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.

**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 detecting operational incidents in banking
- **Web scraping & Monitoring Interface**
- **Natural Language Processing (NLP)**

**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.

**Comparison**

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support Vector Machine</td>
<td>88.23%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>88.38%</td>
</tr>
<tr>
<td>Multi-layer Perceptron</td>
<td>87.91%</td>
</tr>
<tr>
<td>K-nearest Neighbors</td>
<td>87.70%</td>
</tr>
</tbody>
</table>

**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.

**References**
- 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