



# **Text Analytics & Business Application**

Text Analytics in Finance

Qinglai He

Department of Operations and Information Management

Wisconsin School of Business

# Topic Modeling vs. Text Clustering

## Similarity

- Both are unsupervised learning method
- Both can be used to reveal the theme of corpus (a collection of documents)
- Both rely on human interpretation to understand the underlying themes



# Topic Modeling vs. Text Clustering

## Differences

- Text clustering is based on similarity measure to group documents into different clusters, whereas topic modeling is based on several assumptions to discover the latent factor (topics) of a document. The outputs of these two techniques are different.
- Text clustering normally assign a document to a cluster, whereas topic modeling assign a document to various topics with different probabilities.
- When you have large corpus and long documents, topic modeling works better.
- Different assumptions
- Different model structures



# Outline of Today's Class

- NLP applications in finance
- FinBERT
- Challenges and future scope



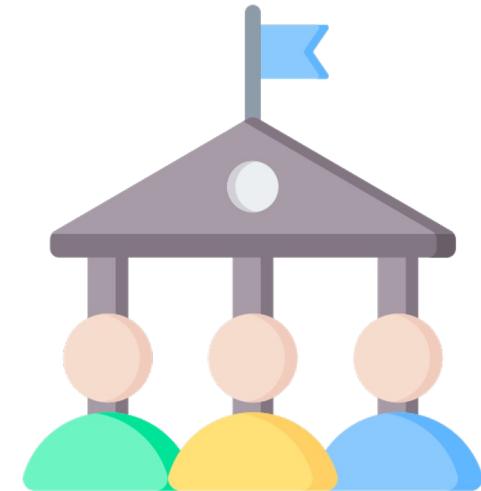
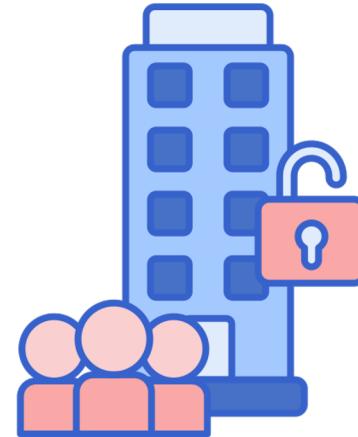
# Text Analytics in Finance

- When integrating and utilizing NLP in the finance operations, reporting, and evaluation, we can look at finance from the following three angles:
  - Organization perspectives
  - Actions
  - Financial context



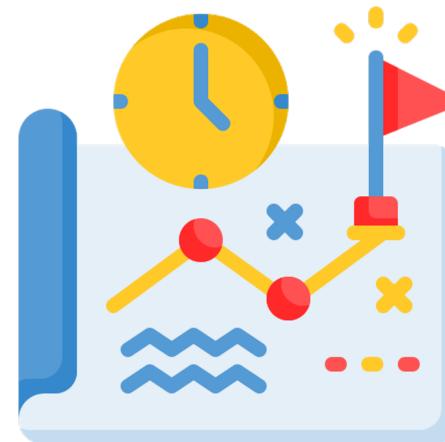
# Organization Perspectives

- Different organization types have different requirements and perspectives that need to be taken into account. These organizations include:
  - Private companies
  - Public companies
  - Non-profit enterprises
  - Governmental organizations



# Actions

- There are different actions that an organization can take, including:
  - Allocating and reallocating funds
  - Accounting and auditing, which includes identifying anomalies and outliers to investigate for both value and risk
  - Prioritization and resource planning
  - Compliance with legal and policy norms



# Financial Context

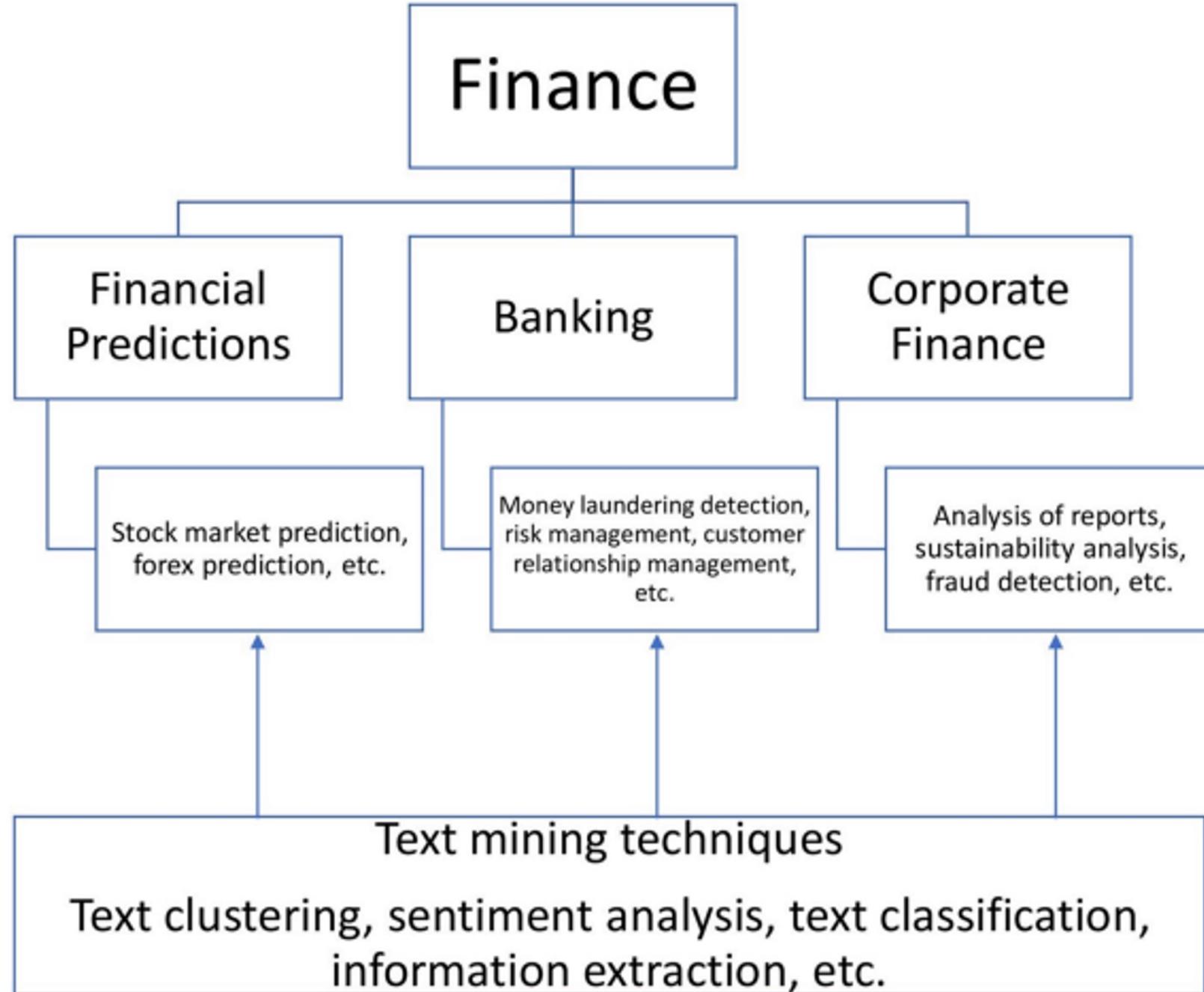
- These actions can have various contexts, including:
  - Forecasting and budgeting
  - Retail banking
  - Investment banking
  - Stock market operations
  - Cryptocurrency operations



# NLP Applications in Finance



# An Illustration for Text Analytics in Finance



# NLP Applications in Finance

- We'll cover some specific applications of NLP in finance:
  - Financial sentiment analysis
  - Loan risk assessments
  - Auditing and accounting problems
  - Predicting stock behavior



# Financial Sentiment

- Stock market trading relies on a set of information about specific companies
- Knowledge about specific companies is crucial for stock market trading
- Analysis can be based on:
  - Quarterly financial reports
  - Analyst reports
  - Social media posts



# Positive, Negative or Neutral?

- “Growth is strong and we have plenty of liquidity.”
- “There is a shortage of capital, and we need extra financing.”
- “Formulation patents might protect Vasotec to a limited extent.”



# Financial Sentiment

- Social media analysis helps in monitoring social media posts and pointing out potential opportunities for trading.
- For example:
  - If a CEO is resigning, that sentiment is often negative, which can negatively affect the company's stock price
  - If the CEO is not performing well and markets welcome their resignation, that could lead to an increase in stock price
- Companies that provide trading information:
  - DataMinr
  - Bloomberg



# Risk Assessments

- Credit risk is a way to quantify the chances of a successful loan repayment.
  - It's generally calculated by an individual's past spending and loan repayment history.
  - However, this information is limited in many scenarios, especially in underprivileged communities.
- Often in personal loan agreements, various information has to be captured from loan documents, which are then fed to **credit risk models**.
- The information captured helps in identifying credit risk, and erroneous data extraction from these documents can lead to flawed assessments.
- Named entity recognition (NER), which we will cover in detail in next week, can improve this.



# Risk Assessments

Loan agreement  
with annotated entities:

## LOAN AGREEMENT

This LOAN AGREEMENT, dated as of November 17, 2014 (this "Agreement"), is made by and among Auxilium Pharmaceuticals, Inc., a corporation incorporated under the laws of the State of Delaware ("U.S. Borrower"), Auxilium UK LTD, a private company limited by shares registered in England and Wales ("UK Borrower" and, collectively with the U.S. Borrower, the "Borrowers") and Endo Pharmaceuticals Inc., a corporation incorporated under the laws of the State of Delaware ("Lender").

### RECITALS

WHEREAS, U.S. Borrower, Endo International PLC ("Endo"), a public limited company incorporated under the laws of Ireland, Endo U.S. Inc. ("HoldCo"), a corporation incorporated under the laws of the State of Delaware and an indirect wholly-owned subsidiary of Endo, and Avalon Merger Sub Inc., a corporation incorporated under the laws of the State of Delaware ("AcquireCo"), are parties to that certain Agreement and Plan of Merger (the "Merger Agreement"), dated as of October 8, 2014, pursuant to which AcquireCo will merge with and into U.S. Borrower, with U.S. Borrower surviving the merger, subject to the terms and conditions of the Merger Agreement;

WHEREAS, pursuant to the terms of the QLT Merger Agreement (as defined in the Merger Agreement), upon the termination of the QLT Merger Agreement in connection with the execution of the Merger Agreement, U.S. Borrower was obligated to pay the QLT Termination Fee (as defined in the Merger Agreement);

WHEREAS, Lender is an indirect wholly-owned subsidiary of Endo;

WHEREAS, on October 9, 2014 (the "Payment Date"), Lender paid the QLT Termination Fee in the amount of \$28,400,000 (the "Payment"), which, in accordance with the terms hereof, the parties have agreed shall constitute a loan from Lender to Borrowers on the terms and conditions set out in this Agreement; and

# Accounting and Auditing

- The global firms Deloitte and PwC now have a significant focus on delivering more meaningful, actionable, and relevant audit conclusions and observations on a company's annual performance.
- While applying NLP and ML to areas like contract document reviews and long-term procurement agreements, Deloitte, for example, has evolved its Audit Command Language into a more efficient NLP application.

**Deloitte.**



# Predicting Stock Behavior

- Stock data is fluctuating and random.
- Long-term and seasonal variations can cause significant mistakes in the evaluation.
- ML and NLP can effectively deal with large amounts of data and assist in predicting the volatility of stock prices and trends, helping with the stock trading decisions.



# FinBERT



# FinBERT

- FinBERT is a financial domain-specific pre-trained language model, based on Google's BERT.
- The goal is to enhance financial NLP research and practice.
- One of the key features of FinBERT is its ability to perform sentiment analysis on financial news and social media posts.
  - This can be particularly useful for traders and investors who are trying to stay on top of market trends and make informed decisions.



# FinBERT

- FinBERT is trained on the following three financial communication corpus. The total corpora size is 4.9B tokens:
  - Corporate Reports 10-K & 10-Q: 2.5B tokens
  - Earnings Call Transcripts: 1.3B tokens
  - Analyst Reports: 1.1B tokens
- FinBERT results in state-of-the-art performance on various financial NLP task, including:
  - Sentiment analysis
  - ESG classification
  - forward-looking statement (FLS) classification



# FinBERT: Other Pre-trained Language Models

There are several other packages that are similar to FinBERT in that they are pre-trained language models designed for specific domains or tasks:

- **BioBERT**: A pre-trained language model designed specifically for **biomedical** text mining tasks.
- **SciBERT**: A pre-trained language model that has been designed specifically for **scientific** text.
- **GPTs**: A pre-trained language model developed by OpenAI that has generated a lot of buzz in the AI community.



# FinBERT - Sentiment

- FinBERT model fine-tuned on 10,000 manually annotated (positive, negative, neutral) sentences from analyst reports.
- This model achieves superior performance on financial tone analysis task.
- **Input:** A financial text.
- **Output:** Positive, Neutral or Negative.



# FinBERT - ESG

- ESG analysis can help investors determine a business' long-term sustainability and identify associated risks.
- FinBERT-ESG is a FinBERT model fine-tuned on 2,000 manually annotated sentences from firms' ESG reports and annual reports.
- **Input:** A financial text.
- **Output:** Environmental, Social, Governance or None.



# FinBERT - FLS

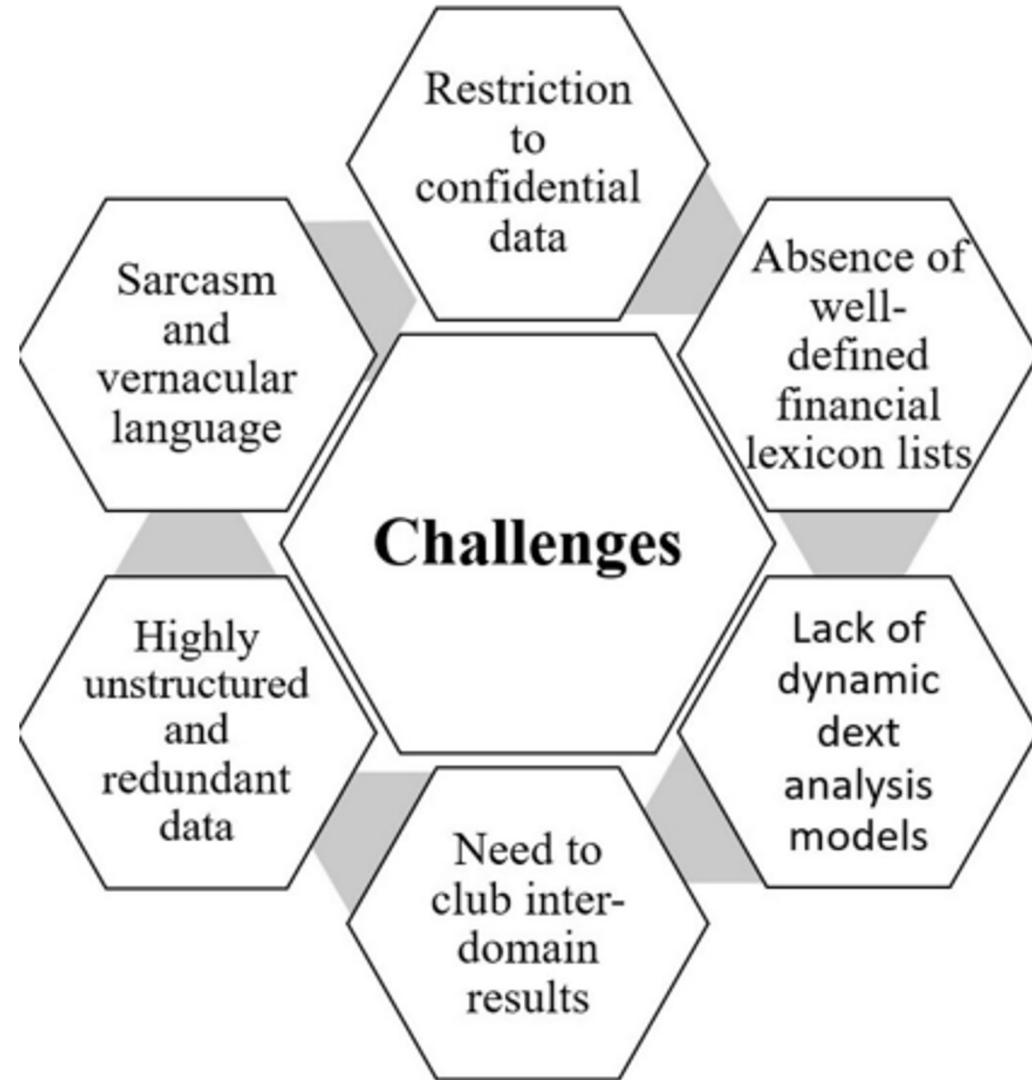
- Forward-looking statements (FLS) inform investors of managers' beliefs and opinions about firm's future events or results.
  - Identifying forward-looking statements from corporate reports can assist investors in financial analysis.
- FinBERT-FLS is a FinBERT model fine-tuned on 3,500 manually annotated sentences from Management Discussion and Analysis section of annual reports of Russell 3000 firms.
- **Input:** A financial text.
- **Output:** Specific-FLS , Non-specific FLS, or Not-FLS.



# Challenges and Future Scope



# Challenges and Future Scope



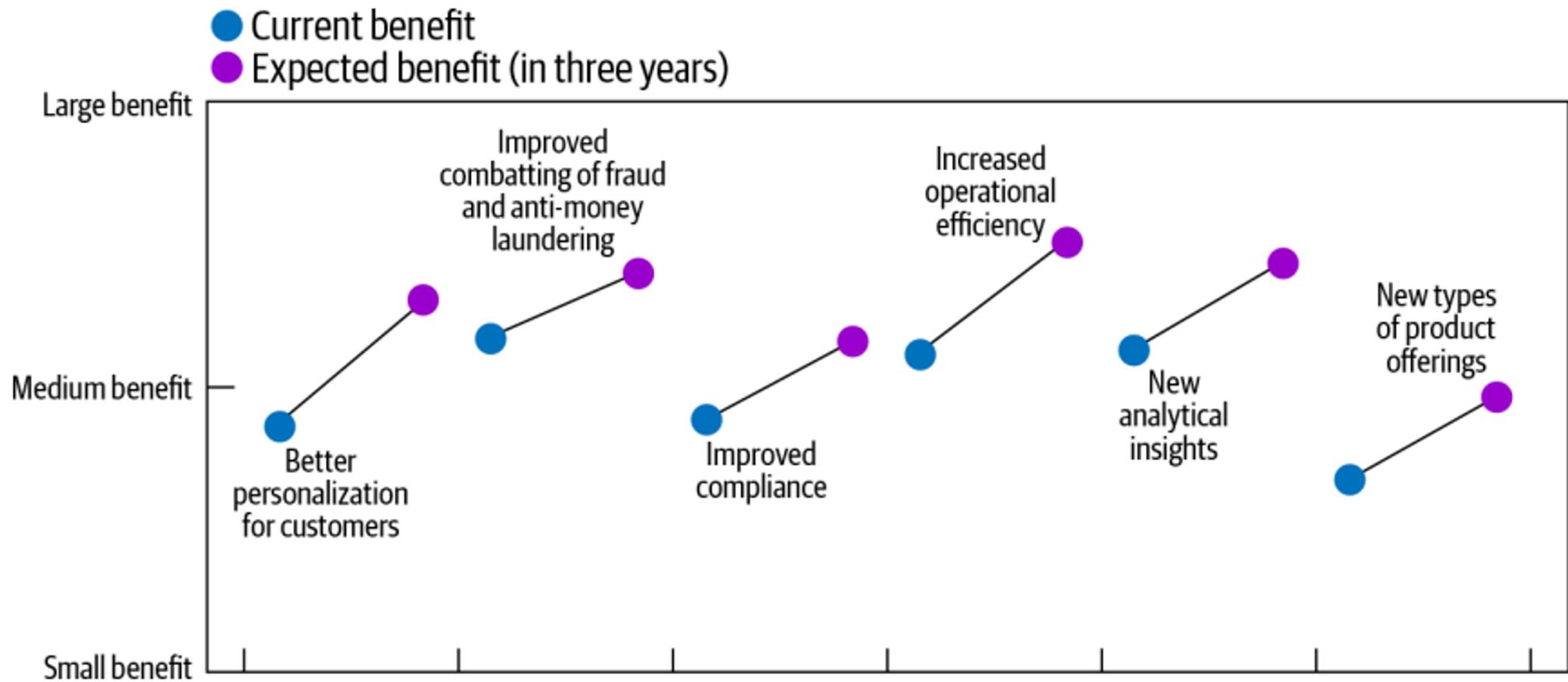
# Challenges and Future Scope

Study	Datasets	Techniques	Evaluation Parameters	Performance
Identifying text patterns for financial performance	Annual reports of US-listed companies	Clustering, sentiment analysis	Correlations between text patterns in company's reports and its sales performance	-
Automatic classification of accounting literature	Articles from EbscoHost online academic database	Bayes classifier, decision tree, rule-based classifier	Accuracy	87.27%
Financial footnote analysis	Income tax footnotes from financial reports	NB, k-means, KNN, SVM, decision tree	Runtime, accuracy, RMSE, absolute error	Accuracy: 82.86% RMSE: 0.414
Competitive analysis from social media	Social media data from Social Mention, SABI database	Social mention tool	Pearson correlation, F-ratio	F-ratio: 3.361 (Good fit)



# ML and NLP Benefits Finance

- The graph shows the estimated ML benefits survey in the UK



**Take 10 minutes break...**



# Exercises using Google Colab

