Изображение выглядит как текст, часы, знак

Автоматически созданное описание

**Deep Learning 2021 course**

**Final Project: Speech Recognition**

Zhapar Zhuldyz

Zhusupbaeva Malika

Omarov Adilkhan

CONTENTS:

1. Motivation and objective

2. Related work and originality

3. Design architecture

4. Detailed algorithm or functions

5. Coding

6. Results and Performance evaluation

7. Conclusion

8. References

9. Roles of members

**Motivation and objective:**

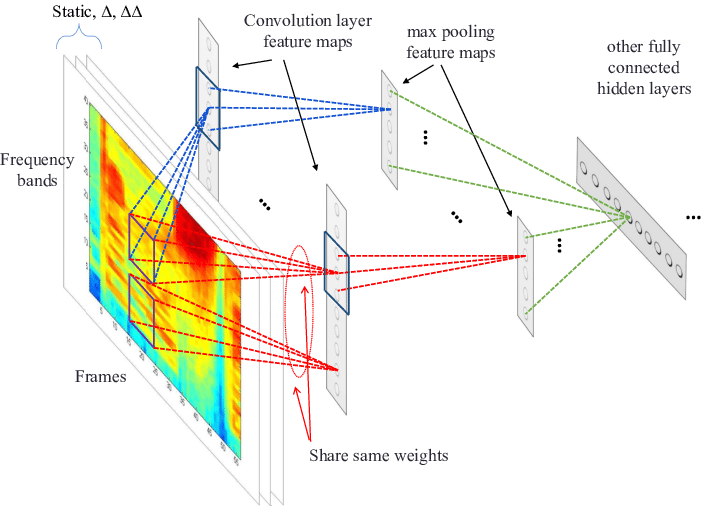
Speech recognition allows documents to be created faster because the software generally produces words as quickly as they uttered, which is usually much faster than a person can type. Our main motivation was to create a voice assistant like SIRI.

**Related Work and Originality:**

1. Kaggle challenge – TensorFlow Speech Recognition Challenge

We have taken main architecture from scripts of challengers and adjusted for ourselves.

**Design Architecture:**

Convolutional neural networks (CNNs) have been recently used for acoustic modeling and feature extraction in speech recognition systems, where their inputs have been speech spectrogram or even raw speech signal.

**4. Detailed algorithm or functions**

import librosa

from sklearn.preprocessing import LabelEncoder

from keras.utils import np\_utils

from sklearn.model\_selection import train\_test\_split

from keras.layers import Dense, Dropout, Flatten, Conv1D, Input, MaxPooling1D

from keras.models import Model

from keras.callbacks import EarlyStopping, ModelCheckpoint

from keras import backend as K

**Coding**

Librosa library in python have a MFCC(Mel-frequency cepstrum )

INPUT DATA – wav file

we read this file with librosa library

Implemetation

CNN model

We found that we can use train model with 2 diffrenet methods

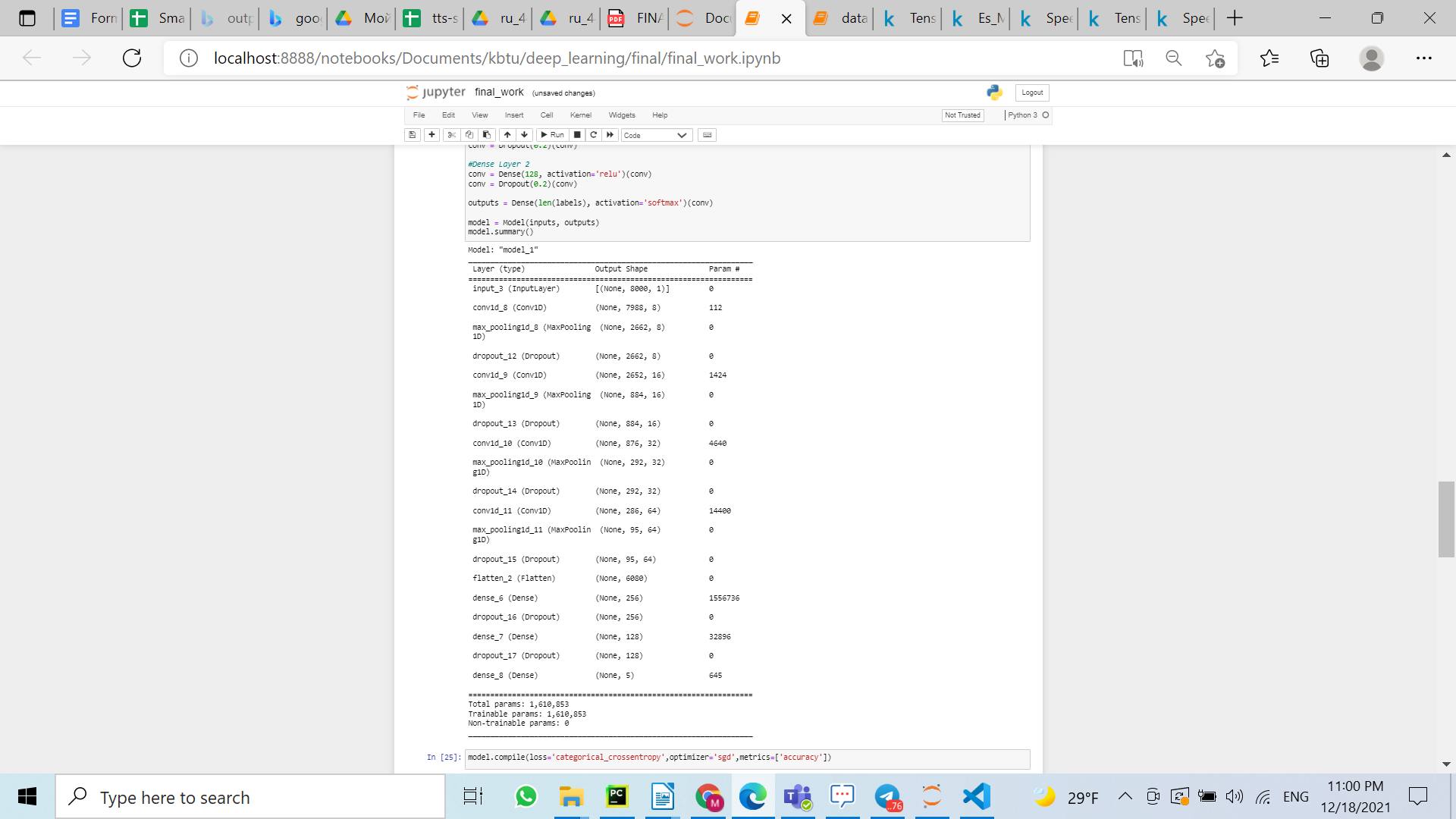
1) MFCC(Mel-frequency cepstrum ) with spectogram , using cv models to classify the images

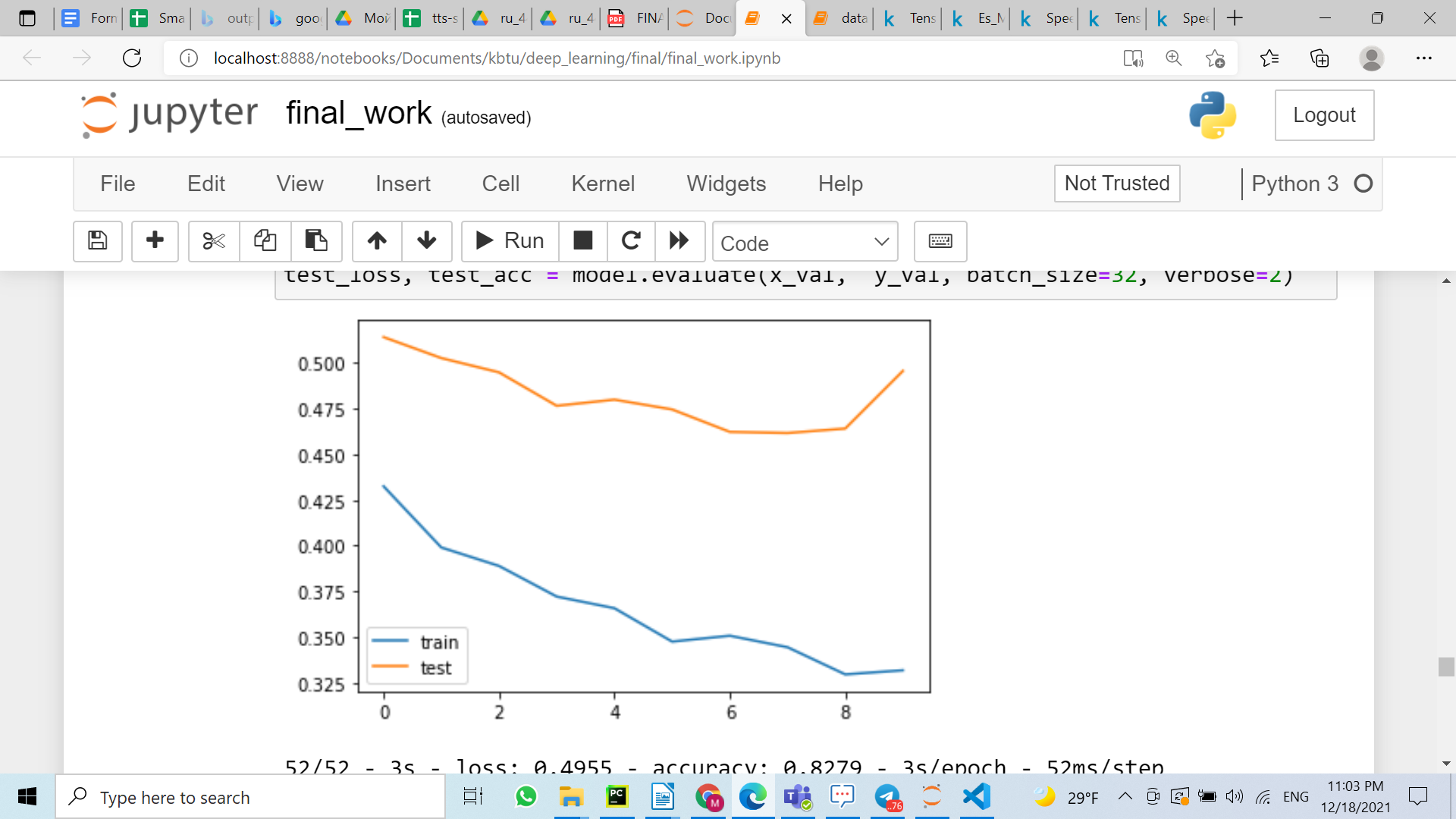
extracting mfcc features

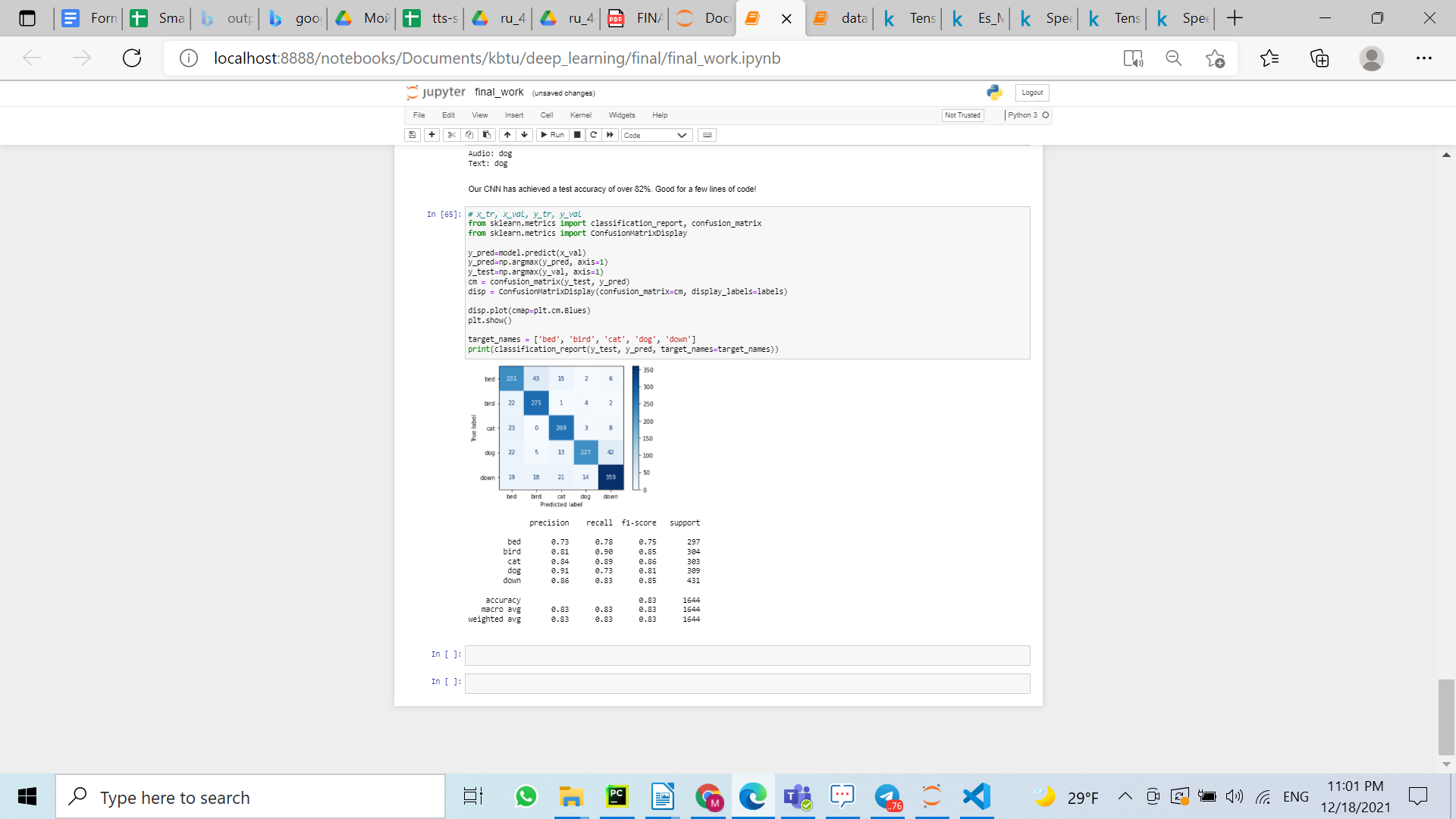
2) Using

label and numpy array of sampels

**6. Results and Performance evaluation**

****

****



**7. Conclusion**

We learned how sound works and what it consists of. It was interesting to explore his processing and methods for extracting payloads to train the model. We did not manage to consider all methods of training models, such as kaldi asr. But we decided to use an inexpensive and simple cnn.

**8. References**

https://keras.io/examples/audio/transformer\_asr/

https://www.youtube.com/watch?v=Qf4YJcHXtcY

[Extracting Mel-Frequency Cepstral Coefficients with Python - YouTube](https://www.youtube.com/watch?v=WJI-17MNpdE)

[Распознавание речи-понимание MFCC - Русские Блоги (russianblogs.com)](https://russianblogs.com/article/1220415980/)

[Introduction to Audio Classification with PyTorch - Learn | Microsoft Docs](https://docs.microsoft.com/en-us/learn/modules/intro-audio-classification-pytorch/)

**9. Roles of members**

**Adil**

preparing dataset, and anylize ways to speech recognition, such as mfcc and spectogram

**Malika and Zhuldyz**

training dataset to with different type that we pass on lecture, eventually using one model 1dConv