

گزارش پروژه پایان دوره (شماره یک) : واکسیناسیون کرونا اردلان دانشور طریق

توضیحات کلی اولیه:

قبل از این که دست به کد بشیم بهتره صورت سوال رو بهتر درک کنیم تا به نتیجه کاربردی برسیم. در حال حاضر هنوز پاندمی کرونا در جهان وجود داره و روند واکسیناسیون کماکان ادامه داره.

برای انجام این پروژه دو هدف وجود دارد:
هدف اول این است که اگه کشور مورد نظر با زیر ساخت و امکانات و عوامل بهداشت خود واکسیناسیون را ادامه دهد چه زمانی همه ی مردم کامل واکسینه می شوند؟
هدف دوم این است که آیا راهی وجود دارد که روند واکسیناسیون را سرعت ببخشند؟ (با تحلیل داده های کشور ها و شرکت های ساخت واکسن)

قدم اول:

مثل همیشه اولین قدم ایمپورت کردن پکیج های مورد استفاده در برنامه هست.

قدم دوم:

برای حل مساله ابتدا هر دو فایل csv را داخل برنامه ایمپورت می کنیم.

قدم سوم:

از آن جایی که در صورت سوال دو دیتاست متفاوت داشتیم به صورت موازی بر روی هر دو دیتاست کار می کنیم. کار را با دیتاست کشورها شروع می کنیم. برای این که یک دید جامع و ساده از دیتاست داشته باشیم ابتدا ۵ مورد از نمونه های داخل آن را به صورت تصادفی نمایش می دهیم و چک می کنیم. در همین مرحله با این فرضیه که داده ها از منبع موثق به دست آمده اند (سازمان بهداشت جهانی) ستون های مربوط را حذف می کنیم.

درون دیتاست کشورهای زیادی وجود دارند که طبق صورت سوال یکی از کشورها را باید به انتخاب خودمان بررسی کنیم. معیار خاصی برای انتخاب کشور وجود ندارد به همین دلیل کشوری رو انتخاب می کنم که بیشترین داده آماری را درون دیتاست داشته باشد (بیشترین تعداد instance). بیشترین نمونه را کشور نروژ دارد. تمامی داده های کشور نروژ را در داخل دیتافریم دیتاست ذخیره می کنیم و برای چک کردن درست بودن داده ها داده های سه روز اخیر را مشاهده می کنیم.

تا این جا دیتاست اولیه خودمون رو به دست آوردیم و حالا نیاز است که داده های null را ور کنیم. ۵ تا از ستون ها دارای داده های null هستند که نمی توانیم بدون تحلیل درست آن ها را حذف کنیم و یا برای پر کردن آن ها تصمیم بگیریم. تا این جا اطلاعات و مراحل اولیه را برای دیتاست اول به پایان رساندیم. الا نوبت دیتاست دوم است که روند واکسیناسیون را بر اساس شرکت های سازنده نشان می دهد.

در مقایسه با دیتاست قبلی دیتاست دوم کوچکتر و ساده تر به نظر می رسد که هم ستون های کمتری دارد هم instance های کمتری دارد. خوشبختانه هیچ داده null در این دیتاست وجود ندارد.

از این دیتاست می توانیم اطلاعات تمامی شرکت ها را استخراج کنیم. لیست تمام شرکت های سازنده واکسن را به دست می آوریم. بر اسا این لیست می توانیم تصمیم های احتمالی سیاست مداران را بررسی کنیم.

قدم چهارم:

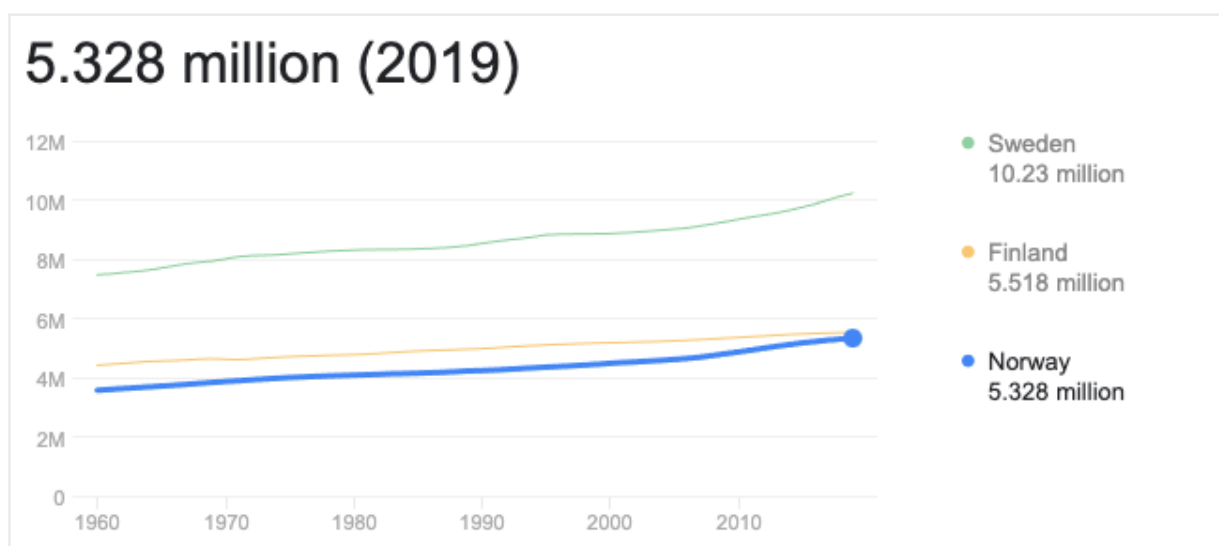
همان طور که به خاطر داریم در دیتاست اول داده null داشتیم که نیاز است آن ها را بررسی کنیم. ستون اول ستون تمامی کسانی است که واکسینه شده اند. از آن جایی که دیتاست به صورت خطی در زمان به جلو حرکت می کند و تعداد افراد واکسینه شده امکان ندارد کاهش یابد پس می توانیم داده های خالی را با داده روز قبل و یا تفاضل بین دو روز ور کنیم. (مانند فراوانی تجمعی در آمار). سپس ستون ساخته شده را به دیتاست بازمی گردانیم.

ستون افراد واکسینه شده در هر صد نفر را حذف می کنیم چون اگر نیاز داشته باشیم قابل محاسبه است. در سایت کگل چک کردم در توضیحات دیتاست دو تا از ستون مثل هم هستند و یکی از آن ها را حذف می کنم. ستون مهم ما ستون تعداد واکسینه های در طول روز می باشد. از آن جایی که زیر ساخت بهداشتی کشور یک شبه تغییر نمی کند می توان در نظر گرفت که میانگین تعداد واکسن زده شده تمام روز ها می تواند معیار خوبی باشد و داده های خالی را پر می کنیم. مانند ستون های

واکسینه در صد در میلیون را نیز حذف می کنیم.
مشاهده می کنیم که تمامی داده های خالی ور شدند.
برای مشاهده بهتر و رسم منطقی تر داده ها اندیس دیتاست را
بر اساس زمان در نظر می گیریم. (اجباری نیست ولی به درک
بهتر داده ها خیلی کمک می کنه)

قدم پنجم:

ما هیچ داده ای برای کیفیت واکسن نداریم به همین دلیل
عملکرد همه ی آن ها را مساوی در نظر می گیریم. هدف ما این
اس که تا جایی که امکان دارد افراد بیشتری را واکسینه کنیم
پس باید بدانیم برای سرعت بخشیدن به روند واکسیناسیون از
چه شرکت هایی می توانیم واکسن بخریم؟ شرکت های فایزر و
مدرنا حدود ۸۰ درصد از واکسن های دنیا را تولید می کنند. بعد
مدرنا شرکت بعدی فقط توان تولید ۴ درصد را دارد. داده های
به دست آمده را بر روی نمودار رسم می کنیم و به نظر می آید
شاید تنها شرکت با پتانسیل بالا شرکت فایزر است.
با توجه به جمعیت کشور نروژ حدود ۲ میلیون نفر باقی مانده اند
که بر روی نمودار رسم می کنیم.



در مرحله بعدی روند کلی واکسیناسیون مردم را روی نمودار رسم می کنیم که مشاهده می شود به جز یک مدت بسیار کوتاه روند کلی واکسیناسیون به سرعت در حال پیشرفت است. در عین حال مهم است تعداد واکسن های زده شده در یک روز هم در قالب زمان مشاهده کنیم. واقعا کشور نروژ عالی عمل کرده و با یک روند صعودی تعداد واکسیناسیون روزانه را افزایش داده است.

با توجه با داده های به دست آمده کشور با همین روند پیش برود همه ی مردم خود را در مدت زمان کوتاهی واکسینه کرده است و نیازی به خرید بیشتر واکسن و یا ... ندارد . همان روند خود را ادامه دهند کارشان عالی پیش می رود.

قدم ششم:

برای ساخت مدل می خواهیم از pycaret استفاده کنیم به این معنی که نیازی به اسکیل کردن داده ها گرید سرچ انکود کردن و ... وجود ندارد. این کتابخانه خود همه ی کارها را انجام می دهد. فقط داده های آموزش و تست رو بر اساس زمان جدا می کنیم.

قدم هفتم:

مدل خود را بر اساس رگرسیون آموزش می دهیم چوه هدف پیش بینی تعداد واکسن هست.

قدم هشتم:

نتیجه مدل در قدم قبل را مشاهده می کنیم و بهترین الگوریتم را از آن ها برگزینیم.

ابتدا نتایج اولین الگوریتم را روی نمودار رسم می کنیم با در نظر گرفتن این که اوور فیت شده است مدل آموزشی بعدی را نیز بر روی نمودار رسم می کنیم. با توجه به اختلاف بالا میان آن دو مدل اول را انتخاب می کنیم و مدل آماده را در یک فایل pickle ذخیره می کنیم. فایل ذخیره شده در فالی زیپ قرار داده شده است.

با تشکر از وقتتون
ازتون ممنونم همیشه برای کارهام نظر بسیار مثبت دارید
برام باعث دلگرمی هستش استاد
ممنون از لطفتون