

به نام خدا



دانشکده مهندسی کامپیوتر

مبانی و کاربردهای هوش مصنوعی ترم پاییز 1402

پروژه چهارم

مهلت تحویل 2 تیر 1402 ساعت 23:55

---

### پیش بینی به اشتراک گذاری دوچرخه ها در شهر لندن

در این پروژه ما با مبانی Data Analysis و نحوه ی استفاده از Bayesian Network در یک مساله ی کاربردی آشنا میشویم.

#### مقدمه

سرویس های اشتراک دوچرخه علاوه بر حفظ محیط زیست در سلامت عمومی جامعه موثر هستند. یکی از مشکلات متداول در مدیریت این دوچرخه ها، کنترل تعداد آنها در بازه های زمانی مختلف است. در این پروژه ما از دیتاست بدست آمده از یکی از سرویس های اشتراک دوچرخه در شهر لندن استفاده میکنیم. (این دیتاست از مجموعه دیتاست های وب سایت کگل است، [لینک برای دسترسی مستقیم](#)). هدف این پروژه تقویت مهارت های تحلیلی شما و یادگیری اصول Data Driven Decision Making است.

فایل pdf گزارش در [این لینک](#) را دانلود و مطالعه کنید و به سوالات زیر پاسخ دهید:

الف) درباره ی پیش پردازش داده ها، لزوم و متد های آن تحقیق کنید. چه تغییراتی بر روی دیتاست ما انجام شده است؟ دلیل اینکار چیست؟

ب) حتما با شبکه های ییزی آشنایی دارید. شبکه های ییزی نوعی مدل گرافیکی احتمالی هستند که از استنتاج ییزی برای محاسبات احتمالی استفاده می کنند. برای ساخت یک شبکه ییزی، اولین گام تعریف ساختار شبکه به وسیله یک مجموعه از متغیرهای تصادفی و وابستگی های آنهاست. سپس، باید برای هر گره، توزیع های احتمالی شرطی (CPD) را که تأثیر والدین بر آن گره را نشان میدهد، تعریف کنیم. کتابخانه pgmpy امکان تعریف دستی یا خودکار هر دو ساختار شبکه و CPD ها را فراهم می کند.

نگاهی به مستندات این کتابخانه بیندازید و با آن آشنا شوید.

پ) یکی از مهمترین مهارت های یک دیتا آنالیزست، تشخیص اولیه و شهودی او درباره ی متغیرهای مختلف دیتاست مورد نظر است. روابط بین متغیرها از این نظر اهمیت دارند که در طول یادگیری مدل ها، سعی میکنیم مجموعه دیتایی داشته باشیم که متغیرهای آن (تأحد امکان) از یک دیگر مستقل باشند. بنظر شما دلایل انجام این کار چه میتواند باشد؟

ج) استفاده از شبکه بیزی هم نیازمند مهارت در تشخیص روابط بین متغیرهاست. تحلیل اولیه خودتان از وابستگی علی متغیرهای دیتاست را ارائه دهید. این کار را بدون نگاه به شبکه ی بیزی در گزارش انجام دهید.

د) ممکن است تشخیص اولیه و شهودی ما به اندازه کافی خوب نباشد. یک ایده خوب استفاده از داده ها برای تشخیص روابط بین متغیرهاست. با توجه به متن گزارش، درباره ی دو روش کلی انجام این کار توضیحاتی ارائه دهید.

ه) نمودارهای بدست آمده از دیتاست در طی گزارش را در نظر بگیرید و تحلیل کوتاهی از هر کدام بنویسید. هدف یادگیری تحلیل نمودارها به عنوان یک دیتا آنالیزست است. با این حال، یکی از چالش های همیشگی دقیق نبودن یا ناقص بودن داده های بدست آمده است پس تلاش کنید دچار Over Analysis نشوید. پس از مطالعه ی گزارش خلاصه ای از نتیجه ی بدست آمده بنویسید.

### توضیحات تکمیلی

- پاسخ به تمرین ها باید به صورت فردی انجام شود. در صورت مشاهده تقلب، برای همه ی افراد نمره صفر لحاظ خواهد شد.
- پاسخ خود را در قالب یک فایل PDF به صورت تایپ شده یا دست نویس (مرتب و خوانا) در سامانه کورسز آپلود کنید.
- فرمت نامگذاری تمرین باید مانند AI\_HW4\_9931099 باشد.
- در صورت هرگونه سوال یا ابهام از طریق ایمیل؟ با تدریس یاران در ارتباط باشید. همچنین خواهشمند است در متن ایمیل به شماره دانشجویی خود اشاره کنید.
- همچنین می توانید از طریق تلگرام نیز با آیدی های زیر در تماس باشید و سوالاتتان را مطرح کنید:
  - @Mohammad\_H915
  - @farhad771380
  - @Arriann
  - @AhooraGorji
- ددلاین این تمرین 2 تیر ۱۴۰۲ ساعت ۲۳:۵۵ است و امکان ارسال با تاخیر وجود ندارد، بنابراین بهتر است انجام تکلیف را به روزهای پایانی موکول نکنید.