

پروژه ۱ – درس مبحث ویژه ۲

موضوع: دریافت داده

سوال اول - برنامه ای به زبان پایتون بنویسید که محتوی اخبار صفحه اول دانشگاه کردستان را لیست کرده و موارد زیر را در خروجی نمایش دهد:

- عنوان خبر
- نوع خبر (اخبار آموزشی، اخبار دانشجویی و ...)
- تاریخ درج خبر
- متن خبر

سوال دوم: برنامه ای به زبان پایتون بنویسید که لینک URL یک محصول را در دیجیکالا دریافت کند و موارد زیر را در خروجی نمایش دهد:

- عکسهای محصول
- ویژگیهای محصول
- نقد و بررسی
- مشخصات فنی
- نظرات کاربران
- قیمت محصول

برای مثال می توانید لینک زیر را ببینید

<https://www.digikala.com/product/dkp-2504378/%DA%A9%D8%A7%D9%88%D8%B1-360-%D8%AF%D8%B1%D8%AC%D9%87-%D8%AC%DB%8C-%DA%A9%DB%8C-%DA%A9%DB%8C-%D9%85%D8%AF%D9%84-gk-a51-20-%D9%85%D9%86%D8%A7%D8%B3%D8%A8-%D8%A8%D8%B1%D8%A7%DB%8C-%DA%AF%D9%88%D8%B4%DB%8C-%D9%85%D9%88%D8%A8%D8%A7%DB%8C%D9%84-%D8%B3%D8%A7%D9%85%D8%B3%D9%88%D9%86%DA%AF-galaxy-a51#/tab-params>

توجه کنید حتما پروژه را خودتان انجام دهید. برای بررسی اینکه پروژه را خودتان انجام داده باشید از شما مصاحبه شفاهی به عمل خواهد آمد. در صورت هر گونه تقلب نمره منفی برای شما و برای کسی که پروژه را در اختیار قرار داده است در نظر گرفته خواهد شد. در صورت تکرار موضع امکان گذراندن درس وجود نخواهد داشت

پروژه ۲ – درس مبحث ویژه ۲

موضوع: پاکسازی داده و بصری سازی داده

سوال اول: مجموعه داده ای که در لینک زیر وجود دارد را در نظر بگیرید. فایل اصلی این مجموعه داده ای در فایل با پسوند data. وجود دارد.

<https://archive.ics.uci.edu/ml/machine-learning-databases/audiology>

/

برنامه ای با پایتون بنویسید و موارد زیر را انجام دهید:

- فایل را از ورودی خوانده و به CSV تبدیل کنید
- تعداد سطرها و ستونها را در خروجی چاپ کنید
- آیا دیتاست دارای مقدار نامعلوم (miss value) است؟ (مقادیر نامشخص با علامت ؟ مشخص شده است)
- تعداد مقادیر Miss value را برای هر ستون در خروجی چاپ کنید.
- مقادیر miss value را با یکی از روشهای موجود تکمیل کنید
- نمودار گرافیکی مربوط به هر صفت (ستون) را بر حسب تعداد اعضای هر طبقه چاپ کنید.
- نمودار کلاس دیتا (ستون آخر) را بر حسب مقادیر آن در خروجی رسم کنید.

سوال دوم: موارد درخواستی سوال اول را برای داده زیر انجام دهید:

<https://archive.ics.uci.edu/ml/machine-learning-databases/adult>

/

توجه کنید حتما پروژه را خودتان انجام دهید. برای بررسی اینکه پروژه را خودتان انجام داده باشید از شما مصاحبه شفاهی به عمل خواهد آمد. در صورت هر گونه تقلب نمره منفی برای شما و برای کسی که پروژه را در اختیار قرار داده است در نظر گرفته خواهد شد. در صورت تکرار موضع امکان گذراندن درس وجود نخواهد داشت

پروژه ۳ – درس مبحث ویژه ۲

موضوع: طبقه بندی داده

مجموعه داده ای زیر را در نظر بگیرید. در این مجموعه داده ای تعداد ۱۷۲۳ داده و ۶ صفت وجود دارد. هدف این مجموعه داده ای ارزیابی یک اتومبیل بر اساس به دسته های (خیلی خوب، خوب، متوسط و ضعیف) است. این مقادیر در دیتا ست با نامهای **unacc, acc, good, vgood** مشخص شده است. اطلاعات زیر در مورد هر اتومبیل ثبت شده است

<https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

Attributes:

buying: vhigh, high, med, low.

maint: vhigh, high, med, low.

doors: 2, 3, 4, 5more.

persons: 2, 4, more.

lug_boot: small, med, big.

safety: low, med, high.

برنامه ای به زبان پایتون بنویسید و موارد زیر را انجام دهید:

- فایل مجموعه داده ای را دانلود کرده و در ورودی بخوانید
- مجموعه داده ای را به دو دسته آزمایش و تست تقسیم بندی کنید. ۷۰٪ داده برای آزمایش و ۳۰٪ برای تست.
- بر اساس داده آزمایشی، مدل های شبکه عصبی (Neural Networks)، درخت تصمیم (Decision Trees) و بیس ساده (Naïve Bayes) را آموزش دهید.
- بر اساس داده تستی، هر یک از مدل ها را ارزیابی کنید و خروجی هر مدل را بر اساس شاخص های زیر نمایش دهید:

- Precision ○
- Recall ○
- AUC ○
- Accuracy
- F1 ○

توجه کنید حتما پروژه را خودتان انجام دهید. برای بررسی اینکه پروژه را خودتان انجام داده باشید از شما مصاحبه شفاهی به عمل خواهد آمد. در صورت هر گونه تقلب نمره منفی برای شما و برای کسی که پروژه را در اختیار قرار داده است در نظر گرفته خواهد شد. در صورت تکرار موضع امکان گذراندن درس وجود نخواهد داشت

پروژه 4 – درس مبحث ویژه ۲

موضوع: خوشه بندی داده

سوال اول: روش خوشه بندی k-means و fuzzy-c means را به صورت کامل شرح دهید

سوال دوم: روشهای ارزیابی خوشه بندی را توضیح دهید. توضیح: فرض کنید که دو روش برای خوشه بندی داده داریم. چه معیارهایی برای ارزیابی این دو روش خوشه بندی وجود دارد (توضیح : معیارهایی مانند NMI ، Rand Index و ...وجود دارد). این معیارها را توضیح دهید

سوال سوم: روشهای k-means و fuzzy-c-means را بر روی داده زیر اجرا کرده و خروجی را بر اساس معیارهای خوشه بندی موجود در scikit-learn (مانند NMI و Rand Index(RI) ، ARI و ...) نشان دهید.

<https://archive.ics.uci.edu/ml/datasets/iris>

سوال چهارم: روشهای k-means و FCM را بر روی داده های مقاله زیر اجرا کرده و بر اساس معیارهای استفاده شده در این مقاله نمایش دهید

<https://www.sciencedirect.com/science/article/abs/pii/S0031320320302521>

برای دانلود مقاله از سایت sci-hub.tw استفاده کنید. از سایت <https://online.semantak.com> نیز میتوانید مقالات را دانلود کنید (نیاز به آی پی دانشگاه دارد). فایل مقاله از طریق گروه تلگرامی هم در اختیار قرار داده خواهد شد