

تحلیل و مدل‌سازی پیشرفته شهرهای جهان بر اساس شاخص‌های اجتماعی-اقتصادی و زیست محیطی جهت سنجش شادی جهانی

هدف

انجام تحلیلی عمیق بر روی مجموعه داده‌های گزارش شادی جهانی با به‌کارگیری روش‌های پیشرفته بهینه‌سازی محدب، به‌منظور مدل‌سازی و پیش‌بینی نمرات شادی بر اساس عوامل مختلف اجتماعی-اقتصادی.

مجموعه داده

<https://www.kaggle.com/datasets/unsdsn/world-happiness>

طرح کلی پروژه

۱. پیش‌پردازش داده‌ها

- پاک‌سازی داده‌ها: رسیدگی به مقادیر گمشده و نقاط پرت برای اطمینان از یکپارچگی داده‌ها.
- مهندسی ویژگی‌ها: ایجاد یا تبدیل ویژگی‌های جدید برای بهبود درک الگوهای پنهان.
- نرمال‌سازی: مقیاس‌بندی ویژگی‌ها برای اطمینان از یکنواختی و تسهیل در بهینه‌سازی.

۲. تحلیل اکتشافی داده‌ها (EDA)

- تحلیل آماری: محاسبه آمار توصیفی برای هر ویژگی به‌منظور درک گرایش‌های مرکزی و پراکندگی‌ها.
- بصری‌سازی: تولید نمودارهای پیشرفته (مانند نمودارهای جفتی، نقشه‌های حرارتی همبستگی) برای بررسی روابط بین متغیرها.

۳. کاهش ابعاد

- تحلیل مؤلفه‌های اصلی (PCA): کاهش ابعاد ویژگی‌ها در حالی که واریانس قابل توجهی حفظ می‌شود.
- بصری‌سازی: ترسیم مؤلفه‌های اصلی برای مشاهده توزیع داده‌ها و خوشه‌های احتمالی.

۴. مدل‌سازی بهینه‌سازی محدب

- فرمول‌بندی: توسعه یک مسئله بهینه‌سازی محدب برای پیش‌بینی نمرات شادی.
- تنظیم منظم‌سازی: استفاده از تکنیک‌های منظم‌سازی (مانند لاسو یا ریج) برای جلوگیری از بیش‌برازش و بهبود تعمیم‌پذیری مدل.
- پیاده‌سازی: حل مسئله بهینه‌سازی با استفاده از روش‌هایی مانند گرادینان نزولی یا الگوریتم‌های نقطه داخلی.
- اعتبارسنجی: ارزیابی عملکرد مدل با استفاده از معیارهایی مانند خطای میانگین مربعات (MSE) و ضریب تعیین (R^2).

۵. گرادینان نزولی تصادفی (SGD)

- پیاده‌سازی: اعمال SGD برای بهینه‌سازی مدل محدب، با آزمایش نرخ‌های یادگیری و اندازه‌های دسته مختلف.
- مقایسه: تحلیل نرخ‌های همگرایی و عملکرد در مقایسه با روش‌های گرادینان نزولی سنتی.

۶. دوگانگی و تحلیل حساسیت

- مسئله دوگان لاگرانژ: فرمول‌بندی مسئله دوگان برای درک عمیق‌تر.
- تحلیل حساسیت: بررسی تأثیر تغییرات در داده‌ها یا محدودیت‌ها بر روی حل بهینه.

بخش اختیاری: خوشه‌بندی و تحلیل

- خوشه‌بندی: گروه‌بندی کشورها بر اساس شاخص‌های اجتماعی-اقتصادی برای شناسایی الگوها با استفاده از روش‌های خوشه‌بندی.
- ارزیابی: ارزیابی عملکرد خوشه‌بندی با استفاده از معیارهایی مانند امتیاز سیلوئت.
- بصری‌سازی: ترسیم خوشه‌ها برای تفسیر گروه‌بندی در زمینه نمرات شادی.
- اهمیت ویژگی‌ها: استفاده از روش‌هایی مانند SHAP یا LIME برای توضیح اینکه کدام ویژگی‌ها در تشکیل خوشه‌ها مؤثر بوده‌اند.

