Analysing Data and Relationship between New Mobile Devices based on Experts and Public Ratings

Kulraj Singh Saini
Department of Computer Science
State University of New York
Binghamton
ksaini1@binghamton.edu

ABSTRACT

In this time of modern world, new technologies is introduced all the time to the consumers, also in the field of mobile devices which is a very competitive market, there are new devices launched every other month and sales are very important for each device to determine whether there is much profitable or has been successful after its launch. Hence with this project the success of each device which has been in the market for the past year which has gained a lot of popularity is determined based on different factors such as how important cameras, performance and usage is and its correlation to the price of the device and how popular the device is across the social media.

RESEARCH OBJECTIVES

The objectives are :1)What are different resources needed to be considered for comparing mobile device specification? 2) Which factors are correlated and mainly consider by the market? 3) Which factors are the most important when it comes to success of the device? 4) Social Media reaction of each devices based on facts as well as their opinions on social media?

The aim is to understand the correlation of devices success and also sentiment analysis of the devices on the social media .

DATA SOURCE

This Project will have Reddit as data source for Social Media sentiment analysis and also since we need experts rating as well, antutu.com and DXOmark.com would be used as a data source for its performance, graphical and camera ranking of each device since this data set is reliable for accurate results. The price, CPU and GPU rankings from antutu is very accurate since this website has been working on these data since a long time. For the camera ranking data was collected from DXO mark since this website only focuses on camera performance. This data set would contain rating of CPU, GPU and DXO in single csv file and another csv file would contains posts from each subreddit of each individual mobile device. Both datasets were collected using BeautifulSoup4 library under python3.

Name	DXO	CPU	GPU	MEM	UX	Total	Price
Xiaomi Redmi K20 Pro	102	140217	183068	60120	68349	451761	400
Xiaomi Mi 9	110	136311	180695	63725	64828	445567	445
Honor 20 pro	113	127930	125978	74776	62718	391410	500
Samsung Galaxy S10	116	121378	166060	59383	64365	411193	590
OnePlus 7T Pro	114	144011	197877	78107	74703	494705	600
OnePlus 7 Pro	113	142000	178641	78724	74619	473991	670
Apple iPhone 11	109	155235	211480	75370	82044	524137	700
Huawei p30	116	129216	131850	67960	65166	394199	750
LG G7 ThinQ	85	97120	103076	44170	53175	297546	760
Apple iPhone 11 Pro	115	167055	213503	83000	83365	546930	1000
Samsung Galaxy Note 10+ 5G	117	128071	171608	89136	65118	453939	1100
Apple iPhone 11 Pro Max	118	166509	215130	82035	83267	546948	1200

Fig.1 Snapshot of Expert Ratings and Price rankings.

The above figure shows a snapshot of one of the datasets which was used and shows all the devices which were taken into consideration, to include devices from all the operating systems as well as different markets. Some of these devices were much more popular in a particular country yet the performance of the devices remains the same hence it creates much more equal comparison between each devices.

clean	
Phone vs iPhone 11	
turned 33 years old two days ago and got this from my parents as a gift	
Phone 11 LRM. Laguna beach delivered last night.	
Newport Beach sunsets with iPhone 11.	
years of android came to an end six months ago. The tables have turned.	
Months In. Still a great device	
lust beautiful	
Same expressions.	
inally upgraded from my 6s!	
And it s here!	
1 gang! After long frustrations with my iPhone X.	
Finally got my iPhone 11	

Fig.2 Snapshot of Subreddit Posts after cleaning

The above figure is the snapshot of how the csv file for the subreddit posts would look like after cleaning after which the keywords are picked up from the texts to determine whether the comments are positive or a negative opinion. For each of the subreddit of devices around 1500-2000 posts of consumers data was collected from reddit since the more number of posts are

collected there is much more clarity in the observation is present from the results of each analysis.

DATA CLEANING

The most important part of this is the cleaning of the data which is being collected from different website and also reddit. Since data is being collected from reddit. Each post on subreddit and it's are not the cleanest and not at the most suitable condition to perform sentiment analysis.

The cleaning process was started from beginning with removal of the empty comments and some indentations, also changing the abbreviations to full forms of it so that the sentiment analysis gets easier and much more accurate results. The next step was to remove all the emojis and also text in the language other than English from the subreddit posts since textBlob sentiment analysis needed all the UNICODE texts to be removed from the texts.

Got the 11 and I'm a happy boy 😇	Got the 11 and I m a happy boy
Mt . Fuji captured using iPhone 11	Mt. Fuji captured using iPhone 11
Took a picture on my iPhone	Took a picture on my iPhone.
Helio! Bought this yellow iPhone 11 after 1.5 years with space gray iPhone 8.	Helio! Bought this yellow iPhone 11 after 1.5 years with space gray iPhone 8.
The wide angle lens is just stunning	The wide angle lens is just stunning
Switched to Apple after 7 dreadful years of Android 🍎 💝	Switched to Apple after 7 dreadful years of Android
Changed from 6S+ 🖖	Changed from 6S
Just joined the gang!	Just joined the gang!
Just got my very first iPhonel	Just got my very first iPhonel
Shot on iPhone 11. It's one of my best photos.	Shot on iPhone 11. It s one of my best photos.
Got this for Christmas 🙂 🙂 🙂	Got this for Christmas
This team makes my days better. 😣	This team makes my days better.
lphone 11	Iphone 11
UPGRADE! From Johone 7 to Johone 11 00 The purple pastel hit my soft soot absolutely love it	UPGRADE! From lphone 7 to lphone 11 The purple pastel hit my soft spot absolutely love it

Fig 3. Cleaned Data

The above figure is snapshot of a part of data for one of the mobile devices where we can see the cleaned section which would be much more accurate for the results for sentiment analysis.

METHODOLOGY

This project was executed mainly using python3 and its libraries such as Beautifulsoup4, textblob, pandas, matplotlib, urlib and sklearn based on the requirements of each part of the project and by the end of it correlating everything and bring a useful observation and results.

For the first part of the project initially data was collected from antutu.com and DXOmark and stored in a csv file all together for further study on it. Initially the whole data which was shown in the figure 1 is normalized to between 0 and 1 so that there is an equal point of reference of each element of study which where CPU, GPU, DXO, and price of each device

```
$\(\text{Constraint}\) $\(\text{Constraint}\)
```

Fig.4 Snapshot of Normalized data

In figure 4 you can see that the data which had very big and irregular values are normalized into one single range which is between 0 and 1 so that the further analysis is fair and even.

Firstly ANOVA(Analysis of Variance) test was performed on the dataset is performed since ANOVA test is generalization of more than 2 groups and to determine whether there are any statistical

significance difference between the means of three or more independent groups. This test was then performed on the total ranking added up and price of the device to study whether each specification is equally important when compared to the price.

```
FVALUE AND PVALUE
316.12405382717657 3.0603169653409916e-37

SUMMARY

sum_sq df F PR(>F)
C(treatments) 1.608840e+11 4.0 316.124054 3.060317e-37
Residual 6.997745e+09 55.0 NaN NaN
```

Fig. 5 Result of ANOVA test

Fig. 5 Shows us the result of the ANOVA test performed on the total ranking and the price of each device and with this result one of the important observation which was noticed is that each specification is very important because the F value obtained is 316.124 which is more than the expected F value that is 205 to satisfy the null hypothesis, hence null hypothesis was reject and hence further studies where to be conducted to observe the importance of relation between which of the two factors out of all of them.

The next test which was performed from was Tukey's HSD (Honest Significant Difference) test which is mainly used to find out significance between two factors or how different they are.

	Multip	le Comparison	of Mea	ans – Tukey H	SD, FWER=0.10	
group1	group2	meandiff	p-adj	lower	upper	reject
CPU	DXO	-137810.4167	0.001	-149434.1187	-126186.7146	True
CPU	GPU	-137920.4571	0.001	-149544.1591	-126296.7551	True
CPU	MEM	-66545.5833	0.001	-78169.2854	-54921.8813	True
CPU	UX	-67778.0	0.001	-79401.702	-56154.298	True
DXO	GPU	-110.0404	0.9	-11733.7425	11513.6616	False
DXO	MEM	71264.8333	0.001	59641.1313	82888.5354	True
DXO	UX	70032.4167	0.001	58408.7146	81656.1187	True
GPU	MEM	71374.8738	0.001	59751.1718	82998.5758	True
GPU	UX	70142.4571	0.001	58518.7551	81766.1591	True
MEM	UX	-1232.4167	0.9	-12856.1187	10391.2854	False

Fig. 6 Snapshot of Tukey's HSD test

The above fig.6 is the results obtained when the Tukey's Test was performed on the csv file with the data shown in Fig.1 but the price factor wasn't included in this test. With this test results few things we able to observe. For the first observation it is seen that camera specifications and graphical performance are not relevant to each other's success since the null hypothesis was rejected. The next observation which was seen is memory management of device is also not associated with user interface since the null hypothesis was rejected.

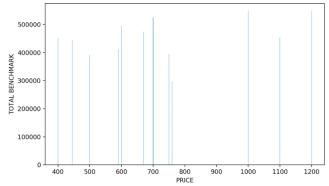


Fig.7 Graph with benchmark and price

This fig.7 was plotted using adding up all the different factors which were DXO, CPU, MEM,UX to add up for a total value and then comparing it to the price of the models available at the market and it is observed that even with low selling prices performance is still up to the mark of the more expensive devices present in the market.

The next section of the project was conducted on the social media, where reddit data was collected from subreddit of each the mobile devices where consumers provide opinions on how they feel about each device hence analysis of this is a very important part of this project.

Data collected in csv files as seen in Fig.2 and sentiment analysis was performed on each of the csv files using TextBlob. TextBlob is a library for processing textual data providing easy API used to Natural Processing Language (NLP). Using sentiment analysis, we were able to plot polarity of the text ranging from positive to negative and also facts to opinions from each of the posts of the subreddits.

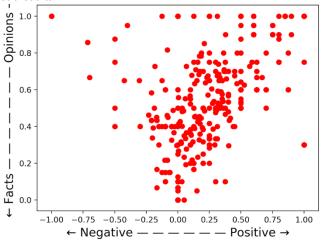
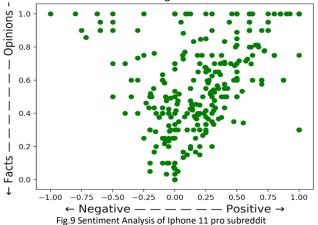


Fig.8 Sentiment Analysis of Iphone11 subreddit

Fig.8 is the result of the sentiment analysis of Iphone11 subreddit where we can see this device was very popular among the consumers and all the dots on this scatter plot represents each post from the subreddit. The majority of these posts are positive when it comes to facts as well as opinions which shows how successful this device was among the audience.



From fig.9 We could see that the subreddit for iPhone 11 pro wasn't as positive as compared to fig.8. Even though it was very popular among audience as we could see the positive opinions yet there were considerable number of negative opinions as well which deduces that there were good sales and success but with a room of improvement for a more better success rate of device.

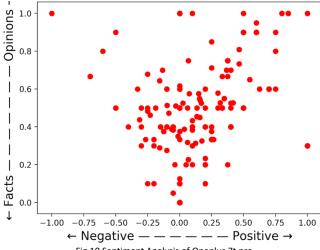


Fig.10 Sentiment Analysis of Oneplus 7t pro

This is one more example of the sentiment analysis of the device Oneplus 7t pro where it is observed that the devices had a lot of neutral reaction being it opinions or even facts which shows there is still a good market for this device yet it isn't a very successful device for the audience only because there are much better price to performance devices available in the market which the audience end up buying compared to this device.

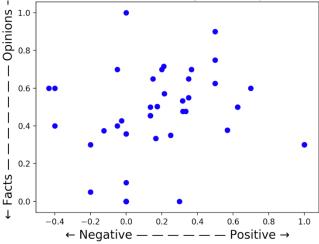


Fig.11 Sentiment Analysis of redmi K20 pro

This results for sentiment analysis of K20 pro displayed in Fig.11 shows us an interesting result where this device benchmark ranking was at the top of the chart competing with all the expensive devices present in the market and still this device doesn't seem to be very popular among the audience. This shows how important the popularity of the device is so that there is comparatively much more success rate of the device, here with all the best specification and performance yet the sentiment analysis

studies show that there is negative approach towards the facts as well as the opinion of the device.

The result of all the other devices were similar from one of these results which brought a few important observations where most of the expensive devices were amongst the most popular around the social media and being also successful as well, but a few improvements needed. The cheaper devices had very good benchmark performance but yet wasn't very popular among the audience. The mid-range devices had average benchmark performance had positive impact on the market but not with big numbers of sales.

ACKNOWLEDGMENT

We would like to thank our professor Jeremey H Blackburn for letting us work on this project and guiding us through and helping us with any problems faced.

REFERENCES

- [1] https://www.antutu.com/en/ranking/rank1.htm
- [2] https://www.antutu.com/en/ranking/ios1.htm
- [3] https://www.dxomark.com/category/mobile-reviews/
- [4] https://smartphonesrevealed.com/the-best-smartphones/?showall=true
- [5] https://www.reddit.com/
- [6] https://textblob.readthedocs.io/en/dev/quickstart.html