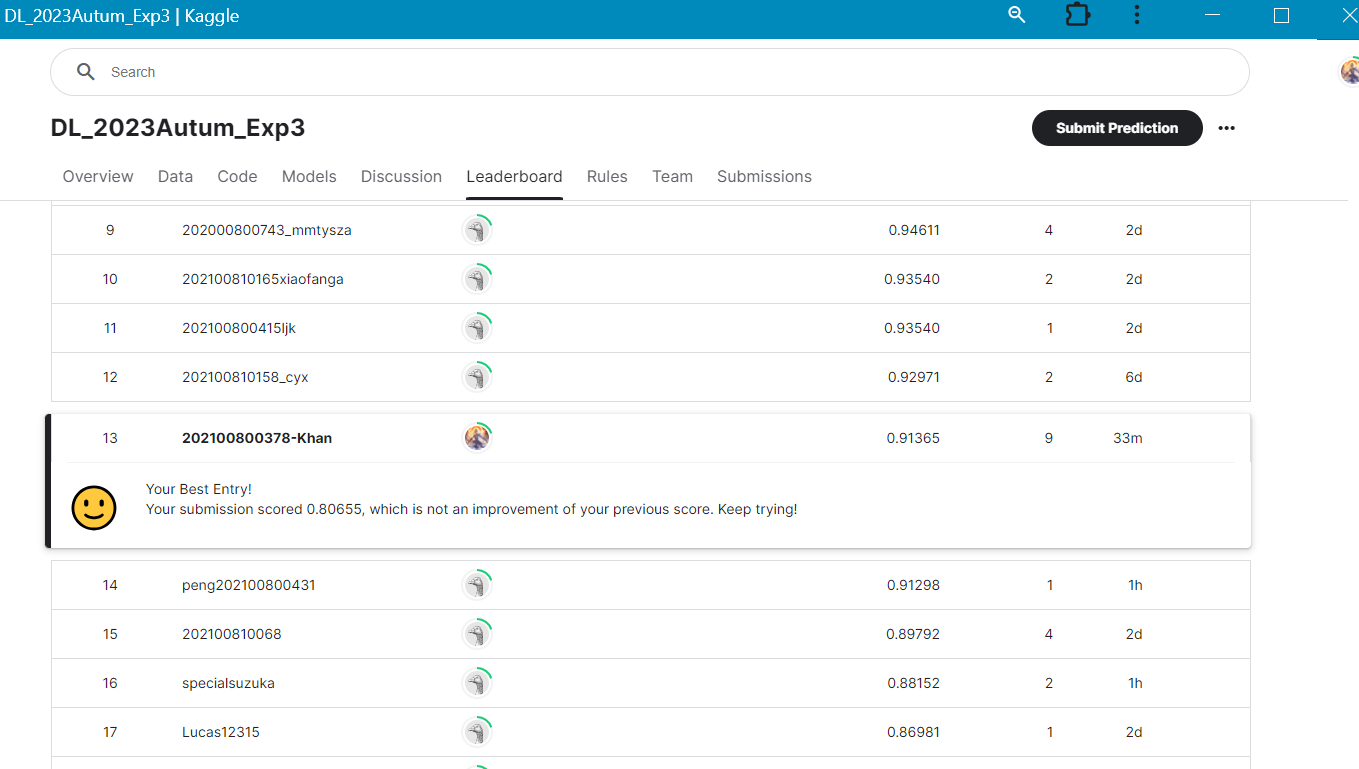
运行结束，获得一个.csv文件



把它交到kaggle上

等待评分

最后在Leaderboard查看自己的排名



# **实验心得体会：**

在本次实验中，我们专注于使用深度神经网络，特别是自注意力机制和变换器（Transformer）架构，来进行说话人分类任务。通过深入探索和实施先进的深度学习技术，我们成功地构建了一个模型，该模型可以从给定的语音中识别出说话人的身份。实验过程中，我们遇到了数据预处理、模型调优和过拟合等挑战，但通过细致的参数调整和模型优化，最终实现了令人满意的分类准确率。这次实验不仅增强了我们对深度学习在音频处理领域应用的理解，也提高了我们解决实际问题的能力。未来，我们期待探索更多复杂的模型和技术，以进一步提升模型的性能和应用范围。