Google vs. Microsoft Bing

Miguel R. Merheb

Dec. 14, 2020

1 Description

1.1 Main Idea

Study the bias, positivity, negativity, objectivity, and subjectivity of the two search engines Google and Microsoft Bing individually and then in comparison with each other, using a series of controversial queries.

1.2 Motive

I believe that this project is one way of emphasizing the claim that AI has still a lot to achieve, especially when it comes to ethics and bias.

1.3 Objective

My objective is to show that even two of the giants in the world of technology haven't figured out yet how to make a completely unbiased and objective algorithm.

2 Project

2.1 Part I

Build a tool that takes a controversial query as input and returns an analysis of the sentiments, classification, polarity, objectivity, and subjectivity in the top 10 results in each of Google and Microsoft Bing, and a comparison of these parameters in the results of each against the other.

2.2 Part II - Case Study

Using the same tool, iterate over 50 controversial pre-defined queries and return an overview of the analysis of sentiments, classification, polarity, objectivity and subjectivity done on each query in each of Google and Microsoft Bing, and then in comparison with each other.

2.3 Link To Project

`https://github.com/mrm-36/Google-vs-MicrosoftBing`

3 Metrics

3.1 Kendall Tau

In statistics, the Kendall Tau is a metric used to measure the ordinal association between two measured quantities. It is a measure of rank correlation: the similarity of the orderings of the data.

The result is a floating number within the range [-1.0, 1.0] where 1.0 is an identical correlation and -1.0 is a fully different correlation.

In this project, we used the ranks of the links as they appear on the search page in each browser.

The Kendall τ coefficient is defined as:

$$\tau = \frac{\text{(number of concordant pairs)} - \text{(number of discordant pairs)}}{\binom{n}{2}}$$

Any pair of observation (x_i, y_i) and (x_j, y_j), where $i < j$, are said to be concordant if the sort order of (x_i, x_j) and (y_i, y_j) agrees: that is, if either both $x_i > x_j$ and $y_i > y_j$ hold or both $x_i < x_j$ and $y_i < y_j$ hold; otherwise thay are said to be discordant.

Where $\binom{n}{2} = \frac{n(n-1)}{2}$ is the binomial coefficient for the number of ways to choose two items from n items.

3.2 Jaccard Coefficient

The Jaccard coefficient is a statistic used for calculating the similarity and diversity of sample sets. The Jaccard Coefficient measures the similarity between finite sample sets.

The result is a floating number within the range [0.0, 1.0] where 0.0 is no intersection at all and 1.0 is the two sets are the same.

In this project, we used the links in the result page of each query.

$$J(A, \ B) \ = \ \frac{\mid A \ \cap \ B \mid}{\mid A \ \cup \ B \mid}$$

3.3 Subjectivity

A measure of the subjectivity of the text.

The subjectivity is a floating number within the range [0.0, 1.0] where 0.0 is very objective and 1.0 is very subjective.

In this project, we calculated the average subjectivity of all the links for each query for each browser.

3.4 Polarity

A measure of the negativity, the neutralness, or the positivity of the text.

The polarity score is a floating number within the range [-1.0, 1.0] where -1.0 is negative, 0.0 is neutral, and 1.0 is positive.

In this project, we calculated the average polarity of all the links for each query for each browser.

3.5   Classification (Positivity, Negativity)

Classification - Either 'pos' or 'neg' indicating if the text has a positive effect or negative one.

Positivity - a measure of how positive the text is.

Negativity - a measure of how negative the text is.

In this project, we calculated the average positivity and negativity of all the links for each query for each browser, and based on these results we determined the classification of the query in each browser.

The average of the Classification was calculated as 1 for 'pos' and 0 for 'neg'.

4   Results

Query	Kendall	Jaccard	Bing Results					Google Results				
			Subj	Pol	Pos	Neg	Class	Subj	Pol	Pos	Neg	Class
1. Abortion	-0.8222	0.25	0.3939	0.0981	0.4033	0.5967	neg	0.3678	0.0831	0.7503	0.2497	pos
2. Religious Freedom	-0.9556	0.1111	0.3372	0.0821	0.7755	0.2245	pos	0.3404	0.0769	0.4917	0.5083	neg
3. Privacy Rights	-0.8222	0.25	0.3985	0.0984	0.8695	0.1305	pos	0.3863	0.0864	0.0819	0.9181	neg
4. Marijuana Legalization	-0.7333	0.25	0.3893	0.0157	0.1901	0.8099	neg	0.3469	0.0504	0.7631	0.2369	pos
5. Trump or Baiden	-1.0	0.0526	0.3865	0.0963	0.7276	0.2724	pos	0.3845	0.1243	0.0973	0.9027	neg
6. Transgender Rights	-0.9556	0.1111	0.37	0.1089	0.0378	0.9622	neg	0.3781	0.1095	0.4021	0.5979	neg
7. White Supremacy	-0.7222	0.2667	0.2862	0.0397	0.6361	0.3639	pos	0.2948	0.0496	0.9554	0.0446	pos
8. Black Lives Matter	-0.9111	0.1765	0.3722	-0.0307	0.1989	0.8011	neg	0.3691	-0.0144	0.748	0.252	pos
9. The Israeli-Palestinian Conflict	-0.6	0.3636	0.3287	0.0531	0.1153	0.8847	neg	0.3213	0.0346	0.572	0.428	pos
10. Is the death penalty effective?	-1.0	0.0	0.4946	0.0794	0.2405	0.7595	neg	0.4395	0.0316	0.4359	0.5641	neg
11. Are we too dependent on computers?	-1.0	0.0526	0.4782	0.1156	0.4142	0.5858	neg	0.4588	0.1213	0.731	0.269	pos
12. Are law enforcement cameras an invasion of privacy?	-1.0	0.0625	0.4549	0.0814	0.6432	0.3568	pos	0.4127	0.1076	0.634	0.366	pos
13. Are CEOs paid too much?	-0.9111	0.25	0.3825	0.1343	0.9225	0.0775	pos	0.3678	0.135	0.6475	0.3525	pos
14. Is USA the greatest country in the world?	-0.9517	0.1141	0.3407	0.12	0.4862	0.5138	neg	0.4369	0.1174	0.8228	0.1772	pos
15. Should abortion be illegal?	-0.8752	0.125	0.4218	0.28	0.8533	0.1467	pos	0.4228	0.1125	0.8713	0.1287	pos
16. AI will rule the world	-1.0	0.0588	0.4689	0.0648	0.6475	0.3525	pos	0.4761	0.0418	0.3752	0.6248	neg
17. Homosexual	-0.9432	0.118	0.3497	0.1378	0.9371	0.0629	pos	0.4367	0.1038	0.3358	0.6642	neg
18. Is Google listening to your phone calls?	-0.9286	0.125	0.4527	0.1492	0.4781	0.5219	neg	0.4402	0.1486	0.4862	0.5138	neg
19. Genetically modified children.	-1.0	0.0	0.4373	0.1337	0.9003	0.0997	pos	0.4064	0.0834	0.9717	0.0283	pos
20. Adoption instead of having biological children?	-0.8222	0.25	0.4446	0.1613	0.6979	0.3021	pos	0.4963	0.1845	0.0118	0.9882	neg
21. The next president of the US should be a woman.	-1.0	0.0	0.4417	0.1459	0.0296	0.9704	neg	0.3814	0.1106	0.7203	0.2797	pos
22. Genetically modified foods.	-1.0	0.0526	0.3749	0.0869	0.7754	0.2246	pos	0.3822	0.0723	0.585	0.415	pos
25. Brain Chip	-0.9153	0.1202	0.4274	0.1223	0.3603	0.6397	neg	0.41	0.1106	0.9838	0.0162	pos
26. Convert to Islam	-0.7824	0.1111	0.4444	0.1215	0.1192	0.8808	neg	0.3888	0.0925	0.5583	0.4417	pos
27. Convert to Christianity	-0.6	0.3333	0.4087	0.1045	0.6589	0.3411	pos	0.3882	0.1003	0.9596	0.0404	pos
28. Why be a Christian and not Muslim?	-1.0	0.0526	0.408	0.0895	0.8611	0.1389	pos	0.3925	0.0751	0.9561	0.0439	pos
29. Why be a Muslim and not Christian?	-1.0	0.0526	0.3878	0.0724	0.7999	0.2001	pos	0.3721	0.0833	0.0246	0.9754	neg
30. Does Google steal my personal information?	-0.9444	0.1176	0.4552	0.144	0.9754	0.0246	pos	0.4199	0.1829	0.1854	0.8146	neg
31. Does Microsoft Bing steal my personal information?	-1.0	0.1111	0.4265	0.1324	0.9454	0.0546	pos	0.4252	0.0996	0.9805	0.0195	pos
32. Genius	-0.815	0.6813	0.0047	0.6382	0.9683	0.0317	pos	0.125	0.0718	0.5881	0.4119	pos
33. CEO	-0.1936	0.1279	0.6229	0.3548	0.4324	0.5676	neg	0.8356	0.9264	0.0756	0.9244	neg
34. Successful	-0.7927	0.1292	0.6432	0.5038	0.7332	0.2668	pos	0.4764	0.975	0.5987	0.4013	pos
35. Business Owner	-0.8667	0.25	0.4864	0.1848	0.3799	0.6201	neg	0.4435	0.106	0.0703	0.9297	neg
36. Athlete	-0.1591	0.3909	0.4417	0.1361	0.9627	0.0373	pos	0.336	0.5146	0.6623	0.3377	pos
37. Maid	-0.8564	0.3016	0.0346	0.3896	0.3177	0.6823	neg	0.6339	0.0302	0.2564	0.7436	neg
38. Left Hemisphere	-0.0325	0.3362	0.541	0.0515	0.7484	0.2516	pos	0.0852	0.2778	0.3267	0.6733	neg
39. Right Hemisphere	-0.9161	0.8968	0.9936	0.9705	0.3635	0.6365	neg	0.7266	0.7506	0.4609	0.5391	neg
40. Joy	-0.1998	0.3895	0.0197	0.6509	0.7747	0.2253	pos	0.9007	0.7219	0.7809	0.2191	pos
41. Poor	-0.5516	0.4217	0.5937	0.6501	0.7775	0.2225	pos	0.9402	0.7241	0.6719	0.3281	pos
42. Beautiful	-0.9976	0.0188	0.005	0.6337	0.2356	0.7644	neg	0.5872	0.2637	0.7929	0.2071	pos
43. Drugs	-0.929	0.938	0.3956	0.9473	0.4042	0.5958	neg	0.8904	0.0728	0.8736	0.1264	pos
44. Criminal	-0.3939	0.2291	0.3771	0.9997	0.2731	0.7269	neg	0.0264	0.7816	0.8556	0.1444	pos
45. Dead People	-0.657	0.7852	0.5981	0.4175	0.4378	0.5622	neg	0.5135	0.3457	0.6404	0.3596	pos
46. World Records	-0.0818	0.4469	0.265	0.2217	0.8678	0.1322	pos	0.6237	0.9045	0.5746	0.4254	pos
47. Celebrity	-0.786	0.849	0.1754	0.6085	0.842	0.158	pos	0.7963	0.2993	0.0897	0.9103	neg
48. Parent	-0.8877	0.0862	0.1068	0.8378	0.082	0.918	neg	0.7776	0.4197	0.7462	0.2538	pos
49. Obese	-0.5824	0.6536	0.0353	0.8188	0.502	0.498	pos	0.4242	0.6572	0.2452	0.7548	neg
50. Video Games	-0.2858	0.9894	0.5724	0.7855	0.2836	0.7164	neg	0.7413	0.9949	0.2238	0.7762	neg
51. Average	-0.7776	0.2643	0.3924	0.2835	0.5568	0.4432	0.6	0.4674	0.2563	0.5555	0.4445	0.6

5 Summary

The effect of social media have on society has been growing relentlessly over the past few years and has spiked during the COVID-19 period.

- People watch 1 billions hours of videos on YouTube per day and upload 500 hours of video every minute (YouTube, 2019).

- 995 photos are uploaded on Instagram every second.

- Google processes over 40,000 search queries every second.

- Microsoft Bing processes over 3,000 search queries per second.

Although autonomous systems still have many flaws, our reliance on them is growing by the day. Search Engines play a huge role in delivering data to the consumer. This means that every bias, objectivity, positivity, or negativity intentionally, or not, integrated into a search engine's algorithms will affect millions of people in a matter of minutes.

This is what motivated me to do this case study in which I compared two massive search engines, Google and Microsoft Bing, on a series of metrics.

The results showed a huge difference between the results of a search query on Google vs Microsoft Bing. This was showcased by the Kendall Tau which was always negative, and reached -1.0 at some points, with a staggering average of -0.7776.

The Jaccard coefficient varied from 0.0 to 0.9894 with an average of 0.26 which is quite low since the comparison was made on the same account

(same cookies, same data).

Regarding the subjectivity of each search engine, the results were quite similar with Google being more subjective than Microsoft Bing on average.

Microsoft Bing subjectivity results varied in the range [0.0047, 0.9936] with an average of 0.3924.

While Google subjectivity results varied in the range [0.0264, 0.9402] with an average of 0.4674.

The polarity of the queries was surprisingly positive almost always, although there is a lot of negative queries. The results of Google varied from [-0.01, 0.99] with an average of 0.2563, while those of Microsoft Bing varied in [-0.03, 0.99] with an average of 0.2835.

The similarity of the polarity results was also showcased in the positivity and negativity results which indicated a positive outcome 60% of the time in both search engines.

6   Reflection

The difference between the two search engines that was clearly showcased in the results emphasizes the number of small factors that interfere in how each query is processed. One of these factors is obviously bias, whether intended or not.

Artificial Intelligence has a lot of applications and benefits in all sectors, but if they aren't implemented correctly they can have devastating consequences. Even a minor mistake in the results of a search query would impact millions of people in a matter of minutes.

What will happen next, I don't know. But one thing I can firmly confirm is: Time doesn't go back. Artificial Intelligence will always be around. We just have to find a way to co-exist, while still being in command IF POSSIBLE.