**Phân công công việc làm việc nhóm**

**1. Thành viên nhóm**

1612374 - Trần Quang Minh

1612780 - Đoàn Quang Tuấn

**2. Phân công công việc**

**15/12/2019:**

+ Chọn bài toán, tìm kiếm trang web để tiến hành thu thập dữ liệu

+ Lấy url các nhóm dữ liệu trên trang vietnamnews.vn (Tuấn)

+ Parse nội dung văn bản theo các url (Minh)

**24/12/2019:**

+ Lập danh sách công việc cần làm, phân chia công việc

+ Lựa chọn các phương pháp giải quyết

**28/12/2019:**

+ Tạo pipeline cơ bản (Minh)

+ Chia dữ liệu train/test (Minh)

**30/12/2019:** cập nhật pipeline

+ Thống kê dữ liệu (Minh)

+ Tiền xử lý dữ liệu (Tuấn)

**04/01/2020:** cập nhật pipeline

+ Chọn và xử lý các đặc trưng: số lượng ký tự, số lượng từ, số lượng các dấu câu, số lượng các từ in thường, in hoa,... (Tuấn)

**05/01/2020:** cập nhật pipeline

+ Chọn và xử lý các đặc trưng: vector tf-idf (Minh)

+ Thiết kế mô hình pipeline hoàn chỉnh (Minh)

+ Huấn luyện và đánh giá các mô hình (Tuấn)

06/01/2020: cập nhật pipeline

+ Chỉnh sửa dữ liệu

+ Tái cấu trúc code

+ Tinh chỉnh mô hình

**07/01/2020 - 08/01/2020:** cập nhật pipeline dựa vào các góp ý

**3. Github**

Link github: <https://github.com/tuandoan998/DS_Final_Project>

Contributor:

