Updated Model Performance Report

Preprocessing Pipeline:

1. Non-ASCII characters are removed
2. URL’s are removed
3. Duplicate tweets are removed
4. Empty tweets are removed

Test Set Description:

Tweets taken from completely different sources. This time, several tweets that contained keywords like “vulnerability” or “exploit” but in non-cybersecurity related terms were purposefully included in order to gauge how well the model learned the correct context.