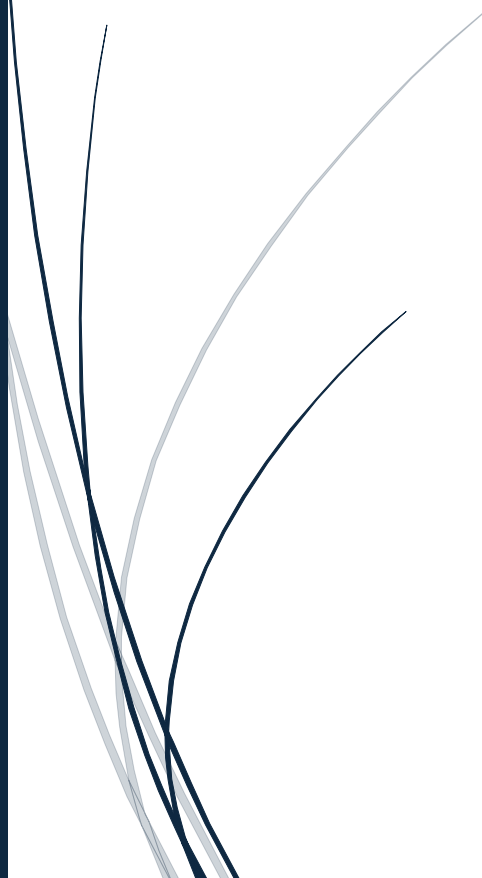


18/03/1404

تحلیل صفات شخصیتی و مصورسازی تعاملی در bokeh



این پروژه با هدف تمرین و تسلط بر تحلیل داده‌های متنی و مصور سازی تعاملی در پایتون انجام شده است . داده شامل نمونه هایی از تیپ های شخصیتی و متون منتشر شده توسط کاربران هستند. در این پروژه با استفاده از کتابخانه bokeh نمودار های تعاملی طراحی شده اند که به کاربران امکان می‌دهند الگو های رفتاری تیپ های مختلف شخصیتی را بررسی کنند.

این داشبورد شامل نمودار نمودار دوناتی برای توزیع انواع شخصیت ها و هیستوگرام تعاملی برای تحلیل ویژگی نوشتاری است. همچنین قابلیت فیلتر گذاری با اسلایدر و چک باکس نیز پیاده سازی شده است.

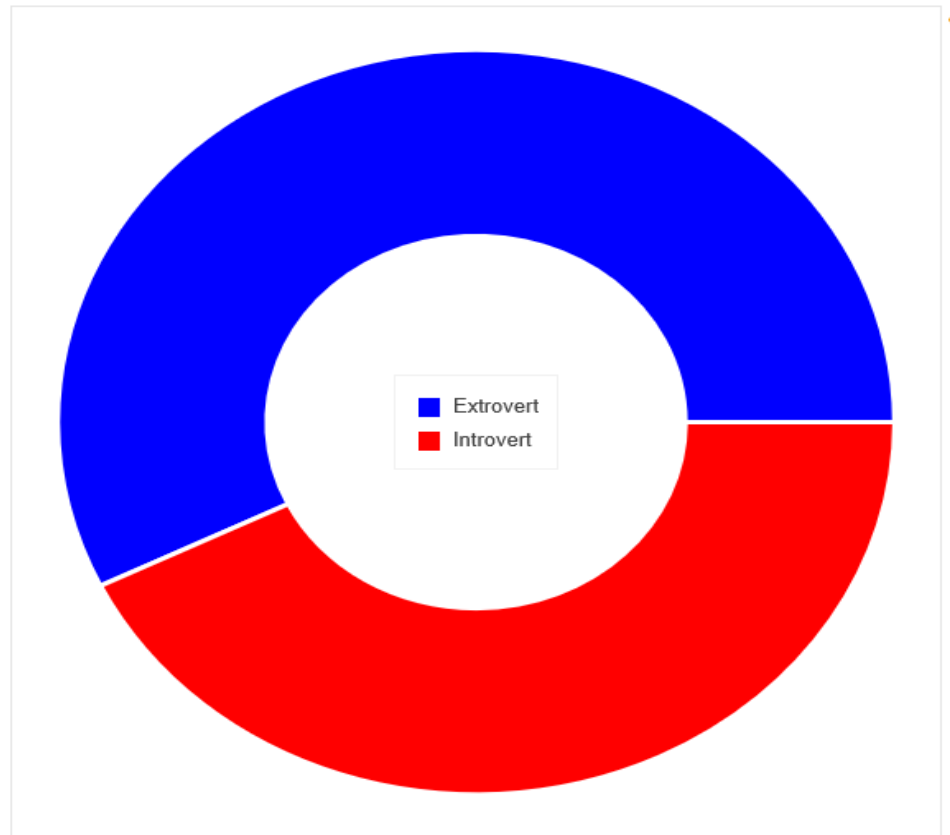
این دیتاست شامل 3000 نمونه از نوع شخصیت کاربران و رفتارهای اجتماعی آن ها بود که پس از حذف داده های ناقص به 2098 نمونه کاهش یافت. ستون هایی که در فایل دیتاست قرار دارد را مورد بررسی قرار می دهیم.

هدف نهایی پروژه پیاده سازی یک داشبورد ساده قابل فهم و توسعه پذیر برای تمرین در میر یادگیری علم داده بوده است.

نحوه نصب کتابخانه bokeh

```
“pip install bokeh”
```

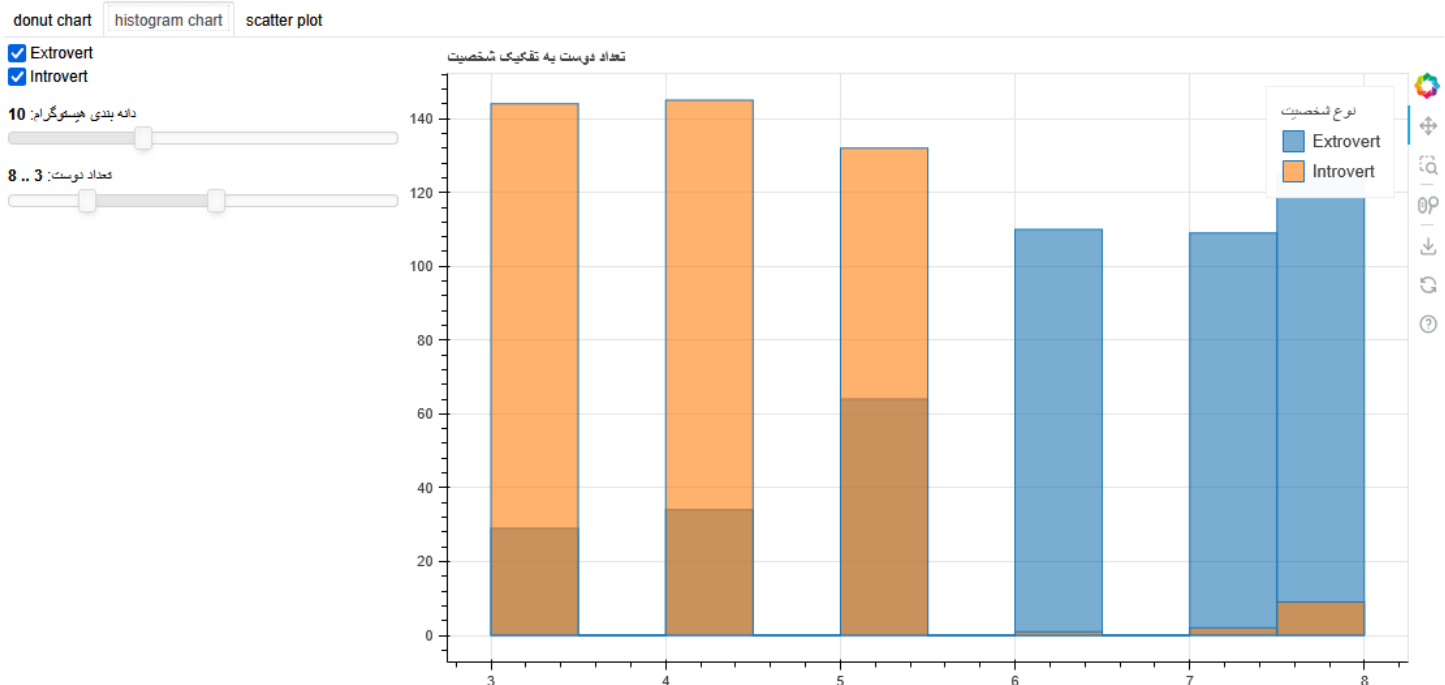
پروژه در آینده میتواند با افزودن تحلیل های عمیق تری مانندی بررسی احساسات متون یا کلمات کلیدی مرتبط با تیپ ها توسعه یابد.



یکی از بخش‌های اصلی این پروژه، طراحی یک نمودار دوناتی برای نمایش توزیع انواع تیپ‌های در دیتاست است. این نمودار به‌گونه‌ای طراحی شده که کاربر بتواند با یک نگاه کلی، MBTI شخصیتی درصد نسبی هر تیپ شخصیتی را در میان کل داده‌ها مشاهده کند. استفاده از نمودار دوناتی به‌جای نمودار ستونی یا جدولی، باعث شده که مقایسه بین گروه‌ها ساده‌تر و بصری‌تر صورت بگیرد. این سبک نمایش سروکار دارند، بسیار رایج و کاربردی است (Categorical) در پروژه‌هایی که با داده‌های دسته‌ای

علاوه بر نمایش کلی، این نمودار با یک اسلایدر تعاملی ترکیب شده که به کاربر اجازه می‌دهد تیپ‌های شخصیتی را بر اساس تعداد دوستان افراد فیلتر کند. این ویژگی باعث می‌شود بینش دقیق‌تری نسبت به ارتباط ویژگی‌های اجتماعی (مثل تعداد ارتباطات فردی) با نوع شخصیت آن‌ها به دست آید. برای مثال، در صورت فیلتر کردن کاربران با تعداد زمانی که تمایل دارند در اجتماع حضور داشته باشند، درصد تیپ‌های درون‌گرا افزایش پیدا می‌کند، که می‌تواند بازتابی از رفتارهای اجتماعی این دسته از شخصیت‌ها باشد.

طراحی شده و به‌صورت تعاملی روی مرورگر اجرا Bokeh این نمودار با استفاده از کتابخانه‌ی می‌شود. طراحی آن به‌گونه‌ای انجام شده که توسعه‌پذیر بوده و می‌توان در مراحل بعدی ویژگی‌های تحلیلی یا فیلترهای بیشتری به آن اضافه کرد. به‌کارگیری مصورسازی تعاملی در این بخش نه‌تنها تجربه‌ی کاربر را بهبود می‌بخشد، بلکه دید بهتری از روندهای موجود در داده‌ها ارائه می‌دهد.



در بخش هیستوگرام تعاملی این پروژه، توزیع یکی از ویژگی‌های عددی داده‌ها (در این پروژه، تعداد دوستان هر کاربر) به تصویر کشیده شده است. هدف از این نمودار، تحلیل دقیق‌تر و کمی پراکندگی این ویژگی و مقایسه آن بین تیپ‌های شخصیتی درون‌گرا و برون‌گراست. برای این منظور، هیستوگرام با از طریق یک اسلایدر پیاده‌سازی شده است تا کاربر بتواند با تغییر دقت (Bins) «امکان تنظیم تعداد «بِن دسته‌بندی، الگوهای توزیع را در سطوح تفکیکی متفاوت مشاهده کند.

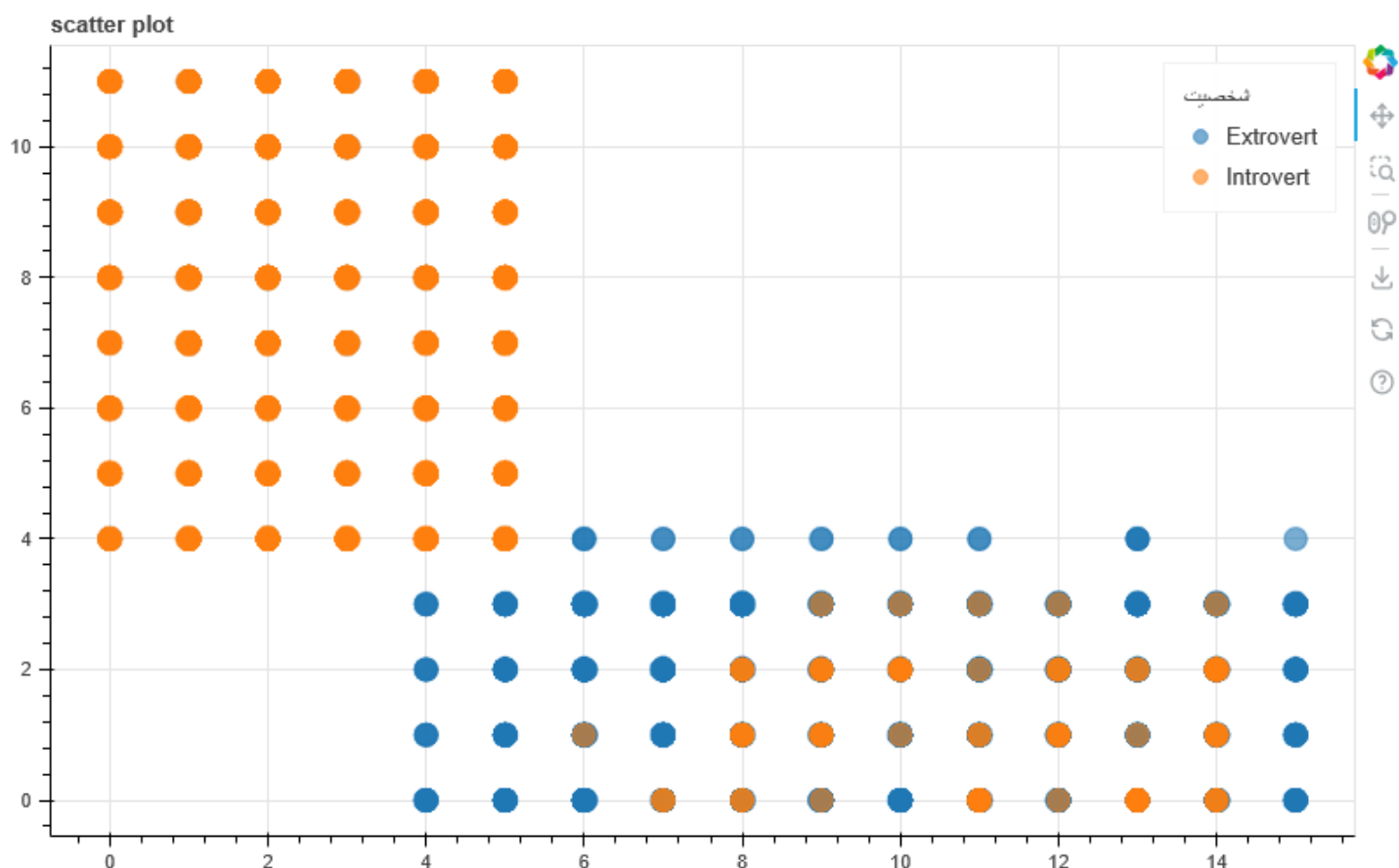
علاوه بر اسلایدر تنظیم بن، دو چک‌باکس تعاملی برای فیلتر تیپ‌های «درون‌گرا» و «برون‌گرا» در نمودار قرار داده شده است. این امکان به کاربر اجازه می‌دهد توزیع تعداد دوستان را به تفکیک هر گروه شخصیتی بررسی کند. به عنوان مثال، با فعال کردن فقط چک‌باکس «درون‌گرا»، می‌توان مشاهده کرد که اغلب افراد درون‌گرا تعداد دوستان کمتری دارند و توزیع آن‌ها متمرکز در بازه‌های پایین‌تر است. در مقابل، فعال‌سازی صرفاً «برون‌گرا»ها، پراکندگی وسیع‌تر و میانگین بالاتری در شمار دوستان نشان می‌دهد.

مانند Bokeh هیستوگرام تعاملی با استفاده از قابلیت‌های >

figure.hist و CheckboxGroup/Slider

مدیریت می‌شوند. داده‌های هر گروه شخصیتی به صورت مجزا از طریق ColumnDataSource

طراحی میشوند. این طراحی امکان افزودن فیلترهای بیشتر (مثلاً سن یا جنسیت) را فراهم می‌کند و می‌تواند در آینده برای تحلیل‌های پیشرفته‌تر گسترش یابد. این بخش از پروژه، نمونه‌ای از ترکیب مصورسازی کمی و تعاملی است که به کاربر کمک می‌کند به صورت فعال در فرآیند تحلیل داده مشارکت کند و بینش‌های عمیق‌تری از داده‌ها به دست آورد.



نمودار پراکندگی طراحی شده در این پروژه با هدف بررسی همبستگی میان دو ویژگی عددی از داده‌ها به ترتیب دو ویژگی انتخاب شده از (Y) و محور عمودی (X) ایجاد شده است. در نسخه فعلی، محور افقی دیتاست را نمایش می‌دهند — برای مثال، تعداد دوستان در برابر طول متن یا سایر ویژگی‌های قابل استخراج. هر نقطه نمایانگر یک کاربر در دیتاست است و موقعیت آن بر اساس مقادیر ویژگی‌ها مشخص شده است.

این نوع نمودار برای تحلیل رابطه‌ی احتمالی بین دو متغیر بسیار کاربردی است. اگر الگوی نقاط به صورت صعودی یا نزولی مشخصی ظاهر شود، می‌توان فرض کرد که بین آن دو متغیر رابطه‌ای مستقیم یا معکوس وجود دارد. در صورت پراکندگی کامل و تصادفی نقاط، معمولاً نشان‌دهنده‌ی نبود رابطه‌ی مشخص بین متغیرهاست.

در مراحل بعدی می‌توان به این نمودار قابلیت‌هایی مانند انتخاب بازه‌ی خاص، فیلتر بر اساس تیپ Scatter شخصیتی یا افزودن رنگ‌بندی بر اساس گروه‌ها اضافه کرد تا تحلیل‌ها دقیق‌تر شوند. استفاده از Plot در این پروژه، قدمی مهم در جهت تحلیل‌های آماری و تصویری دقیق‌تر داده‌هاست.