

WK 11 Report

Optimal Random Forest Model

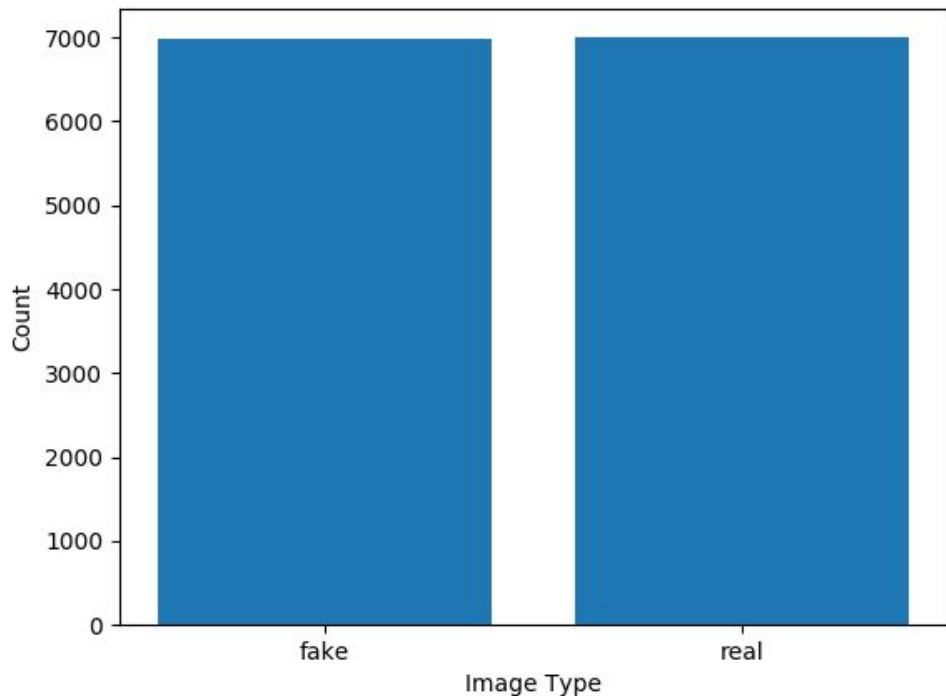
Yunting Chiu

A decorative light blue triangle is located in the bottom right corner of the slide, pointing towards the top right.

Discussed with the client

1. Still keep the number of features (image size: 320 x 320 x 3)
2. Keep images as RGB channels instead of gray scale
3. Not necessary to normalize data for a random forest model

Data Visualization



- **Variables: 307200 of X + 1 of y**
- **Observations: 13984**
- **Supervised Learning**

```
[array([134, 131, 116, ..., 68, 60, 71], dtype=uint8) 'fake']  
[array([133, 130, 115, ..., 71, 59, 71], dtype=uint8) 'fake']  
[array([117, 113, 112, ..., 43, 31, 45], dtype=uint8) 'fake']  
...  
[array([ 33,  20,  66, ..., 188, 155, 172], dtype=uint8) 'real']  
[array([ 51,  28,  46, ..., 116,  52,  50], dtype=uint8) 'real']  
[array([174, 140, 102, ...,  23,  39,  98], dtype=uint8) 'real']]
```

Random Forest

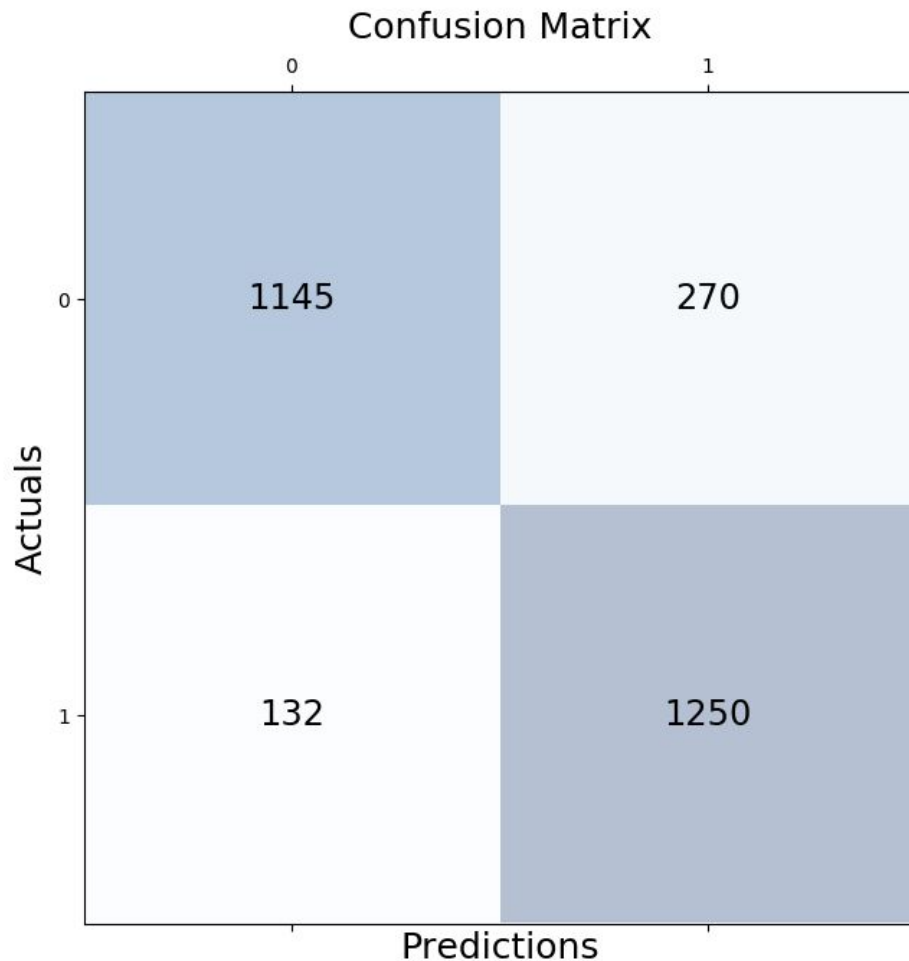
Before grid search, cv

Running time: 576 sec

Avg memory: 15481 MB

Accuracy: 85.6%

Num of tree: 100



Random Forest

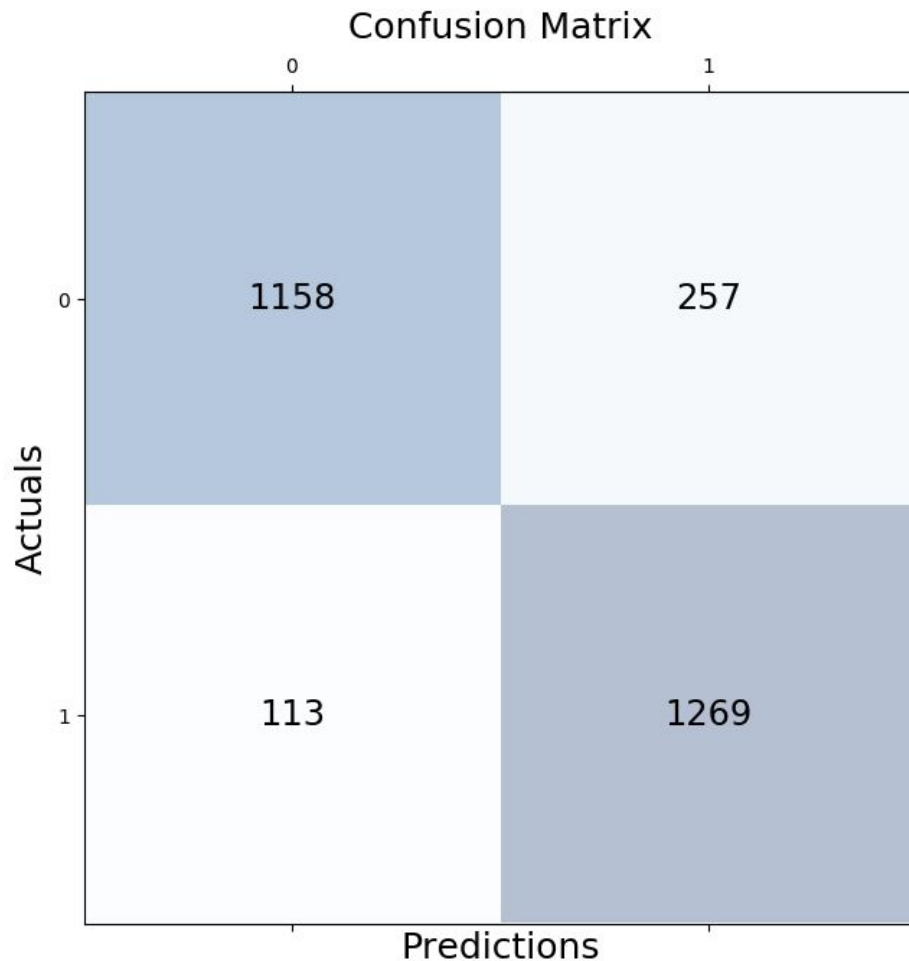
After grid search, cv

Running time: 45047 sec

Avg memory: 13219 MB

Accuracy: 86.8%

Num of trees: 2000



Improvement

- HPC run time limit