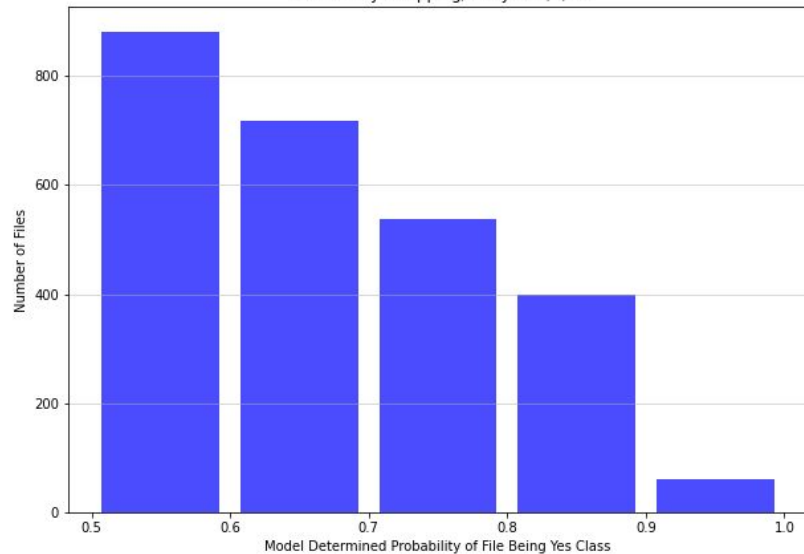
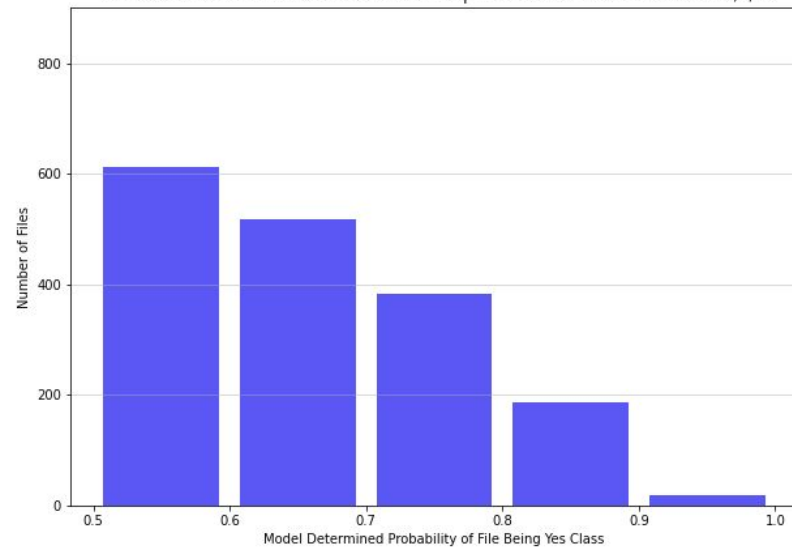


# Model Results After Manual Review

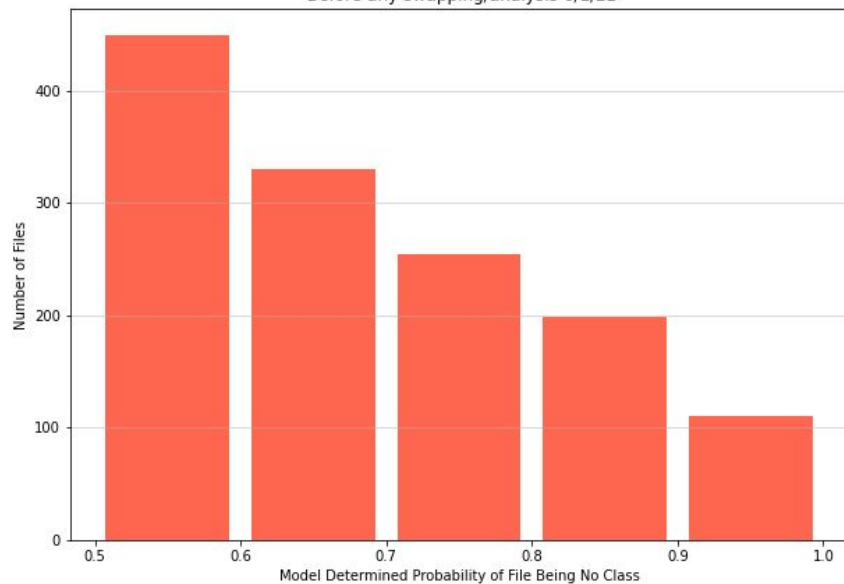
TUFF Robot Probability of Correct Classification With Files Misclassified as Yes  
Before any swapping/analysis 6/1/22



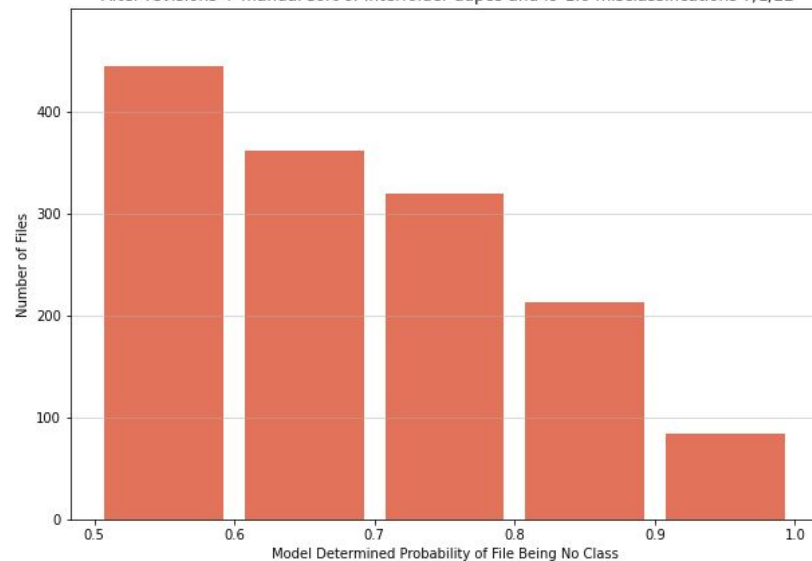
TUFF Robot Probability of Correct Classification With Files Misclassified as Yes  
After revisions + manual sort of interfolder dupes and .9-1.0 misclassifications 7/1/22



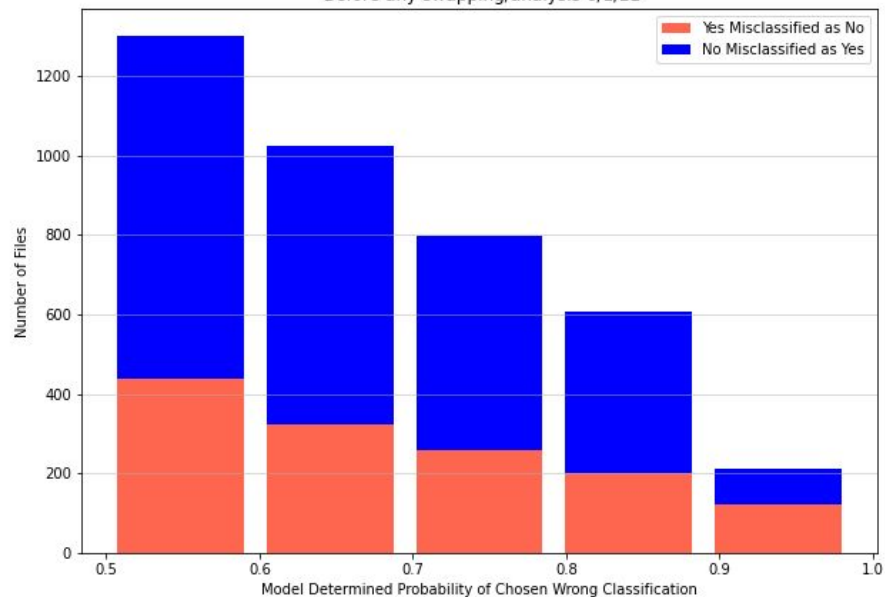
TUFF Robot Probability of Correct Classification With Files Misclassified as No  
Before any swapping/analysis 6/1/22



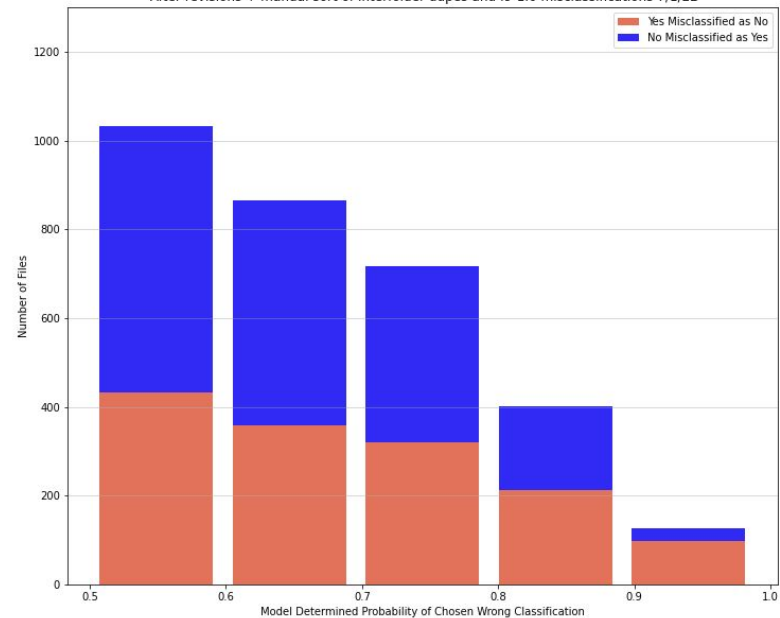
TUFF Robot Probability of Correct Classification With Files Misclassified as No  
After revisions + manual sort of interfolder dupes and .9-1.0 misclassifications 7/1/22



TUFF Robot Probabilities of All Incorrect Classifications Being Correct  
Before any swapping/analysis 6/1/22



TUFF Robot Probabilities of All Incorrect Classifications Being Correct  
After revisions + manual sort of interfolder dupes and .9-1.0 misclassifications 7/1/22



# Example Yes Confidently Misclassified as No (text941.txt)

news.10563

Companies/Markets

Live Help

Display Options

+ Related Dow Jones Intelligent Indexing™

Listen to Article

| Translate Article | Share

Domestic news items from Xinhua -- Sept. 8

573 words

9 September 2005

Xinhua News Agency

XNEWS

English

(c) Copyright 2005 Xinhua News Agency

BEIJING, Sept. 8 (Xinhua) -- The following are domestic news items released by Xinhua on Thursday:

Market exchange rates -- Sept. 8

Hungarian PM starts China visit

Highlights of major Beijing-based newspapers -- Sept. 8

Chinese president leaves for two-nation state visit

Hong Kong stocks open higher -- Sept. 8

2 Drug dealing suspects nabbed in Macao

Ninth China International Fair for Investment and Trade

raises curtain in Xiamen

Symposium on low-cost-carrier held in Macao

CPC representative leaves for France, Sweden, Finland

UPDATE: Chinese president leaves for two-nation state visit

Hong Kong stocks finish lower at midday

Senior Chinese official to visit four European countries

ADB revises up China's economic growth to 9.2 percent for 2005

URGENT: Six-party talks to resume Sept. 13

Air quality of major Chinese cities -- Sept. 8

Chinese scientists start expedition to country's eighth largest desert

Taiwan stocks close higher -- Sept. 8

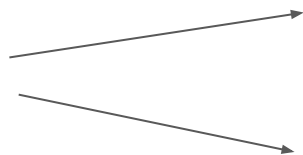
Hushen 300 index up -- Sept. 8

Shanghai stock index mixed -- Sept. 8

---

# New Yes/No Dedup Results

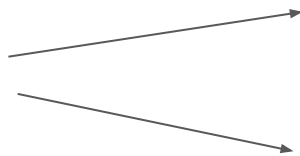
“Yes” Training Data  
6,468 total files



6463 root documents

5 “dupe” documents

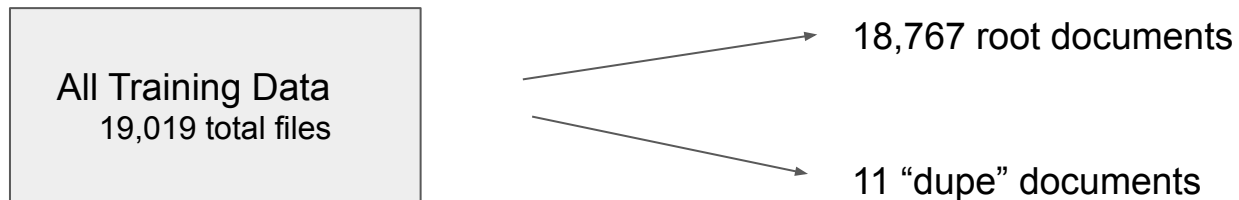
“No” Training Data  
12,424 total files



12,420 root documents

4 “dupe” documents

# New Interfolder Dedupe Results



# Score Changes

- F1 Score: 0.759  $\rightarrow$  0.7673
- Balanced Accuracy: 0.8215  $\rightarrow$  0.82959
- MCC: 0.644  $\rightarrow$  0.660