

## **BÁO CÁO ĐỒ ÁN CUỐI KỲ**

# **HỆ THỐNG GỢI Ý ĐỊA ĐIỂM DỰA TRÊN DỰ ĐOÁN ĐIỂM ĐÁNH GIÁ NHẬN XÉT CỦA GOOGLE**

**MÔN HỌC: IE212.011 - CÔNG NGHỆ DỮ LIỆU LỚN**

**GVHD: TS. Đỗ Trọng Hợp**

**GV HDTH: Nguyễn Hiếu Nghĩa**

Nhóm sinh viên thực hiện: Nhóm 7

# DANH SÁCH THÀNH VIÊN

**21522436      Đặng Huỳnh Vĩnh Tân**

21522436      Nguyễn Thị Hồng Nhung

21520949      Nguyễn Viết Kha

21522093      Nguyễn Huy Hoàng

# NỘI DUNG

**1**    **GIỚI THIỆU**

**2**    **DỮ LIỆU - MÔ HÌNH**

**3**    **DEMO**

**4**    **KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

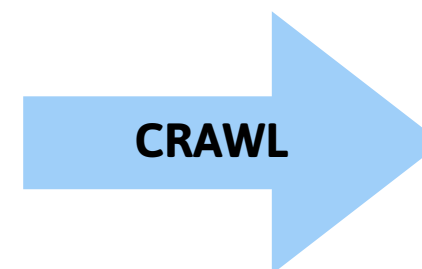
# 1. GIỚI THIỆU

IE212.011.Group7

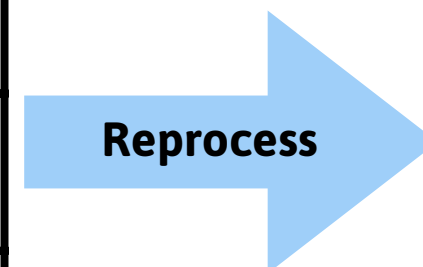
Dự án nghiên cứu này tập trung vào việc dự đoán điểm đánh giá của nhận xét địa điểm trên Google, nhằm cải thiện khả năng đề xuất các địa điểm cho người dùng dựa trên dữ liệu lịch sử và tình cảm trong các đánh giá.



# 1. GIỚI THIỆU



Tên bộ dữ liệu	Mô tả	Số lượng	Kích thước tệp
Places Data	Mô tả	8,210 địa điểm	3.24MB
Review Data	Mô tả	1,737,239 đánh giá	482MB



Tên bộ dữ liệu	Mô tả	Số lượng
Places Data	Mô tả	6,298 địa điểm
Review Data	Mô tả	456,792 đánh giá

TÓM TẮT BỘ DỮ LIỆU

# 1. GIỚI THIỆU

Attribute	Meaning	Datatype
placeId	Id of a place	String
title	Name of place	String
categoryName	Place category	String
location/lat	Latitude of a place	Float
location/lng	Longitude of a place	Float
address	Place address	String
permanentlyClosed	Operation status	Bool
url	Url to GoogleMap	String
categories	List of all categories of place	String

## DATA PLACE

Attribute	Meaning	Datatype
reviewId	Id of a review	String
placeId	Id of a place	String
title	Name of place	String
location/lat	Latitude of a place	Float
location/lng	Longitude of a place	Float
reviewerId	Id of reviewer	String
name	Name of reviewer	Bool
stars	Rating point	Int
text	Reviewer text	String
publishedAtDate	Review time	Datetime
categories	List of all categories of place	String
categoryName	Place category	String

## DATA REVIEW

## 2. DỮ LIỆU - MÔ HÌNH

## 2. DỮ LIỆU - MÔ HÌNH

Công thức mô hình:

$$\text{Rating} = \theta_0 + \theta_1 \times \text{ProportionOfPositive} + \theta_2 \times \text{ProportionOfNegative}$$

$$\text{Rating} = \theta_0 + \theta_1 \times \text{ProportionOfPositive} + \theta_2 \times \text{ProportionOfNegative} \\ + \theta_3 \times \text{ReviewHour} + \theta_5 \times \text{ReviewLength}$$

Kết quả:

Model	Train's MSE	Test's MSE
Initial model	4.759583	4.793305
Explanation model	3.466059	3.497937



# 2. DỮ LIỆU - MÔ HÌNH

IE212.011.Group7

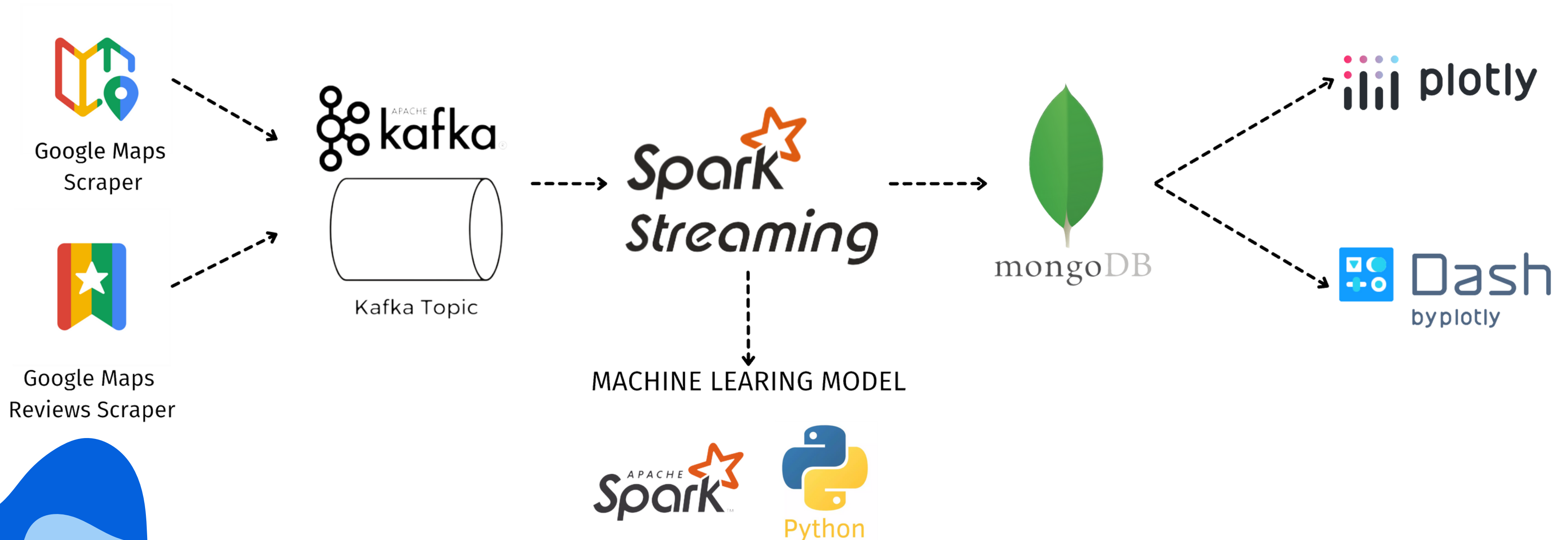
DATA

COLLECT

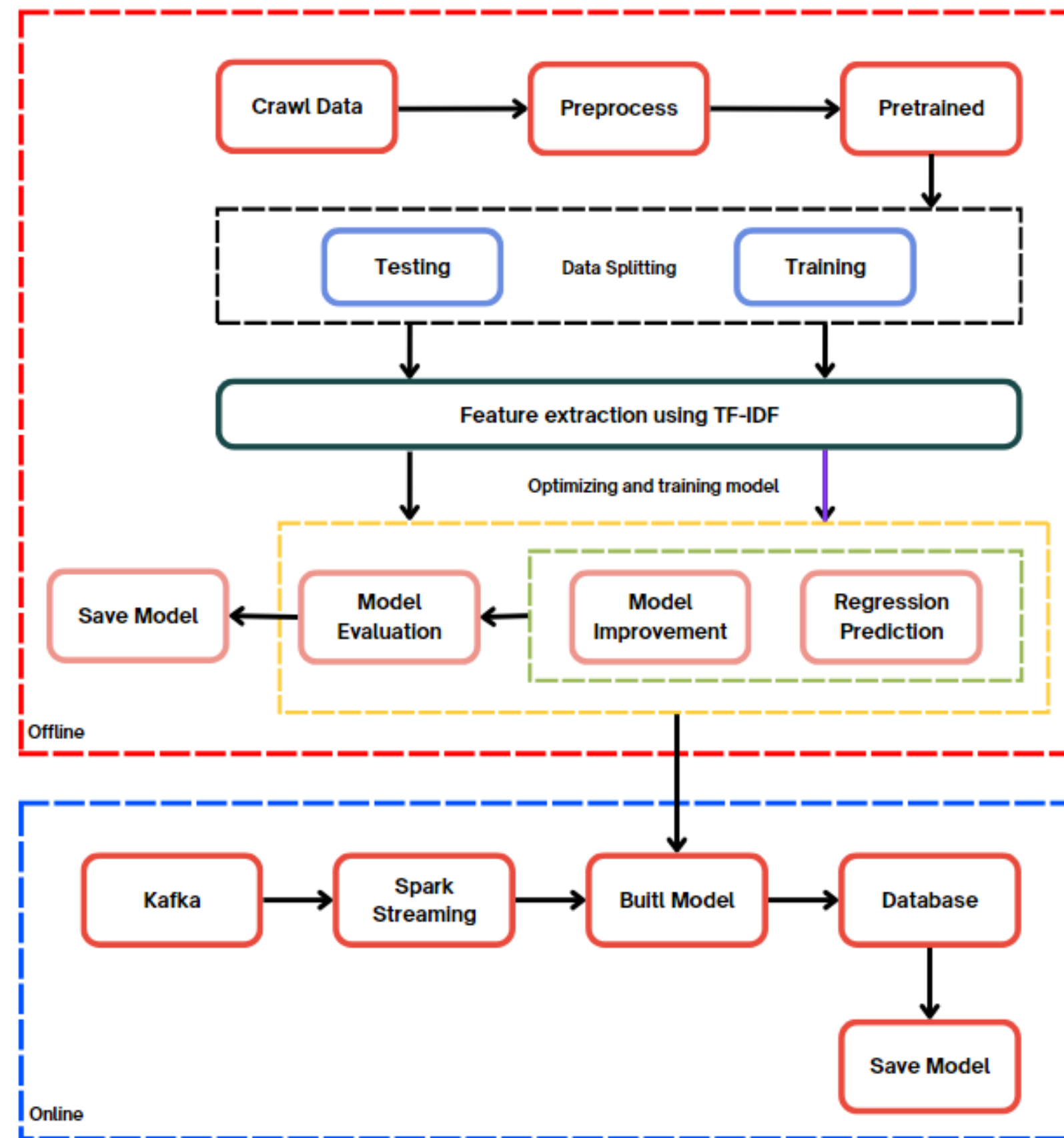
PROCESS

STORE

REALTIME  
VISUALIZATION



# 2. DỮ LIỆU - MÔ HÌNH



# 3. DEMO

# 4. KẾT LUẬN - HƯỚNG PHÁT TRIỂN

# 4. KẾT LUẬN - HƯỚNG PHÁT TRIỂN

## KẾT LUẬN

Theo những kết quả đã thể hiện trong báo cáo này việc sử dụng Spark Streaming một phần mở rộng từ lõi của Spark Api đã giúp xử lý tốt được luồng dữ liệu liên tục và giúp kết nối giữa các thiết bị với Spark trở nên dễ dàng.

## PHÁT TRIỂN

Hướng phát triển tương lai, tích hợp thêm các nguồn dữ liệu từ các ứng dụng khác, hỗ trợ xử lý nhiều ngôn ngữ hơn, tích hợp các kỹ thuật xử lý nâng cao chính xác.

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**XIN CẢM ƠN THẦY VÀ CÁC BẠN  
ĐÃ CHÚ Ý LẮNG NGHE**

**MÔN HỌC: IE212.011 - CÔNG NGHỆ DỮ LIỆU LỚN**

**GVHD: TS. Đỗ Trọng Hợp**

**GV HDTH: Nguyễn Hiếu Nghĩa**

Nhóm sinh viên thực hiện: Nhóm 7