

## گزارش پروژه‌ی تحلیل احساسات

تحلیل احساسات روی توییت‌های مردم در بازه‌ی یک ماهه تا وقوع Brexit

دیتا:

۷۰۰ هزار توییت در بازه‌ی یک ساعته از یک ماه قبل از برگزیت.

نکته: توییت‌ها تمام در مورد این موضوع و اروپا می‌باشد.

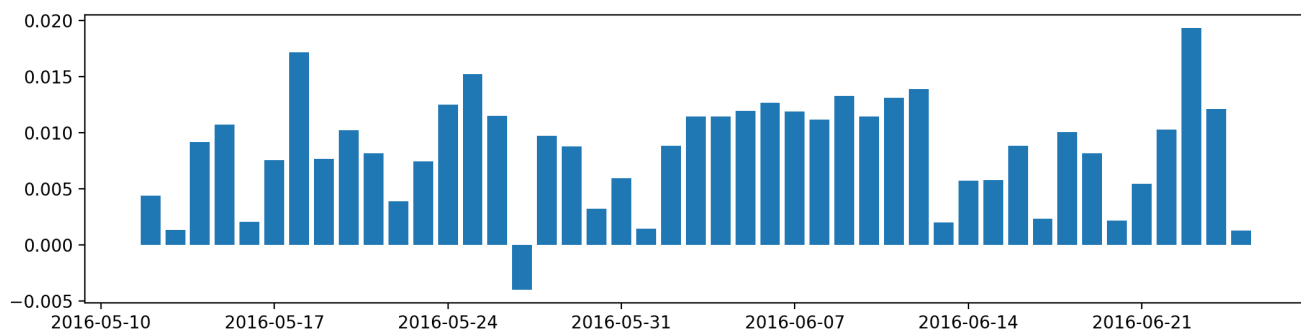
فقط توییت‌ها به زبان انگلیسی هستند.

در ابتدا از کل دیتاست که حاوی ۵ میلیون توییت‌آیدی بود، ۱ میلیون سмпل انتخاب شد. (سمپلینگ به صورت یکنواخت انجام می‌شود)

سپس از توییت‌آیدی‌های جمع شده متن توییت‌های آن‌ها استخراج شد که تعدادی به خاطر پاک شدن توییت و غیره از دست رفتند به تعداد به ۷۰۰ هزار رسید.

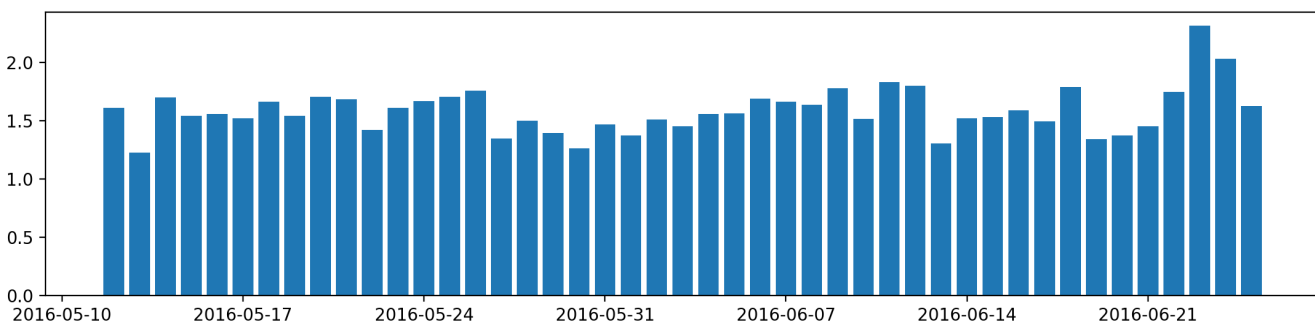
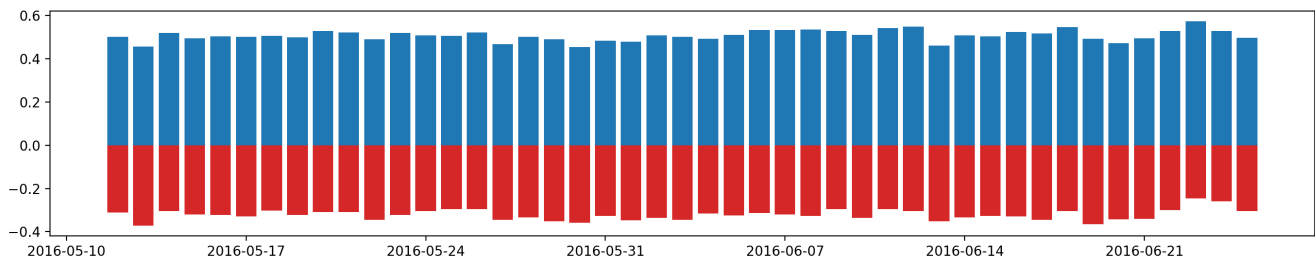
نتایج:

هر کدام میانگین احساسات در روز است. عددهای مثبت احساسات مثبت هستند و عدد های منفی احساسات منفی.

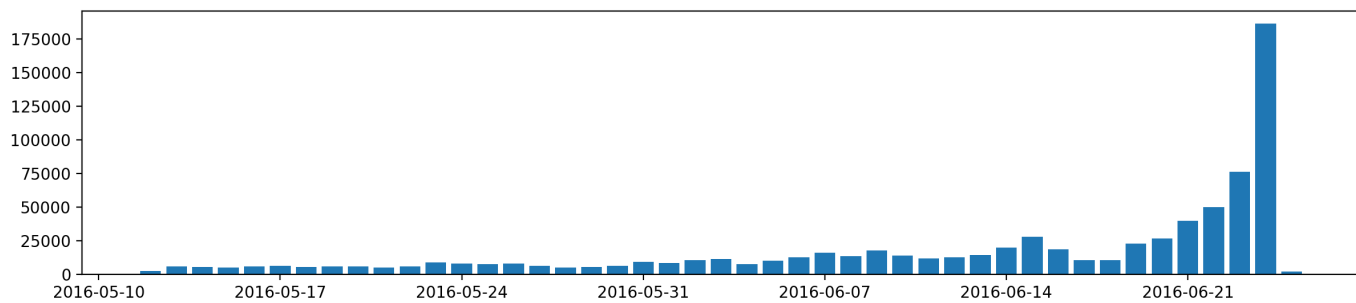


همان‌طور که دیده می‌شود میزان مثبت بودن احساسات به مرور زمان تا برگزیت به صورت نسبی بالا می‌رود.

### نسبت مثبت به منفی:



### تعداد توییت‌ها:



### **دقت:**

دقت با استفاده از لیبیل‌زدن تعدادی سمپل، به دست آمده‌است.

precision	recall
70.58	57.14

## مراحل کار:

بعد از جمع‌آوری متن توییت‌ها نوبت به تمیزکاری دی‌تا است. در این مرحله با استفاده از رجکس تمام منشن‌ها و انتیتی‌های توییت‌ر (مانند : RT) حذف می‌کنیم. سپس هت‌شگ‌ها را به کلمه تبدیل می‌کنیم.

برا به دست‌آوردن عدد احساسات هر توییت، دو روش می‌توان انجام داد. روش اول استفاده از sentiwordnet است. در این روش ابتدا توییت‌را توکنایز می‌کنیم و سپس پس از تگ‌زدن POS آن، به بررسی وجود هر کلمه در این دیتاست می‌پردازیم. برای هر کلمه اعداد به دست آمده را به مجموع اضافه می‌کنیم و در نهایت با کم کردن مجموع مثبت از مجموع منفی، عدد احساسات این توییت. را به دست می‌آوریم.

در روش دوم، مثبت و منفی بودن عبارات را می‌خواهیم. به این صورت که ابتدا عبارات دو کلمه‌ای را از تمام توییت‌ها استخراج می‌کنیم (انتخاب دو کلمه‌ی کنار هم به عنوان عبارت با توجه به برچسب pos آن انجام می‌شود). حال نوبت به آن است که ببینیم این عبارات مثبت هستند یا منفی، برا این کار، تعداد دفعاتی که این عبارات در یک توییتی که در آن کلمات مثبت از دیتاست sentiwordnet محاسبه می‌کنیم و هم‌چنین همین کار را هم برای کلمات منفی تکرار می‌کنیم. سپس با توجه به فرمول PMI مثبت یا منفی بودن آن حساب می‌شود.