資料分析

Gender Recognition by Voice

組員：

00557103游俊弘

00657207林弈呈

00457131丁兆文

TABLE OF CONTENTS

1 INTRODUCTION
About our project

2 Dataset
Data categories

3 Our Thought
What we want to do analysis

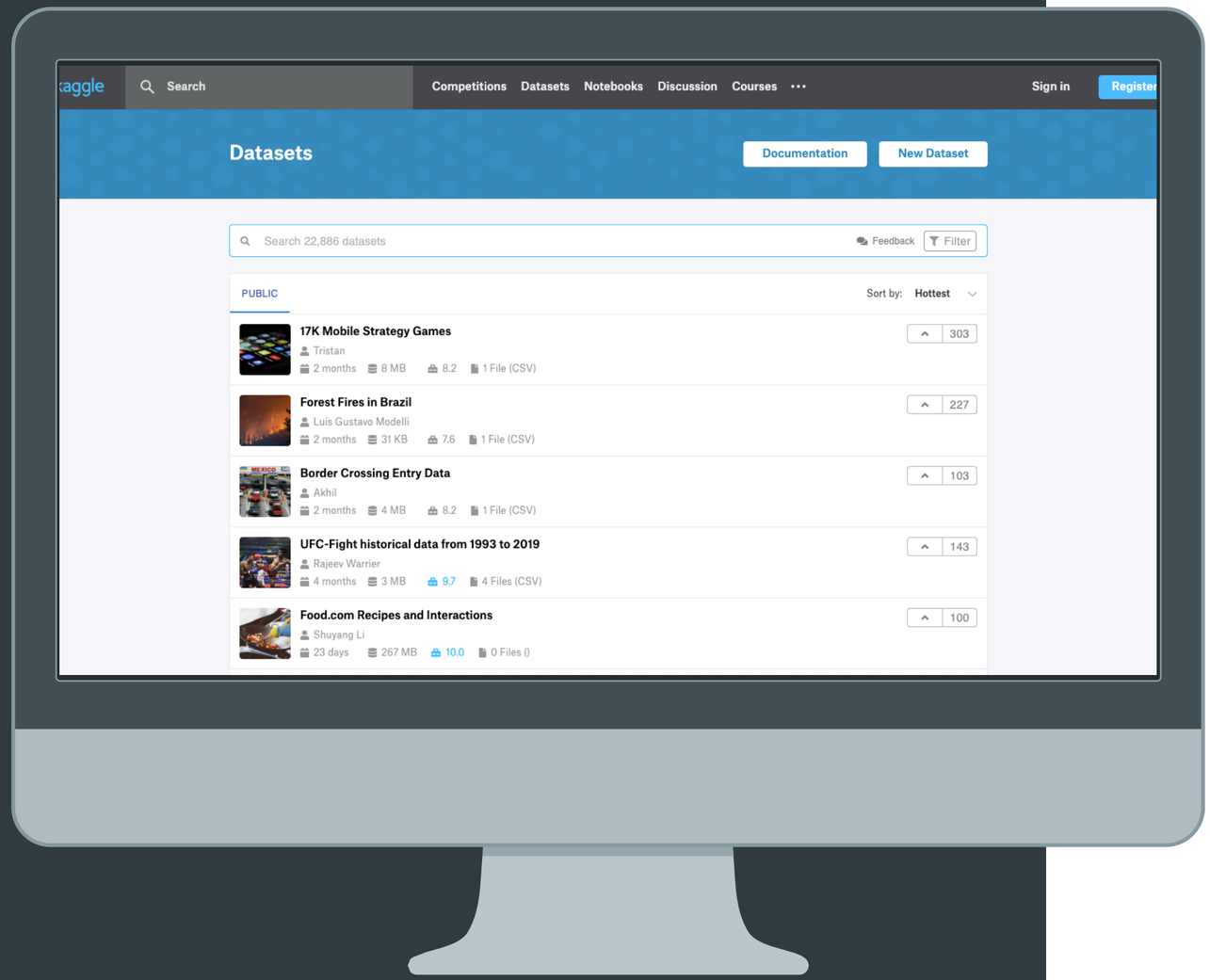
SECTION 1

INTRODUCTION

Gender Recognition by Voice and Speech Analysis

Identify a voice as male or female

- This database was created to identify a voice as male or female, based upon acoustic properties of the voice and speech.
- The dataset consists of 3,168 recorded voice samples, collected from male and female speakers.
- The voice samples are pre-processed by acoustic analysis in R using the seewave and tuneR packages, with an analyzed frequency range of 0hz-280hz



Use the dataset By Kaggle

https://www.kaggle.com/primaryobjects/voicegender

SECTION 2

DATASET

Dataset

data categories

- Dataset: 3619*22
- Filename: voice.csv
- Data based on acoustic properties of the voice and speech (frequence:kHz)
  - meanfreq、sd、median、Q25、Q75、IQR、skew、kurt、sp.ent、sfm、mode、centroid、meanfun、minfun、maxfun、meandom、mindom、maxdom、dfrange、modindx、class

Question

■主要

1．男性和女性聲音之間還有哪些其他特徵？

2．可以發現男性和女性聲音在共鳴(resonance)上有區別嗎？

■次要

1．可以從正常聲音中識別假音(falsetto)嗎？ （為此可能需要單獨的數據集）

2．數據中還有其他有趣的功能嗎？

SECTION 3

OUR THOUGHT

Analysis

Use different algorithm

- 期望找到最適合的演算法，達到最精準的分析
- 用不同的演算法進行分析，例如
  a. Logistic Regression
  b. CART algorithm
  c. SVM
  d. etc..
- 對dataset進行feature selection，例如
  a. PCA
  b. LDA

“以人耳看來，用聲音決定性別能否依靠簡單的頻率來決策呢？”

References

- https://www.kaggle.com/primaryobjects/voicegender
- http://www.primaryobjects.com/2016/06/22/identifying-the-gender-of-a-voice-using-machine-learning/

THANK YOU!