**OUR GOALS:**

In this project, our focus will be on tablets. Specifically, the Samsung tablet S9 FE. We will also be looking at their competitors to compare against.

The goals for our project is a versatile programme that only takes in a valid tablet name and outputs the pros and cons for each tablet, recommendations for design improvements, and a visual representation of all the information gathered.

For example, our input looks something like this,

And our output will looks something like.

It has the pros of the tablet,

The cons of the tablet,

And an overall summary of the tablet.

It also infers what the direction of the company is

So that in the recommended design changes, what is changed or improved will align with the goals of the tablet and the company.

We also aim to have graphs that visually represent the information gathered for further analysis.

Before I talk about the program created, I would like to preface the presentation with our design process.

The thought process behind the decision making for the program is done by thinking about how the manual process would be done.

If I was a consumer planning to buy a tablet, I will first look at online reviews, amazon reviews, and the specifications of the tablet.

However, this is from the perspective of a consumer.

If we are to complete the objectives of the assignment, we need to look at it from the producer perspective.

As such, we will also need to include the data from news and patents that reflect the business direction. This is to ensure that whatever design recommendations are provided, the recommendations align with the company’s direction.

With this, we will now explain the process behind the scenes from our input to our output.

Our program has the data architecture as follows

We start with our product name and search online for its competitors. We then scrape for information regarding all the tablets. Our data sources include amazon reviews, youtube comments, and youtube captions. These data are then cleaned and passed through ChatGPT.

When getting our competitors, we first search through gsmarena for the tablet. Gsmarena is a website with information on all phones and tablets.

We then get to the page with the information of the tablet to get the year of release and the dimensions.

Next, we search for tablets with the same year of release and with similar dimensions. We limit the search result to the big players in the tablet scene such as Samsung, apple, and Xiaomi.

We then retrieve the top 5 results of the competitor tablets from this webpage.

To scrape all these information, we first looked into using beautifulsoup. Its fast, lightweight, and easy to use. However, its easily detected by anti scraping measures.

In the case of gsmarena, an issue of encryption arose when trying to get the tablet names from the search result page. As such, we decided to use selenium.

Selenium works by directly opening a browser window and interacting with the page as a human user would. This way, encryption will not be an issue when trying to retrieve the specifications of the tablet.

Getting the reviews from amazon worked similarly. We first type in the search term,

Clicked on the first result,

See all reviews about the tablet,

And collate the reviews. Then, we click on the next page until there is no more next page.

We attempted to use beautifulsoup but it was easily detected by amazon. As such, they hit us with a captcha.

This captcha could not be solved through beautifulsoup. Even if we managed to do it, we will be faced with

Amazon dogs.

To work around this, we scripted a login to the page.

Next, we collated the youtube reviews. This is similar to the exercise done in class.

To get the youtube captions, we used a library called youtube transcript api.

We looked into google’s youtube data API but found out that authorization from the creator has to be granted. Thus, we ended up using the transcript api.

We all know the term garbage in garbage out. If we feed our model garbage, we will only receive garbage outputs. Therefore, cleaning the data appropriately before processing is crucial.

After cleaning a comment, we want to ensure that all natural language nuances are maintained while removing non human language like codes and links.

As such, we have adjusted the function provided in class to suit this need. This ensures that the nuances are retained while non human language is removed.

With all the data cleaned, we can now implement retrieval augmented generation or RAG. This is similar to what was done in class. The question now is what data are we passing to ChatGPT.

From a tablet designer point of view, we first have to examine the user opinion on the tablet. This is done by processing the reviews and comments to get a summary of the user opinions.

The response will look like this with the

Pros,

Cons,

And summary.

Next, we have to get the business direction to constraint our design recommendations to align with the company. Due to the lack of reliable information on business directions, we employ the use of ChatGPT to infer the business direction from the youtube captions. Our assumption here is that the reviewers are more tech savvy and up to date with the latest news.

The response will look something like this, where the general direction of the product is inferred.

Lastly, we feed the user opinion and the business direction, back into ChatGPT.

This is to get our recommended design improvement,

while ensuring its alignment with the company and product.

This is how we start with a single input and end with our tailored design recommendations.

We also explored another method that the goal of providing information that helps guide decision making alongside the recommendations provide witih ChatGPT.

This is the data architecture for the use of exploratory data analysis EDA. The key difference between this and RAG is the use of the data. Only the comments and reviews are used in EDA for the analysis of the consumer preferences.

EDA has been covered in class. The question now is what our classifications are.

We figured this out by searching what people look for in a tablet and what factors are considered before buying them.

We ended up with these categories. Ideally, a tablet will hit all of these criteria.

It will have good ergonomics, have a premium feel, high performance, low cost, things like that.

With this in mind, we set up our program to classify each comment 8 times, each time classifying if a comment is related to a category or not.

For example, if this is the review being processed,

It will pick up that the review is related to display, connectivity and audio.

For sentiment analysis, it is similar to the lesson in class, using the positive neutral and negative sentiments.

Overall, the review has a positive sentiment.

As such, with all these information,

We collate them to have these classifications and sentiment. We then repeat this process for all the comments.

Collating the information gathered, we end up with this graph.

In the case of our previous example, the review contributes to the relevant classifications and the relevant sentiment as highlighted in the slides.

Looking at the total number of reviews for each component, we can tell that the audio is the most mentioned. We assume that each mention represents the user’s interest in the component. After all, you talk about the audio if it is deemed important to you.

Putting them in a pie chart, we can get an overall view of what consumers are looking for in the tablet model. In this case, it can be inferred that 35% of consumers get the galaxy tab s9 for the sound quality. If for example, I wish to improve the build quality of the galaxy tab without sacrificing something important, I can look at the pie chart and determine that connectivity can be sacrificed because it is the least important to my audience. Repeating this graphing for all tablets, we get this.

This collation of graphs help the business have an overview of what the users of competing tablets are interested in. If I want to steal the users of Xiaomi pad 6 and get them to use my galaxy tab s9, I can focus my efforts on improving my audio to compete against Xiaomi.

Going back to the graph of reviews,

We can see that the ratio of good reviews to bad reviews about the cost of the tablet is high. If we use this ratio for all components, we get a radar graph like this.

Each component is rated from 1 to 5, 5 being the best. If we repeat that for all the tablets, we get this.

This graph can help us understand how each of our component is compared against the competing tablet components. For example, if my business direction is focused on playing towards my strength, I can focus on leading budget friendly options to push for my cost to be the most competitive, winning the market of budget tablets.

If we recall previously, our RAG model mentions that the display and performance are good while the cons are the cost.

If we look at our graph, we can see that our radar graph shows the cost being good while the display is mediocre.

And if we look at the pie chart, it shows that most consumers go for the galaxy tab for the audio. Yet, the chatgpt has no mention about the audio. So what is going on?

To answer that, we have to benchmark our model.

Looking at the graph of all the reviews, we can see that all of them have a similar shape. This should not be the case.

This is the result of our manual benchmarking of the EDA model.

Our sentiment analysis shows a very decent performance.

That is not the case for our classification.

Remember how I mentioned garbage in garbage out?

Lets take a look at some reviews from our dataset.

This review talks about display and performance. The rest of the categories are not related to this review. As such, 6 classifications are correct and we give it a 6 out of 8.

Next, we have a review that only talks about performance. Our model correctly classifies that this is only performance related. Therefore, we give it an 8 out of 8. These are good data sources. But what about bad ones?

The way you said AKG is so hot I bet your wife must be so proud. Related to everything except for cost. 1 out of 8. Next,

Meessiiiii. Everything is related. 0 out of 8.

This issue can be mitigated by adjusting the code. We can try out different classifiers. Maybe remove the term “related” or include the term “tablet”

We could also try out different models available on huggingface.

Or better yet we can fine-tune our model to be specialised in everything related to tablets.

Unfortunately, we were not able to try these options as running the classification is very time consuming. I can start the classification, shop for groceries, cook for a family of 5, finish my food and come back, and it will still be classifying. Therefore, we leave this as one of our…

Limitations. With our EDA being inaccurate, we could not use it to supplement our analysis. We have also lost a reliable method to benchmark our RAG. This is temporarily mitigated by manually checking the data sources used by chatgpt.

The google translator api is imperfect as well. This could lead to unforeseen issues. Fortunately, it does not affect our model as of now but it can cause problems in the future.

Sometimes amazon hits us with the captcha. As of our current version, we have to manually input the captcha to proceed. It is possible to automate the process of filling this captcha. However, it is out of the project scope and does not directly affect the performance of our program.

Next, our sample size is limited to 100 as of our current version. In hindsight, other teams have managed to bypass this by scraping reviews sorted by star rating. This can be implemented in the near future. Furthermore, the total amount of reviews on amazon on a particular product is small compared to the sheer amount of comments in the youtube comments. This means that the higher quality reviews can easily be overshadowed by the irrelevant youtube comments if the amount of comments gathered are too high.

Last but not least, our existing model is prone to failures. This is because the response from chatgpt is fed into chatgpt again. If chatgpt gives a nonsense reply on the first pass, it will more likely give another nonsense reply on the second pass.

Most of the time it works. For the business direction portion, we can see that the relevant sources are selected, taking captions that review the tablet.

However, in the pros and cons, chatgpt took in a portion of the caption that was a sponsored segment of the creator’s video.

This can be mitigated by adjusting the search type, kwargs, and chunk overlap and sizes. As for the portion of sending chatgpt’s response back into chatgpt, this can be mitigated by having the code check the response from chatgpt and send another request if the data sources used are irrelevant. However, this requires extensive testing.

Testing is not free.

With all these limitations and its mitigations in mind, what other ways can we improve our model to bring it to its full potential?

At the start of the presentation,

We mentioned that the direction of the business is important in the decision making of the design improvement. As such, the model can have an improved reliability and accuracy by getting reliable external data sources about the tablet industry.

Furthermore, the model can take in and consider the cost and effect of different parts for the components of the tablets. For example, it can consider the trade offs and collateral design impact of the change of a component. It can also provide the recommended components based on the goal of the tablet model released.

We can also expand our data sources to other social media platforms. This allows us to get a more complete picture of the consumers. This also means that our zero shot classification model must be robust enough to handle a variety of natural language nuance.

Lastly, our model would be able to consider the ecosystem support and external factors surrounding a tablet. This would mean that our dataset must be diverse enough such that it covers the demands of consumers that the consumers do not even specify in reviews or comments.

Overall, our group subscribes to the notion of “don’t know don’t anyhow say”. The project sounds like it is riddled with holes and does not work. To a small extent, that is correct. However, it is difficult to create a useful and functional model within 5 weeks. As such, we focused our efforts on creating a proof of concept and researching how these issues can be fixed of mitigated. With this mentality, even if our current model fails, we leave room for our future model to succeed. Thank you.