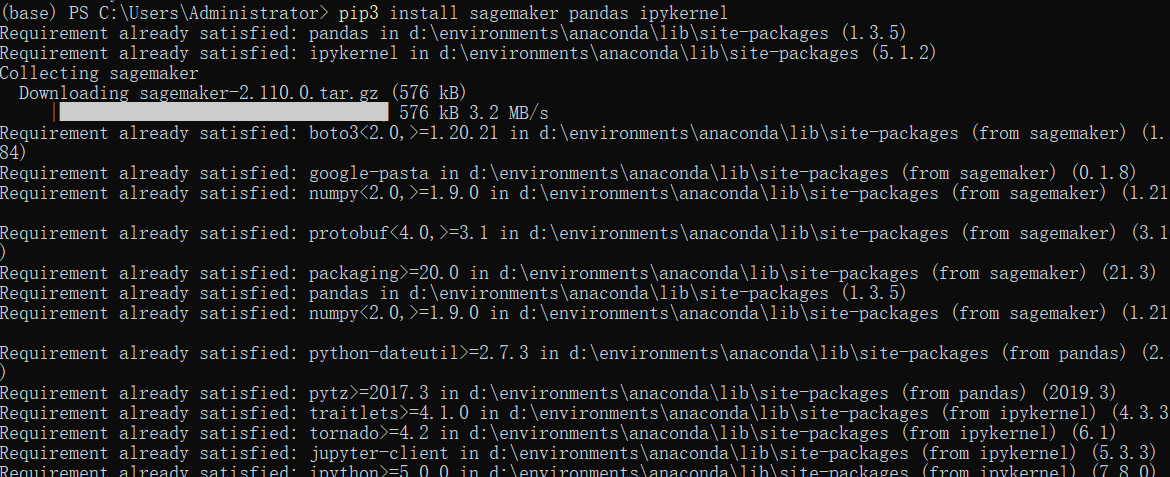
**Lab 8 Notes**

Student ID: 22994257

Name: Gaoyuan Zhang

### [Step 1] Install missing libraries



### [Step 2] Deal with dataset on Jupyter Notebook

图片包含 图形用户界面

描述已自动生成

图片包含 表格

描述已自动生成

表格

描述已自动生成

图片包含 图形用户界面

描述已自动生成

### [Step 3] Create a S3 bucket and copy files

Create a bucket named “22994257-lab8” and then copy the dataset to my bucket.

文本, 信件

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

### [Step 4] Setup Hyperparameter Optimization

文本

描述已自动生成

文本, 日程表

描述已自动生成

Finally, it can run but fail because of the resource limit on AWS.

图形用户界面, 文本, 应用程序

描述已自动生成

### [Step 5] Answer the questions.

**a) In your S3 bucket, how many folders were created using the script (under the "{student\_id}-hpo-xgboost-dm" folder)? List their name.**

**Answer:** Three folders. Their name is “train” , “validation” and “output”.

**b) How many Hyperparameter tuning jobs were created using the script?**

**Answer:** Two.



c) What metric was used in this script to evaluate the training results?

**Answer:** Auc(Area under the ROC Curve)

图片包含 文本

描述已自动生成

**d) What strategy was used in the tuning job?**

**Answer:** Bayesian network.

