**کسری فرخی**

1. من کلا سوال اول را با این دید نگاه کردم که در حین حل بقیه سوالات اگر به مشکلی خوردم اعلام کنم

* اولین مشکل را در خط 26 کامنت کردم. طبق گفته شما ستون source\_event\_id باید شناسه کوئری کلیک یا پست را نمایش دهد اما بعضی کلیک­ها هستند که کوئری پست متناظر آن­ها مشخص نشده. چون تعداد آیتم­های کلیک شده متمایز بیشتر از تعداد پست­ها است. یعنی کلیکی صورت گرفته بدون آن که پست شده باشد که کمی عجیب است. در چاپ خط 40 هم به این مسئله اشاره کرده­ام البته ممکن است فرد لینک را نگه داشته باشد و یا به کس دیگری داده باشد(البته نمی­دانم این حالت را هم حساب می­کنید یا خیر).
* بعد از پرینت load\_post\_df در خط 32 من متوجه شدم که بعضی کوئری ها device\_id برابر NaN دارند که مسئله عجیبی از نظر من بود.

1. دلایل به صورت زیر است:

* میانگین فاصله بین رتبه کلیک‌های یک کاربر (مثلا اگر کاربر از بین نتایج کوئری روی آگهی‌های دهم و شانزدهم کلیک کند، این متریک برابر ۲/(۶+۱۰) یعنی ۸ خواهد بود): این معیار شاید دومین معیار خوب باشد اما اگر کاربر فقط روی اولین اگهی کلیک کند انوقت این مقدار 1 میشود و اگر روی تعداد زیادی اگهی پشت هم نیز کلیک کند عدد بیشتر یا کمتر نمی­شود. منظور من این است که به تعداد آگهی­های کلیک شده حساس نیست.
* اینکه آیا روی یکی از ۳ نتیجه اول کوئری کلیک شده یا نه: باز هم چون آگهی­ها بر اساس زمان پست شدن نمایش داده می­شوند از نظر من معیار خوبی نیست. چون احتمال اینکه سه تای اول خیلی به سلیقه کاربر نخورد ولی بقیه مطابق سلیقه او باشد زیاد است. همین دلیل آمدن پست­ها روی یک کوئری بر اساس زمان برای " رتبه اولین کلیک کاربر" نیز صادق است چون زمان تعیین کننده بالا بودن یک آگهی است پس معیار خوبی نمی­تواند باشد.
* در نتیجه انتخاب من "درصد آگهی‌های کلیک شده نسبت به آگهی‌های لود شده" است

1. فرمول توزیع برنولی به صورت زیر است:

این فرمول در واقع مدلسازی احتمال وقوع و عدم وقوع یک اتفاق است که متغیر x وقوع و یا عدم وقوع این احتمال را نشان می­دهد و p احتمال متناظر آن است. یعنی اگر این اتفاق بیوفتد مقدار p(x) برابر p و در غیر این صورت 1-p می­شود. حالا در پایین توضیح خواهم داد که چرا درصد آگهی‌های کلیک شده نسبت به آگهی‌های لود شده با این توزیع می­تواند مدلی از بقیه پارامترها به ما بدهد:

رتبه اولین کلیک کاربر:

طبیعتا به صورت میانگین احتمال کلیک شدن هر پست p است پس برای اینکه رتبه کلیک شدن پست iام را حساب کنیم داریم:

اینکه آیا روی یکی از ۳ نتیجه اول کوئری کلیک شده یا نه:

خب با توجه به فرمول بالا اگر فرمول را برای مقادیر i در حالتی که برابر 1 و 2 و 3 است حساب کرده و با هم جمع کنیم عملا احتمال وقوع را بدست خواهیم آورد.

میانگین فاصله بین رتبه کلیک‌های یک کاربر (مثلا اگر کاربر از بین نتایج کوئری روی آگهی‌های دهم و شانزدهم کلیک کند، این متریک برابر ۲/(۶+۱۰) یعنی ۸ خواهد بود):

حال چون احتمال وقوع هر کلیک را به صورت داریم با محاسبه امید ریاضی می­توانیم میانگین را هم به صورت احتمالاتی حساب کنیم.