FINANCIAL STATEMENT ANALYSIS DATABASE

TOMMY CHEUNG (DS STREAM MAR2021)

AIM

1. Apply ML skills to F/S
2. Predict from pre-existing data
3. Detect fraud from latest F/S
4. Summarize annual report info

BUSINESS OBJECTIVE

Very few F/S analysis database based on ML

No convenient database to gather info

Useful for analysis:

Accountants
Auditors
Investors
Media

DATA

- Financial statements from selected regions (Pre: US)
- All relevant obligatory reporting document (8-K, 10-K, etc.)
- Pre-trained models (to be researched)

MAJOR TASKS

Week 1		Week 2-3	Week 3-4
Gathering	Cleaning	Training	Building
Gathering data Using API and webscraping	Data cleaning Building up live database (may have API keys for public)	Training models for following predictions: • Ratios increase/decrease • Investment decision (e.g. NPV calculation) • Fraud risk detection • Sentiment analysis from text reports	Building a website with simple UI for input (BIG CHALLENGE!)

REFERENCE

1. Amel-Zadeh, Amir and Calliess, Jan-Peter and Kaiser, Daniel and Roberts, Stephen, Machine Learning-Based Financial Statement Analysis (November 25, 2020). Available at SSRN: https://ssrn.com/abstract=3520684 or http://dx.doi.org/10.2139/ssrn.3520684
2. Perols, Johan. (2010). Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms. Auditing A Journal of Practice & Theory. 30. 10.2308/ajpt-50009. SSRN: https://www.researchgate.net/publication/256045322