AI Final Project

成員：

王力得、朱哲毅、吳維誠

INTRODUCTION

交大學生外出不便，時常利用 Youbike 來當作交通工具。
然而，常常在需要自行車時，學生卻只能看到空無一車的租借站，因此想利用所學來嘗試解決共享單車租借預測的問題。

LITERATURE REVIEW

THEM

- dataset歪斜 → Log
- 太多極端值 → 重新預估

US

- 標準化
- 把一個[0:3]的欄位拆成4x[0:1]

DATASET

原計畫

利用爬蟲程式來爬取新竹市 Youbike 租借站狀態以及當天天氣

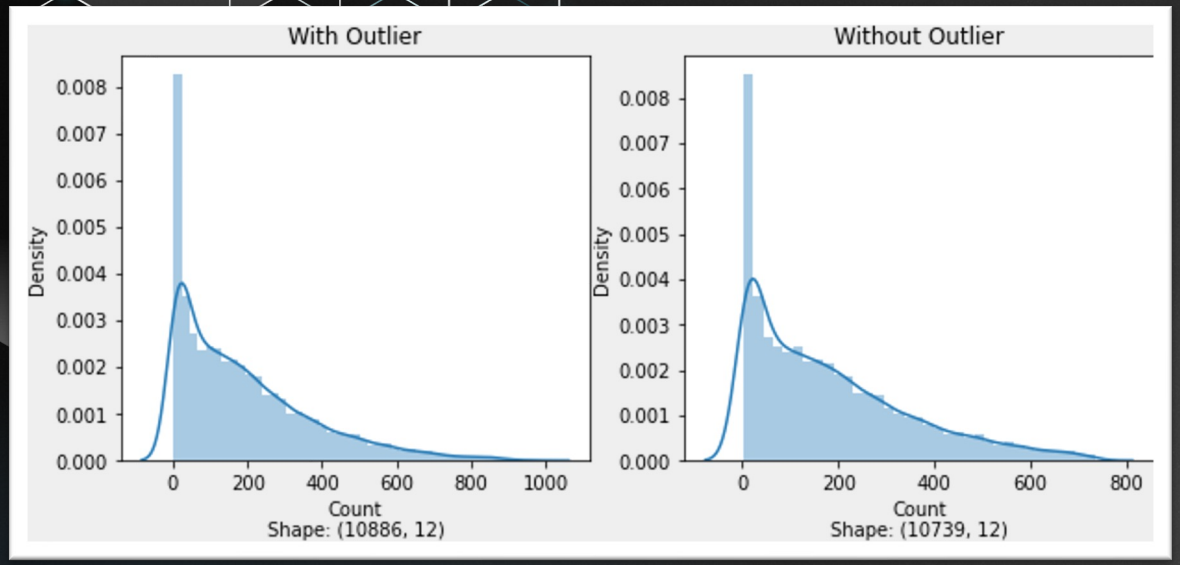
實作

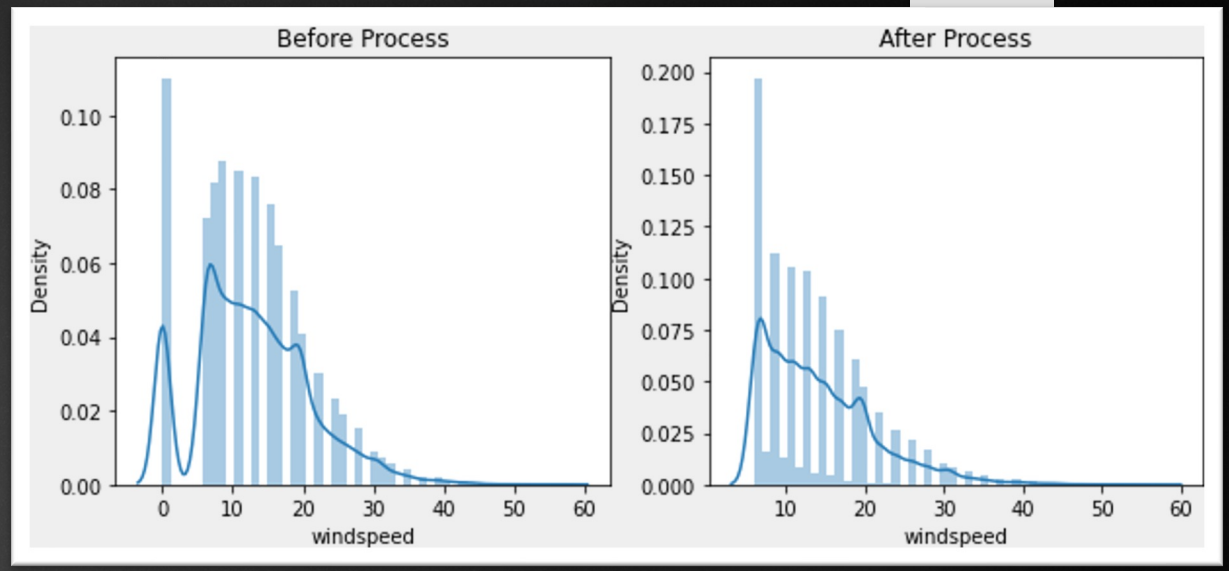
取用 Kaggle 上類似題目的 DATA

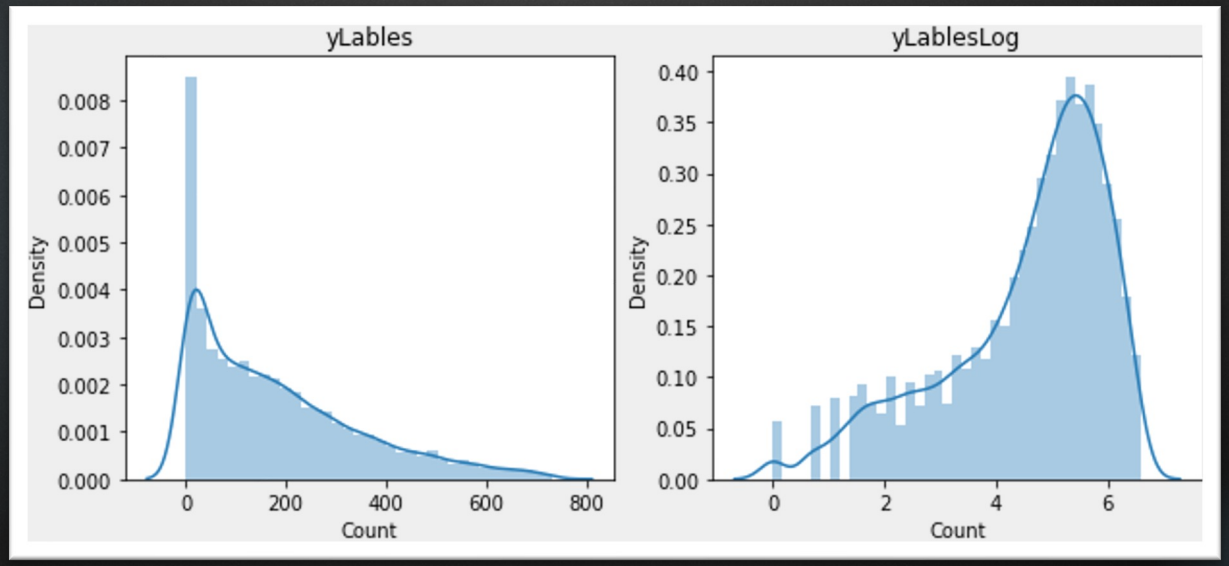
Preprocess

- 標準化 (standardization)
- 去除極端值
- 細分datetime欄位

DATASET


With Outlier
Density
Count
Shape: (10886, 12)
Without Outlier
Density
Count
Shape: (10739, 12)


Before Process
Density
windspeed
After Process
Density
windspeed


yLables
Density
Count
yLablesLog
Density
Count

BASELINE

在輸入一個時間後，程式能夠判斷該時間點是否有自行車可以租借

MAIN APPROCH

KNN

找尋最相近的k個點

Bike Sharing

Random Forest

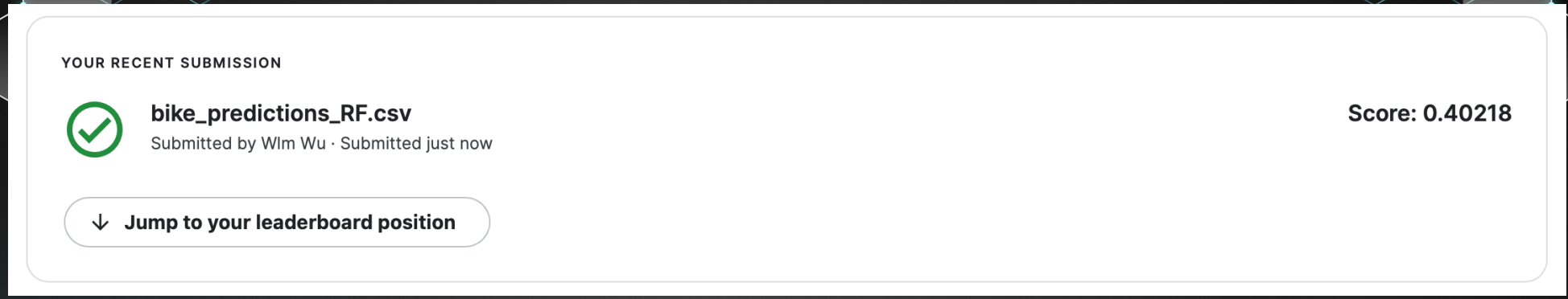
多次地隨機抽取其中n筆資料

EVALUATION METRIC

RMSLE(root mean squared log error)

$$RMSLE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(log(\hat{y}_i + 1) - log(y_i + 1))^2}$$

RESULT & ANALYSIS



Random Forest

KNN

ERROR ANALYSIS

Random Forest

n_estimators = 500

n_estimators = 1000

n_estimators = 1500

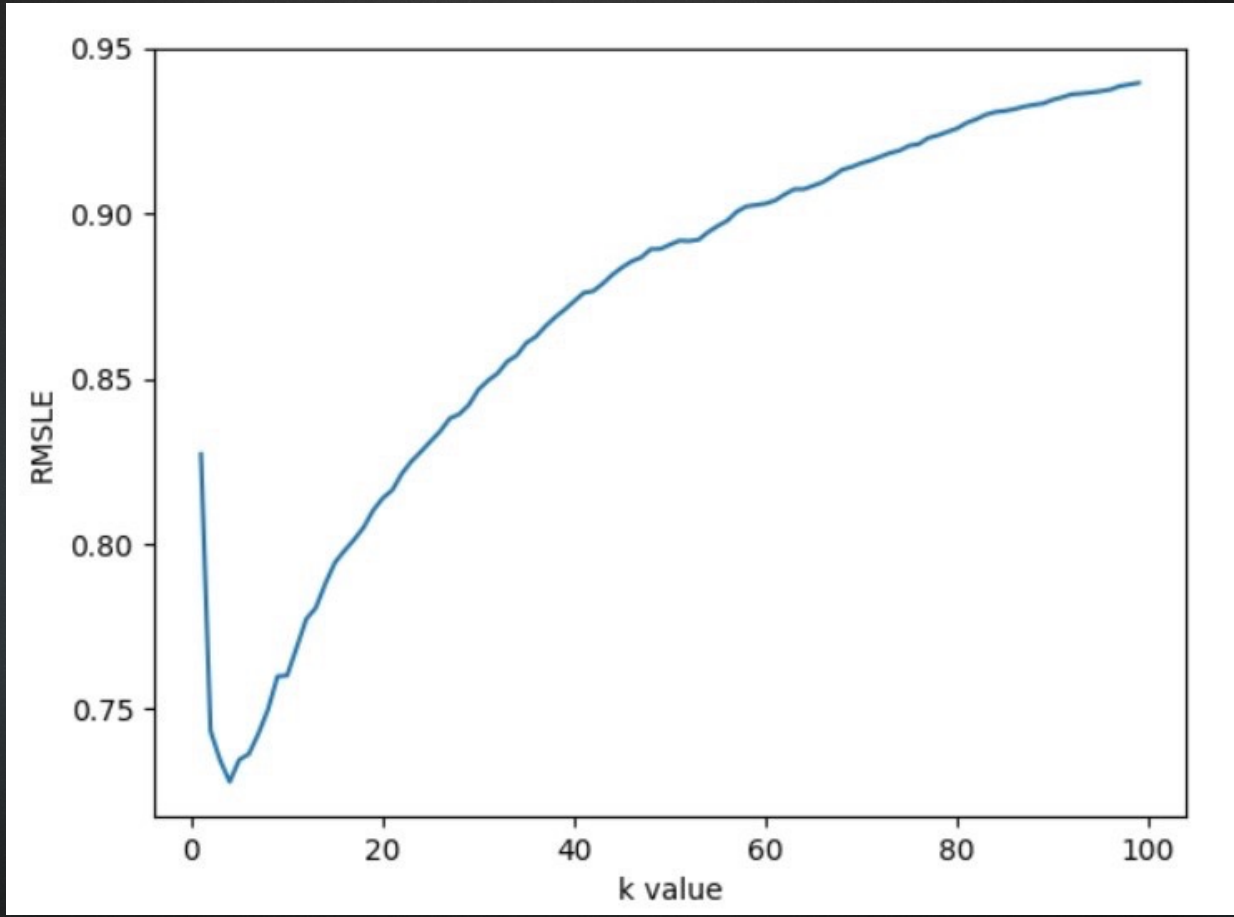
ERROR ANALYSIS

Random Forest

Without Preprocess

ERROR ANALYSIS

KNN

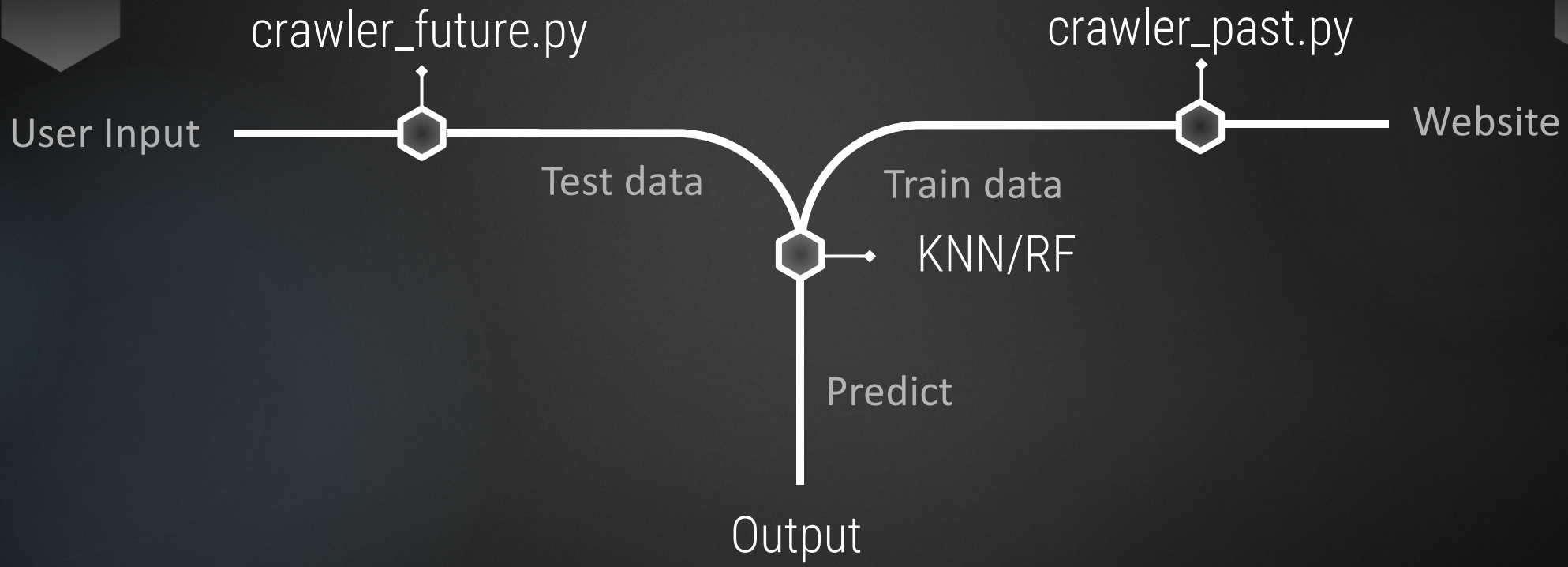


ERROR ANALYSIS

KNN

Without Preprocess

FUTURE WORK


crawler_future.py
crawler_past.py
User Input
Website
Test data
Train data
KNN/RF
Predict
Output

CODE

Our Github:
https://github.com/jerry910525/AI_project

CONTRIBUTION OF EACH MEMBER

33%

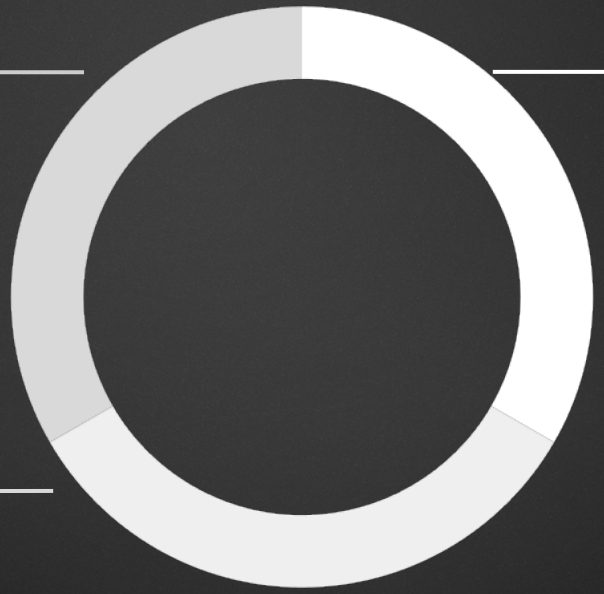
朱哲毅

爬蟲、資料預處理

33%

王力得

KNN演算法



33%

吳維誠

Random Forest演算法

REFERENCES

- https://levelup.gitconnected.com/random-forest-regression-209c0f354c84
- https://towardsdatascience.com/basic-ensemble-learning-random-forest-adaboost-gradient-boosting-step-by-step-explained-95d49d1e2725
- https://medium.datadriveninvestor.com/ensemble-learning-and-random-forest-7430ebf3da7e
- https://medium.com/@maggieliao.cm04g/ml技術筆記-3-1-ensemble-learning-bagging-and-the-gradient-boosting-model-family-intro-250c59b9ab60

THANKS

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik