Real-time monitoring of social media sentiment for detecting operational incidents in banking

Background and Motivation

Background

- The rise in operational incidents and the need for timely responses necessitate the **adoption of real-time and data-driven solutions** in operational risk management (Eceiza et al., 2020).
- HKMA has emphasized the urgent concern posed by technological advancements (HKMA, 2022), conducting various case studies on news monitoring, ML, and sentiment analysis.

HK$223B (+8.7%)
In operational costs in 2021

US$210B
Cost of operational incidents from 2011 to 2016
(Huber and Funaro, 2018)

Obtain information from customers and staff via
unofficial channels

Lower costs, more control
Establish trust with general public and regulators

Why social media?

- Operational incidents pose risks for banks, with **social media becoming a preferred platform for customers to voice complaints and frustrations**, as seen in SCB's FB page flooding with complaints about unauthorized transactions (Ma, 2022; Blanco, 2022).

Objectives and Deliverables

Real-Time Monitoring of Social Media Sentiment for

Web scraping & Monitoring Interface

Natural Language Processing (NLP)

Detecting Operational Incidents in Banking

Methodology

Phase 1: Data Collection

- Comments made between Jan 1, 2021, and Aug 31, 2023 from LIHKG and official Facebook pages of four major banks in HK (i.e. HSBC, Hang Seng Bank, Standard Chartered Bank, Citibank) were sourced.
- LIHKG: XPath and BeautifulSoup were used to extract information such as textual comments, original posts, timestamps, and emojis to provide context, analyze trends, and enhance sentiment analysis.
- Facebook: Selenium was used to automatically log into Facebook using provided credentials, navigated to the desired bank pages, and retrieved the necessary information such as comments, comment date, post content, post date, and retrieval time.

Phase 2: Machine Learning and NLP Models

- 20% of the data collected from LIHKG Is labelled. The titles of LIHKG threads and the comments themselves were carefully labeled to indicate whether the content is related to potential incidents or not. This involved reviewing the content and assigning appropriate labels manually based on our understanding of operational risks and incident-related discussions.
- The labelled data serves as the training data for the NLP and machine learning models. By feeding this labelled dataset into the models, they can learn the underlying patterns and relationships between the text features and the incident labels.
- A variety of machine learning algorithms is trained and evaluated such as Support Vector Machine, Random Forest, Multi-layer Perceptron, K-nearest Neighbors.

Phase 3: Dashboard Interface

- The dashboard and alert system is served as a front-end interface for banks and regulators to access insights and trends derived from social media mentions of banks.
- It not only provides graphical presentations to highlight and visualize anomalies, but also generates ratings to indicate whether the banks are at "Low", "Middle", or "High" risk of operational incidents.
- It notifies relevant parties of abnormal social media behavior so that they can conduct immediate investigations and mitigate potential incidents in a timely manner.

Model Performance

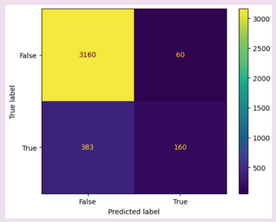
Comparison

Model	Accuracy Rate
Support Vector Machine	**88**.23%
Random Forest	**88**.38%
Multi-layer Perceptron	**87**.91%
K-nearest Neighbors	**87**.70%

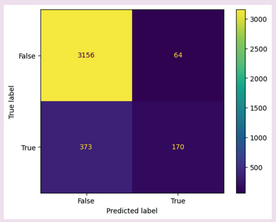
Confusion Matrix

- High in false negative
- Possibly caused by biased and imbalanced dataset, or inadequate data preprocessing

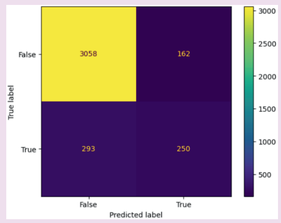
1. Support Vector Machine



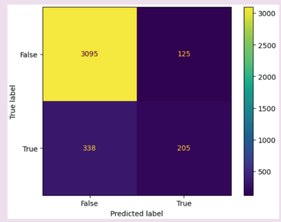
2. Random Forest



3. Multi-layer perceptron



4. K-nearest Neighbors

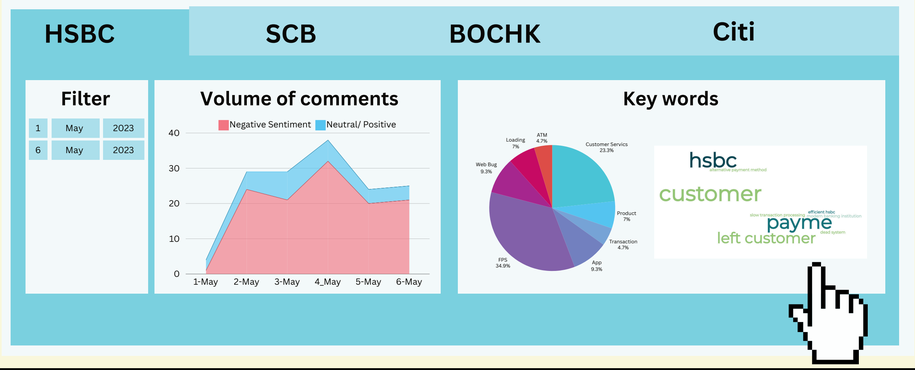


Front End

Python Web App

- Streamlit - common Python package and free cloud platform for deployment
- Packed all codes together with the dashboard to reduce time in front-end programming

Dashboard



Conclusion

- The system provides a robust solution for detecting abnormal activities and addressing operational risks in banks.
- It enables efficient risk management through data-driven insights, visualizations, and alert systems.
- It also empowers banks to proactively manage risks and maintain a stable banking environment

Miscellaneous

FITE 4801 Final Year Project 2023-24
Project No.: FYP 23071
Supervisor: Prof. Yiu SM
Team: Sin Hoi Yan Melanie, Wong Nicole, Yao Yi Tung

References