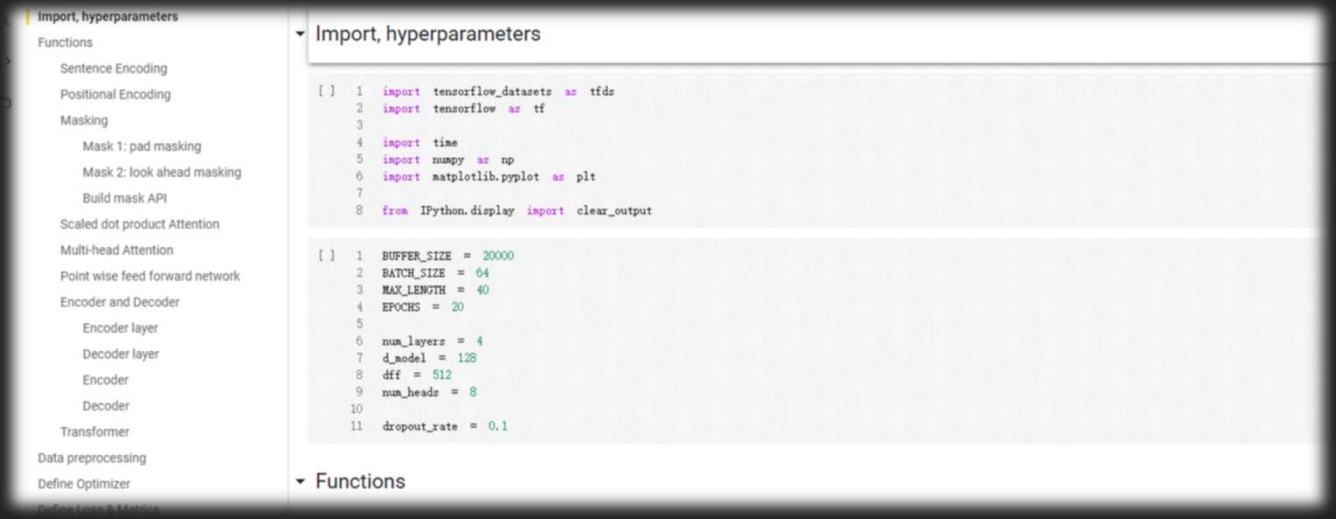
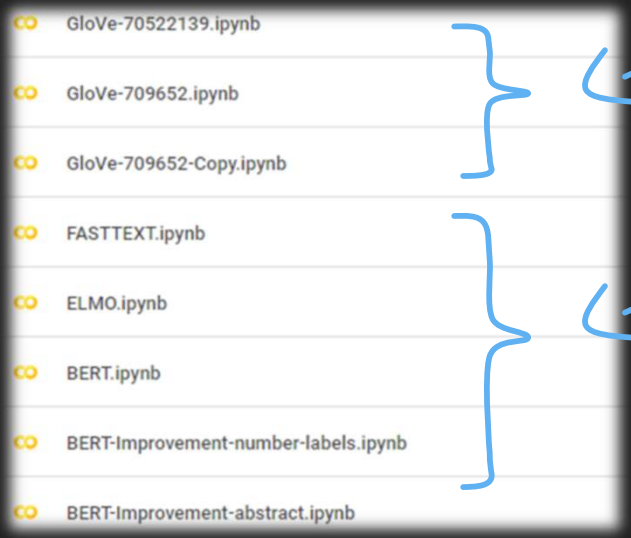
重構動機

問題一　處理 DS, ML, DL 專案時，依賴 JUPYTER NOTEBOOK 的特性
一條龍式地完成一個方法、專案，而沒有使用到任何軟體設計技巧



重構動機

問題二　雖然是解決同一個專案問題，但為了想要保存每個 NOTEBOOK 的特徵、成果，卻用最土法煉鋼的方式，寫出好幾個差不多的程式



專案語言

Python

與 JAVA 的不同

動態類型語言 (~~OVERLOADING~~)

```python
class Some:
    pass

s = Some()

s.name = "some"
print(s.name)  # "some"

s.hello = lambda: print("hello")
s.hello()  # "hello"
```

一切皆為物件

```python
def apply(func, element):
    print(func(element))

def mul(x): return x * 2
def div(x): return x // 2

apply(mul, 2)  # 4
apply(div, 2)  # 1
```

DESIGN PATTERN 參考

faif / python-patterns ★ 26.2k

zedr / clean-code-python ★ 2.4k

AICUP 2020

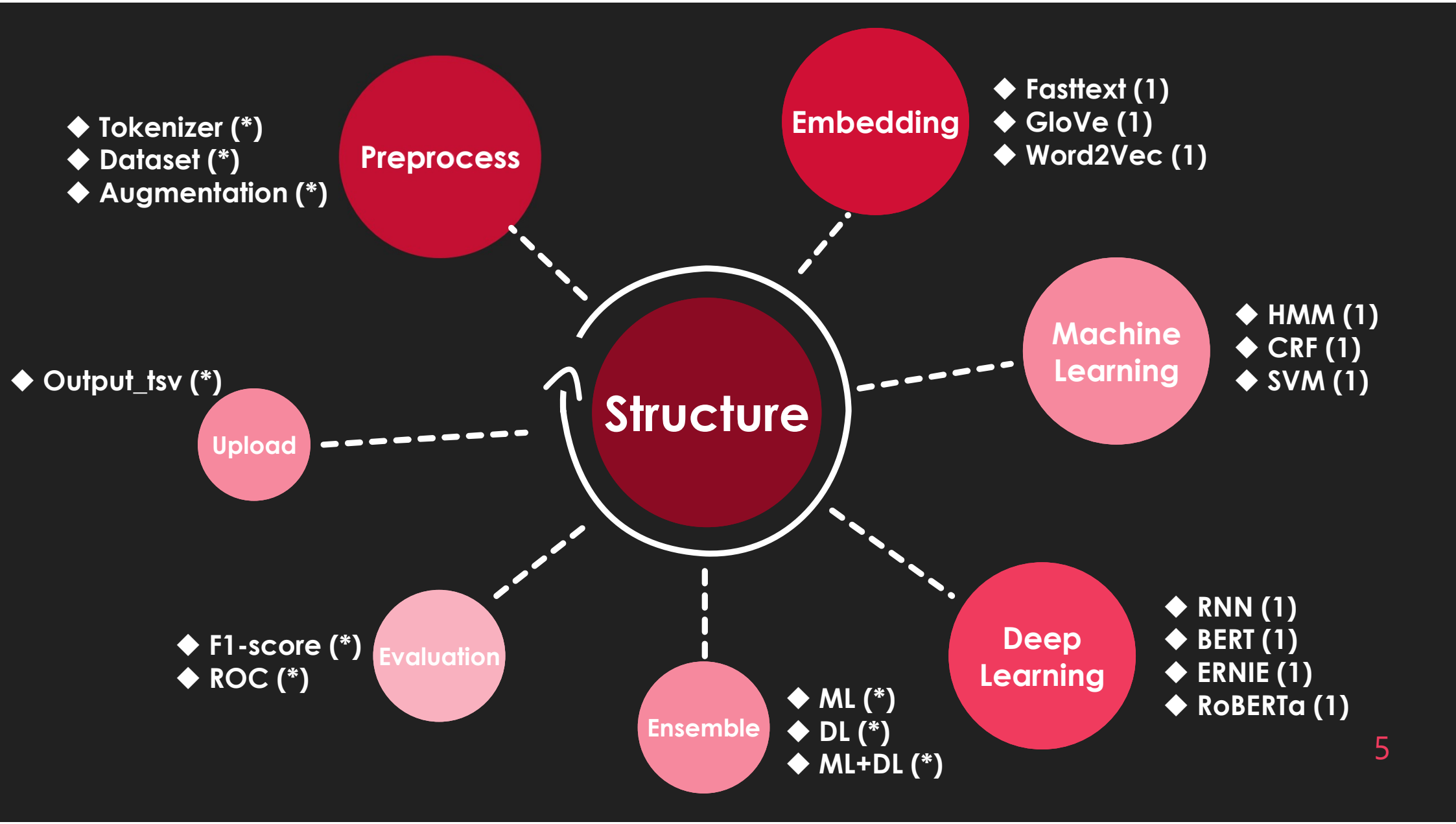
專案主題

- Deidentification of medical data
- Named Entity Recognition

醫師：你有做超音波嘛，那我們來看報告，有些部分有紅字耶。民眾：紅字是甚麼意思？醫師：就是肝功能有比較高，肝功能68，就是這個ALP是68，這樣比較高，正常應是50以下。

專案架構

- Data preprocessing method
- Embedding method
- Machine learning method
- Deep learning method
- Emsemble method
- Evaluation method
- Upload method


Structure
Preprocess
Tokenizer (*)
Dataset (*)
Augmentation (*)
Embedding
Fasttext (1)
GloVe (1)
Word2Vec (1)
Machine Learning
HMM (1)
CRF (1)
SVM (1)
Upload
Output_tsv (*)
Evaluation
F1-score (*)
ROC (*)
Ensemble
ML (*)
DL (*)
ML+DL (*)
Deep Learning
RNN (1)
BERT (1)
ERNIE (1)
RoBERTa (1)