\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*(Slide 2 Logistics)

Qua phần 2 này mục tiêu chính của chúng ta là giải quyết một vấn đề đơn giản bằng cách sử dụng Logistic Regression. Nếu như dự đoán những biến định lượng và định tính như giá nhà, thời tiết, cổ phiếu …. Tất cả những trường hợp này giá trị dự đoán là liên tục nên đa số thường dùng thuật toán Linear Regression. Nhưng xuất phát một số loại dự đoán có vấn đề khác như là Khác hàng có mua bảo hiểm nhân thọ hay không ? , Trời hôm nay có mưa hay không ? hay là Cử tri sẽ bỏ phiếu cho bên Trump hay Biden ? tất cả những vấn đề này nó là giá trị dự đoán mang tính phân loại vì thứ mà bạn đang cố gắng chỉ đơn giản là có hoặc không. Trong khi trong trường hợp Linear Regression thì ví dụ về giá cả có thê thấy rằng giá trị dự đoán có thể lad bất kỳ số nào, nó không có phạm vi xác định do đó với trường hợp thứ 2 này được gọi là bài toán classfication

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*(Slide 3 Logistic)

Và Logistics Regression là một kỹ thuật được sử dụng để giải các bài toán phân loại dạng này, bây h trong các ví dụ phân loại mà chúng ta đã thấy có 2 loại vì vậy ví dụ đầu tiên là liệu khách hàng có mua bảo hiểm nhân thọ hay không. ở đây kết quả có thể trả về thật sự rất đơn giản. chỉ có hoặc không đây được gọi là phân loại nhị phân, mặt khác khi bạn có nhiều hơn 2 giá trị trả về thì sẽ gọi là phân loại đa lớp trong trường hợp này nên dùng Linear Regression,

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*(Slide 4 Logistic)

Giả sử bạn đang làm việc với tư cách là nhà khoa học dữ liệu trong một công ty bảo hiểm nhân thọ và sếp của bạn giao cho bạn nhiêm vụ dự đoán khả năng xảy ra cho việc 1 khách hàng sẽ mua sản phẩm bạn và những gì bạn đang thấy ở đây là Dataset được dựng thủ công dựa trên độ tuổi mà thông tin bạn có là liệu khác hàng có mua bảo hiểm hay không. Bây h ở đây bạn có thể thấy một số mẫu bất thường như người trẻ tuổi thì không mua bảo hiểm quá nhiều, bạn có thể thấy như có những người ở độ tuổi 20-25 mà bằng 0 có nghĩa là họ không mua bảo hiểm trong khi độ tuổi của họ tăng lên thì khả năng họ mua bảo hiểm lại nhiều hơn để nắm rõ được mối quan hệ hơn thế nữa bạn muốn xây dựng một mô hình học máy có thể thực hiện dự đoán dựa trên độ tuổi của khác hàng, so với tư cách là nhà khoa học dữ liệu, đây là công việc bạn bạn phải làm được

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*(chuyển sang slide 5 Logistics)

Ok và công việc đầu tiên khi có được dữ liệu này là vẽ một biểu đồ phân tán trông giống như slide khi bạn đã nghiên cứu các bài toán