

# Bài tập: Thu thập dataset cho bài toán Sarcasm detection in news headline

## 1. Kết quả

- Kết quả bao gồm 2 file json
- Is\_sarcastic.json: chứa 7613 headline mang tính mỉa mai, châm biếm, thu thập từ 3 trang website:

[Chaser](#): 1754 headlines

[Beaverton](#): 3338 headlines

[Clickhold](#): 2521 headlines

- Is\_sarcastic.json: chứa 108605 headline lấy từ 3 trang báo uy tín

[Breitbart](#) (us): 97117 headlines

[Deadline](#) (uk): 2030 headlines

[Crikey](#) (australia): 9456 headlines

[Data format:](https://www.kaggle.com/rmisra/news-headlines-dataset-for-sarcasm-detection)( <https://www.kaggle.com/rmisra/news-headlines-dataset-for-sarcasm-detection>)

```
{"article_link": string "https://chaser.com.au/world/fuuuuuuuck-says-newspaper-tech-guy/",
```

```
"headline": string "\n‘FUUUUUUUCK’ says newspaper tech guy ",
```

```
"is_sarcastic": 1 }
```

Thành viên nhóm:

- [Đặng Văn Minh – 195219832](#)
- [Tạ Huỳnh Đức Huy – 19521634](#)
- [Trang Trương Giang - 19521450](#)