Custom View Settings

Question #53 *Topic 1*

Your team is working on an NLP research project to predict political affiliation of authors based on articles they have written. You have a large training dataset that is structured like this:

```
AuthorA:Political Party A
    TextA1: [SentenceA11, SentenceA12, SentenceA13, ...]
    TextA2: [SentenceA21, SentenceA22, SentenceA23, ...]
    ...
AuthorB:Political Party B
    TextB1: [SentenceB11, SentenceB12, SentenceB13, ...]
    TextB2: [SentenceB21, SentenceB22, SentenceB23, ...]
    ...
AuthorC:Political Party B
    TextC1: [SentenceC11, SentenceC12, SentenceC13, ...]
    TextC2: [SentenceC21, SentenceC22, SentenceC23, ...]
    ...
AuthorD:Political Party A
    TextD1: [SentenceD11, SentenceD12, SentenceDl3, ...]
    TextD2: [SentenceD21, SentenceD22, SentenceD23, ...]
    ...
...
```

You followed the standard 80%-10%-10% data distribution across the training, testing, and evaluation subsets. How should you distribute the training examples across the train-test-eval subsets while maintaining the 80-10-10 proportion?

A. Distribute texts randomly across the train-test-eval subsets: Train set: [TextA1, TextB2, ...] Test set: [TextA2, TextC1, TextD2, ...] Eval set: [TextB1, TextC2, TextD1, ...]

B. Distribute authors randomly across the train-test-eval subsets: (*) Train set: [TextA1, TextA2, TextD1, TextD2, ...] Test set: [TextB1, TextB2, ...] Eval set: [TexC1,TextC2 ...]

C. Distribute sentences randomly across the train-test-eval subsets: Train set: [SentenceA11, SentenceA21, SentenceB11, SentenceB21, SentenceC11, SentenceD21 ...] Test set: [SentenceA12, SentenceA22, SentenceB12, SentenceC22, SentenceC12, SentenceD22 ...] Eval set: [SentenceA13, SentenceA23, SentenceB13, SentenceC23, SentenceC13, SentenceD31 ...]

D. Distribute paragraphs of texts (i.e., chunks of consecutive sentences) across the train-test-eval subsets: Train set: [SentenceA11, SentenceA12, SentenceD11, SentenceD12 ...] Test set: [SentenceA13, SentenceB13, SentenceB21, SentenceD23, SentenceC12, SentenceD13 ...] Eval set: [SentenceA11, SentenceA22, SentenceB13, SentenceD22, SentenceC23, SentenceD11 ...]

Question #54 *Topic 1*

Your team has been tasked with creating an ML solution in Google Cloud to classify support requests for one of your platforms. You analyzed the requirements and decided to use TensorFlow to build the classifier so that you have full control of the model's code, serving, and deployment. You will use Kubeflow pipelines for the ML platform. To save time, you want to build on existing resources and use managed services instead of building a completely new model. How should you build the classifier?

A. Use the Natural Language API to classify support requests.

B. Use AutoML Natural Language to build the support requests classifier.

C. Use an established text classification model on AI Platform to perform transfer learning.

D. Use an established text classification model on AI Platform as-is to classify support requests.

Question #55 *Topic 1*

You recently joined a machine learning team that will soon release a new project. As a lead on the project, you are asked to determine the production readiness of the ML components. The team has already tested features and data, model development, and infrastructure. Which additional readiness check should you recommend to the team?

A. Ensure that training is reproducible.

B. Ensure that all hyperparameters are tuned.

C. Ensure that model performance is monitored.

D. Ensure that feature expectations are captured in the schema.

Question #56 *Topic 1*

You work for a credit card company and have been asked to create a custom fraud detection model based on historical data using AutoML Tables. You need to prioritize detection of fraudulent transactions while minimizing false positives. Which optimization objective should you use when training the model?

A. An optimization objective that minimizes Log loss

B. An optimization objective that maximizes the Precision at a Recall value of 0.50

C. An optimization objective that maximizes the area under the precision-recall curve (AUC PR) value

D. An optimization objective that maximizes the area under the receiver operating characteristic curve (AUC ROC) value

Previous Questions Next Questions