

1. Quản lý tiến độ

- Gitlab: lưu code.

- Drive: Lưu nội dung bài viết.

- Trello: Chia task.

2. Chia việc

2.1. Crawler

STT	Member	Page
1	Hoàng	cafef.vn vietnambiz tinnhanhchungkhoan.vn vietstock.vn vietnamnews.vn ndh.vn
2	Phúc	www.bidv.com.vn www.sbv.gov.vn finance.vietstock.vn www.vpbank.com.vn www.stockbiz.vn portal.vietcombank.com.vn
3	Bảo (Optional)	www.vietinbank.vn finance.vietstock.vn vingroup.net forbesvietnam.com.vn qandme.net

Phase 1: 3 ngày đầu.

Phase 2: 7 ngày để crawl hết data

Phase 3: 2 ngày để lọc data

10 days

2.2. Gán nhãn – 10 days

Classify + Cluster: Lọc + tiền xử lý: 5 ngày (tùy data) – Hoàng + Bảo

Summarization: Tóm tắt bằng tay mỗi danh mục con khoảng 5 bài

Hoàng: Danh mục 1 – 7 của tin vĩ mô việt nam + tin Vingroup (8 danh mục)

Bảo: Danh mục 1 – 3 tin quốc tế + danh mục 1 – 3 tin thị trường chứng khoán + danh mục 1 – 2 tin khối năng lượng (8 danh mục)

Phúc: Danh mục 1 – 3 tin khối ngân hàng + danh mục khối bất động sản + khối bán lẻ + ngành khác + ngành chứng khoán + xu hướng thể hệ Y (8 danh mục)

2.3. Gợi ý tin – 15 days

a) Classify – Phúc - 15

2 model: bản tin 1 (daily) và bản tin 2 (morning news).

b) Cluster - ranking (bé) Hoàng + Bảo (Bảo làm phần core - 10, Hoàng training)

Model 1 (Tin vĩ mô): 6 clusters

Model 2 (Tin quốc tế): 3 clusters

Model 3 (Tin chứng khoán): 3 clusters

Model 4 (Khối ngân hàng): 3 clusters

Model 5 (Khối năng lượng): 2 clusters

2.4. Tóm tắt – Hoàng - 10

Extractive summarization

2.5. Web frontend – Đức

2.6. Backend – Hoàng – 3

- 2 days for db api