محمد دهقانی

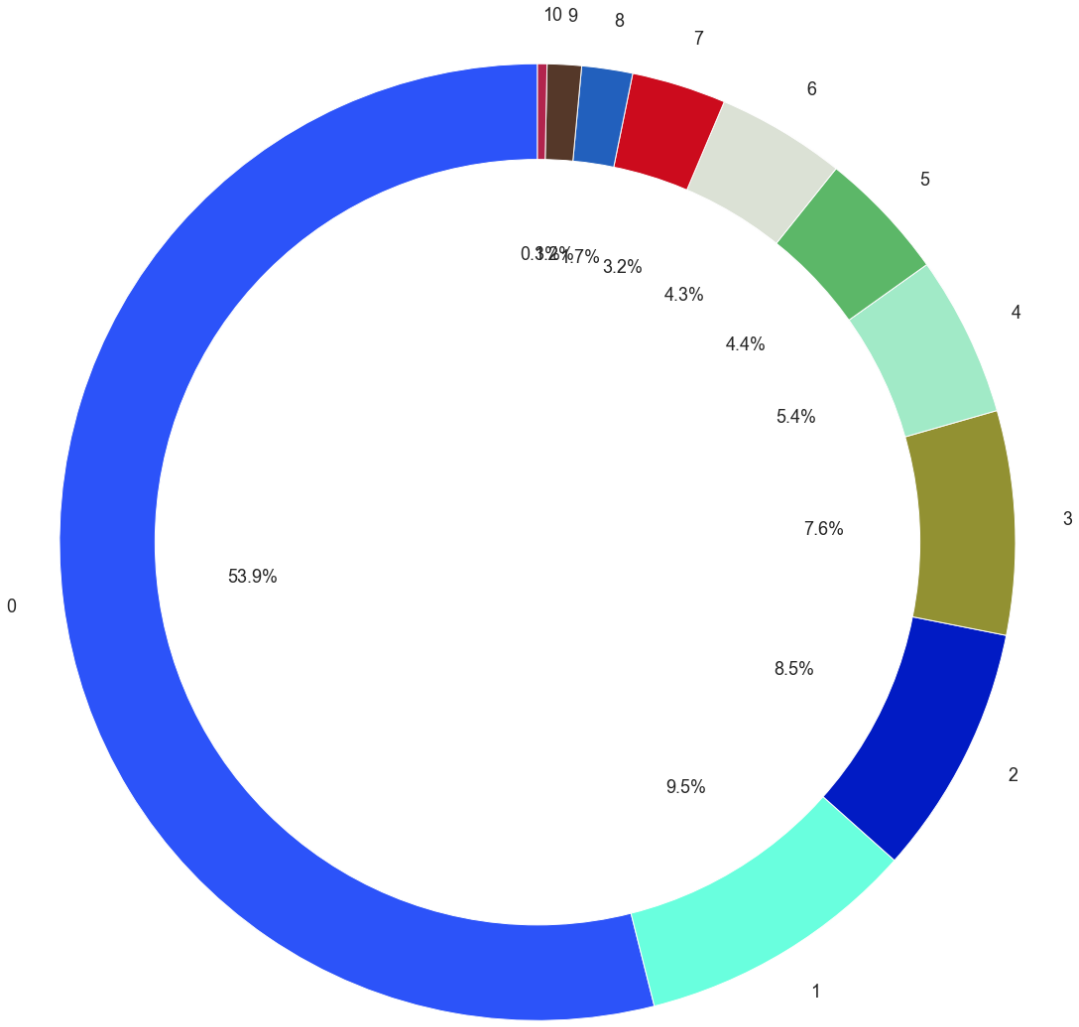
@data_hub_ir

در این پروژه از داده های مربوط به فایل top_views_13981022-13981012 استفاده شد. این فایل مربوط به پیام های پربازدید در کانال های تلگرام در بازه 12 تا 22 دیماه 98 است. در این بازه شاهد اتفاقات مثل ترور سردار سلیمانی، حمله به پایگاه عین الاسد و ماجرا هواپیمای اوکراین بودیم. در زمینه ورزشی خبر داغ مربوط به حواشی سرمربی استقلال، استراماچونی بود.

از ویژگی های بارز داده های تلگرامی، کثیف بودن داده هاست. از طرفی این داده ها ظاهرا به صورت تصادفی انتخاب نشده و به شدت بایاس به موضوعات خاصی هستند.

در مرحله اول به نظر رسید دسته بندی موضوعی انجام دهیم. این دسته بندی فقط مبتنی بر متن پیام ها بوده و از مدل های خوشه بندی استفاده شد.

پس از پیش پردازش های متنوع به یک دیتاست به اندازه 6575 سطر رسیدیم. سپس پس از بررسی روش های مختلف، داده به صورت زیر خوشه بندی موضوعی شدند.



در شکل زیر ایندکس مربوط به هر موضوع مشخص شده است.

0 : سیاست خارجی ، ترور و وقایع پیرامونی
1 : سیاست خارجی
2 : هواپیمای اوکراینی و وقایع پیرامونی
3 : ورزشی عمومی
4 : انتقام سخت ، سیاست خارجی ، ترور و وقایع پیرامونی
5 : حوادث طبیعی
6 : ورزشی ، استقلال
7 : ورزشی ، پرسپولیس
8 : حمله به سفارت امریکا ، سیاست خارجی ، ترور و وقایع پیرامونی
9 : موسیقی ، تفریح ، تبلیغ
10 : پزشکی

با توجه به اینکه در بازه ذکر شده مهم ترین تحولات از جنس سیاسی بوده و از طرفی کرال دیتاها بیشتر از کانال های خبری صورت گرفته، پس بیشترین فراوانی مربوط به پیام های با موضوعات سیاسی است.

در شکل های زیر بعضی از ابرکلمات خوشه ها قابل مشاهده است.

ابر کلمات مربوط به خوشه (ورزشی عمومی)

ابر کلمات مربوط به خوشه (انتقام سخت ، سیاست خارجی ، ترور و وقایع پیرامونی)

قابل تصور بود که کلماتی مثل پایگاه، عین الاسد، فرودگاه فراوانی بالایی داشته باشند

ابر کلمات مربوط به خوشه (پزشکی)

ابر کلمات مربوط به خوشه (حوادث طبیعی)

این خوشه مربوط به بلایای طبیعی بوده و طبیعتا اسامی شهرها و استان ها در آن بیشتر به چشم می خورد.

ابر کلمات مربوط به خوشه (هواپیمای اوکراینی و وقایع پیرامونی)

ابر کلمات مربوط به خوشه (ورزشی ، استقلال)

ابر کلمات مربوط به خوشه (ورزشی ، پرسپولیس)

ابر کلمات مربوط به خوشه (حمله به سفارت امریکا ، سیاست خارجی ، ترور و وقایع پیرامونی)

این خوشه مربوط به وقایع حمله به سفارت امریکا در عراق است

ابر کلمات مربوط به خوشه (سیاست خارجی)

ابر کلمات مربوط به خوشه (سیاست خارجی ، ترور و وقایع پیرامونی)