データ

本研究では，分析の対象期間を 2007 年から 2023 年までとし，ニュースデータとマーケットデータ（始値、高値、安値、終値、出来高）を 用いて分析を行った。2022年10月31日TOPIX Core30（トピックス コア30）の構成銘柄（3382 セブン&アイ・ホールディングス、4063 信越化学工業4502 武田薬品工業4503 アステラス製薬4568 第一三共6098 リクルートホールディングス6273 SMC6367 ダイキン工業6501 日立製作所6594 ニデック6758 ソニーグループ6861 キーエンス6954 ファナック6981 村田製作所7203 トヨタ自動車7267 本田技研工業7741 HOYA7974 任天堂8001 伊藤忠商事8031 三井物産8035 東京エレクトロン8058 三菱商事8306 三菱UFJフィナンシャル・グループ8316 三井住友フィナンシャルグループ8411 みずほフィナンシャルグループ8766 東京海上ホールディングス9432 日本電信電話9433 KDDI9434 ソフトバンク9984 ソフトバンクグループ）分析対象とした。

TOPIX Core30（トピックス コア30）は、[TOPIX](https://ja.wikipedia.org/wiki/%E6%9D%B1%E8%A8%BC%E6%A0%AA%E4%BE%A1%E6%8C%87%E6%95%B0)ニューインデックスシリーズの一つで、[東京証券取引所](https://ja.wikipedia.org/wiki/%E6%9D%B1%E4%BA%AC%E8%A8%BC%E5%88%B8%E5%8F%96%E5%BC%95%E6%89%80)の市場第一部全銘柄のうち、[時価総額](https://ja.wikipedia.org/wiki/%E6%99%82%E4%BE%A1%E7%B7%8F%E9%A1%8D)、[流動性](https://ja.wikipedia.org/wiki/%E6%B5%81%E5%8B%95%E6%80%A7_(%E7%B5%8C%E6%B8%88%E5%AD%A6))の特に高い30銘柄で構成された[株価指数](https://ja.wikipedia.org/wiki/%E6%A0%AA%E4%BE%A1%E6%8C%87%E6%95%B0)のことをいう。国内30大企業を意味し、 全て世界的な一流企業である。

ニュースデータとして，日経新聞、朝日新聞、Japanese Newsの検索窓に各社名を入力し指定期間内に出力された全てのニュースを取得した。2005年 1 月 1 日から 2023 年 8月までのニュース約5万件を取得した。ニュースデー タには，タイトルと本文があり，タイトルは本文内の重要な内容を要約したテキストデータである。ニュースには配 信された日時のタイムスタンプもある。