****

دانشکده فنی و مهندسی

گروه مهندسی کامپیوتر

**پروژه یادگیری ماشین**

**نام دانشجو**

**شماره دانشجویی**

**استاد درس**

**دکتر فاطمه باقری**

نیمسال اول سال تحصیلی 1403-1404

1. **مجموعه داده**

لطفاً داده های خود را **در یک پاراگراف** معرفی کنید، این پاراگراف باید شامل توضیحات در رابطه با موارد زیر باشد.

• موضوع داده های شما (موضوع، زمان و مکان و ... نیز ذکر شود)

• تعداد کل رکوردها

• تعداد ویژگی ها

• بیان ویژگی ها

• اگر داده دارای مقادیر null باشد

• اگر داده ها دارای نقاط پرت باشد

و سایر مواردی که نیاز است در مورد مجموعه داده خود بیان کنید.

در پاراگراف بعدی بیان کنید که از چه ابزاری برای توصیف داده خود استفاده کردید. تمامی کد های خود را در گیت هاب بارگزاری کرده و آدرس آن را در اینجا بیان کنید

* 1. **مشخصات مجموعه داده**

جدول توصیف مجموعه داده

|  |  |  |
| --- | --- | --- |
| ردیف | ویژگی | توصیف |
|  |  |  |

هر کدام از ویژگی ها را در یک ردیف قرار دهید

* 1. **نمونه هایی از مجموعه داده**

در این جدول 6 ردیف مختلف از مجموعه داده خود را، قبل از انجام هر کاری بر روی آن، به عنوان نمایش مجموعه داده قرار دهید.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ویژگی 1 | ویژگی 2 | ویژگی 3 | ویژگی 4 | ویژگی 5 | ... |
|  |  |  |  |  |  |

* 1. **مشخصات اولیه ویژگی های مجموعه داده**

اگر برخی از ویژگی های مجموعه داده به صورت اسمی هستند، در ردیف های انتهایی جدول زیر توصیفی به صورت نمونه قرار دهید. برای هر کدام از ویژگی های اسمی نیز یک ردیف در انتهای جدول قرار دهید.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ویژگی | تعداد | ماکزیمم مقدار | مینیمم مقدار | مقدار میانه | تعداد داده های ناموجود |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| جنسیت | 1000 | 600 مورد 🡪 زن  400 مورد 🡪 مرد | | | 20 |

* 1. **نمودارهای توصیف داده**

در این بخش نمودارهایی که به توصیف داده ها کمک می کنند قرار می گیرند. به عنوان مثال برای هر کدام از ویژگی های اسمی نموداری طراحی شود که تعداد نمونه های هر کدام از مقادیر اسمی را نشان دهد و در عین حال قابل مقایسه با تعداد مقادیر دیگر باشد.

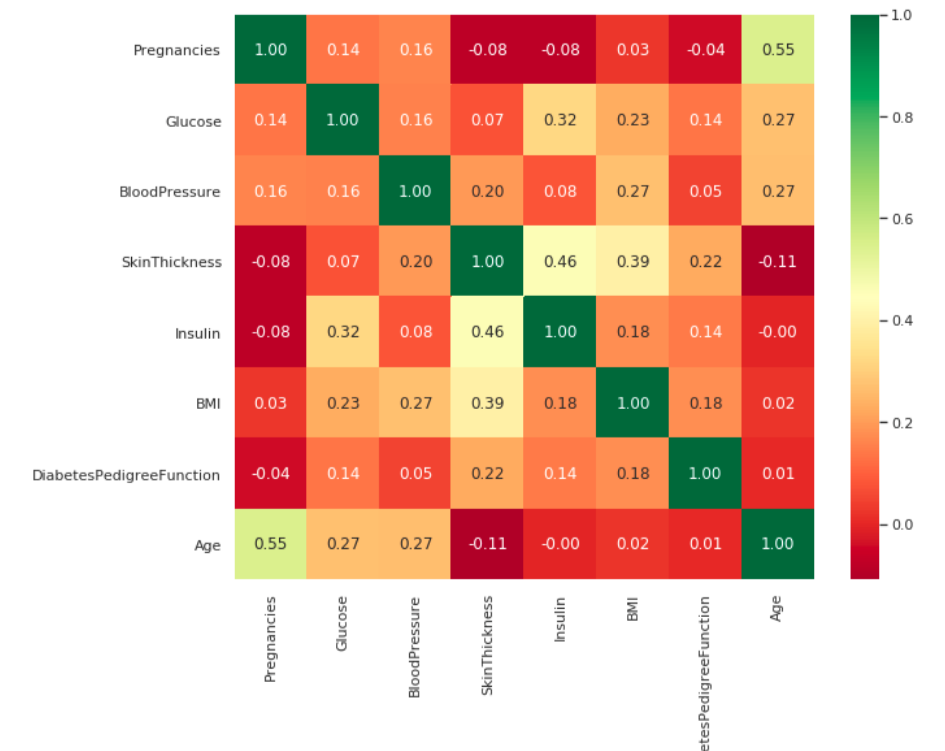
به عنوان مثال برای جنسیت در مجموعه داده بالا می توان نمودار زیر را رسم کرد.

هر کدام از اشکال می بایست دارای عنوان باشد و در یک پاراگراف مجموعه اشکال توصیف داده شوند و در حین توصیف به شکل مربوطه ارجاع داده شوند.

در شکل 1 فراوانی جنسیت در مجوعه داده نشان داده شده است.

شکل 1: فراوانی ویژگی جنست در مجموعه داده

و در پایان اشکال، نمودار وابستگی ویژگی های مجموعه داده (heatmap) را قرار دهید. به عوان مثال شکل زیر این نمودار را برای مجموعه داده PIMA نشان می دهد.



شکل ؟؟؟: نقشه حرارتی ضرایب همبستگی پیرسون برای همه ویژگی های دیابت.

سایر موارد در تمارین بعدی به این فایل اضافه خواهد شد.