International Journal of Advances in Production Research

Insights from Dynamic Pricing Scenarios for Multiple-generation Product Lines with an Agent-based Model using Text Mining and Sentiment Analysis

Atefeh Anisi[a,1], Gül E. Okudan Kremer[b,2], Sigurdur Olafsson[a,3*]

[a] Department of Industrial and Manufacturing Systems Engineering, 3004 Black Engineering, Iowa State University, USA

[b] College of Engineering, University of Dayton, USA

[1]anisi@iastate.edu. [2]gkremer2@udayton.edu. [3]olafsson@iastate.edu.

***Abstract*— Corporations must constantly upgrade and improve their offerings due to changes in customer preferences. It is a common strategy for firms in technology-intensive markets to use online reviews as a source of product information to inform such changes. This user-generated information is valuable since it provides companies with valuable and low-cost input. In this paper, we propose an agent-based model for simulating potential cannibalization situations with respect to customer satisfaction throughout consecutive generations of a product line. The level of customer satisfaction is regarded as a parameter in the model, which is conceptualized to affect the product price. The proposed model provides insights into different pricing strategies regarding customer satisfaction levels affect the total lifecycle profitability of multiple-generation product lines, and how they can be used to assist organizations in developing appropriate dynamic pricing strategies.**

***Keywords*— text mining, sentiment analysis, multiple-generation product lines, cannibalization, agent-based modeling, dynamic pricing scenarios.**

I. INTRODUCTION

We are surrounded by multiple-generation products (MGP), from simple ones like NIVEA skin care products to the most complicated ones like Apple Mac Books. In MGP lines, a company will introduce the first version of a product to the market and periodically introduce the following versions. Different versions of an MGP have the same core functionality, each with higher technologies or more features than the previous versions [1]. Choosing multiple-generation product strategies has advantages for businesses but may also create cannibalization. Market cannibalization is a loss in sales, revenue, or market share of a company caused by introducing a new product similar to another product in the market. For MGPs, cannibalization refers to the competition that may happen between different generations of a product. When the company releases its new generation, the price of previous generations may be discounted, and therefore, they may attract customers due to their lower price. This, in turn, leads to less profit for the company since the latest generation typically has the highest profit margin.

Despite the risk of cannibalization, research shows that having a multiple-generation product line can be 26% to 40% more cost-effective than introducing a single generation of a product or sequentially introducing a multiple-generation product [2]. Choosing multiple-generation product strategies can benefit companies in different aspects. For example, with multiple-generation product line thinking, companies should have a longer product lifespan to allow more time to develop new products, and therefore, they can use resources better, keep their market share, and finally earn an optimal profit level in the long run [3].

Multiple-generation products differ from product families where the company produces different products for the customers with different priorities to cover all market segments. In contrast, multiple-generation products are introduced to the market over time, and all versions aim to cover the same market segment [4,5]. For example, Apple iPhone 12 series are different generations in a product line; although different, they are all designed for the same market niche and target the same customer group. Their core functionality is that they are all high-performance cell phones. In product families, the same product will be offered in different sizes, flavors, colors, and textures, but always the same product. The goal is to provide various products to the market for different market niches.

[*] Corresponding Author

In prior literature, several approaches were used to model the long-term behavior of multiple-generation products, including the Dynamic State Variable Model (DSVM). DSVM can be successfully used for MGPs and effectively forecast the sales performance and the introduction timing of each generation regarding the cannibalization. In their related work, Lin and Okudan applied DSVM using the Apple Inc. iPhone product line with its sales data [3,6]. Lin and Okudan applied DSVM to generate full performance predictions for a new multiple-generation product line, namely Apple iPads, using the historical data from a similar product line if Apple iPhones [6].

To cope with cannibalization, companies need to choose the best pricing strategy to balance the sales from different generations present in the market at each period. Accordingly, the company needs a model that can simulate the cannibalization scenarios within a product line with multiple generations and generate pricing strategies. Lin, Kilicay-Ergin, and Okudan proposed an agent-based model (ABM) that can help companies choose the best pricing strategies to maximize their profit in the long run [7]. They used this model along with a DSVM based on a two-phase methodology. In the first phase, the sales forecast and introduction timings are acquired through the DSVM model. In the second phase, the DSVM outputs are used as the input data for the agent-based model to generate optimal pricing strategies during the product life cycle. Kilicay-Ergin, Lin, and Okudan used their suggested two-phase methodology for Apple Inc.'s iPhone line [8].

With the technology improvement, online reviews are one of the best sources of information on products. This user-generated information is valuable for manufacturers because they can give companies vital feedback for free while achieving this information through surveys or focus groups can cost companies. Companies need to continuously update their decisions and strategies because of the continuous changes in customer preferences to make items attractive or remove product deficiencies.

Parallel simulation models, which can evaluate all the distributed pieces and their complicated interactions concurrently, have numerous advantages over traditional modeling tools for large adaptive systems with multiple distributed interacting parts. Particularly, the agent-based model is a framework capable of simulating these parts (agents), their decisions, their interactions, and the system's overall behavior. The ABM proposed herein considers MGPL as a complex adaptive system and analyzes the whole system's behavior, enabling analysis of pricing strategies [8]. MGPL is considered a complex adaptive system since each agent (generation) in this system is authorized to change its price and adapt to the market considering the current market situation. Each product generation's strategic decisions influence the overall system behavior in this system.

Natural language processing (NLP) enables computers to automatically comprehend text and words as humans do, and sentiment analysis is the use of NLP to identify, extract and study subjective information. Sentiment analysis can be widely applied to analyze reviews or surveys. There is prior research that uses text mining techniques to evaluate customer satisfaction towards products from relevant reviews. There are also papers that use ABM to model MGP pricing scenarios. The main contribution of the present work is to use text mining techniques and pricing scenarios simultaneously to generate appropriate pricing strategies that inherently take customers' opinions into account. We extend the work in [8] by considering the consumer's opinion towards different product generations as a factor using text mining techniques on the same case study (Apple iPhones). Although there are numerous related works, to our knowledge, text mining techniques have never been used in combination with ABMs. In this work, we propose a two-phase framework based on the ABM model to analyze pricing scenarios for companies. In the first phase, we apply text mining and sentiment analysis techniques to extract user reviews from the Apple community forum and evaluate the consumer's satisfaction level for each product generation in each time-period. In the second phase, we use the outputs from the previous step in the agent-based model when generating the best pricing strategies during the product lifecycle. Figure 1 shows these steps respectively.

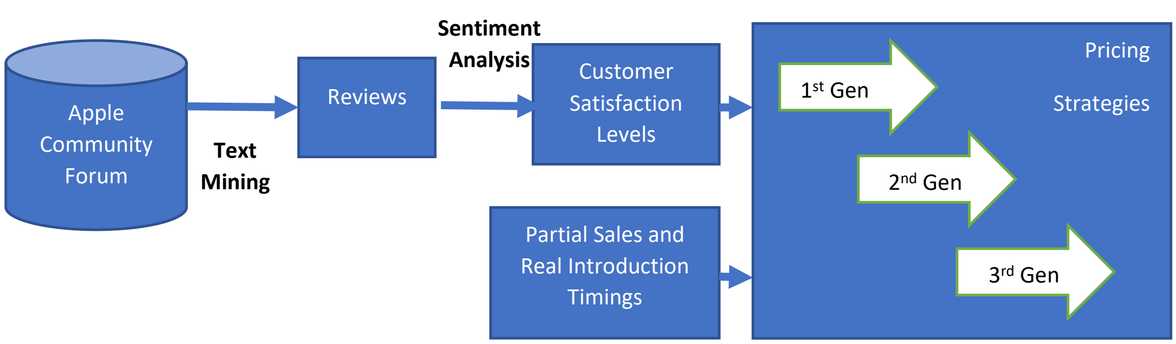

FIGURE 1. OVERVIEW OF AGENT-BASED DYNAMIC PRICING MODEL FOR MULTI-GENERATION PRODUCT LINES (MGPL) WITH THREE GENERATION

II. Literature review

A. Multiple Generation Product Lines and Quantitative Models to Analyze Their Behavior

Due to today's technology-intensive and rapidly changing market environments, applying an MGP strategy becomes a favorable tool for companies. For example, Apple Inc. became one of the most profitable companies in the world with its three well-known MGP lines: The Apple iPods, iPhones, and iPads. Ofek and Sarvary studied the competition between market leaders and followers, created a multi-period Markov game model, and evaluated the effects of innovative advantage, reputation advantage, and advertising on R&D for leaders through their model [9]. Moreover, they investigated appropriate strategies for followers. They discovered that the market leader concentrates on strong demand for its existing product by investing more in advertising, whereas the follower focuses more on R&D. Their analysis gives useful insights for developing a marketing strategy in fast-paced, high-tech business environments. Microsoft (MS) focused its R&D on successive generations of products and applied forward-looking MGP strategies on all its product lines. For instance, when MS launched its Windows 7, its developers had the idea of Windows 8 in their minds. MS looks back on its strategies periodically to review and adjust them if needed [10]. The quantitative models used for multiple-generation products can be categorized into behavioral, dynamic competition, and pricing models. Behavioral models try to interpret the sales behavior of the product during its lifecycle, while dynamic competition models consider the market as a competitive environment and try to generate competitive scenarios for companies. Pricing models tend to create pricing policies and determine optimal pricing strategies for companies. Norton and Bass utilized the bass diffusion model to analyze the MGP sales behavior with respect to the substitution effect among consecutive generations to predict the changes that may happen to future demands of MGPs [11]. In the substitution effect, a portion of the demands from the current product generation is replaced with successive ones. Mahajan and Muller expanded this work by proposing a new demand model that considers the adoption and substitution effects, not only between two consecutive generations but also across different generations, named the "leapfrog" effect [12]. Their model is capable of generating optimal timing strategies regarding cannibalization. Morgan et al. evaluated the quality versus time-to-market trade-offs for MGPs [2]. They proposed an innovative model regarding multiple-generation product lines with different factors, including additional costs, the focal firm's quality, competitive quality, and market share with an active competitor to maximize the total profit.

Krankel et al. used a dynamic programming technique to model MGPs [1]. They proposed a multi-stage decision model to investigate successive introduction timing strategies and forecast future demands while the technology level is additive, and the new generation of the product completely replaces the previous one. Bardhan and Chanda proposed a new behavioral model incorporating the Bass diffusion model and extended [12] by dividing the cumulative adaptors into two categories: First-time purchasers and repeat purchasers and modeling them, respectively [13]. Huang and Tzeng suggested a new two-stage fuzzy piecewise regression for predicting product lifetime and annual MGP shipments [14]. First, the lifetime of each product will be predicted based on the historical data through the proposed regression model, and then the annual shipments of each product will be determined.

Dobson and Kalish developed a tool to help managers in the process of product line design and pricing by providing three types of information: (1) what kind of product customers want, (2) the cost to produce each type of product, and (3) information about their current and future products [15]. They determined the introduction timings and pricing strategies using a heuristic algorithm to maximize the profit. Arslan et al. proposed solutions for firms to manage introducing and pricing their product generations in the competitive market [16]. They analyzed both monopolistic and duopolistic (competitive) environments. They provided introduction timings and optimal pricing policies for two successive product generations under complete replacement and when there are coexisting generations. They proposed a competition model between two firms under complete replacement.

Some pricing models for MGPs are based on the multinomial logit (MNL) models. MNL models assume that the probability of a customer's purchase depends on the customer's utility function for each product [17]. These models perform admirably when the products are distinct from one another. Nonetheless, a number of studies demonstrate the disadvantages of the MNL models when there is a link between product alternatives [17,18]. Kim et al. created a purchase timing and generation choice model for initial and recurring purchases of multiple-generation products [19]. Their model has a logit formulation and models the likelihood of purchasing a generation given prior purchases. Their approach uses individual purchase histories to explain repeating purchase behavior. Schön considered an MNL model with price discrimination among customers [20]. The authors used probabilistic customer choice models to make pricing decisions for MGPs.

Pricing is an effective tool for companies to prevent or at least mitigate problems during an inter-generational product transition since uncertainty in the introduction of a new product may lead to mismatches between demand and supply. The authors considered a transition in which a new generation product replaces an old one, assuming the new one has better features and performance than the old one. They formulated a dynamic pricing problem and derived the optimal pricing policies for both the old and new product generations. In addition to product replacement, they considered several dynamics, such as substitution, external competition, scarcity, and inventory, as well as how these factors affect pricing policies. They also determined the optimal initial inventory for each product [17]. Chen and Chang suggested a dynamic programming model to manage new and remanufactured products by studying their pricing behavior over their life cycle length [21]. The authors considered the new product and the remanufactured product as subsequent generations. The primary purpose of their formulation is to examine the pricing behavior of the product under different parameter settings such as manufacturing and remanufacturing costs, market growth rate, return rate, and substitutability. Fruchter et al. employed genetic algorithms, a mathematical heuristic mimicking the process of biological evolution, to the problem of optimal product line design to generate pricing decisions [22]. Special operators were used to help genetic algorithms mitigate cannibalization. The authors considered the manufacturer's profit as the criteria for fitness-evaluating chromosomes.

Agent-based models (ABM) consist of entities (agents) and a framework to simulate agent decisions and interactions. In a multiple-generation product line, each generation acts as an independent agent and will adjust its sales price according to market demand. These agents' decisions and interactions generate the general behavior of the system. The advantage of using an Agent-based model compared to MNL is that ABM lets decision-makers analyze cannibalization scenarios over the MGP life cycle, while it is challenging to study cannibalization scenarios using MNL since there is a correlation between different generations [8]. Also, the ABM lets the decision-makers understand the implications of varying pricing decisions on the firm's overall profit. We use ABM as the core of our two-phase framework to evaluate different pricing scenarios for Apple iPhones with respect to the importance of consumers' feedback. We need sentiment analysis techniques to gather, integrate, and interpret users' reviews; then, we will use the information extracted from them in our framework.

B. Text Mining and Sentiment Analysis

In the last few decades, a huge amount of information has been generated in text format. Text mining refers to the process of extracting knowledge and information from unstructured text [23]. Zhan et al. analyzed the consumers' reviews using text mining to extract the customers' concerns and summarize topics based on their rankings using an automatic text summarization approach. [24]. The authors compared their method with other approaches, such as opinion mining. Thorleuchter et al. used text mining and text classification (tokenization, term filtering method, Euclidean distance measure, etc.) techniques and a novel heuristic measure for idea mining [25]. The process of extracting new and useful ideas from unstructured text data is known as idea mining. Their evaluated approach is implemented as a web-based application titled "Technological Idea Miner".

Sentiment Analysis (SA) is a method to identify and categorize opinions expressed in a text to determine the sentiment behind it. Using fundamental sentiment analysis, a program can determine whether a text's sentiment is positive, negative, or neutral regarding a special product or service [26]. Specifically, it is a process of analyzing people's opinions and emotions in some special piece of text. There are two primary applications of sentiment analysis [27]. First, sentiment analysis has been applied to documents to differentiate between positive and negative reviews [28-31]. Second, it has been applied at the sentiment level to accomplish some tasks such as multi-perspective question answering and summarization, opinion-oriented information extraction, and customer review mining [27, 32-35].

Sentiment analysis starts with sentiment expressions existing in the specified object and then recognizes positive from negative words and phrases [36]. Lexicons can be categorized into three types: Positive polarity (e.g., excellent, great, perfect), negative polarity (e.g., bad, terrible, awful), and contextual polarity (i.e., words with different meanings in different contexts). Turney used a simple method to classify reviews into two types: recommended and not recommended, according to their average semantic orientation value of phrases containing adjectives or adverbs [31]. Pang and Li used a machine learning technique to apply text categorization methods to the subjective part of any document based on the minimum number of cuts [29]. Beineke et al. considered the traditional sentiment classification method but as a Naïve-Bayes model [28].

Some other works focused on sentiment analysis at the sentence or phrase level. Wilson et al. proposed a new approach to classify the expressions into neutral or polar and disambiguate the polar expressions at the phrase level [27]. Recently, Täckström and McDonald developed two semi-supervised latent variable models for sentiment analysis at the sentence level [37]. Mostafa used text mining techniques on 3516 tweets during a specific time period to evaluate the consumers' sentiment towards brands such as Nokia, T-Mobile, IBM, KLM, and DHL [38]. The author showed that there is positive sentiment toward some famous brands. Kontopoulos et al. suggested an ontology-based method for sentiment analysis of Twitter posts [39]. In their approach, each distinct notion in a post will receive a sentiment grade instead of assigning a total sentiment score to the post. This can finally lead to a more detailed analysis of post opinions regarding a specific topic. There are some works about the details of the sentiment analysis approach. Deng et al. introduced a strategy that gives terms specific weights to improve sentiment analysis performance [40]. They proposed a supervised term weighting scheme based on two main factors: The importance of a term in a document (ITD) and the importance of a term for expressing sentiment (ITS). The authors introduced seven statistical functions that learn the sentiment importance of a term through its statistical distribution in positive and negative documents.

Although the studies mentioned above focused on sentiment analysis at document and sentence levels, they are only about techniques, not their usage in combination with different frameworks. These studies cannot derive useful information about the usage of Text mining for customer satisfaction evaluation. Kang and Park proposed a sentiment-analysis-based framework for measuring customer satisfaction using the VIKOR approach [42]. They applied their framework to customer reviews of mobile application services as a case study. We propose a sentiment analysis-based framework in the first phase of our innovative two-phase framework to measure the customer satisfaction level of different generations of Apple iPhones. In the second phase, we will generate pricing strategies for the company using the outputs from the first step.

Aspect-level or aspect-based sentiment analysis (ABSA) lets companies conduct a comprehensive analysis of their customer's feedback data, enabling them to gain a deeper understanding of their customers and develop products and services that better meet their needs. ABSA is a method for categorizing data by aspect and identifying the sentiment associated with each. Aspects are the attributes or components of a product or service2. Since deep learning approaches have emerged as potential methods for achieving objectives in ABSA and their ability to capture both syntactic and semantic features of text without the need for high-level feature engineering, a comparative review of deep learning for ABSA has been provided [43]. The concerns and challenges associated with extracting distinct sentiments from various aspects and establishing relational mappings between aspects, dependencies, and interactions were highlighted in a survey. A comprehensive summary of recent developments has been provided along with their performance outcomes, demonstrating the quantitative assessment of the proposed methodology [44]. In our case study, aspects are some iPhones' features, such as the camera, operating system speed, and stylishness. Features were categorized into three groups. Software-related aspects, hardware-related aspects, and those that do not fit into one of the two groups. In other words, the aspects that can be categorized into both groups.

III. METHODOLOGY

We present a two-phase methodology for determining the lifetime profitability of each generation in a multiple-generation product line in this study. The proposed model's purpose is to generate pricing strategies that can give the company a good prediction of the actual prices and maximize overall profit over the life of MGPL. Figure 2 depicts the objectives of each step.

In the second phase of our framework, we used available partial sales data to get the sales of generations during time periods. According to Lin and Okudan [3,6], when a new generation enters the market, we allocate 60% of the sales for that quarter to the new generation and the remainder to previous generations. Prior generation sales are declining at a rate of 20% across subsequent quarters. Table I displays sales data3 for all generations within specified time periods.

TABLE I
SALES PER PRODUCT GENERATION FOR FIVE GENERATIONS OF IPHONES

Period	**Quarter**	**Gen. 8** (8/ 8 Plus)	**Gen. 9** (XS, XS Max)	**Gen. 10** (XR)	**Gen. 11** (11/ 11 Pro/ 11 ProMax)	**Gen. 12** (12/ 12 Pro/ 12 ProMax)	**Gen. 13** (13/ 13 Pro/ 13 ProMax)
1	**2019 Q1**	10,944,000	16,416,000	41,040,000			
2	**2019 Q2**	1,427,200	13,132,800	21,840,000			
3	**2019 Q3**	3,013,760	10,506,240	20,280,000			
4	**2019 Q4**		2,416,000	16,224,000	27,960,000		
5	**2020 Q1**		11,808,000	17,712,000	44,280,000		
6	**2020 Q2**			14,680,000	22,020,000		
7	**2020 Q3**			15,040,000	22,560,000		
8	**2020 Q4**				16,640,000	24,960,000	
9	**2021 Q1**				36,040,000	54,060,000	
10	**2021 Q2**				22,080,000	33,120,000	
11	**2021 Q3**				17,680,000	26,520,000	
12	**2021 Q4**					20,160,000	30,240,000

[2] https://monkeylearn.com/blog/aspect-based-sentiment-analysis/
[3] https://en.wikipedia.org/wiki/IPhone

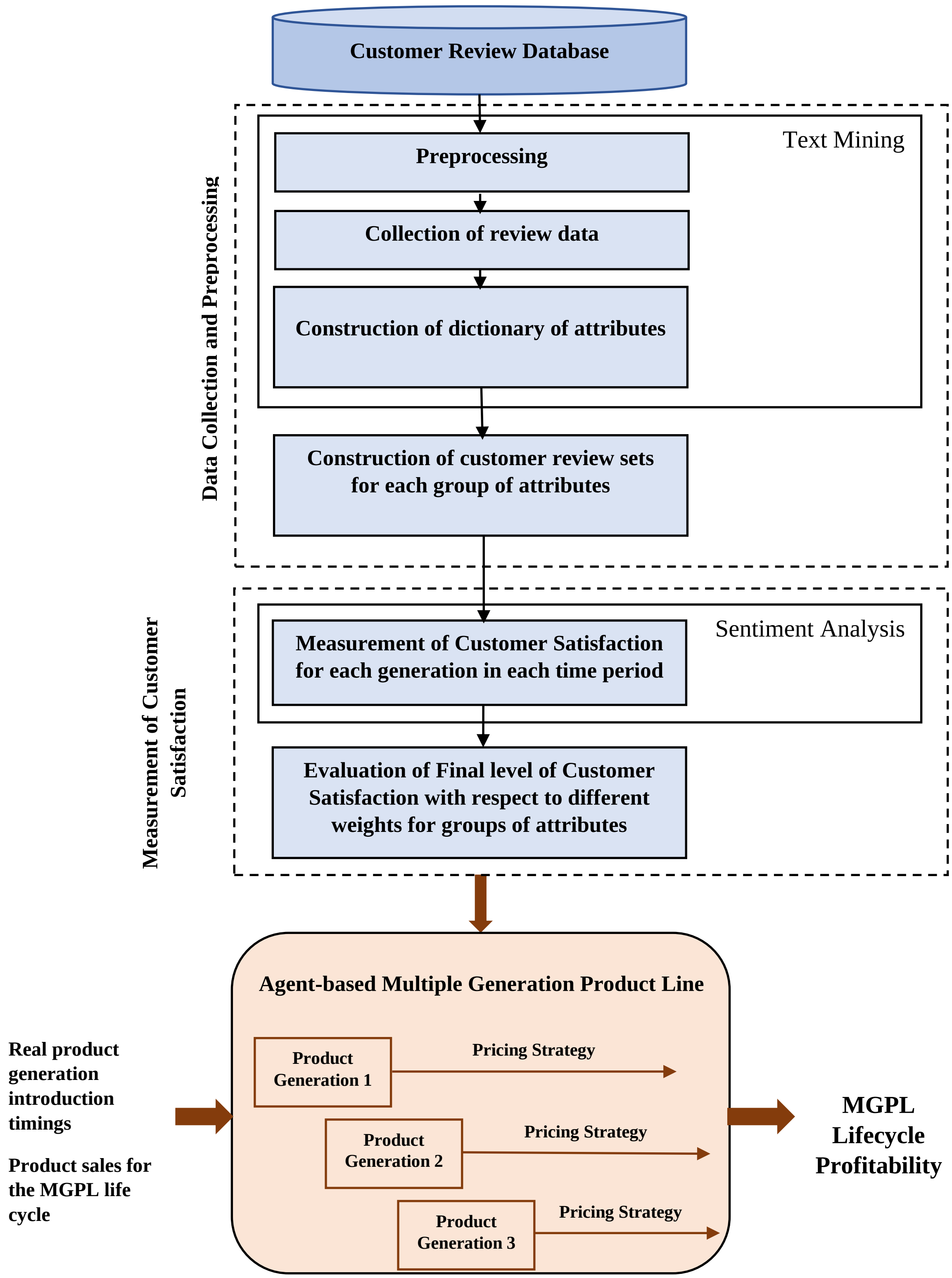

FIGURE 2. FLOW OF THE PROPOSED TWO-PHASE FRAMEWORK THAT COMBINES TEXT MINING AND SENTIMENT ANALYSIS TO PROVIDE INPUT TO AN AGENT-BASED MODEL OF A MULTIPLE GENERATION PRODUCT LINE, RESULTING IN ANALYSIS AND INSIGHTS INTO ITS LIFECYCLE PROFITABILITY

Sentiment analysis is a popular text-mining technique that reads emotional content from massive amounts of data and translates it into relevant consumer feedback [41]. Sentiment analysis has grown in popularity due to its easy-to-use results, and it has been used in various industries where user opinion is crucial [32,45]. Web-based information gives the same amount of information as sales statistics, surveys, and focus groups; however, web-based information is less structured [46]. Herein, we use sentiment analysis algorithms to analyze customer feedback from each iPhone generation and apply them to the Apple Community Forum for iPhone. The differences between iPhone versions are then investigated to trace the advancement of iPhone devices. These attempts are intended to provide an answer to the following research question: Can we improve pricing strategies by using novel sentiment analysis algorithms? Can we assist businesses in determining which aspects of their business need to be improved in the eyes of their customers?

We map sentiment-loaded words to the typical sentiment scale. The results will show the sentiment analysis's preliminary efficacy for application in multi-generation product lines. In the first stage, we use the R Programming Language to do sentiment analysis on customer reviews. The Apple community forum4 is where consumer reviews in 2021 are gathered. The iPhone reviews are sorted and categorized into several groups based on their time and generation. To find the sentiments associated with reviews, sentiment analysis techniques are applied at the word level. When the reviews are split down into words, stop words are removed. Each review was converted into a vector of words and sentiments. As a result, at this point, a technique for assessing feelings is required.

There are numerous approaches for assessing emotion in text data. The R programming environment's "tidytext" package provides access to numerous sentiment lexicons. AFINN5, bing6, and nrc7 are three general-purpose lexicons. All these lexicons are based on single words, and words are scored for positive/negative sentiment as well as emotions such as joy, anger, sadness, and so on. The "nrc" lexicon divides words into binary ("yes"/"no") categories: positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. The "bing" lexicon classifies terms into positive and negative categories on a binary basis. The "AFINN" lexicon gives a score to each term ranging from -5 to 5, with negative scores representing negative sentiments and positive ones representing positive sentiments8. Our sentiment analysis methodology takes advantage of the "bing" lexicon results. This lexicon tells us how many positive and negative terms are linked with each group.

The second phase involves the implementation of an agent-based model (ABM) with numerous agents representing different generations of the MGPL. In ABM, each generation of the MGPL is regarded as an independent agent with the ability to set its own sales price at any time. The following assumptions explain this phase: (1) When the product first hits the market, each generation has its initial price. (2) Lowering the sales price has a beneficial impact on product sales. (3) The price change's effect on product sales is believed to be known. Table II contains all the parameters and variables used in this phase. Each product generation in this model follows predetermined stochastic principles. Each agent monitors if product sales are increasing or decreasing at each time t and whether lowering the sales price would result in more profit. If the response is yes, the agent will decide to lower its price. When the product's sales price falls, the product's sales volume adjusts at a randomly chosen rate. The products of competitors are not included in this model.

TABLE II
PARAMETERS USED IN THE PROPOSED MODEL

Symbol	Meaning
$p_n^{(t)}$	Product unit price of product generation n
a_{inc}	Product sales adjustment rate when product sales are in an increasing manner
a_{dec}	Product sales adjustment rate when product sales are in a decreasing manner
C	Product unit cost
$S_n^{(t)}$	Product sales forecast of product generation n at time t
n	Product generation number, n=1, 2, …, N
d_g	Product price discount rate for general case

[4] https://discussions.apple.com/community/iphone/iphone_hardware?page=1
[5] From Finn Arup Nielsen
[6] From Bing Liu and collaborators
[7] From Saif Mohammad and Peter Turney
[8] https://www.tidytextmining.com/sentiment.html

d_n	Product price discount rate when a new generation of product is introduced to the market
$Pr_n^{(t)}$	Expected profit for product generation n at time t
θ	Cannibalization threshold for the product price difference, θ<1
$Can(t)$	Cannibalization sales reduction rate
$TP_m^{(t)}$	M different conditions of total profit for product generation (n-1) and n at time period t
$TM_n^{(t)}$	Customer satisfaction parameter for generation n at time t
$f1_n^{(t)}$	Number of Positive terms in reviews associated with generation n at time t
$f2_n^{(t)}$	Number of Negative terms in reviews associated with generation n at time t

If a new generation joins the market, all prior generations, except for the last, will reduce the price of their items at a predetermined rate. Furthermore, the overall volume of all product sales should not exceed the company's production capacity within any given time. Finally, we anticipate gaining some useful pricing strategies for each generation of the multiple-generation product line at each stage of its existence. We anticipate that in comparison to Lin et al.'s work [7], we will find superior Pricing Strategies for each generation of the MGPL. We expect higher sales prices for generations with more positive reviews and lower prices for generations with fewer positive reviews.

We considered $f1_n^{(t)}$ and $f2_n^{(t)}$as the number of positive and negative terms that come from using the Bing lexicon on the related reviews. Related reviews include generation *n* of iPhone and submitted at time *t*. $TM_n^{(t)}$ will be as follows:

$$TM_n^{(t)} = a_1 * f1_n^{(t)} + a_2 * f2_n^{(t)}. \quad (1)$$

A simple linear model was used to create the $TM_n^{(t)}$ formula. This method employs direct counts of comments rather than the ratio of positive versus negative comments or the ratio of positive comments overall. a_1 and a_2 are coefficients (weights) of $f1_n^{(t)}$ and $f2_n^{(t)}$ in the linear model. Different values for two parameters in the linear model will be tested and evaluated by RMSE to find the best values. As a result, the more comments about a product, the greater the upward adjustment to the price predicted by ABM. Moreover, the number of negative reviews can increase the price on their own. We must emphasize that as more data becomes available, these weights will not only need to be adjusted, but it also implies the need for a more complicated formula in the future to avoid above mentioned problems. For example, one generalization could be:

$$TM_n^{(t)} = a_1 * \min\left\{B_1, f1_n^{(t)}\right\} + a_2 * \min\left\{B_2, f2_n^{(t)}\right\}. \quad (2)$$

Negative comments could never increase the price on their own if $B_2 < 1/a_2$. This above formula necessitates fitting values for four parameters. We leave such a more complex formula as future work due to the lack of enough data at the present time. We also emphasize that the main point of this work is to determine whether sentiment analysis explains most of the differences between the ABM and the real prices. With additional data, it may be possible to improve this formula.

Decision rules for previous generations (k-2, k-3, …, 1)

In time period t, with the latest generation k:

Step 1: sales are increasing.

$$S_n^{(t)} - S_n^{(t-1)} \geq 0$$

For n< k-1 (k-2, k-3, …, 1) if:

$$\left[\boldsymbol{P_n^{(t-1)}} - \boldsymbol{C}\right] \times \boldsymbol{S_n^{(t)}} \leq \{\left[\boldsymbol{P_n^{(t-1)}} \times \boldsymbol{d_g}\right] - \boldsymbol{C}\} \times \boldsymbol{S_n^{(t)}} \times \boldsymbol{a_{inc}}$$

And $\boldsymbol{P_n^{(t-1)}} \times \boldsymbol{d_g} > \boldsymbol{C}$

Then,

$$\boldsymbol{P_n^{(t)}} = \boldsymbol{P_n^{(t-1)}} \times \boldsymbol{d_g} \times \boldsymbol{TM_n^{(t-1)}} \tag{3}$$

Otherwise,

$$\boldsymbol{P_n^{(t)}} = \boldsymbol{P_n^{(t-1)}} \times \boldsymbol{TM_n^{(t-1)}} \tag{4}$$

Step 2: sales are decreasing.

$$S_n^{(t)} - S_n^{(t-1)} < 0$$

For n< k-1 (k-2, k-3, …, 1) if:

$$\left[\boldsymbol{P_n^{(t-1)}} - \boldsymbol{C}\right] \times \boldsymbol{S_n^{(t)}} \leq \{\left[\boldsymbol{P_n^{(t-1)}} \times \boldsymbol{d_g}\right] - \boldsymbol{C}\} \times \boldsymbol{S_n^{(t)}} \times \boldsymbol{a_{dec}}$$

And $\boldsymbol{P_n^{(t-1)}} \times \boldsymbol{d_g} > \boldsymbol{C}$

Then,

$$\boldsymbol{P_n^{(t)}} = \boldsymbol{P_n^{(t-1)}} \times \boldsymbol{d_g} \times \boldsymbol{TM_n^{(t-1)}} \tag{5}$$

Otherwise,

$$\boldsymbol{P_n^{(t)}} = \boldsymbol{P_n^{(t-1)}} \times \boldsymbol{TM_n^{(t-1)}} \tag{6}$$

Step 3: When a new generation k+1 comes to the market, older ones n< k+1 adjust their prices:

$$P_n^{(t)} = P_n^{(t)} \times d_n \tag{7}$$

For two recent generations (k and k-1)

We assume that cannibalization has just occurred for the last two generations. Three new parameters have been added to the model. The cannibalization threshold for the product price difference Ɵ. The cannibalization sales reduction rate $Can(t)$ and $TP_m(t)$, where m denotes various conditions of total profits for the last two product generations at time t. In other words, m states are indicative of market circumstances. There are four different cases:

1) Discount on previous generation (k-1)
2) Discount on current generation (k)
3) Discount on both generations (k-1, k)
4) No discount on both generations.

The case with the maximum profit will be selected and continued. In each case, we should check whether the sales price ratio between the previous generation and the current one is above Ɵ (cannibalization threshold).

Step 1: the sales are in an increasing manner.

$$S_n^{(t)} - S_n^{(t-1)} \geq 0$$

For $n \geq k-1\ (k-1, k)$

Case 1: No price or sales volume changes across two generations, and cannibalization may occur based on Ɵ.

If:

$$[P_{k-1}^{(t-1)} < P_k^{(t-1)}\ and\ (\frac{P_{k-1}^{(t-1)}}{P_k^{(t-1)}}) < \theta]$$

Or,

$$[P_{k-1}^{(t-1)} > P_k^{(t-1)}\ and\ (\frac{P_k^{(t-1)}}{P_{k-1}^{(t-1)}}) > \theta]$$

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + [P_k^{(t-1)} \times TM_k^{(t-1)} - C] \times S_k^t,$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{8}$$

Otherwise,

$$(\frac{P_{(k-1)}^{(t-1)}}{P_k^{(t-1)}}) < \theta$$

Then:

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)],$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{9}$$

Or:

$$(\frac{P_k^{(t-1)}}{P_{k-1}^{(t-1)}}) < \theta$$

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{10}$$

Case 2: Discount on the previous generation (k-1) with the discount rate d_g. (The previous generation may be so much cheaper as customers prefer the previous generation to the new one, so the new generation will be cannibalized by the previous one.)

$$(\frac{[P_{k-1}^{(t-1)} \times d_g]}{P_k^{(t-1)}}) > \theta$$

Then:

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)},$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{11}$$

Otherwise,

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \tag{12}$$

Case 3: Discount on the new generation (k) with the discount rate d_g. (The new generation may be so much cheaper as customers prefer the new generation to the previous one, so the previous generation will be cannibalized by the new one.)

If:

$$\left(\frac{P_{k-1}^{(t-1)}}{[P_k^{(t-1)} \times d_g]}\right) < [\frac{1}{\theta}]$$

Then:

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{13}$$

Otherwise,

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{14}$$

Case 4: In this case, there is a discount on both recent generations (k, k-1), and they will both enter cannibalization circumstances.

If:

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) > \theta \text{ and } \left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) > \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{15}$$

Otherwise,

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) < \theta$$

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} \times [1 + can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{16}$$

Or:

$$\left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) < \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{inc} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{inc} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{17}$$

The case with Max $TP_m(t)$, m= 1 to 4 will be chosen.

Step 2: the sales are in a decreasing manner.

$$S_n^{(t)} - S_n^{(t-1)} < 0$$

For $n \geq k - 1 \; (k - 1, k)$

Case 1: No price or sales volume changes across two generations, and cannibalization may occur based on ϴ.
If:

$$[P_{k-1}^{(t-1)}/P_k^{(t-1)}] > \theta$$

Then:

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + [P_k^{(t-1)} \times TM_k^{(t-1)} - C] \times S_k^t,$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \quad (18)$$

Otherwise,

$$(\frac{P_{(k-1)}^{(t-1)}}{P_k^{(t-1)}}) < \theta$$

Then:

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)],$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \quad (19)$$

Or:

$$(\frac{P_k^{(t-1)}}{P_{k-1}^{(t-1)}}) < \theta$$

$$TP_1(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \quad (20)$$

Case 2: Discount on the previous generation (k-1) with the discount rate d_g. (The previous generation may be so much cheaper as customers prefer the previous generation to the new one, so the new generation will be cannibalized by the previous one.)

$$(\frac{[P_{k-1}^{(t-1)} \times d_g]}{P_k^{(t-1)}}) > \theta$$

Then:

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)},$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \quad (21)$$

Otherwise,

$$TP_2(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} \times [1 + can(t)] + \left[P_k^{(t-1)} \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times TM_k^{(t-1)} \quad (22)$$

Case 3: Discount on the new generation (k) with the discount rate d_g. (The new generation may be so much cheaper as customers prefer the new generation to the previous one, so the previous generation will be cannibalized by the new one.)

If:

$$\left(\frac{P_{k-1}^{(t-1)}}{[P_k^{(t-1)} \times d_g]}\right) < [\frac{1}{\theta}]$$

Then:

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{23}$$

Otherwise,

$$TP_3(t) = \left[P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{24}$$

Case 4: In this case, there is a discount on both recent generations (k, k-1), and they will both enter cannibalization circumstances.

If:

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) > \theta \text{ and } \left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) > \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec}$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{25}$$

Otherwise,

$$\left(\frac{P_{k-1}^{(t-1)} \times d_g}{[P_k^{(t-1)} \times d_g]}\right) < \theta$$

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} \times [1 + can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec} \times [1 - can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{26}$$

Or:

$$\left(\frac{P_k^{(t-1)} \times d_g}{[P_{k-1}^{(t-1)} \times d_g]}\right) < \theta$$

Then:

$$TP_4(t) = \left[P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)} - C\right] \times S_{k-1}^{(t)} \times a_{dec} \times [1 - can(t)] + \left[P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} - C\right] \times S_k^{(t)} \times a_{dec} \times [1 + can(t)]$$

$$P_{k-1}^{(t)} = P_{k-1}^{(t-1)} \times d_g \times TM_{k-1}^{(t-1)}, P_k^{(t)} = P_k^{(t-1)} \times d_g \times TM_k^{(t-1)} \tag{27}$$

The case with Max $TP_m(t)$, *m*= 1 to 4 will be chosen.

Step 3: When a new product generation (*n*= *k*+1) comes to the market, the previous ones (*n*< *k*+1) adjust their unit sales price to be:

$$p_n^{(t)} = P_n^{(t)} \times d_n \tag{28}$$

IV. RESULTS

A. *Text Mining Result*

In the first phase we extract 1591 reviews for 2021 from the Apple community forum. The reviews were first filtered based on containing the word iPhone, as we do not want to include reviews associated with other products such as MacBooks and iPods. The reviews were then split based on their time into four groups according to yearly quarters. Each of the four groups is further split into different subgroups based on the product generations. We thus have different groups of reviews with time and generation. Reviews are preprocessed and tokenized. Using the Bing lexicon for the reviews in each group, we can get the number $f1_n^{(t)}$ of positive and $f2_n^{(t)}$ of negative terms. Recall that we define a sentiment analysis factor $TM_n^{(t)}$ as a linear combination of these counts: $TM_n^{(t)} = a_1 * f1_n^{(t)} + a_2 * f2_n^{(t)}$. Here a_1 and a_2 are weights determined by minimizing the RMSE of the difference between model predictions of price and the observed price. In this case, the best weights were obtained as $a_1 = 0.1269$ and $a_2 = 0.0340$, resulting in the following linear expression for $TM_n^{(t)}$:

$$TM_n^{(t)} = 0.1269 * f1_n^{(t)} + 0.0340 * f2_n^{(t)}$$

Table III reports the TM parameter value for each time and generation. Note that when there are no reviews, the parameter is automatically set equal to one, as no adjustments should be made to the price according to reviews if no reviews are available.

TABLE III

NUMBER OF POSITIVE AND NEGATIVE TERMS AND THE SENTIMENT ANALYSIS PARAMETERS FOR DIFFERENT GENERATIONS IN DIFFERENT TIMES

	2021-Q1		**2021-Q2**		**2021-Q3**		**2021-Q4**	
	Gen 11	**Gen 12**	**Gen 11**	**Gen 12**	**Gen 11**	**Gen 12**	**Gen 12**	**Gen 13**
$\boldsymbol{f1_n^{(t)}}$	-	-	3	-	3	8	58	54
$\boldsymbol{f2_n^{(t)}}$	-	-	21	-	12	3	95	80
$\boldsymbol{TM_n^{(t)}}$	1	1	1.09	1	1.16	1.11	10.59	9.57

Table IV illustrates the sensitivity of the prediction RMSE to the different values of weights. We note that how the text mining results are used greatly influences the quality of the predictions. By tuning the weights, the RMSE can be reduced by two orders of magnitude.

TABLE IV

RMSE FOR DIFFERENT VALUES OF THE WEIGHTS $\boldsymbol{a_1, a_2}$ USED TO CALCULATE THE SENTIMENT ANALYSIS PARAMETER

RMSE		$\boldsymbol{a_2 =}$				
		0.014	**0.024**	**0.034**	**0.044**	**0.054**
$\boldsymbol{a_1 =}$	**0.0269**	504.8	441.7	389.9	354.6	312.2
	0.0769	330.2	290.6	258.8	236.8	191.3
	0.1269	191.4	99.9	9.1	24.4	32.1
	0.1769	193.8	165.8	138.7	172.6	204.1

Further analysis of the reviews provides additional insights. Topic modeling was done on the reviews with the different number of topics, and five is the number of topics that fits the data best. We experimented with the number of topics, and five was chosen because the frequent terms in each topic were mostly related, and the result was more meaningful. From the most frequent terms in each topic, we removed terms such as verbs and adverbs and concentrated on terms that refer to features of iPhones. Figure 3 shows the most frequent terms in each topic.

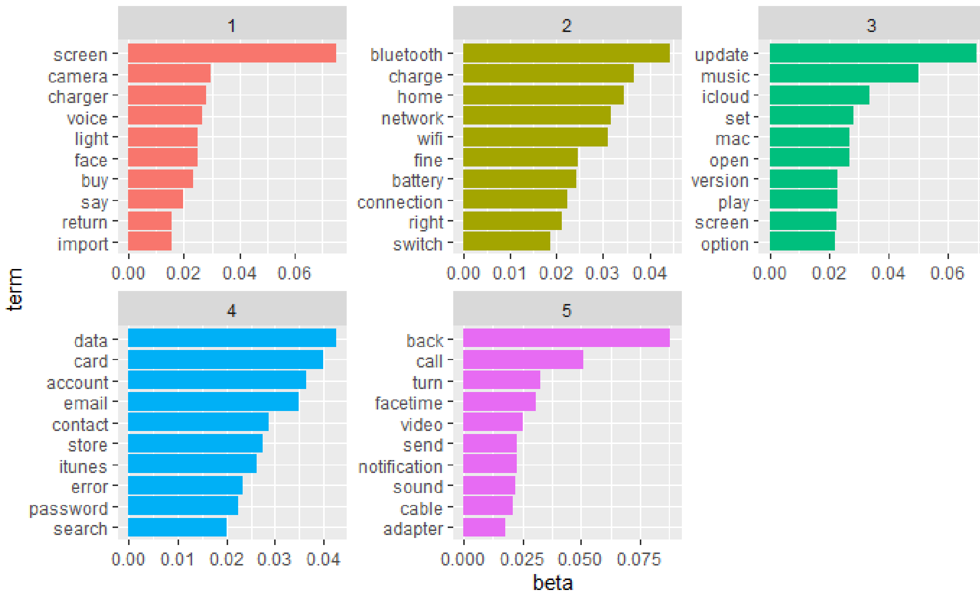

FIGURE 3. TOPIC MODELLING OUTPUT WITH 5 TOPICS

These most frequent terms (features) have been categorized into three groups: (1) hardware-related features, (2) software-related features, and (3) between-group features. The first and second group contains features related to iPhone hardware and software respectively. The terms in the third group are those that refer to both the hardware and software of iPhones, like Bluetooth, which can declare something about hardware or software. The reviews are split into these three groups and further divided based on the product generations. This results in the number of positive and negative terms in each group for different generations. Figures 4-6 show the details from our analysis for generations in each of the three groups.

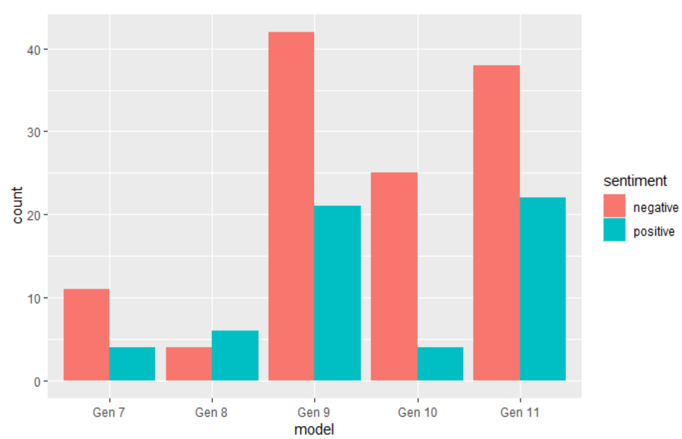

FIGURE 4. NUMBER OF POSITIVE AND NEGATIVE TERMS FOR DIFFERENT GENERATION EXTRACTED FROM REVIEWS ASSOCIATED WITH HARDWARE FEATURES

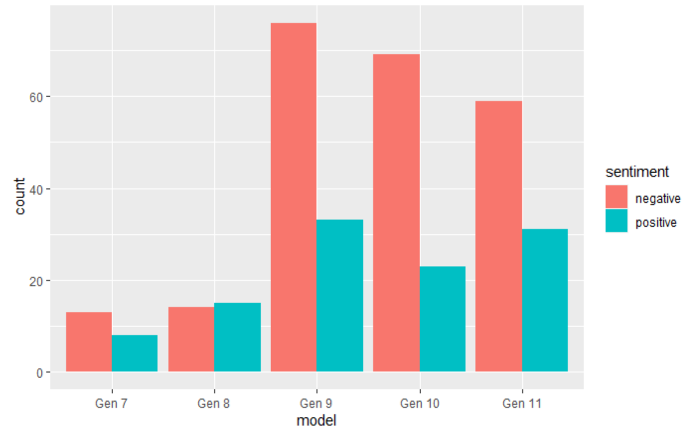

FIGURE 5. NUMBER OF POSITIVE AND NEGATIVE TERMS FOR DIFFERENT GENERATION EXTRACTED FROM REVIEWS ASSOCIATED WITH SOFTWARE FEATURES

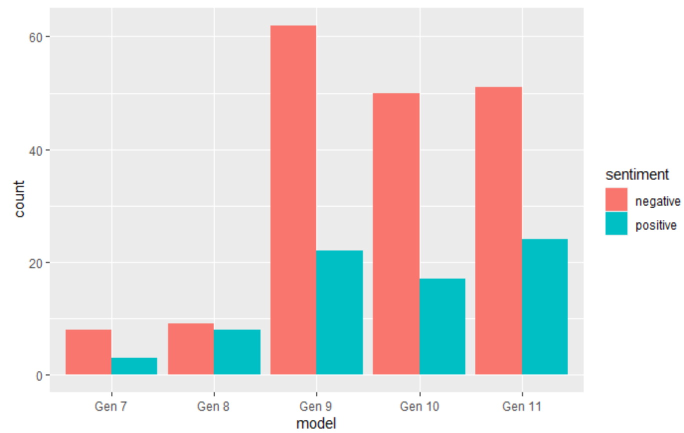

FIGURE 6. NUMBER OF POSITIVE AND NEGATIVE TERMS FOR DIFFERENT GENERATION EXTRACTED FROM REVIEWS ASSOCIATED WITH BETWEEN-GROUP FEATURES

After this step, with the help of Bing lexicon results, we can get three *F* Parameters for hardware, software, and between-group of each generation. For each of these three groups, the *F* parameter is the ratio of positive reviews to all reviews in that relative group:

$$F_{(n)} = \frac{\sum_t f1_n^{(t)}}{\sum_t f1_n^{(t)} + \sum_t f2_n^{(t)}}. \tag{29}$$

Table V shows the values of *F* parameters for different generations in 2021. We also suggest a total *F* parameter, which will be calculated below. Since we do not have a basis for giving different weights to different components, we use an unweighted average:

$$F_{total(n)} = \frac{1}{3} \times F_{Hardware(n)} + \frac{1}{3} \times F_{Software(n)} + \frac{1}{3} \times F_{Between-group(n).} \tag{30}$$

TABLE V
ALL F PARAMETERS FOR DIFFERENT GENERATIONS IN 2021

	$F_{Hardware(n)}$	$F_{Software(n)}$	$F_{Between-group(n)}$	$F_{total(n)}$
Gen 7	0.27	0.38	0.27	0.307
Gen 8	0.6	0.52	0.47	0.53
Gen 9	0.33	0.3	0.26	0.297
Gen 10	0.14	0.25	0.25	0.213
Gen 11	0.37	0.34	0.32	0.343

These results show the company how much consumers in 2021 were satisfied with each iPhone generation. The values reported in Table V show that generation 8 has the highest F_{total} parameter. So, this generation has the highest level of customer satisfaction in comparison with other generations. After that, generations 11, 7, 9, and 10 have the next levels of customer satisfaction, respectively. It can help companies find which generations need more improvement or how much customers are satisfied with each generation of iPhones. Moreover, F parameters related to hardware, software, and between-group features show the company which of these three different areas needs more improvement. For example, in generation 10 parameters, $F_{Hardware}$ is the lowest one in comparison with the others. It can help companies to find that for generation 10; they need to focus on hardware more than software and between group features depending on customer's opinions.

B. ABM Result

With all the available agent-based-model (ABM) inputs, we can get the pricing strategies for different generations in each time period. Using the proposed ABM, companies can get the best pricing strategies with the highest profits for different generations in each quarter of 2021. We used the NetLogo software to simulate the model. We used the sales data in Table I in the Agent-based model, and one sale unit equals 1,000,000 sale volume in real data. Table VI summarizes the parameters used in the simulation experiment.

TABLE VI
PARAMETERS USED IN THE SIMULATION EXPERIMENT.

Symbol	**Meaning**	**Simulation Value**
$p_n^{(t)}$	Product unit price of product generation n	-
a_{inc}	Product sales adjustment rate when product sales are in an increasing manner	0.5
a_{dec}	Product sales adjustment rate when product sales are in a decreasing manner	0.7
C	Product unit cost	-
d_g	Product price discount rate for general case	0.7
d_n	Product price discount rate when a new generation of product is introduced to the market	0.8
θ	Cannibalization threshold for the product price difference, θ<1	0.4
$Can(t)$	Cannibalization sales reduction rate	0.6

Real initial prices were used for $p_n^{(t)}$ when a new generation came to the market for the first time. The costs of different generations were assumed to be known based on a linear relation with the initial prices. Different values for d_g and d_n were tested and evaluated by considering the fit between the prices predicted by the model and the observed prices. The values that minimized the RMSE were found to be 0.7 and 0.8 for d_g and d_n, respectively. In the simulation experiment, the total profit from the current and previous generations is assessed under the four cannibalization conditions. In the first experiment, we used the base ABM model with the above parameters to monitor the total profit of the two latest generations under four cannibalization scenarios. Figure 7 illustrates the total profit from the two latest generations under different scenarios.

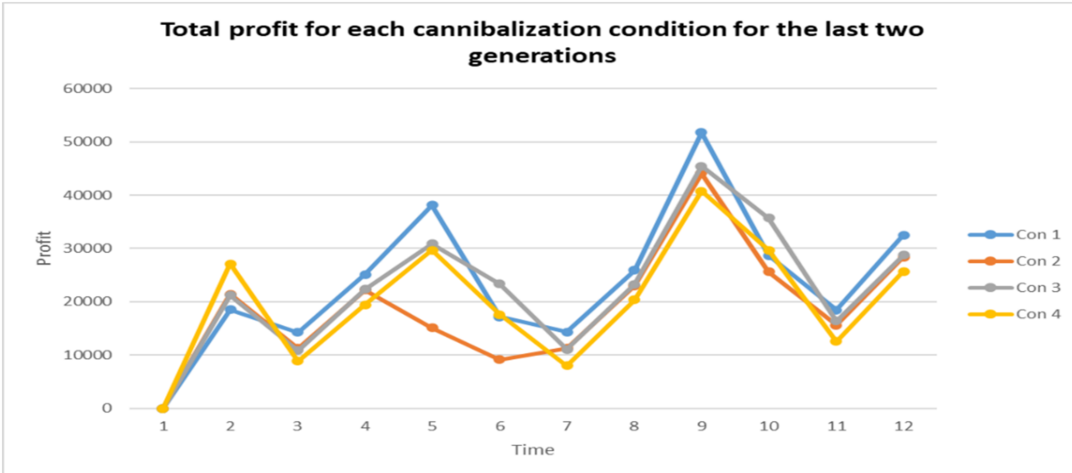

FIGURE 7. TOTAL PROFIT FOR EACH CANNIBALIZATION CONDITION FOR THE LAST TWO GENERATIONS

In simulation experiment 2, we simulated the model with the defined parameters in Table VI, and we added the TM parameter to the model with the related values from Table III. We considered $TM_n^{(t)} = 1$ whenever we did not have a determined value for the TM parameter. Figure 8 shows the results of this second experiment.

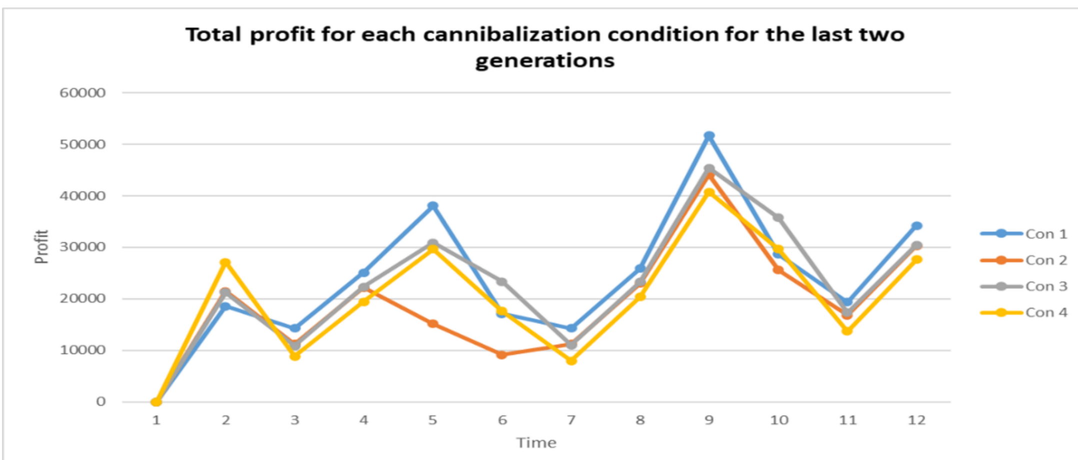

FIGURE 8. TOTAL PROFIT FOR EACH CANNIBALIZATION CONDITION FOR THE LAST TWO GENERATIONS

The results show that two cannibalization conditions (conditions 1 and 3) generate more profit in comparison with others (conditions 2 and 4). Condition 1 is when there is no discount on both generations, but cannibalization occurs based on the cannibalization threshold. Condition 3 is the case where there is a discount on the most recent generation, which results in the new generation cannibalizing the previous one.

After adding the defined parameters to the model, results show that at the latest periods, with the help of the text mining parameter, the profit was increased. Moreover, the simulation gives us price changes based on choosing the best strategy with the maximum profit in each time period. In the table below, the price changes for generations have been listed from the first simulation experiments (without using the text mining parameter) and compared with the actual prices over time.

TABLE VII
PRICE CHANGES DURING TIME FROM SIMULATION EXPERIMENTS IN COMPARISON WITH REAL ONES

Price changes													
Time		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Gen 8	Predicted price without TM	650	650	650									
	Actual Price	650	650	650									
Gen 9	Predicted price without TM	1049	1049	1049									
	Actual Price	1049	1049	1049									
Gen 10	Predicted price without TM	749	749	749	599.2	599.2	599.2	599.2					
	Actual Price	749	749	749	600	600	600	600					
Gen 11	Predicted price without TM				932	932	932	932	745.6	745.6	745.6	745.6	
	Actual Price				932	932	932	932	830	830	830	830	
Gen 12	Predicted price without TM								976	976	976	976	780.8
	Actual Price								976	976	976	976	880
Gen 13	Predicted price without TM												976
	Actual Price												976

Regarding the availability of text mining parameters for some generations in the latest periods, we can see the changes for Gen 5 and 6 in Table VIII. The actual prices were extracted from the GSMArena9 website.

TABLE VIII
PRICE CHANGES DURING THE LATEST TIME PERIODS FROM TWO SIMULATION EXPERIMENTS IN COMPARISON WITH REAL ONES

Price changes			
Time		**11**	**12**
	Predicted price without TM	**745.6**	
Gen 11	Predicted price with TM	**816.2**	
	Actual price	**830**	
Gen 12	Predicted price without TM	**976**	**780.8**
	Predicted price with TM	**976**	**872.3**
	Actual price	**976**	**880**
Gen 13	Predicted price without TM		**976**
	Predicted price with TM		**976**
	Actual price		**976**

Table VIII shows that the text mining parameter explains most of the difference between the ABM prediction and the actual prices. These modifications are demonstrated by the reduction in RMSE from 75.19 to 9.13 or an almost 88% reduction in RMSE. Furthermore, as we carefully tuned the key parameters to reduce the RMSE of the ABM model on its own, we claim that this type of decrease could not be accomplished merely by better implementing ABM without the TM and that combining the ABM with text mining resulted in the improvement.

[9] https://www.gsmarena.com/charts_show_the_evolution_of_iphone_prices_over_the_years_and_show_how_well_older_phones_held_their_-news-51097.php#image0

V. Discussion

This paper builds on the two-phased method proposed by Kilicay-Ergin, Lin, and Okudan's study, where the authors developed pricing strategies for MGPs to maximize their lifetime profitability using their proposed two-phase methodology [8]. Although they have good pricing strategies, there are several circumstances that can compel corporations to adjust the price of their products, affecting the company's profit. Customer satisfaction is one of them. This additional parameter was employed in our suggested ABM to simulate cannibalization conditions in a more realistic manner. The outcome indicates that the price will be changed by this new parameter. The customer opinion toward each generation of an MGPL in each time period might affect the price of that generation in the next time period, as well as the profit. In comparison to [8], we have more data input for our suggested ABM, and adding a new effective parameter will provide us with more realistic pricing strategies. As a result, online evaluations and their sentiments can assist businesses in gaining a better understanding of their customers and how they interpret successive generations of the MGPL.

VI. Conclusion and Future Work

The MGPL planning problem is a multidimensional problem with several dimensions. Behavioral models are generally concerned with sales forecasts and generational introduction timing, whereas dynamic competition models are concerned with pricing dynamics under competition to determine optimal pricing options for the most recent generations. To investigate the entire MGPL planning challenge, we present a theoretical two-phase approach. We assess relevant reviews to determine customer satisfaction and ultimately obtain dynamic pricing strategies for the full multiple-generation product line. The model views a multiple-generation product line as an adaptive living creature whose behavior is affected by the strategic decisions made by each product generation. In this research, an agent-based model of MPGL pricing decisions is created to assess the dynamics of alternative pricing strategies on the overall profitability of the MGPL. Experiments in simulation provide insights into the dynamics of pricing strategies under various cannibalization scenarios. For pricing decisions, four alternative cannibalization scenarios are investigated in this study. Analyses of the results show that higher customer satisfaction leads to higher prices and more profit, whereas lower customer satisfaction forces companies to drop prices and improve the product.

It is vital to highlight that the results are valid in circumstances when the company operates in a monopoly market and new generations are introduced to answer varied customer value preferences and expectations. Early in the MGPL lifecycle, customer product selection is based on product performance evaluation. Now, customer valuation moves to product quality-price preferences. As a result, profits under cannibalization should be considered in the later lifecycle states of the MGPL. In this work, we used partial sales data and real introduction timings to get the pricing strategies for different time periods. In future work, we will use the DSVM model from [8] to predict sales and introduction timings of generations in the whole MGPL life cycle.

If there are several competitors in the market, the ABM outcomes may alter since sales dynamics would change because of competition. Future research will investigate the dynamics of four cannibalization scenarios under a competitive market environment. The current model is the initial step in analyzing the pricing dynamics of MGPL. There are other factors that influence an MPGL's sales volume and cannibalization circumstances. For instance, the products launched by competitors, market conditions such as technological developments, and product quality all influence the demand dynamics of each generation. In future works, we intend to develop this model and integrate other aspects into the study, such as many competitors and technology-changing factors. It is hoped that this form of MPGL agent-based modeling will serve as a test bed for analyzing various aspects of the complex cannibalization dynamics of an MPGL planning problem.

[1] Krankel, R. M., Duenyas, I., & Kapuscinski, R. (2006). Timing successive product introductions with demand diffusion and stochastic technology improvement. Manufacturing & Service Operations Management, 8(2), 119-135.

[2] Morgan, L. O., Morgan, R. M., & Moore, W. L. (2001). Quality and time-to-market trade-offs when there are multiple product generations. Manufacturing & Service Operations Management, 3(2), 89-104.

[3] Lin, C.-Y., & Kremer, G. E. O. (2014). Strategic decision making for multiple-generation product lines using dynamic state variable models: The cannibalization case. Computers in Industry, 65(1), 79-90.

[4] Jiao, J., & Tseng, M. M. (1999). A methodology of developing product family architecture for mass customization. Journal of Intelligent Manufacturing, 10(1), 3-20.

[5] Simpson, T. W. (2004). Product platform design and customization: Status and promise. Ai Edam, 18(1), 3-20.

[6] Lin, C.-Y., & Okudan, G. E. (2013). Planning for multiple-generation product lines using dynamic variable state models with data input from similar products. Expert systems with applications, 40(6).

[7] Lin, C.-y., Kilicay-Ergin, N. H., & Okudan, G. E. (2011). Agent-based modeling of dynamic pricing scenarios to optimize multiple-generation product lines with cannibalization. Procedia Computer Science, 6, 311-316.

[8] Kilicay-Ergin, N., Lin, C.-y., & Okudan, G. E. (2015). Analysis of dynamic pricing scenarios for multiple-generation product lines. Journal of Systems Science and Systems Engineering, 24(1), 107-129.

[9] Ofek, E., & Sarvary, M. (2003). R&D, marketing, and the success of next-generation products. Marketing Science, 22(3), 355-370.
[10] Edelheit, L. S. (2004). The IRI Medalist's Address: Perspective On GE Research & Development. Research-Technology Management, 47(1), 49-55.
[11] Norton, J. A., & Bass, F. M. (1987). A diffusion theory model of adoption and substitution for successive generations of high-technology products. Management science, 33(9), 1069-1086.
[12] Mahajan, V., & Muller, E. (1996). Timing, diffusion, and substitution of successive generations of technological innovations: The IBM mainframe case. Technological Forecasting and Social Change, 51(2), 109-132.
[13] Bardhan, A., & Chanda, U. (2008). A model for first and substitution adoption of successive generations of a product. International Journal of Modelling and simulation, 28(4), 487-494.
[14] Huang, C.-Y., & Tzeng, G.-H. (2008). Multiple generation product life cycle predictions using a novel two-stage fuzzy piecewise regression analysis method. Technological forecasting and social change, 75(1), 12-31.
[15] Dobson, G., & Kalish, S. (1988). Positioning and pricing a product line. Marketing Science, 7(2), 107-125.
[16] Arslan, H., Kachani, S., & Shmatov, K. (2009). Optimal product introduction and life cycle pricing policies for multiple product generations under competition. Journal of Revenue and Pricing Management, 8(5), 438-451.
[17] Li, H., & Graves, S. C. (2012). Pricing decisions during inter-generational product transition. Production and Operations Management, 21(1), 14-28.
[18] Gallego, G., & Wang, R. (2014). Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. Operations Research, 62(2), 450-461.
[19] Kim, N., Srivastava, R. K., & Han, J. K. (2001). Consumer decision-making in a multi-generational choice set context. Journal of Business Research, 53(3), 123-136.
[20] Schön, C. (2010). On the optimal product line selection problem with price discrimination. Management Science, 56(5), 896-902.
[21] Chen, J.-M., & Chang, C.-I. (2013). Dynamic pricing for new and remanufactured products in a closed-loop supply chain. International Journal of Production Economics, 146(1), 153-160.
[22] Fruchter, G., Fligler, A., & Winer, R. (2006). Optimal product line design: Genetic algorithm approach to mitigate cannibalization. Journal of optimization theory and applications, 131(2), 227-244.
[23] Hotho, A., Nürnberger, A., & Paaß, G. (2005). A brief survey of text mining. Ldv Forum,
[24] Zhan, J., Loh, H. T., & Liu, Y. (2009). Gather customer concerns from online product reviews–A text summarization approach. Expert Systems with Applications, 36(2), 2107-2115.
[25] Thorleuchter, D., Van den Poel, D., & Prinzie, A. (2010). Mining ideas from textual information. Expert Systems with Applications, 37(10), 7182-7188.
[26] Nasukawa, T., & Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. Proceedings of the 2nd international conference on Knowledge capture,
[27] Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of human language technology conference and conference on empirical methods in natural language processing,
[28] Beineke, P., Hastie, T., & Vaithyanathan, S. (2004). The sentimental factor: Improving review classification via human-provided information. Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04),
[29] Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. arXiv preprint cs/0409058.
[30] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.
[31] Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. arXiv preprint cs/0212032.
[32] Hu, M., & Liu, B. (2004a). Mining and summarizing customer reviews. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,
[33] Hu, M., & Liu, B. (2004b). Mining opinion features in customer reviews. AAAI,
[34] Kamps, J., & Marx, M. (2001). Words with attitude.
[35] Zhang, W., Xu, H., & Wan, W. (2012). Weakness Finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Systems with Applications, 39(11), 10283-10291.
[36] Kang, D., & Park, Y. (2012). Measuring customer satisfaction of service based on an analysis of the user generated contents: Sentiment analysis and aggregating function based MCDM approach. 2012 IEEE International Conference on Management of Innovation & Technology (ICMIT).
[37] Täckström, O., & McDonald, R. (2011). Semi-supervised latent variable models for sentence-level sentiment analysis. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,
[38] Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. Expert systems with applications, 40(10), 4241-4251.
[39] Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. Expert systems with applications, 40(10), 4065-4074.
[40] Deng, Z.-H., Luo, K.-H., & Yu, H.-L. (2014). A study of supervised term weighting scheme for sentiment analysis. Expert Systems with Applications, 41(7), 3506-3513.
[41] Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. Proceedings of the 12th international conference on World Wide Web,
[42] Kang, D., & Park, Y. (2014). based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications, 41(4), 1041-1050.
[43] Do, H. H., Prasad, P. W., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. Expert systems with applications, 118, 272-299.
[44] Nazir, A., Rao, Y., Wu, L., & Sun, L. (2020). Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. IEEE Transactions on Affective Computing, 13(2), 845-863.
[45] Zhuang, L., Jing, F., & Zhu, X.-Y. (2006). Movie review mining and summarization. Proceedings of the 15th ACM international conference on information and knowledge management.

[46] Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. Marketing Science, 31(3), 521-543.