**Industry problem recap**

Financial compliance regulations necessitate organizations to monitor all communication channels used by their regulated employees. Demonstrating awareness of any unlawful actions by traders requires thorough examination of all sent and received communications, promptly identifying any signs of malicious behaviour. The sheer volume of communication data generates numerous alerts, many of which turn out to be false alarms. By implementing machine learning (ML), the number of false positives (innocent communications causing unnecessary work for reviewers) can be reduced, and false negatives (potential regulatory breaches leading to substantial fines and sanctions) can be minimized. The alerts are triggered through a scoring system that employs lexicon searches with proximity parameters. Multiple scored criteria contribute to the overall score of a given scenario applied to a particular communication. A measure of predictive performance of the applied models could be calculated using any one of multiple of the metrics spoken about on this course so far.

#Focus more on the role of predictive performance

**What information would you need in order to make the most accurate predictions?**

* Frequency of alerts
* Frequency of communication
* Surrounding communications
* Parties involved in the conversation
* The means of communication, i.e. the channel through which they were communicating
* HR records
* Location of the sender and recipient
* The sentiment and emotion of the message body
* The time and day of the message
* The baseline emotion of the sender
* The context / topic of conversation
* Position / role of the sender and recipient within the organisation
* Market news and alerts
* Trade and market data

**How would you aggregate the information?**

**Of the machine learning techniques, you have learnt so far, which approaches would you try?**