From POC to Prod

EPF - P2023
Yassine Idrissi - yassine.idrissir@gmail.com
Ismail Chafai - ismail.chafai@teads.com

EPF
ÉCOLE D'INGÉNIEUR-E-S

From POC to Prod - Lab1

- In this lab we will construct a model to automatically tag stackoverflow.com questions by using their title

Python list function or list comprehension **← Input**

Asked 3 years, 3 months ago   Active 3 years, 3 months ago   Viewed 5k times

12

Can anybody please explain to me the difference between these two ways of creating a list. Are they the same thing ? If not which one should I use ?

```
squares1 = [x**2 for x in range(1, 11)]
squares2 = list(x**2 for x in range(1, 11))
```

1

python   python-3.x

Share Improve this question Follow

Output

asked Aug 15 '18 at 15:16

Oppac
135 ● 1 ● 7

- This is a multi-class classification task

From POC to Prod - Lab1

- First approach: use scikitlearn's SGDClassifier combined with a HashingVectorizer to transform the input

Read data

1. Download data from Moodle
2. Read data from notebook

Preprocessing

1. Keep only the examples where tag_position is equal to 0
2. Filter the example where the labels are seen less than 100 occurrences
3. Keep only the title as feature and use HashingVectorizer

Training

1. Use a SGDClassifier from sklearn
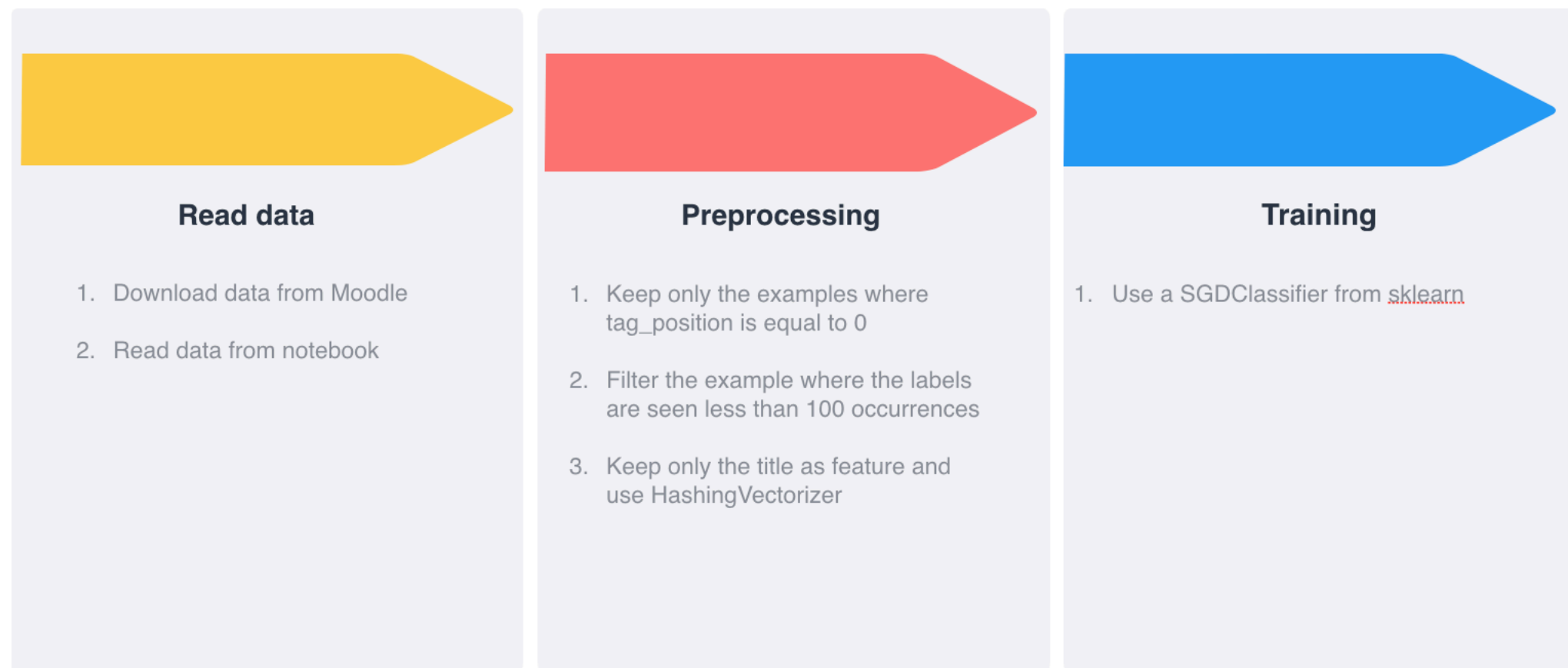
From POC to Prod - Lab1

- Second approach: use Keras sequential model combined with a pre-trained BERT model to embed the input

Read data

1. Download data from Moodle
2. Read data from notebook

Preprocessing

1. Keep only the examples where tag_position is equal to 0
2. Filter the example where the labels are seen less than 100 occurrences
3. Keep only title as feature and use a pre-trained BERT model to title text into embeddings

Training

1. Use a Sequential Keras model with the following architecture
   1. A dense layer with the dimension of your choice and ReLu activation
   2. Another dense layer with dimension as the number of unique labels and sigmoid
   3. Loss should be binary cross entropy, optimizer should be SGD and metric is accuracy

EPF
ÉCOLE D'INGÉNIEUR-E-S