

IRIS News Curator: Intelligent Risk and Incident Sharing – A Proof of Concept

TANDIAWAN, Cathleen Trixie
Supervised by Prof. SHEN, Jiasi



Introduction

With the emerging trend of **Big Data**, having the power to harness insights from these high-volume, high-veracity data opens opportunities to uncover insights necessary for the industry. Similarly, for the **anti-money laundering industry**, our client aims to leverage latest news data to identify early warning signs of money laundering or terrorist funding activities. By putting in analytics to these textual data, we can **perform supervisory activities** to monitor threats and do early prevention by investigating the entities involved. Hence, with that the solution proposed is a **one stop portal, which we call IRIS: Intelligent Risk and Incident Sharing**. IRIS acts as a proof of concept to better understand whether the implementation of **NLP** and **LLM** can help identify these threats better.

Objectives



Reducing client's manual workload, redirecting the division's resources



The platform encompasses a wide range of information scope from different sources



Implement an effective way to identify highly relevant news articles



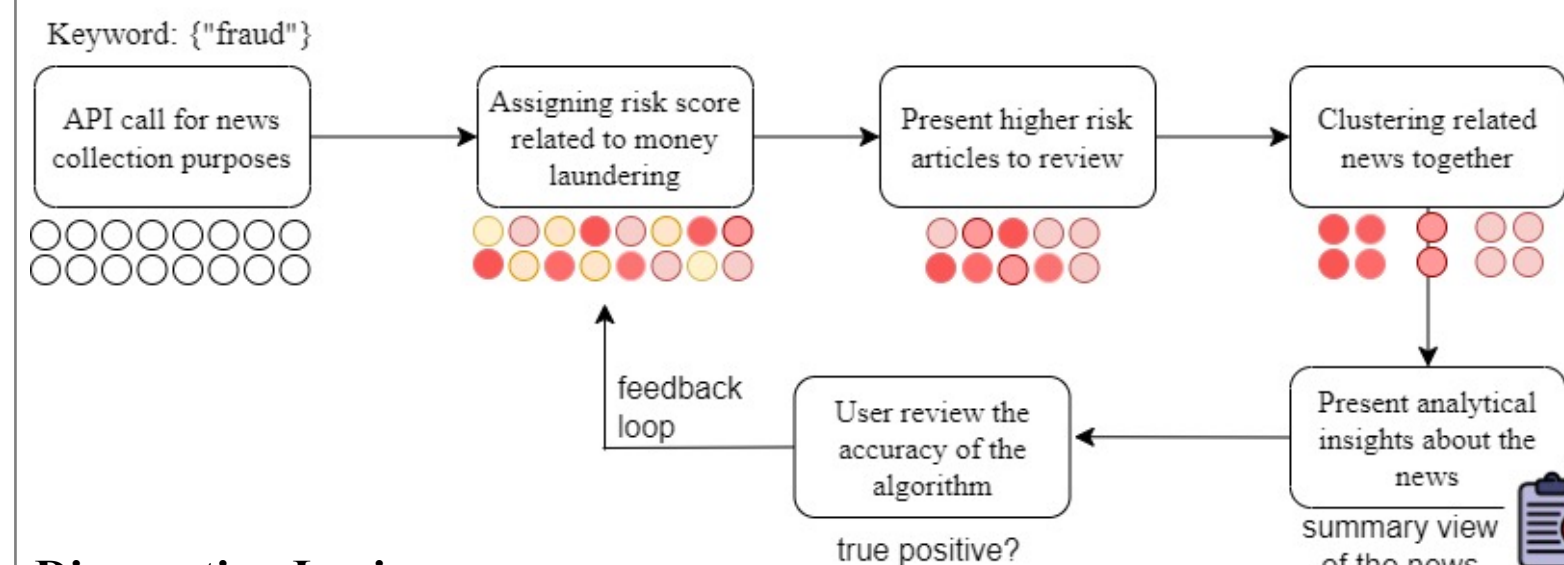
Accurate identification of the potential emerging threats



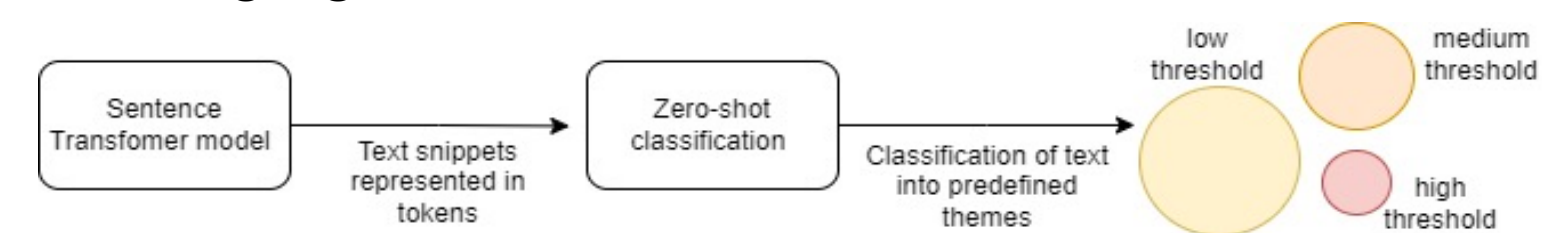
Rapid identification of emerging threats

Implementation

Below is the **high-level design** of how the IRIS workflow operates:

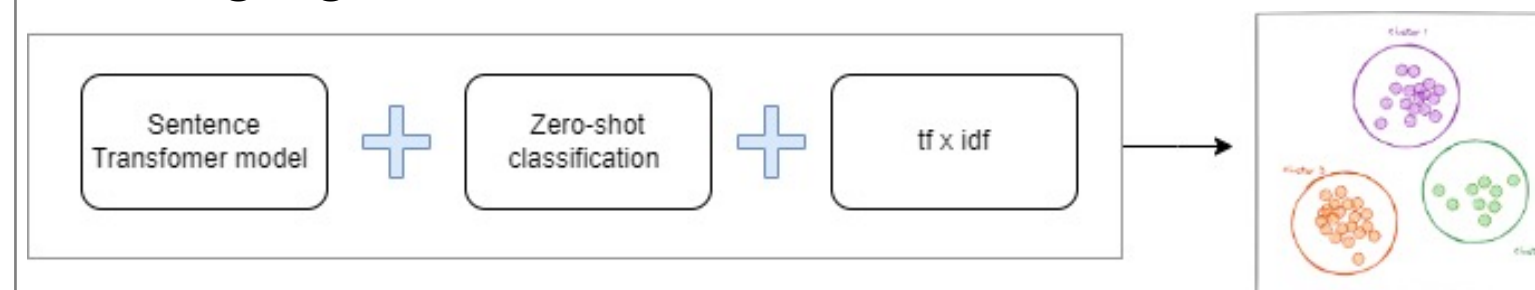


Discounting Logic



The filtering news leverages on **BERT model** to use **sentiment analysis** on each sentence, outputting an array of **tokens**. It goes through a second layer of filtering through **zero-shot classification**, which outputs a certain score indicating its **similarity to a predefined theme** before. If the score is bigger than a set threshold, then the news will be escalated.

Clustering Logic



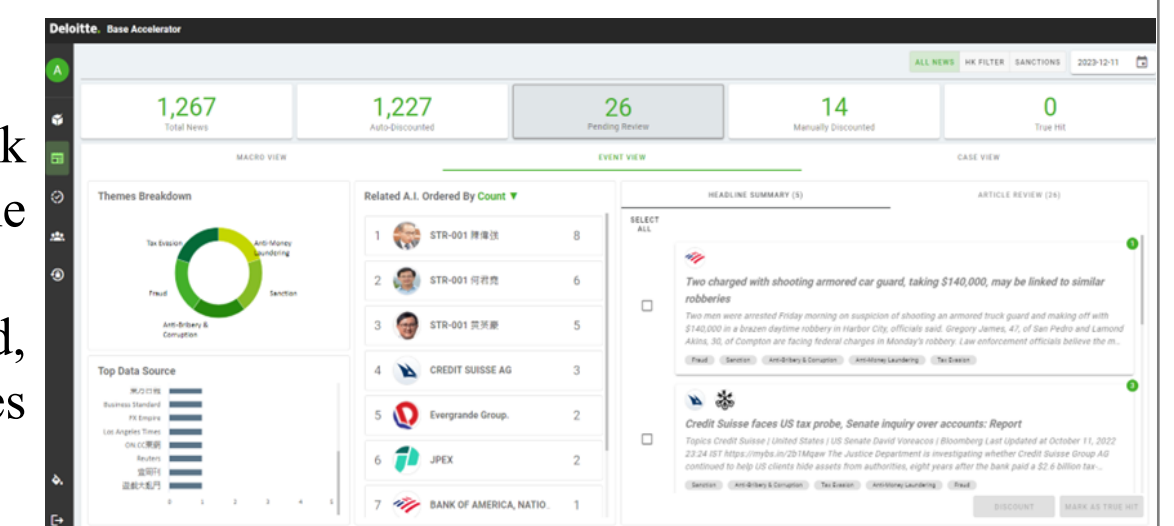
The logic tries to approach the in 3 level or depths:

1. Understanding the **general theme** of the news which is done through the **zero-shot classification** model.
2. Capturing the **contextual meaning** of the sentences which is implemented using the **sentence transformer (BERT)** model.
3. Comprehending the news **word by word** using **tfidf** model.

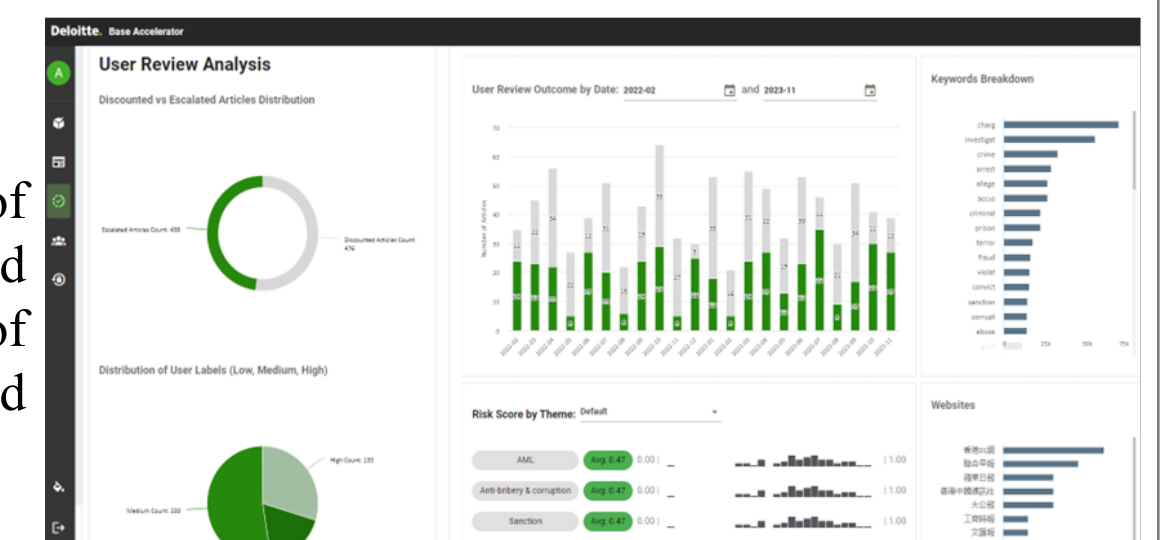
The final score is a linear combination of different weightings for these 3 models. We put more emphasis on the **word level** model.

Result

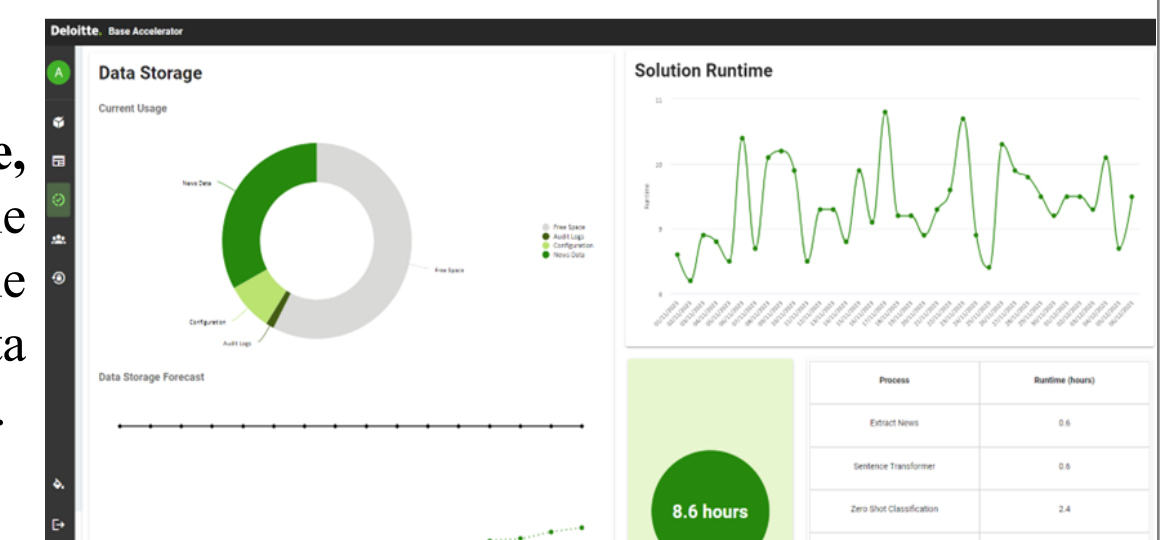
Main dashboard, contains high risk level news, theme breakdowns, entities involved, and data sources news are extracted.



User dashboard, contains analysis of risk keywords used and breakdown of the filtered and escalated articles.



Data storage page, which gives the runtime data of the script and data storage maintained.



Conclusion

Done as a co-op intern in Deloitte Advisory Hong Kong in partnership with HKMA's Anti Money Laundering division, the team built a proof-of-concept of a holistic solution to create this one stop portal to provide quick insights and analytics of current events happening in the financial crime industry. Project will be further developed to be used commercially.