

ĐẠI HỌC QUỐC GIA TP HCM
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN



Đồ án 1

Môn: Toán Ứng dụng và thống kê

GVHD: ThS. Vũ Quốc Hoàng

ThS. Trần Thị Thảo Nhi

Thầy Nguyễn Văn Quang Huy

SVTH:

1753104 – Trần Thuận Thành

1753116 – Đặng Đức Trung

1753118 – Nguyễn Thanh Trường

Lớp: 17CLC - KTPM

TP. HCM, Tháng 06 năm 2020

Mục lục

1. *Lựa chọn mô hình* 3
2. *Lý do lựa chọn mô hình:*..... 3

1. Lựa chọn mô hình

- Mô hình được lựa chọn để giải quyết bài toán dự đoán giá xe là Linear Regression của thư viện sklearn. Phần tiền xử lý sử dụng lớp PolynomialFeatures của sklearn.preprocessing.

2. Lý do lựa chọn mô hình:

- Sử dụng sklearn.linear_model.LinearRegression: đã được thiết kế sẵn cho mô hình dự đoán. Thư viện sklearn được xây dựng trên NumPy, SciPy và matplotlib giúp tập trung vào giải quyết bài toán nhanh hơn vì mô hình được thiết kế sẵn. Tự thiết kế có thể không tính toán được tất cả trường hợp và xử lý không tối ưu, hoặc tốn nhiều thời gian.
- Dùng lớp Polynomial để sinh ra các thuộc tính đa thức và có tương tác. Chia các thuộc tính thành 2 phần, với phần đầu tiên là các biến features True/False và biến odometer, phần thứ 2 là các biến còn lại. Việc lựa chọn PolynomialFeatures(degrees =6) cho các thuộc tính ở phần đầu tiên . Với phần thứ 2, lựa chọn interaction_only sinh thêm dữ kiện là các thuộc tính có tương tác, và degree=1 tránh trường hợp overfitting.