**بسم الله الرحمن الرحیم**

**آنالیز علامت ها و ایموجی**

**برای محصولات دیجیکالا**

تیم ارائه دهنده: داده کاوانِ فونیکس

ناظر و سرپرست تیم: دکتر علیرضا وفایی صدر

IPMدرخواست دهنده: ستاد علوم شناختی

زمستان 1398

تعریف پروژه

در این گزارش، به بررسی تعداد علامت ها و سمبل های شناختی و تعداد ایموجی ها در داده های دیجیکالا پرداختیم. این کار توسط *آرمیتا رضوی* بر روی داده های اصلی و پیش پردازش نشده و توسط *خانم ستاره* ترنج بر روی داده های پاکسازی شده انجام گرفت.

بررسی ایموجی و سمبل ها

در ستون های کامنت، عنوان، مزایا و معایب محصول، شروع به پشتن علامت های مختلف و ایموجی کردیم و نتایج به صورت زیر حاصل شد:

برای داده های اصلی و پردازش نشده دیجیکالا با دیتاست Digi\_2.xlsx :

علامت های مورد بررسی:

؟ ? $ & # % ! و ایموجی ها

نمودارهای زیر نشان دهنده تعداد کامنت های موجود که شامل علائم مختلف می باشند است:

نمودارهای زیر نشان دهنده تعداد عنوان کامنت موجود که شامل علائم مختلف می باشند است:

همینطور برای ستون های مزایا و معایب نیز این پلات ها کشیده شده است. جدول زیر تعداد داده های در ستون مختلف که شامل علامت ها و ایموجی هستند را نشان می دهد:

فایل این آنالیز در مسیر زیر قرار دارد:

Report2->Armita Razavi -> Find\_Emojies-Symbols-ipynb

همچنین تمامی عکس ها وپلات ها در فولدری در مسیر زیر موجود است:

Report2->Armita Razavi -> Symbols\_Not\_Clean

همانطور که مشاهده می شود بیشتر سطرها در ستون های مختلف علامت تعجب و کترین علامت مربوط به علامت $ می باشد. با توجه به آمار، کمتر از 10% از داده های و سطرها شامل علایم شناختی و ایموجی ها هستند که اصلا به چشم نمیاد.

در ستون های کامنت، عنوان، مزایا و معایب محصول، شروع به پشتن علامت های مختلف و ایموجی کردیم و نتایج به صورت زیر حاصل شد:

برای داده های پاکسازی شده ولی با حفظ علائم و ایموجی ها دیجیکالا با دیتاست digi\_clean\_final.xlsx :

علامت های مورد بررسی:

؟ ? $ & # % ! و ایموجی ها

نمودارهای زیر نشان دهنده تعداد کامنت های موجود که شامل علائم مختلف می باشند است:

. جدول زیر تعداد داده های در ستون مختلف که شامل علامت ها و ایموجی هستند را بعد از پردازش داده ها نشان می دهد:

فایل و پلات های ابن بخش در مسیر Report2->Setare Toranj -> Symbol قرار گرفته است.

کلیه فایل های مربوط به این گزارش در گیت هاب زیر و در فولدر Report2 موجود است.

https://github.com/phoenix-dataminers/Digikala2

پایان