

¿Quién y qué dice?

Laura Villalba

Proyecto final | Ironhack | Data Analytics

Proyecto: ¿Quién y qué dice?

Objetivo

Generador de subtítulos para ‘podcasts’.

Motivación

Facilitar el formato ‘podcast’ a las personas con discapacidad auditiva.

Estructura de la API

1. Tratamiento de audios → `cleanOriginals.py`
2. Generación de Features → `featuresExtract.py`
3. Entrenamiento del modelo → `trainModel.py`
4. Generador de mezcla de audios → `audioGenerator.py`
5. Predicciones (¿quién lo dice?) → `predictions.py`
6. Transcripción (¿qué dice?) → `transcription.py`

En detalle

— — —

- Dataset → OpenSRL (test-other.tar.gz)
- Red Neuronal:

```
model = models.Sequential()
model.add(layers.Dense(512, activation='relu', input_shape=inshape))
model.add(layers.Dense(256, activation='relu'))

model.add(BatchNormalization())
model.add(layers.Dense(128, activation='relu'))
model.add(layers.Dense(128, activation='relu'))
model.add(layers.Dense(number_classes, activation='softmax'))
model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy', 'mse'])

model.fit(X_train,
        y_train,
        epochs=50,
        batch_size=50,
        validation_data=(X_test, y_test))
```

MFCC:

- loss: 3,5138
- acc: 0,10000
- mse: 27,6369

15':

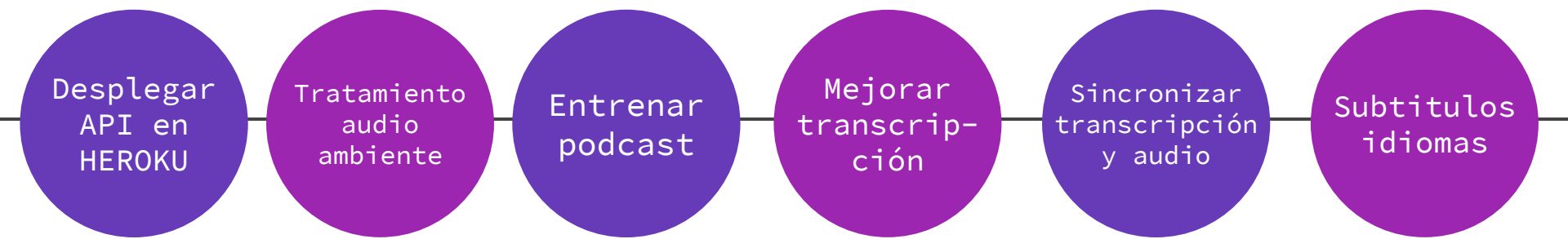
- loss: 1,6733
- acc: 0,5639
- mse: 27,6811

FFT:

- loss: 0,00240
- acc: 1.0
- mse: 27,6996

DEMO

Próximos pasos



Desplegar
API en
HEROKU

Tratamiento
audio
ambiente

Entrenar
podcast

Mejorar
transcrip-
ción

Sincronizar
transcripción
y audio

Subtítulos
idiomas

Gracias!

Laura Villalba