Coursework

Individual Research Project Preparation(7000CEM)

Beauty Product Development in Cosmetic Industry Based on Online Reviews Using Big Data Techniques.

Name: Jusel Justin
SID: 12646627
Mail ID: juseljustj@uni.coventry.ac.uk
Module Leader: Mr. Peter Every

Table of Contents

Component 1: Project Proposal:

Project Title:

Beauty product development in cosmetic industry based on online reviews using big data techniques.

Project Topic:

The cosmetic business includes skin care, hair care, personal care and perfumes was at its peak in the year 2020. Cosmetics sales had a downfall of 33% worldwide, while the overall sales in the industry downfall of 15% (Kim & Jung, 2022). But the business has proven to be robust during the past. Experts foresee a steady growth in the year 2022.(The Beauty Industry-Encompassing Skin Care, n.d.).Beauty products are goods used to keep the body beautiful, appealing and clean, to enhance the complexion, and to maintain the health of largest organ of the body that is the skin. Cosmetic products and beauty industries are now included in the five major consumer goods and are regarded as essential multi-million-dollar industry. (Kim & Jung, 2022).

The growth of technology has paved way to the rise of digital customers who purchase beauty items online and create and sell information through digital media thereby generating a huge amount of data. The changing digital platforms increase the volume of digital information while decreasing transaction costs and thereby improving the availability and convenience for customers (Kim & Jung, 2022) Following that, people turn to internet platforms to look for different products, compare prices, read reviews, keep informed about products, and have various conlusions before making final judgments on the beauty products they consider buying.(Kim & Jung, 2022)

The topic choice for this coursework is to develop a beauty product (for example a lipstick) in the cosmetic industry based on the customer reviews available on platforms like Amazon.com or MakeupAlley.com using text mining which is an elaboration of data mining. Text mining is a process of deriving patterns and insights to find trends and models from unstructured text document collections which are usually unstructured.(Haddara et al., 2020) Data analytics tools and techniques like Topic modelling analysis and Sentiment analysis are used to derive valuable insights to the problem.

Research Questions:

Based on the project title, some of the important research questions that have been developed are-
Question 1:
Which consumer related criteria should be evaluated during the new product development stage?
Question 2:
How can data analytics approaches be used to discover and predict reinforcers?

Motivation and Expected Outputs:

A key motivation is value extraction from Big Data. Data-derived information and expertise can be applied to product creation.(IEEE Singapore Section et al., n.d.) The experts stressed Big Data's significant potential for the idea generation process. An analysis of use phase data, according to experts, can indicate customer needs and preferences about features of the product and evaluate ideas for products.

To understand customers' relations in respect to the reviews submitted towards features like moisture, odour, colour and taste, the results of sentiment analysis should be carefully evaluation. The model should show the customer preferences based on brands and items.

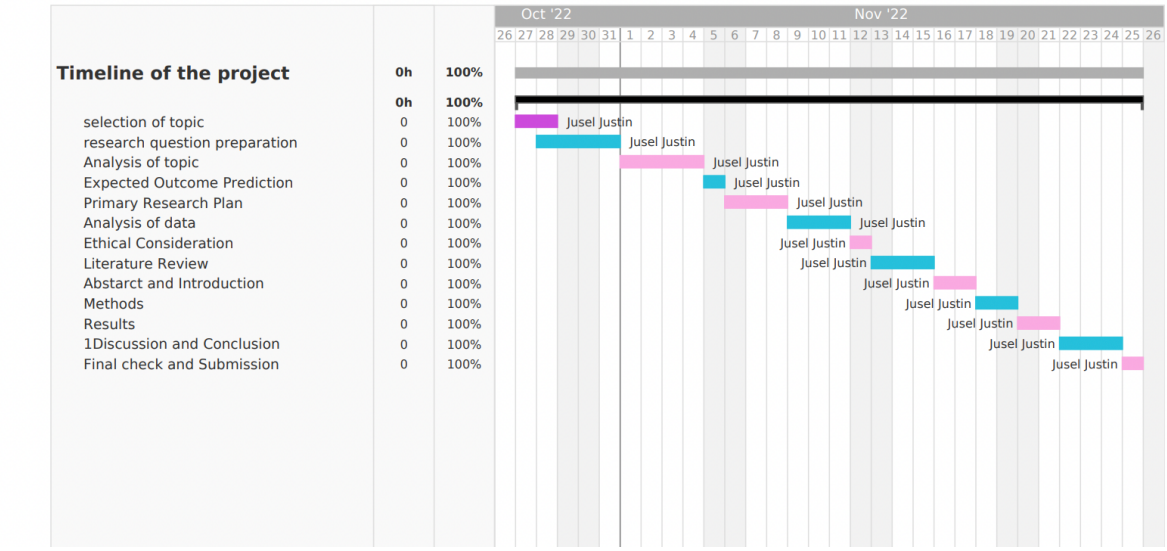
Primary Research Plan:

Gantt Chart:

A Gantt chart is a project management tool that helps with the planning and scheduling of projects of all kinds, but they are especially beneficial for simplifying complex tasks.

The tasks involved in the project are thought about and divided into:
Selection of topic: A large topic gives writers several options to investigate in their hunt for a valid research question. Narrowing down the large topic can provide a research title

1. Research question preparation: A research question is a question that is intended to be answered by a study or research effort. This question frequently discusses an issue or problem that is addressed in the study's conclusion through data analysis and interpretation.
2. Analysis of topic: It is a broad term, but it simply refers to the process of investigating and fine-tuning your topic before beginning your search for information.
3. Expected Outcome Predictions: The outcomes of an experiment or trial are the results of an experiment or trial. It is feasible to create predictions using outcomes. A prediction is an accurate assumption about what will occur.
4. Primary research plan: It is the primary plan as to how to go about the project, scheduling the time scale and planning the research ahead.
5. Analysis of data: The discrete process of applying logical or statistical approaches to describe the data, summarize the data, and assess data is called as analysis of data.
6. Ethical Considerations: Ethical considerations are concerned with preventing harm to the society as a result of their participation in your organization's decision-making processes.
7. Literature Review: A literature review is a piece of research that demonstrates knowledge and understanding of the academic papers on a particular topic.
8. Abstract and Introduction: An abstract is comparable to a summary, except it is shorter and more straightforward. Paper's introduction goes into greater detail.
9. Methods to be used in Data analytics: Choosing the right data analysis methods, here text mining is used which is an extension of data mining.
10. Results: The results section is where we present the conclusions of the study.
11. Discussion and Conclusion: The results should be interpreted in the Discussion and Conclusion sections.
12. Final check and submission: Finally, the coursework is rechecked and submitted



Gantt-chart representing the timeline of the project from 27 Oct to 25 Nov

Ethical Considerations:

What are the major ethical issues?

- Privacy can be assured by reducing the volume of data collected.
- Preserve data privacy, it is advocated that data be altered in a way that makes it less revealing, as well as data access be regulated.
- Informed consent, in which consumers are made fully aware of the aim of the intended present and future data uses.

Five principles cited in Abrams' (2015) Unified Ethical Framework for Big Data Analytics:
Fair – Are the potential consequences of the data collected and its use aware to all parties involved?
Sustainable- Are the insights we uncover through data sustainable?
Respectful - Have we been welcoming and open about the policies?
Progressive- Do we have a progressive culture of continual development and data minimization?
Beneficial - Does our data use benefit customers as much as it does us?

The conducted literature review does not require any legal or ethical agreement.

Component 2: Literature review

Abstract:

The global cosmetics market is growing steadily. This study analyses online reviews to understand customer preferences with respect to cosmetic products. The study extracts and collects customer data reviews by using Big Data techniques and technologies to identify the factors affecting the cosmetic industry. The report focuses on mainly on three factors. Firstly, in understanding the customers' expectations and their sentiments towards the products. Secondly, the customers expectation on the new product based on the reviews. Thirdly, understanding the customer needs and the in-trend and most up to date collection of beauty product. And thereby launching new beauty product that compliments the customers skin tone and age.

Introduction:

Introducing new products is the backbone of companies since they can generate new enterprises, business initiatives, grow existing firms, create income, or raise living standards. (Haddara et al., 2020), Hence In the cosmetic sector, introducing new items that meet the needs of customers is critical to the organization's growth. According to the existing literature, firms that foster data oriented and analytics cultures are likely to have an advanced product design and innovative advancement (Haddara et al., 2020).

It should be taken into consideration that customer reviews are viewed as data and can be used to gain a competitive edge, as a consumer purchasing the product writes a comment or opinion about the product, and potential customers read the review before making a purchase. This study aims to understand consumers' demands by mining online customer feedback to gain more detailed and independent information on customers' product choices. If handled appropriately, the organization's multiple consumer ties with its customers could improve marketing, boost performance, and eventually raise profits and revenues. Even though studies showcase the use of text mining techniques in the production of beauty products, the majority simply calculate data acquired through surveys rather than directly scraping customer feedback and opinions. This study seeks to find consumers' interest by mining online consumer feedbacks in order to get more thorough and factual information on consumers choices toward cosmetic products. The goal is to present an overview of prior studies on the cosmetic industry and the incorporation of online customer feedbacks in beauty product development (Haddara et al., 2020) Numerous studies have been undertaken in recent years as people have become more aware of the value of text mining in customer reviews.

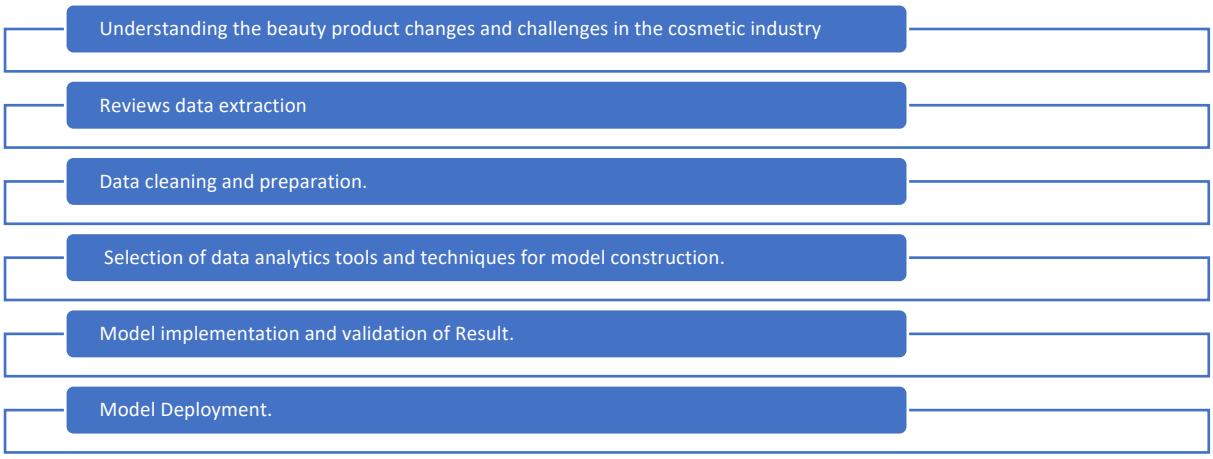
Before proceeding with this project, the literature on text mining was studied to gain a basic understanding. The goal is to create an analytical model that explains the relationship among review, customer features, and customer choices for cosmetic products (Haddara et al., 2020). It's fascinating to see if customers with varied features like age and skin type, will submit reviews from a variety of experiences and viewpoint, the model could also reflect the different preferences of different customer categories. We could also investigate which aspects of cosmetic products influence customer preferences. We could also make new

beauty product development recommendations based on the research. The main goal is to provide a good knowledge of online consumer reviews and to improve the cosmetic industry's Beauty product Development process by navigating and understanding customers' demands in a more accurate way.

Literature Review:

Cosmetic items, often known as beauty products or makeup, they are chemical mixtures used to improve the appearance or odour. (Rajput, 2016). Skin care, hair care, deodorant, cosmetics, and perfumes are all examples of cosmetic items that are widely utilised by the general public. (Haddara et al., 2020). Exclusive brand outlets, pharmacies, and internet shopping centres are the key distribution channels for beauty products. Despite the economic instability, the cosmetic business continues to thrive, with a growth rate of about 5% (Kumar, 2005). In 2017, the global market was estimated at $532.43 billion, with a projected increase to $805.61 billion by 2023. (OrbisResearch.com, 2018). Even while huge corporations create over 70% of total income in the industry, there is still potential for small businesses that can supply customized items to target a specific niche market. Aside from research on the chemicals utilised as new ingredients in cosmetic products, some research has been conducted on the marketing strategy in this industry. Customers can share their product reviews on retail websites or online platform, generally in the form of ratings and reviews.

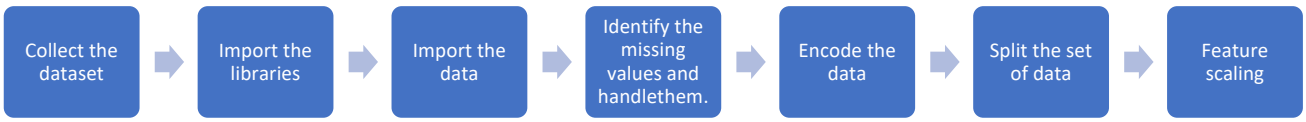
According to research, customer reviews and feedbacks have a beneficial impact on sales (Li, Wu, & Mai, 2019). Amazon, for instance, has generated an increase of 2.7 billion dollars by introducing "Was this review helpful to you?" Reviews that are highly ranked in such order are considered helpful and are likely to boost sales. It is also noted that negative reviews may also have an impact on brand loyalty, trust, and customer purchasing intentions. Several properties of online reviewers have piqued the interest of researchers, including subjectivity classification, sentiment classification, and opinion summarization.



(Source: Self-Created) Steps to understand the research problem

Data pre-processing:

The first phase of data collection is mainly used to collect a huge number of cosmetic reviews data from billions of different sources. The data collected may be in different formats and data types. The data can be either sent to the system for further analysis or stored in databases or a data warehouse to analyse later. After Data cleaning, if the stored data is not usable, the data can be passed down through the full data analysis phase, thereby, increasing the processing error (Anna University & Institute of Electrical and Electronics Engineers, n.d.). Big data storage refers to a computer and storage system that collects and handles enormous data volumes while also allowing for real-time data analytics. Big data storage is used to compute and store the data architecture. A big data storage system is a collection of computers linked to high-capacity disc that is a hard disk to support analytic applications designed to process a massive amounts of data.(Haddara et al., 2020)



(Source: Self-Created) data pre-processing flow

Methods:

Text mining:

Text mining is an extension of data mining. It is a process of deriving meaningful and interesting patterns (Haddara et al., 2020)from unstructured text files in order to construct models or uncover trends and patterns (Elragal & Haddara, 2014; Schelén, 2015). Other terms for data mining are Intelligent Text Analysis, Text Data Mining, or Text Knowledge Discovery (Gupta & Lehal, 2009). Text mining is considered to have a higher scope than that of data mining as it was found that around 80% of the industries information is stored in unstructured format which is handled at ease using text mining techniques. According to few research findings it was concluded that text mining is an interdisciplinary study topic that draws on research topics and methods. Text mining has expertise in machine learning, statistical reasoning, information retrieval, strategic leadership, natural language processing and text analytics. When comparing text mining to data mining, it is noticed that text mining is performing better in extracting implicit and explicit concepts. It is also understood that text mining in sematic relations performs better in sematic relations among concepts, from

unstructured or semi structured text data, such as email, full text documents and HTML. Text summarization approaches include neural networks, decision trees, semantic graphs, regression models, fuzzy logic, and swarm intelligence. (Gurcan & Cagiltay, 2022)Text mining is data driven. Text mining enables businesses to evaluate online customer reviews in order to understand their competitors and monitor customer feedback, but also to identify prospective new customers and build the image of the organization.(Haddara et al., 2020)

Topic Modelling Analysis:

Various topic modelling algorithms are available for text mining and natural language processing (NLP)
They are:
Latent Dirichlet Allocation (LDA)
Hierarchical Latent Dirichlet Allocation (HLDA)
Non-Negative Matrix Factorization (NMF)
Dirichlet Multinomial Regression (DMR)
Hierarchical Dirichlet Process (HDP)
Dynamic Topic Model (DTM0
Correlated Topic Model (CTM). (Gurcan & Cagiltay, 2022)

Topic modelling is an unsupervised machine learning technique used to find hidden semantic patterns known as "topics". (Gurcan & Cagiltay, 2022)LDA is the underlying theory of topic modelling and is frequently used among all topic modelling algorithms. As a result, LDA is extensively used in numerous studies and applications
(Gurcan & Cagiltay, 2022).LDA's main logic is built on the notion that every document in the dataset includes multiple topics. The LDA employs an iterative approach based on Dirichlet distribution to compute as per the document topic distribution.

Sentiment Analysis:

Sentiment analysis is the application of text analysis and natural language processing techniques to categorise text based on the sentiment expressed in the text. (Yadav & Saleena, 2020). Sentiment analysis determines the speaker's sentiment toward a certain issue. It is the process of determining people's feelings or attitudes toward things, movies, and so on, as conveyed in the text. Sentiment analysis is widely utilised in business, including Social Media Analysis, Market Analysis, Product Analytics, and Customer Analysis.

Results:

This study will derive data - driven results from online customer evaluations to better understand customer preferences on cosmetic items. (Haddara et al., 2020)We can determine the results by considering any one cosmetic product like a lipstick or foundation. According to the findings we could evaluate the customer preferences for instance, the shade of the best-selling lipstick based on the colour that compliments the wide range of skin tone. The coverage and colour match of the foundation on various skin tones etc... Various observations can be made with this analysis like based on the age group of the customer, the lipstick shade might vary. Thereby, shades according to age may be marketed to the customer based on their skin complexion. Hence new suggestions and

recommendations can be given to the cosmetic company to develop a new beauty product suiting the needs of the customers.
Without the capacity of data analytics techniques, this unprecedented high number of samples would have been nearly impossible to obtain, gather, and evaluate.

Discussion and Conclusion:

This study demonstrates the scope of combining data analytics techniques and the development of new beauty products, cosmetic product reviews can provide organisations with real and genuine reviews from their customers, which in turn leads to a better understanding of the customers' reinforcers without any need for conducting monitored experiments, as well as facilitating the firm's efforts in appropriately managing the marketing strategies. (Haddara et al., 2020)

Furthermore, this study not only provide an overview of past studies conducted in the cosmetic sector and online customer feedback, but it also advances text mining application techniques in research in the beauty product Development process.(Haddara et al., 2020)

In addition, this study develops a model to explain the close relationship between review features, customer traits,(Haddara et al., 2020) and consumer trends for cosmetic items. Overall, this research highlights the significance and limitless opportunities of implementing text mining tools
To the cosmetic and other industries. This understanding may also assist cosmetic companies in improving the quality of their products in order to increase sales by exactly fulfilling the clients' needs.

Bibliography:

Anna University, & Institute of Electrical and Electronics Engineers. (n.d.). *10th International Conference on Advanced Computing : 13-15 December 2018, Department of Computer Technology, Anna University, MIT Campus, Chennai*.

Gurcan, F., & Cagiltay, N. E. (2022). Exploratory Analysis of Topic Interests and Their Evolution in Bioinformatics Research Using Semantic Text Mining and Probabilistic Topic Modeling. *IEEE Access*, *10*, 31480–31493. https://doi.org/10.1109/ACCESS.2022.3160795

Haddara, M., Hsieh, J., Fagerstrøm, A., Eriksson, N., & Sigurðsson, V. (2020). Exploring customer online reviews for new product development: The case of identifying reinforcers in the cosmetic industry. *Managerial and Decision Economics*, *41*(2), 250–273. https://doi.org/10.1002/mde.3078

IEEE Singapore Section, IEEE Technology and Engineering Management Society. Singapore Chapter, IEEE Technology and Engineering Management Society. Hong Kong Chapter, & Institute of Electrical and Electronics Engineers. (n.d.). *IEEE IEEM2017 : 2017 IEEE International Conference on Industrial Engineering & Engineering Management : 10-13 Dec, Singapore*.

Kim, Y., & Jung, H. (2022). *Beauty Industry's Strategic Response to Metaverse Evolution: Focused on Generation MZ*. 259–264. https://doi.org/10.1109/bcd54882.2022.9900806

The beauty industry-encompassing skin care. (n.d.).

Yadav, S., & Saleena, N. (2020, October 14). Sentiment analysis of reviews using an augmented dictionary approach. *Proceedings of the 2020 International Conference on Computing, Communication and Security, ICCCS 2020*. https://doi.org/10.1109/ICCCS49678.2020.9277094