

# **Classification & Analysis of Car Evaluation**

**Professor : Dr.shakiba**

**By : Maryam Mohammadabadi**

(97143045)

**سوال: ما می‌خواهیم سیستمی را طراحی کنیم که ماشین‌های دست دوم را ارزیابی کند و به مشتریان در خرید خودروهای دست دوم بر اساس نیازشان کمک کند.**

**معرفی:**

هنگام خرید ماشین دست دوم ویژگی‌های مختلفی وجود دارد. این ویژگی‌ها وضعیت نهایی خودرو را تعیین می‌کند. با این حال، اهمیت یک ویژگی ممکن است از اهمیت ویژگی دیگر بیشتر باشد و تصمیم‌گیری به عنوان یک کل و ارزیابی یک خودرو دشوار شود.

**بیان مسأله:**

هدف این پروژه کمک به مشتریان در خرید خودروهای دست دوم بر اساس نیازشان با شناسایی ویژگی‌های کلیدی و انجام تجزیه و تحلیل طبقه‌بندی بر روی مجموعه داده ارزیابی خودرو برای ساخت یک مدل یادگیری ماشینی با استفاده از داده‌های آموزشی برای طبقه‌بندی موثر نمونه‌های آزمایشی بر اساس طبقه‌بندی‌کننده است. مدل.

الگوریتم‌های طبقه‌بندی مختلف مورد استفاده عبارتند از:

- Logistic Regression
- K-Nearest Neighbors
- Decision Tree

خودروها در یکی از دسته‌های غیر قابل قبول، قابل قبول، خوب و بسیار خوب (unacceptable, acceptable, good, and very good) طبقه‌بندی می‌شوند.

ما می‌توانیم با مقایسه دقت‌های به دست آمده از هریک از الگوریتم‌های بکار رفته بهترین مدل را براساس نیاز انتخاب کنیم.

داده‌ها را به مجموعه‌های آموزشی و آزمایشی تقسیم می‌کنیم و مدل‌های طبقه‌بندی‌کننده را آموزش می‌دهیم. در نهایت مدل‌ها تست می‌شوند و دقت آن‌ها را محاسبه می‌کنیم.

## توضیحات مجموعه داده:

مجموعه داده مورد استفاده در این پروژه مجموعه داده ارزیابی خودرو (Car Evaluation Dataset) است.

پایگاه داده شامل 1728 نمونه با 6 ویژگی و یک ویژگی کلاس/رده است. این مدل مقبولیت یک خودرو را بر اساس 6 ویژگی 'buying', 'maint', 'doors', 'persons', 'lug\_boot', 'safety' ارزیابی می‌کند. ویژگی‌های تعیین‌کننده وضعیت خودرو:

- buying {vhigh, high, med, low} – قیمت خرید خودرو از نظر بسیار بالا، بالا، متوسط و پایین
- maint {vhigh, high, med, low} – قیمت تعمیر و نگهداری خودرو از نظر بسیار بالا، بالا، متوسط و پایین
- doors {2, 3, 4, 5more} – تعداد درهای ماشین یعنی 2، 3، 4 و 5 یا بیشتر
- persons {2, 4, more} – ظرفیت بر حسب افرادی که خودرو در آن حمل می‌کند یعنی 2، 4 و بیشتر
- lug\_boot {small, med, big} – اندازه صندوق از نظر کوچک، متوسط و بزرگ
- safety {low, med, high} – برآورد ایمنی خودرو از نظر کم، متوسط و زیاد

مقادیر ویژگی کلاس که مقبولیت یک خودروی ارزیابی شده را بر حسب دسته بندی‌های مختلف تعریف می‌کنند عبارتند از:

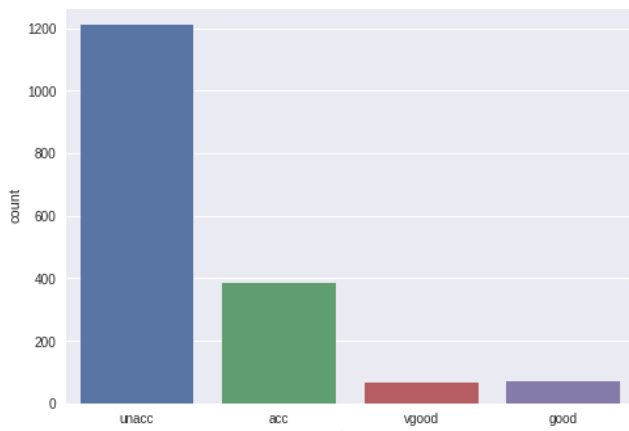
- unacc – unacceptable
- unacc – unacceptable
- good – good
- vgood – very good

## فرضیه:

ما ویژگی‌های مختلف یک خودرو را در نظر خواهیم گرفت: قیمت خرید، تعمیر و نگهداری، درها، افراد، صندوق بار و ایمنی، و سپس آن را در یک کلاس از آن ارزیابی خواهیم کرد. قابل قبول بودن در اینجا ما فرض می‌کنیم که قیمت خرید و ایمنی یک خودرو مهمترین عنصر در ارزیابی یک خودرو است.

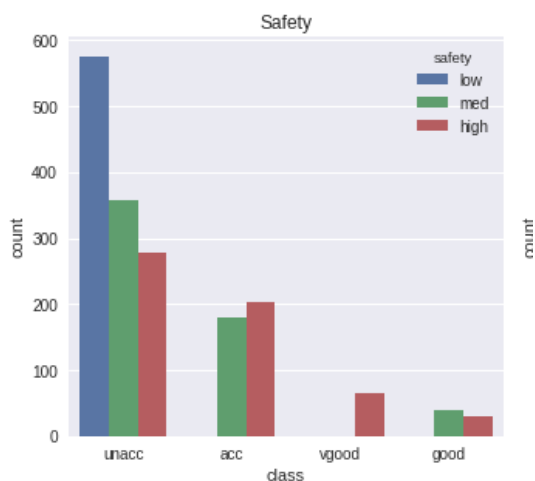
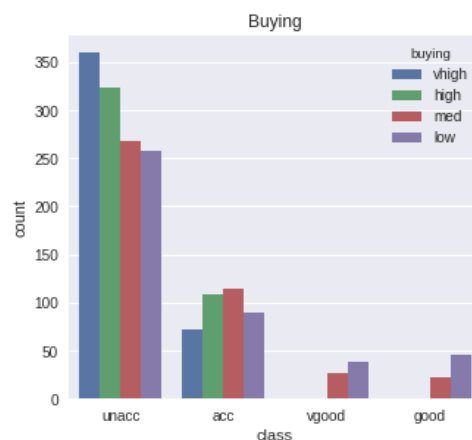
قیمت خرید پایین و ایمنی بالا مطلوب ترین ویژگی در ارزیابی خودرو هستند.

## تحلیل داده ها:



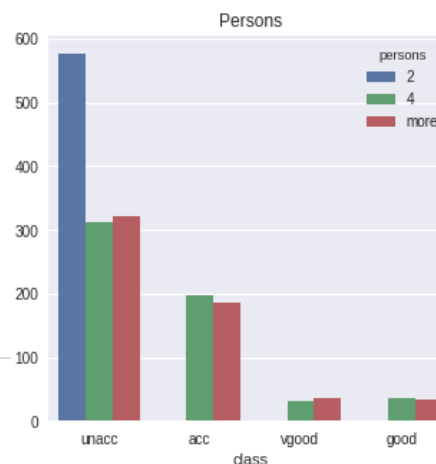
نمودار شمارش کلی برای انواع مختلف خودروهای ارزیابی شده در مجموعه داده

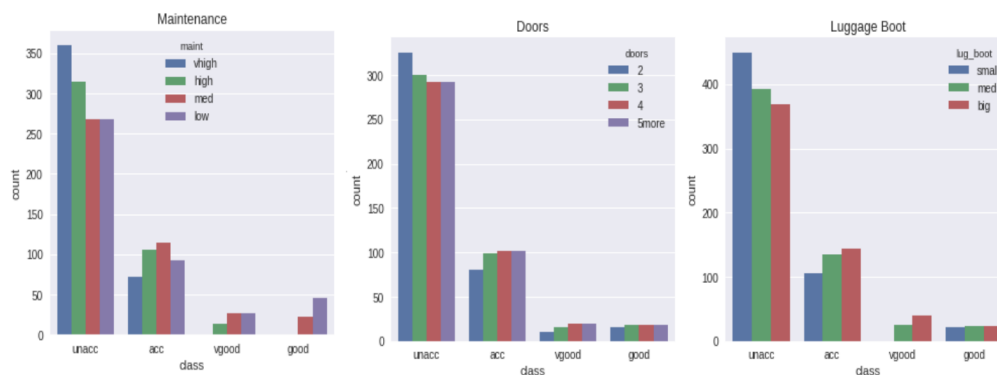
در اینجا در نمودار، می توان مشاهده کرد که خودروهایی با قیمت های خرید «بالا» و «بسیار بالا» در دسته های «غیرقابل قبول» و «قابل قبول» قرار می گیرند. در حالی که قیمت های خرید خودرو «پایین» و «متوسط» به عنوان دسته بندی «خوب» و «بسیار خوب» خودروها در نظر گرفته می شوند



در اینجا در نمودار، می توان مشاهده کرد که خودروهایی با ویژگی ایمنی «کم» به طور خودکار به عنوان خودروهای «غیرقابل قبول» طبقه بندی می شوند. و آنهایی که ایمنی «بالا» دارند بیشتر مورد توجه قرار می گیرند.

از این نمودار، مشاهدات جالب دیگری که می توانیم انجام دهیم این است که خودروهایی با ظرفیت تنها 2 نفر نیز به طور خودکار در گروه غیرقابل قبول طبقه بندی می شوند.





از این سه نمودار، می توان استنباط کرد که ویژگی ها: قیمت تعمیر و نگهداری، تعداد درها، و ظرفیت صندوق بار در هنگام تصمیم گیری در مورد مقبولیت و وضعیت یک خودروی دست دوم خیلی مهم نیستند.

## نتایج طبقه بندی:

مقایسه مدل های مختلف یادگیری ماشین پیاده سازی شده

### **Logistic Regression:**

Accuracy: 0.8410404624277457

Confusion Matrix:

```
[[ 59  2 18  0]
 [ 15  0  0  2]
 [ 11  0 229  0]
 [  7  0  0  3]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.64	0.75	0.69	79
1	0.00	0.00	0.00	17
2	0.93	0.95	0.94	240
3	0.60	0.30	0.40	10
avg/total	0.81	0.84	0.82	346

**Accuracy = 84.10%**

## K-Nearest Neighbors:

Accuracy: 0.9479768786127167

Confusion Matrix:

```
[[ 74  1  4  0]
 [ 7  9  0  1]
 [ 2  0 238  0]
 [ 3  0  0  7]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.94	0.90	79
1	0.90	0.53	0.67	17
2	0.98	0.99	0.99	240
3	0.88	0.70	0.78	10
avg/total	0.95	0.95	0.95	346

Accuracy = 94.79%

## Decision Tree:

Accuracy: 0.9884393063583815

Confusion Matrix:

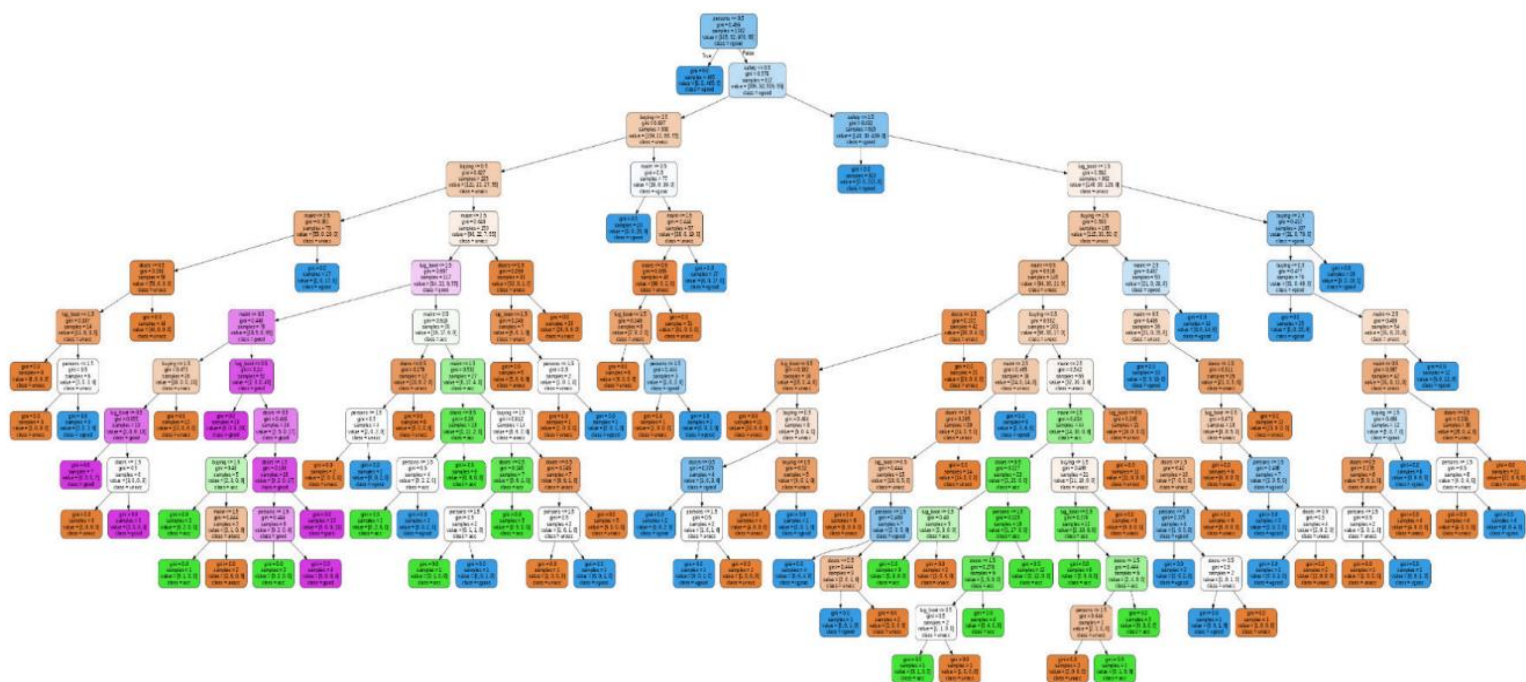
```
[[ 76  2  1  0]
 [ 0 17  0  0]
 [ 0  0 240  0]
 [ 1  0  0  9]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.99	0.96	0.97	79
1	0.89	1.00	0.94	17
2	1.00	1.00	1.00	240
3	1.00	0.90	0.95	10
avg/total	0.99	0.99	0.99	346

Accuracy = 98.84%

✓ همانطور که می بینیم، الگوریتم درخت تصمیم به بالاترین دقت برای بهترین طبقه بندی نمونه های آزمایشی دست یافت.



نمودار بالا مدل درخت تصمیم تولید شده را نشان می دهد.

**نتیجه:**

بنابراین نتیجه می گیریم که طبق تحلیل ما، فرضیه اولیه با موفقیت به اثبات رسیده است. **قیمت خرید و ایمنی خودرو** دو مورد از مهم ترین ویژگی ها در ارزیابی خودرو هستند. و اینکه قیمت پایین تر با ایمنی بالاتر به معنای مقبولیت بهتر خودروی دست دوم است. از بین سه الگوریتم پیاده سازی شده، الگوریتم **درخت تصمیم** به بالاترین دقت دست یافت و با توجه به ویژگی های آن، بهترین مدل برای پیش بینی کلاس خودرو بود.