

پروژه تیم دیتا مزرعه

داده های اقتصادی، اجتماعی، زیست محیطی و نهادی ایران در بازه زمانی ۱۹۶۰-۲۰۲۲ در فایل اکسل پیوست آورده شده است. در این فایل ۱۴۷۷ شاخص از کشور ایران مرتب شده است. در این پروژه دیتا هدف بر آن است تا با استفاده های از ابزارهای علم داده، تصویری از روابط بین این شاخص ها ارائه و مدلسازی شود. از شما به عنوان کارشناس علم داده موارد زیر مطلوب است:

۱- شناسایی مهمترین شاخص های موثر بر صادرات ایران (Exports of goods and services)
(constant 2015 US\$)) و ارائه مدل مناسب برای پیش بینی آن.

یک pipeline طراحی کنید تا کلیه این اقدامات را در برگیرد.
در این بخش اجرای تمام گام ها و مراحل یک پروژه علم داده را مدنظر داشته باشید و در صورت نیاز از Feature engineering، Feature extraction، Feature selection به صلاحدید خود استفاده کنید.

۲- سهم ارزش افزوده بخش کشاورزی از GDP (Agriculture, forestry, and fishing, value added)
(% of GDP)) را بر اساس روش های سری زمانی (یا سایر روش های دلخواه) برای ۵ سال آینده پیش بینی کنید.

۳- یک کارشناس سیاسی اظهار میکند که تحریم های اقتصادی علیه ایران که از سال ۲۰۱۲ اعمال شده است، نه تنها تاثیر منفی بر کشور نداشته است بلکه باعث رشد و شکوفایی کشور شده است. از سوی دیگر یک کارشناس دیگر برخلاف گفته ایشان، بیان میکند که وضعیت فعلی بدتر دوران جنگ تحمیلی است.

از شما به عنوان کارشناس داده انتظار می رود تا با بررسی وضعیت شاخص های درآمد سرانه (GDP per capita (current US\$))، سرمایه گذاری (Gross capital formation (constant (LCU))، تورم ((Inflation, consumer prices (annual %)) در بازه های زمانی تحریم و جنگ تحمیلی، بر اساس ابزارها و روش های علم داده، نظر فنی خود را نسبت به این اظهارات بیان کنید.

*** لطفا پاسخ های خود را در قالب داکيومنت با تشریح مراحل انجام شده و ذکر دلیل دیتایی خود برای انجام هر اقدام بیان کنید.