

گزارش پروژه پایانی درس داده کاوی

نام و نام خانوادگی : محمد گرامی فر

توضیحات کلی: برای هر سرویس فایل مورد نظر را آپلود نمایید و برای دیدن خروجی دکمه مربوط به سرویس مورد نظر را کلیک کنید. خروجی آن نمایش داده می شود و همچنین گزینه دانلود فایل خروجی وجود دارد. برای رفتن به صفحه اول پروژه دکمه homepage را کلیک کنید. برای هر کدام از سرویس ها اگر فرمت فایل json درست نباشد، پیغامی برای کاربر چاپ می شود. برای سرویسهای شماره ۱ و ۲ روش های درونیایی spline و polynomial اضافه شده است.

همچنین آدرس سرویس ابری <https://fandoghi-service-shahlla.fandogh.cloud> است

سرویس شماره ۱: در ابتدا برای سرویس شماره ۱ فایل json آپلود می شود و بر اساس قسمت فایل config این فایل که نشان می دهد داده ها به صورت ماهانه یا روزانه است، قسمت time داده ها بررسی می شود که فرمت مورد نظر رعایت شده است. فرمت time برای داده های ماهانه به صورت y/m و برای داده های روزانه به صورت y/m/d است. سپس اگر زمان داده ها به اساس زمان شمسی باشد به زمان میلادی تبدیل می شود و روزها یا ماههایی که برای آن مقداری تعیین نشده به دیتافریم اضافه می شود و درونیایی بر اساس روشی که در فایل json آمده است انجام می شود. برای درونیایی از تابع interpolate استفاده می شود. زمانی که مقدارهایی که گم شده است، درونیایی شود، زمان ها به فرمت اولی که بودند تبدیل می شود و سپس به صورت یک فایل json کاربر نمایش داده می شوند.

سرویس شماره ۲: این سرویس همانند سرویس شماره یک است با این تفاوت که در پایان زمان ها به زمان شمسی تبدیل می شود و به کاربر نمایش داده می شود.

سرویس شماره ۳: روش های مورد استفاده در این سرویس برای تشخیص داده های پرت به صورت زیر است :

- روش شماره یک : standard Deviation
- روش شماره دو : Boxplots
- روش شماره سه : Isolation Forest
- روش شماره چهار : DBScan Clustering

سرویس شماره ۴: برای مدیریت داده‌های پرت از کتابخانه imblearn استفاده شده است و روش‌های زیر به این سرویس اضافه شده است.

۱. BorderlineSMOTE

۲. SVMSMOTE

۳. ADASYN

۴. NearMiss(version 1, version 2, version 3)

۵. CondensedNearestNeighbour

۶. TomekLinks

۷. EditedNearestNeighbours

۸. OneSidedSelection