

Mobile Price Classification

Lương Khắc Mạnh

Ngày 7 tháng 3 năm 2021

Nội dung chính

- 1 Giới thiệu bài toán
- 2 Các model phân loại
- 3 Giải quyết bài toán và kết quả thực nghiệm

Giới thiệu bài toán

Ngày nay, điện thoại di động là một trong những thiết bị được bán và mua nhiều nhất. Mỗi ngày điện thoại di động mới với phiên bản mới và nhiều tính năng hơn được tung ra. Hàng trăm và hàng nghìn thiết bị di động được bán và mua hàng ngày. Vì vậy, ở đây dự đoán loại giá giúp khách hàng có thể lựa chọn được sản phẩm tối ưu phù hợp với túi tiền. Công việc tương tự có thể được thực hiện để ước tính giá thực của tất cả các sản phẩm như ô tô, xe đạp, máy phát điện, động cơ, thực phẩm, thuốc,...



Các model phân loại

- Support Vector Machine (Kernel='linear')
- Support Vector Machine (Kernel='rbf')
- Gaussian Naive Bayes
- Decision Tree Classifier
- Random Forest Classifier

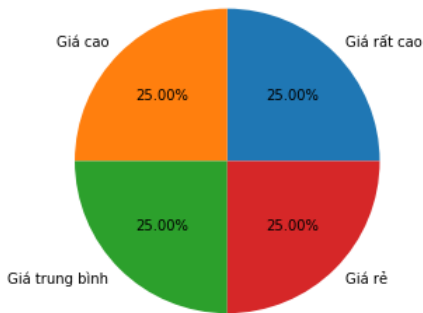
Giải quyết bài toán và kết quả thực nghiệm

- Thu thập dữ liệu
- Tiền xử lý
- Đề xuất giải pháp
- Kết quả thực nghiệm

Thu thập dữ liệu

- Bộ dữ liệu được thu thập từ <https://www.kaggle.com/>
- Bộ dữ liệu huấn luyện gồm có 2000 mẫu. Mỗi mẫu có 20 thuộc tính và trường `price_range`
- Nhãn gồm 4 loại được đánh số 0, 1, 2, 3 lần lượt là giá rẻ, giá trung bình, giá cao và giá rất cao.

Nhóm giá điện thoại



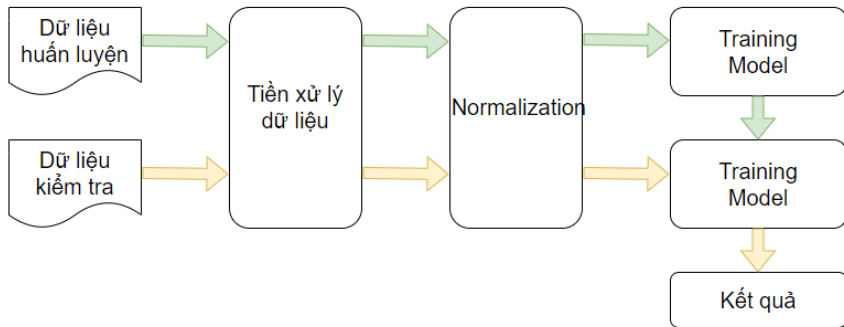
Tiền xử lý dữ liệu

- Các feature 'blue', 'dual_sim', 'four_g', 'three_g', 'touch_screen', 'wifi', 'price_range' không bị mất cân bằng dữ liệu.
- Các feature còn lại chuẩn hóa theo công thức:

$$x' = \frac{x - \bar{x}}{\sigma}$$

với \bar{x} , σ lần lượt là kỳ vọng và phương sai (standard deviation) của thành phần đó trên toàn bộ training data.

Đề xuất giải pháp



Quy trình thực hiện

Kết quả đạt được

| STT | Traning model | Training accuracy | Testing accuracy |
|-----|--------------------------|-------------------|------------------|
| 1 | SVM (Kernel = Linear) | 0.980625 | 0.9525 |
| 2 | SVM (Kernel = rbf) | 0.98 | 0.9175 |
| 3 | Gaussian Naive Bayes | 0.81625 | 0.835 |
| 4 | Decision Tree Classifier | 1.0 | 0.825 |
| 5 | Random Forest Classifier | 0.996875 | 0.7825 |

→ SVM (Kernel = 'linear') cho kết quả tốt hơn so với các model còn lại. Decision Tree và Random Forest bị overfitting.

Source code: https://github.com/ManhLk/Project_III/tree/master/Mobile_Price_Classification

Thank you for your attention!