

T/R

Kaggle dataset



01

LANGUAGE IDENTIFICATION DATASET

WiLI-2018, the Wikipedia language identification benchmark dataset, contains 235000 paragraphs of 235 languages.

Each language in this dataset contains 1000 rows/paragraphs.

After data selection and preprocessing I used the 22 selective languages from the original dataset Which Includes following Languages

- English
- Turkish
- Arabic
- Estonian
- French
- Russian
- Hindi
- Romanian
- Urdu
- Chinese
- Portuguese
- Swedish
- Persian
- Latin
- Pushto
- Indonesian
- Spanish
- Dutch
- Korean
- Japanese
- Tamil
- Thai

<https://www.kaggle.com/zarajamshaid/language-identification-datasst>

02

FACE RECOGNITION

The data set contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured.

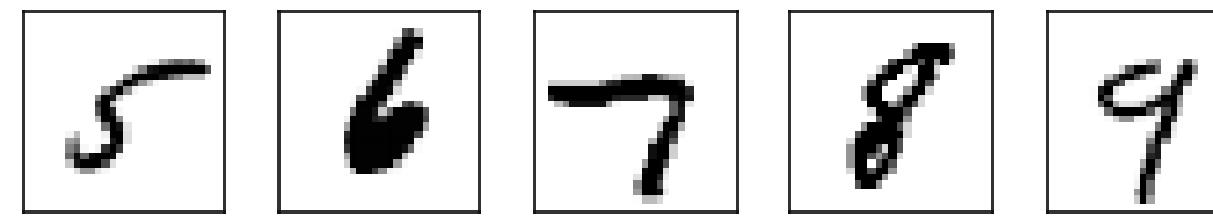
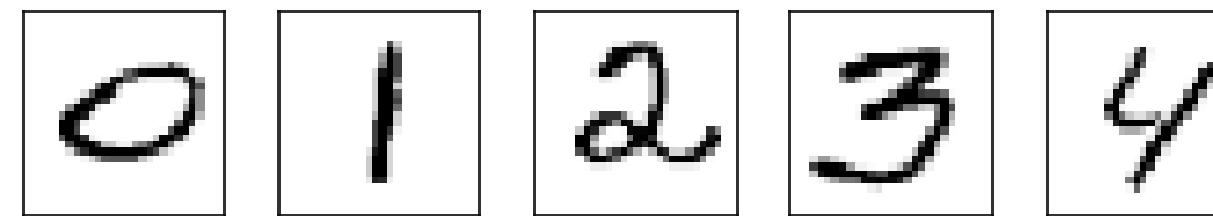
The only constraint on these faces is that they were detected by the Viola-Jones face detector.



<https://www.kaggle.com/atulanandjha/lfwpeople>

03

DIGIT RECOGNIZER



<https://www.kaggle.com/c/digit-recognizer/data>

04

PRODUCT RECOMMENDATION SYSTEM FOR E-COMMERCE

This recommendation system will help the users get a good recommendation to start with and once the buyers have a purchased history, the recommendation engine can use the model based collaborative filtering technique.

<https://www.kaggle.com/shawamar/product-recommendation-system-for-e-commerce/notebook>

<https://www.kaggle.com/skillsmuggler/amazon-ratings>

T/R

Thank you!

Now, let's get back to work.