



```
// You'll need  
// com.google.  
listRef.listAll  
.addOn  
prefixes.  
// All  
// You  
}  
it  
ach { item  
the items  
}  
}
```

devfest

제주도 사투리 번역 프로그램 만들어보기

대구가톨릭대학교 AI빅데이터공학과
None Labs 대표 | 임미르



Google Developer Groups

GDG Cloud Korea, GDSC Yonsei & Ewha

0: 발표자 소개

{

Name: Mireu Lim,

Email: limmireu1214@gmail.com,

HomePage: www.mireu.xyz,

Git: mireu-lab,

Group: [

[대구가톨릭대학교 , AI빅데이터공학과],

None_Labs

]

}



Mireu Lim

Mireu-Lab

전기는 국산이지만 서버는 수입입니다.

```
Text(  
  'Section Title',  
  style: TextStyle(  
    color: Colors.blue[200],  
  ),  
),  
),  
s.star,  
r: Colors.blue[500],  
Text('23'),
```

devfest



Google Developer Groups

GDG Cloud Korea, GDSC Yonsei & Ewha



1.1: 구현 계?기

1.1: 구현 계?기

일단은 제주사람이 '육지' 대학교에 오게되면서 많이 듣는 말 리스트

1.1: 구현 계?기

일단은 제주사람이 ‘육지’ 대학교에 오게되면서 많이 듣는 말 리스트

1. 제주도 사투리 할줄 알아?

1.1: 구현 계?기

일단은 제주사람이 ‘육지’ 대학교에 오게되면서 많이 듣는 말 리스트

1. 제주도 사투리 할줄 알아?
2. 너네집 감귤농사해?

1.1: 구현 계?기

일단은 제주사람이 ‘육지’ 대학교에 오게되면서 많이 듣는 말 리스트

1. 제주도 사투리 할줄 알아?
2. 너네집 감귤농사해?
3. 너네 제주대 감귤포장학과 있어?

1.1: 구현 계?기

일단은 제주사람이 ‘육지’ 대학교에 오게되면서 많이 듣는 말 리스트

1. 제주도 사투리 할줄 알아?
2. 너네집 감귤농사해?
3. 너네 제주대 감귤포장학과 있어?

1.1: 구현 계?기

일단은 제주사람이 ‘육지’ 대학교에 오게되면서 많이 듣는 말 리스트

1. 제주도 사투리 할줄 알아?
2. 너네집 감귤농사해?
3. 너네 제주대 감귤포장학과 있어?

1.1: 구현 계?기

일단은 제주사람이 ‘육지’ 대학교에 오게되면서 많이 듣는 말 리스트

1. 제주도 사투리 할줄 알아?
2. 너네집 감귤농사해?
3. 너네 제주대 감귤포장학과 있어?

1.1: 구현 계?기

일단은 제주사람이 ‘육지’ 대학교에 오게되면서 많이 듣는 말 리스트

1. 제주도 사투리 할줄 알아?
2. 너네집 감귤농사해?
3. 너네 제주대 감귤포장학과 있어?

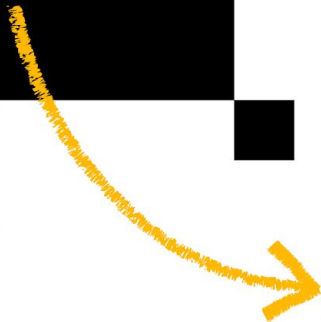
```
Text(  
  'Section Title',  
  style: TextStyle(  
    color: Colors.blue[200],  
  ),  
),  
),  
s.star,  
r: Colors.blue[500],  
Text('23'),
```

devfest



Google Developer Groups

GDG Cloud Korea, GDSC Yonsei & Ewha



1.2: GCP Vertex AI 이란?

1.2: GCP Vertex AI 이란?

구글에서 만든 AI 종합 선물세트(?)!

사용자는 데이터셋하고 모델을 선택하기만 해도 된다!



1.2: GCP Vertex AI 이란?

PaLM같은 모델을 선택해서
Fine-Tuning (파인튜닝)을 할수있는 Gen AI Studio



1.2: GCP Vertex AI 이란?

해당 제품들은 거의 B2B (비즈니스) 제품이라고 봐도 무방합니다
그래서 일반 사용자들에게도 홍보를 할려고 노력을 하고있고요

아 그리고 Bard (PaLM2) 기능을 제공하고있기도 하고요

프롬프트를 커스텀 SDK / API를 사용자에게 제공을
하고있기도합니다

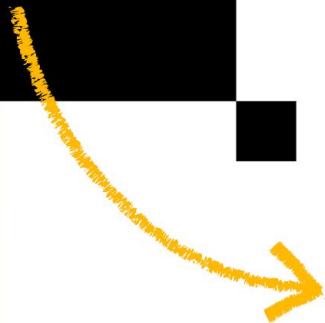
```
Text(  
  'Section Title',  
  style: TextStyle(  
    color: Colors.blue[200],  
  ),  
),  
),  
s.star,  
r: Colors.blue[500],  
Text('23'),
```

devfest



Google Developer Groups

GDG Cloud Korea, GDSC Yonsei & Ewha



2: 구현

2: 구현

구현 방법

1. Model Pipeline 학습

2. PaLM2 파인튜닝

2.0: 구현 (데이터셋)

AIHub에서 제공하는 한국어 방언 발화를 이용해서
데이터셋을 구비



#지능형플랫폼 구축 #AI 돌봄 서비스 #스마트시티 데이터 허브 #스마트팜

한국어 방언 발화(제주도)

분야 한국어 유형 오디오, 텍스트

구축년도: 2020 갱신년월: 2021-11 조회수: 5,894 다운로드: 1,071 용량: 617.09 GB

다운로드 ↓ 샘플 데이터 ?

관심데이터 등록 12



2.0: 구현 (데이터셋)

많은 데이터중 음원 데이터를 제외하고
라벨링 데이터만 가지고 데이터 처리

이중 데이터 전처리 방식을
두가지로 처리함

1. 단어별 텍스트마이닝

2. 문장별 텍스트마이닝



Directory	File Name	File Size	Key	File Key Copy	Download
1.Training	(new3)라벨링데이터				
	제주도_학습용데이터_1.zip	342.53 MB	37329	파일키 복사	다운로드
(new3)원천데이터	제주도_1.zip	23.25 GB	37330	파일키 복사	다운로드
	제주도_2.zip	23.21 GB	37331	파일키 복사	다운로드
	제주도_3.zip	21.64 GB	37332	파일키 복사	다운로드
	제주도_4.zip	18.82 GB	37333	파일키 복사	다운로드
	제주도_5.zip	13.67 GB	37334	파일키 복사	다운로드
	제주도_6.zip	18.60 GB	37335	파일키 복사	다운로드
	제주도_7.zip	18.18 GB	37336	파일키 복사	다운로드
	제주도_8.zip	18.46 GB	37337	파일키 복사	다운로드

2.0: 구현 (단어별 텍스트 마이닝)

중복처리 **안한** 데이터 개수는
16,438,613개

중복처리 **한** 데이터 개수는
1,534,703개

```
dataframe = pd.read_csv("./data.csv")  
dataframe = dataframe.drop_duplicates()  
  
dataframe
```

	source	target
0	{laughing}	{laughing}
1	진짜	진짜
2	나	나
3	캡처	캡처
4	해냈었는데	해냈었는데
...
16438571	띄워놔두고	띄워놔두고
16438579	서로(막)/(아주)	서로(막)/(아주)
16438593	맞져.	맞지.
16438596	없주게.	없지 그래.
16438604	사분.	사분.

1534703 rows × 2 columns

2.0: 구현 (문장별 텍스트 마이닝)

문장별 데이터는 9,341,090개

```
> ~
import pandas as pd

pd.read_csv("./data.csv")

[1] ✓ 15.8s
```

	eojeol	standard
0	{laughing} 진짜	{laughing} 진짜
1	나 캡처 해놨었는데	나 캡처 해놨었는데
2	아 재 못말리크라	아 재 못말리겠다
3	이에스에프피	이에스에프피
4	아 완전 이에스에프피 재질이다 너	아 완전 이에스에프피 재질이다 너
...
9341085	그냥 잠만 꼬질꼬질(행)/(해서) 자면 할망이 그자 (())하면 아기 기저귀도 갈고...	그냥 잠만 꼬질꼬질(행)/(해서) 자면 할머니가이 그저 (())하면 아기 기저귀도 ...
9341086	애기 낳다하면 난 몰라 할망만 다 허여	애기 낳았다고 하면 난 몰라 할머니가 다 해
9341087	경행이제 우리 서방은 일 갔당으네 집이 들어오면 방에 들어왕	그렇게 해서이제 우리 서방은 일 갔다가 집 들어오면 방에 들어와서
9341088	얼굴만 영 바뻥 지는 오썩 할망 방에 잠자불곡	얼굴만 이렇게 봐두고 자기는 후딱 할머니 방에 잠자버리고
9341089	방에도 오지도 안 해 또 나 나갈 때 되면 얼굴만 왕 영 봐뵈 아침이 봐뵈 그냥	방에도 오지도 안 해 또 나 나갈 때 되면 얼굴만 와서 이렇게 봐두고 아침에 봐두고 그냥

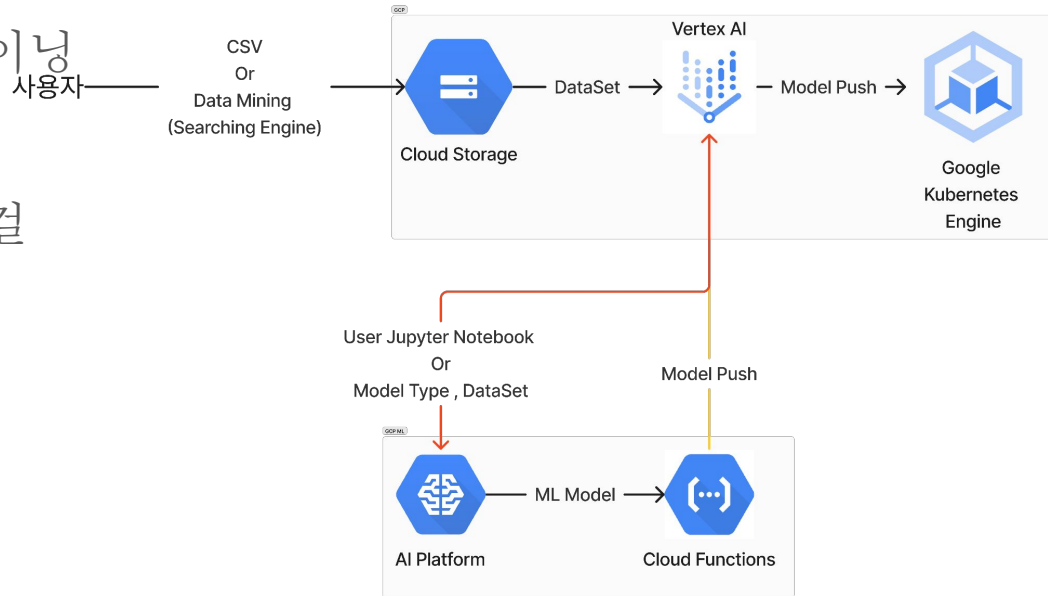
9341090 rows × 2 columns

2.2: 구현 (Model Pipeline 학습)

해당 기능은 GCP에서 ML Pipeline를 사용하는것으로
개발자가 모델 선정부터 데이터 관리까지
모두 전부다 할수있습니다.

구글 제품을 이용해서 데이터 마이닝
및 전처리도 가능합니다

사용자가 직접 데이터 전처리 한걸
올릴수도있고요



2.2: 구현 (Model Pipeline 학습)

일단은 이것을 해봤습니다... 에 ...

Error Messages: There are too many AnnotationSpecs in the dataset. Up to 5000 AnnotationSpecs are allowed in one Dataset. Check your csv/jsonl format with our public documentation.

2.2: 구현 (Model Pipeline 학습)

일단은 이것을 해봤습니다... 에 ...

에 ... 데이터 5000개 제한이라서 더 넣질 못해가지고
5000개 짜리로 넣어봤습니다

2.2: 구현 (Model Pipeline 학습)

일단은 이것을 해봤습니다... 에 ...

에 ... 데이터 5000개 제한이라서 더 ...
5000개 짜리로 넣어봤습니다



Vertex AI <noreply-vertexai@google.com>

나에게 ▼

Hello Vertex AI Customer,

Vertex AI partially imported data into dataset "untitled_1701047507672", but some resources were unable to be imported.

Additional Details:

Operation State: Partially failed with errors

Resource Name:

projects/39311999199/locations/us-central1/datasets/61331561240833556

Error Messages: OK

To view the error on Cloud Console, go back to your dataset using

<https://console.cloud.google.com/vertex-ai/locations/us-central1/datasets/61331561240833556>

Sincerely,

The Google Cloud AI Team

2.2: 구현 (Model Pipeline 학습)

일단은 이것을 해봤습니다... 에 ...

에 ... 데이터 5000개 제한이라서 더 ...
5000개 짜리로 넣어봤습니다

???: 뭐가 OK인데?



Vertex AI <noreply-vertexai@google.com>

나에게 ▼

Hello Vertex AI Customer,

Vertex AI partially imported data into dataset "untitled_1701047507672", but some resources were unable to be imported.

Additional Details:

Operation State: Partially failed with errors

Resource Name:

projects/39311999199/locations/us-central1/datasets/61331561240833556

Error Messages: OK

To view the error on Cloud Console, go back to your dataset using

<https://console.cloud.google.com/vertex-ai/locations/us-central1/datasets/61331561240833556>

Sincerely,

The Google Cloud AI Team

2.2: 구현 (Model Pipeline 학습)

쫓든 이제 Google에서 제공하는 AutoML를 이용해서 Model를 학습했습니다

2.2: 구현 (Model Pipeline 학습)

쫘든 이제 Google에서 제공하는 AutoML를 이용해서 Model를 학습했습니다... 만...

2.2: 구현 (Model Pipeline 학습)

쫘든 이제 Google에서 제공하는 AutoML를 이용해서 Model를 학습했습니다... 만...

EFT)

BERT (PEFT)

Finetune and deploy BERT with PEFT.

[배포](#) [CO](#) [노트북 열기](#)

[OVERVIEW](#) [USE CASES](#) [DOCUMENTATION](#) [PRICING](#)

Overview

Bidirectional Encoder Representations from Transformers (BERT) is a deep learning model published in the paper "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (2018). Unlike previous language representations, BERT can be fine-tuned with just one additional input layer to answer a wide range of tasks such as question classification, sentence classification, and named entity recognition.

Parameter-efficient fine-tuning (PEFT) is a technique for fine-tuning large pre-trained models for downstream applications with a small number of trainable parameters. PEFT methods include LoRA, prefix-tuning, and P-tuning.

This model card demonstrates how to use PEFT to fine tune the BERT model on Vertex AI and deploy it as a serving endpoint.

엔드포인트 배포가 실패했습니다.

오류 메시지: 'The following quotas are exceeded: CustomModelServingT4GPUsPerProjectPerRegion'

[의견 보내기](#) [닫기](#)

리소스 ID
BERT-PEFT-001

태그

태스크

[생성](#)



2.2: 구현 (Model Pipeline 학습)

쫘든 이제 Google에서 제공하는 AutoML를 이용해서 Model를 학습했습니다... 만...

BERT (PEFT)

Finetune and deploy BERT with PEFT.

배포 CO 노트북 열기

OVERVIEW USE CASES DOCUMENTATION PRICING

Overview

Bidirectional Encoder Representations from Transformers (BERT) is a deep learning model that can be used for a variety of tasks, including text classification, question answering, and natural language inference. Unlike previous models, BERT is trained on a large corpus of unlabeled text by using a self-supervised learning objective. This allows BERT to learn rich representations of text that can be used for a wide range of downstream tasks.

Parameter-efficient fine-tuning (PEFT) is a technique that allows you to fine-tune a pre-trained model on a specific task without having to update all of the model's parameters. This is done by using a small number of additional parameters to adapt the pre-trained model to the new task. PEFT is a powerful technique that can be used to fine-tune a wide range of models, including BERT, on a variety of tasks.

This model card demonstrates how to use PEFT to fine tune the BERT model on Vertex AI and deploy it as a serving endpoint.

엔드포인트 배포가 실패했습니다.

오류 메시지: The following quotas are exceeded:
CustomModelServingT4GPUsPerProjectPerRegion

의견 보내기 닫기

태스크

생성

To downstream... ber of (extra)... for PEFT include

Error Messages: There is insufficient training data labeled '아 그래?' in Dataset (371389278670, 8348047531447418880) for automatic splitting. Please manually assign more items with the label into the training data item set, or import more data.



2.1: 구현 (PaLM2 파인튜닝)

PaLM은 간단하게 Bard!

Google에서 만든 Bard의 Model를
Fine-Tuning을 할수있게 서비스를 구현

해당 서비스 말고도 GPT를 가지고
Fine-Tuning을 할수있다

또한 해당 기능들을 사용할수있는 서비스를
Generative AI Studio이라고 이야기한다



2.1: 구현 (PaLM2 파인튜닝)

일단은 우리가 만든 데이터셋을 좀더
가공해야합니다

Input와 output를 관리를 이렇게 해서
Cloud Storage를 저장하거나 구현하거나

Python Vertex AI SDK를 이용해서
구현하거나

Gcloud (Shell)를 이용해서 구현하거나
Colab가지고 구현하거나

JSON 파일 예시

JSON 키-값 쌍에 대한 자세한 내용은 [문서](#)를 참조하세요.

```
[
  {
    "inputs": [
      "Strong",
      "Hot"
    ],
    "outputs": [
      "Weak | Cold"
    ]
  }
]
```


2.1: 구현 (PaLM2 파인튜닝)

▼ Test ?

입력 입력 프리픽스 추

애는 또 왜 안되는건데

2.1: 구현 (PaLM2 파인튜닝)

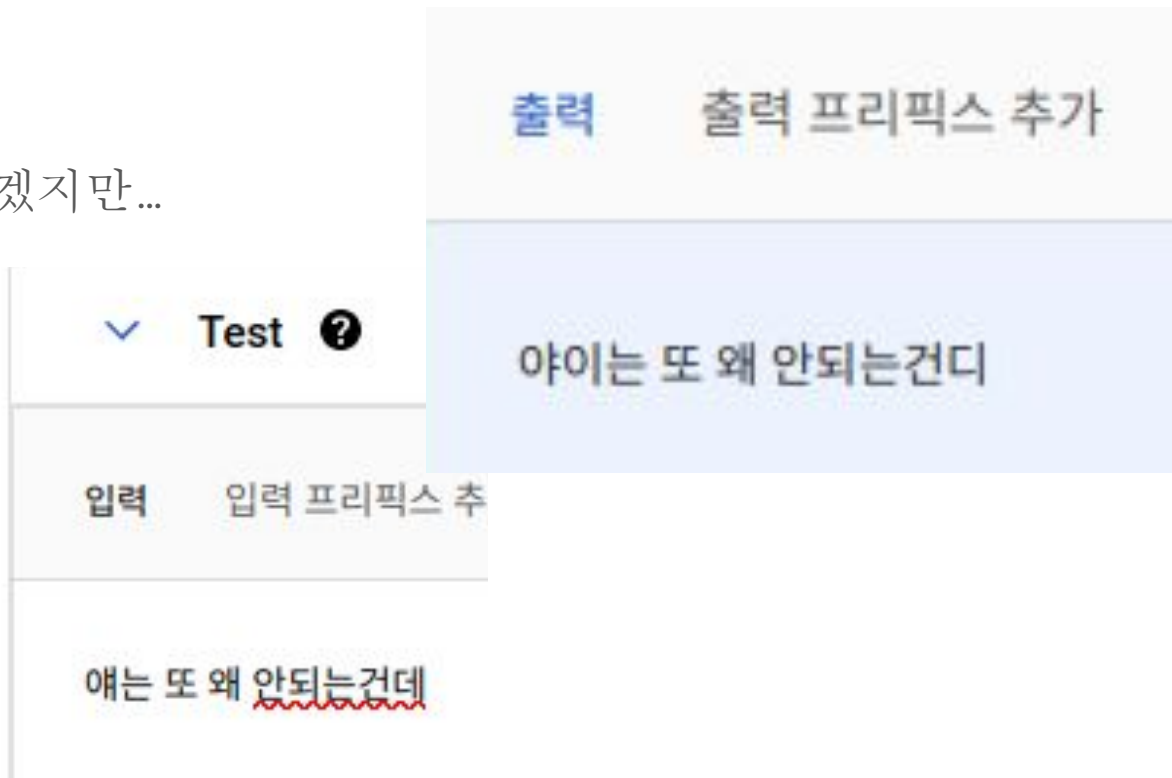
	출력	출력 프리픽스 추가
▼ Test ?	야이는 또 왜 안되는건디	
입력	입력 프리픽스 추	
애는 또 왜 <u>안되는건데</u>		

2.1: 구현 (PaLM2 파인튜닝)

결국은 됐습니다.

데이터를 더 넣어서

Fine-Tuning하면 되겠지만...



2.1: 구현 (PaLM2 파인튜닝)

결국은 됐습니다.

데이터를 더 넣어서

Fine-Tuning 하면 더 개기만.



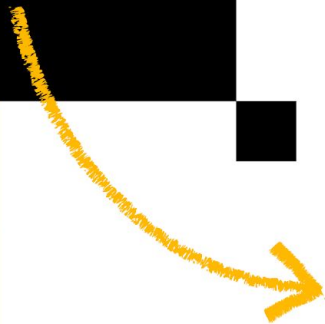
```
Text(  
  'Section Title',  
  style: TextStyle(  
    color: Colors.blue[200],  
  ),  
),  
),  
s.star,  
r: Colors.blue[500],  
Text('23'),
```

devfest



Google Developer Groups

GDG Cloud Korea, GDSC Yonsei & Ewha



3: 결과

3: 결과

네... ML Pipeline 프로그램을 신박하게 볼수있었습니다.

Google에서 KubeFlow를 어떻게 쓰는지
기본 스탠다드를 볼수있어서 매우 좋았습니다

만약에 좀 급박하고 Toy프로젝트를 할때
해당 기능을 쓰시면 될것 같습니다

3: 결과

네... ML Pipe

Google에서
기본 스탠다드

만약에 좀 급
해당 기능을

[Web발신]
[카카오뱅크] 카드결제실패
임*르(2361)
12/02 03:24
75,233원
구글클라우드코리아
잔액부족

3:24 AM



devfest



```
// You'll need  
// com.google.  
listRef.listAll  
.addOn  
prefixes.  
// All  
// You  
}  
it  
ach { item  
the items  
}  
}
```

제주도 사투리 번역 프로그램 만들어보기

대구가톨릭대학교 AI빅데이터공학과 23학번,
None Labs 대표 | 임미르



Google Developer Groups

GDG Cloud Korea, GDSC Yonsei & Ewha

