Predicting merger decision outcomes of the European Commission:
A Natural Language Processing and Machine Learning approach

BACKGROUND

Research on data science in competition economics is limited. Building a **text-based prediction of merger decision outcomes** remains **unexplored**.

Official EC merger decision reports provide rich textual information, which can serve as a **proxy** for the facts and conditions of a proposed merger in the absence of confidential merger filings.

Advancements in data science provide **new opportunities** for analyzing large volumes of unstructured data to **aid antitrust review** and **flag anticompetitive mergers.**

EC Merger Review Process

The EC reviews a high volume of notifications within a limited period.

- Notification
- 25 days — **Phase 1** → clearance → **conditionally** OR **unconditionally**
- 75 days — **Phase 2**
  - → clearance → **conditionally** OR **unconditionally**
  - → prohibition

OBJECTIVE

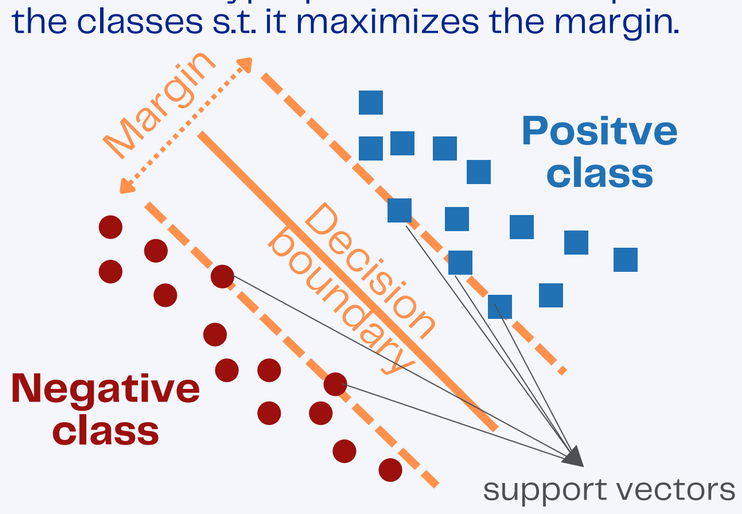
To predict merger decision outcomes using a Support Vector Machine (SVM) linear classifier. The decisions were classified into two binary labels:

$$p2 = \begin{cases} 1, \text{Phase 2 review} \\ 0, \text{Phase 1 review only} \end{cases}$$

$$wc = \begin{cases} 1, \text{approved conditionally} \\ 0, \text{approved unconditionally} \end{cases}$$

Linear Support Vector Machine

To find the hyperplane that best separates the classes s.t. it maximizes the margin.

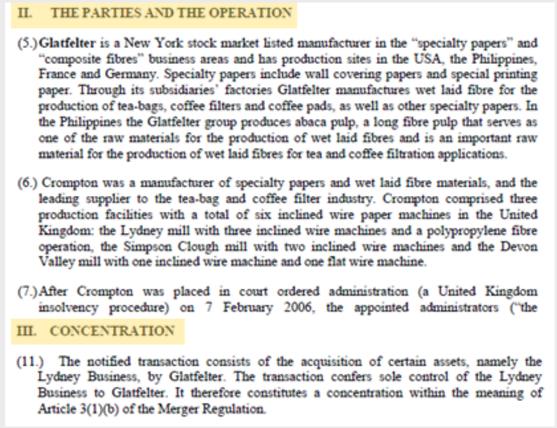


METHODOLOGY

1. **Collecting** decisions (**PDFs**) from the EC search tool (2004–2022)

Web scraping case number, title, notification date, and NACE

2. **Parsing** by section: text b/w **bold** and **ALLCAPS**


II. THE PARTIES AND THE OPERATION

(5.) **Glatfelter** is a New York stock market listed manufacturer in the "specialty papers" and "composite fibres" business areas and has production sites in the USA, the Philippines, France and Germany. Specialty papers include wall covering papers and special printing paper. Through its subsidiaries' factories Glatfelter manufactures wet laid fibre for the production of tea-bags, coffee filters and coffee pads, as well as other specialty papers. In the Philippines the Glatfelter group produces abaca pulp, a long fibre pulp that serves as one of the raw materials for the production of wet laid fibres and is an important raw material for the production of wet laid fibres for tea and coffee filtration applications.

(6.) Crompton was a manufacturer of specialty papers and wet laid fibre materials, and the leading supplier to the tea-bag and coffee filter industry. Crompton comprised three production facilities with a total of six inclined wire paper machines in the United Kingdom: the Lydney mill with three inclined wire machines and a polypropylene fibre operation, the Simpson Clough mill with two inclined wire machines and the Devon Valley mill with one inclined wire machine and one flat wire machine.

(7.) After Crompton was placed in court ordered administration (a United Kingdom insolvency procedure) on 7 February 2006, the appointed administrators ("the

III. CONCENTRATION

(11.) The notified transaction consists of the acquisition of certain assets, namely the Lydney Business, by Glatfelter. The transaction confers sole control of the Lydney Business to Glatfelter. It therefore constitutes a concentration within the meaning of Article 3(1)(b) of the Merger Regulation.

Pre-processing: lowercase, lemmatize, stop words removal, retained four sections

3. Data **balancing**

Data **splitting**

Training	Test
80%	20%

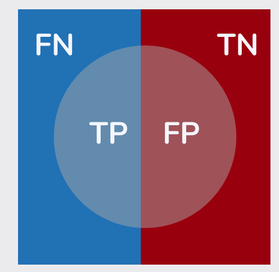
4. Feature **extraction**: **tfidf** of **n-grams**

Parameter tuning: **Grid search** 5-fold cross-validation (**CV**)

1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5
1	2	3	4	5

Model **training** 10-fold CV

5. Model **evaluation**

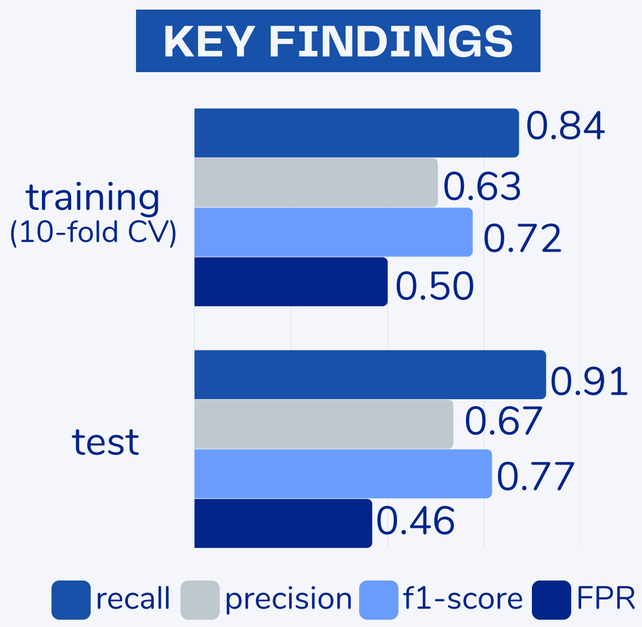


$$\textbf{Recall} = \frac{TP}{TP+FN}$$

$$\textbf{Precision} = \frac{TP}{TP+FP}$$

KEY FINDINGS

An SVM linear classifier was found to be effective in predicting mergers that were either approved unconditionally or subject to conditions. The classifier was trained using the full text from four sections (Parties and Operation, Concentration and Dimension, Market Definition, and Competitive Assessment). It achieved an **84% recall** for identifying cases with serious anticompetitive effects, and a **63% precision** for avoiding unnecessary conditions on cases without anticompetitive effects.



The model picked up specific **keywords** on **relevant markets** as the most important features, indicating its ability to capture classification nuances.

POLICY IMPLICATIONS

Two events in a merger review must be considered: **Event A**, approving a case unconditionally despite anticompetitive effects, and **Event B,** approving a case with conditions despite minimal anticompetitive effects. Event A leads to harm to competition and consumers (e.g., higher prices and decreased quality of goods and services), while Event B incurs compliance costs but prevents potential harm. This study prefers **Event B, prioritizing high recall** to avoid overlooking problematic transactions.

Practical implementation requires a thorough cost-benefit analysis to determine an optimal recall-precision trade-off based on the EC's specific objectives and priorities. This study provides valuable insights into the **potential** of building a **text-based predictive model for antitrust decisions**, but emphasizes the model's limitations and the **importance of expert judgment** in ensuring a **comprehensive** and **reliable decision-making process**.

Ma. Adelle Gia Toledo Arbo (216132) | Master of Data Science for Public Policy 2023 | The Hertie School, Berlin, Germany | giarbo.28@gmail.com