PHÂN TÍCH VÀ THIẾT KẾ

DAMH-02: XỬ LÍ NGÔN NGỮ TỰ NHIÊN

BÁO CÁO TIẾN ĐỘ

Tên nhóm: **RESTUDY**

STT MSSV Họ và tên Email

1 18127158 Lê Thành Nam 18127158@student.hcmus.edu.vn

2 18127121 Nguyễn Đăng Khoa 18127121@student.hcmus.edu.vn

3 18127138 Nguyễn Duy Long 18127138@student.hcmus.edu.vn

4 18127107 Lý Đăng Huy 18127107@student.hcmus.edu.vn

5 18127134 Lê Huỳnh Long 18127134@student.hcmus.edu.vn

Ngày báo cáo: 26/3/2019

**Nội dung công việc:**

* Tìm hiểu về Natuaral Language Programming.
* Tìm hiểu cơ bản về Vector hóa văn bản, Bag-of-words và TF-IDF.

**Kết quả:**

* Lọc được các dấu thừa trong văn bản.
* Vector hóa được dữ liệu nhập vào.
* Tính được TF-IDF.
* Tính được độ tương đồng.

**Đánh giá:**

* Hoàn thành tốt kế hoạch.
* Giải quyết sơ bộ một số yêu cầu cơ bản cho đồ án.

**Khó khăn:**

* Ban đầu chưa hiểu sâu mô hình Bag-of-words
* Chỉ giải quyết được 1 folder