








# Workflow

## Final Project Data Science Programming Tasks

# STT	Aa Task's Name	📋 Requirements	☰ Difficulties	☰ Implementor	# Progress	📅 Completion Date
1	<u>Find a public data source and start crawling</u>	<ul style="list-style-type: none"> <li>- Tập dữ liệu có tính đa dạng</li> <li>- Không mang tính phổ thông</li> <li>- Hơn 10 thuộc tính, hơn 1000 dòng</li> </ul>	<ul style="list-style-type: none"> <li>- Có quá nhiều tập dữ liệu</li> <li>- Tốn nhiều thời gian để tìm</li> </ul>	Bùi Kim Phúc	100%	@October 7, 2024
2	<u>Preprocessing</u>	<b>Xử lý các vấn đề của dataset:</b> <ul style="list-style-type: none"> <li>- Kiểm tra có dòng nào lặp không?</li> <li>- Khử các giá trị trống</li> <li>- Xử lý kiểu</li> </ul>	<ul style="list-style-type: none"> <li>- Dữ liệu có thể thiếu nhiều thông tin về các thuộc tính như thể loại, ngày phát hành</li> </ul>	Bùi Kim Phúc	100%	@October 13, 2024

# STT	Aa Task's Name	 Requirements	 Difficulties	 Implementor	# Progress	 Completion Date
		<p>dữ liệu của các thuộc tính cho phù hợp</p> <ul style="list-style-type: none"> <li>- Xem xét sự phân bố của các dữ liệu dạng số , và dữ liệu không phải dạng số</li> </ul>	<ul style="list-style-type: none"> <li>- Một số thuộc tính trong dataset có thể không ở định dạng đúng, ví dụ như giá trị ngày tháng được lưu dưới dạng chuỗi thay vì kiểu datetime</li> </ul>			
3	<u>Exploring data</u>	<ul style="list-style-type: none"> <li>- Trả lời được các câu hỏi đơn giản về dataset</li> <li>- Nêu các thông tin cơ bản về bộ dữ liệu</li> </ul>		Bùi Kim Phúc	100%	@October 14, 2024
4	<u>Asking meaningful questions</u>			Bùi Kim Phúc	100%	
4.1	<u>Untitled</u>	Nồng độ các chất ô nhiễm thay đổi như thế nào qua các năm?		Bùi Kim Phúc	100%	@December 14, 2024
4.2	<u>Untitled</u>	Mức độ tương quan của các chất ô nhiễm và AQI		Bùi Kim Phúc	100%	@December 14, 2024
4.3	<u>Untitled</u>	Nồng độ PM2_5 và PM10 theo		Bùi Kim Phúc	100%	@December 15, 2024

# STT	Aa Task's Name	 Requirements	 Difficulties	 Implementor	# Progress	 Completion Date
		từng giờ trong ngày				
4.4	<u>Untitled</u>	Chỉ số chất lượng không khí (AQI) theo từng giờ trong tuần?		Bùi Kim Phúc	100%	@December 15, 2024
4.5	<u>Untitled</u>	Nồng độ khí CO thay đổi như thế nào theo mùa?		Bùi Kim Phúc	100%	@December 17, 2024
5	<u>Data Modeling</u>	Đặt ra được bài toán		Bùi Kim Phúc	100%	
5.1	<u>Decision Tree</u>	Sử dụng decision để giải quyết bài toán đã đặt ra. Nêu ưu điểm và nhược điểm. Tính chính xác để ra được mô hình tốt nhất.		Bùi Kim Phúc	100%	@December 19, 2024
5.2	<u>Random Forest</u>	Sử dụng mô hình random forest để giải quyết các yêu cầu của bài toán. Nêu ưu điểm và nhược điểm. Tính chính xác để ra được mô hình tốt nhất.		Bùi Kim Phúc	100%	@December 19, 2024
5.3	<u>SVM model</u>	Sử dụng mô hình SVM để giải quyết các yêu cầu của bài toán. Tính		Bùi Kim Phúc	100%	@December 19, 2024

# STT	Aa Task's Name	 Requirements	 Difficulties	 Implementor	# Progress	 Completion Date
		chính tham số. Nêu ưu điểm và nhược điểm				