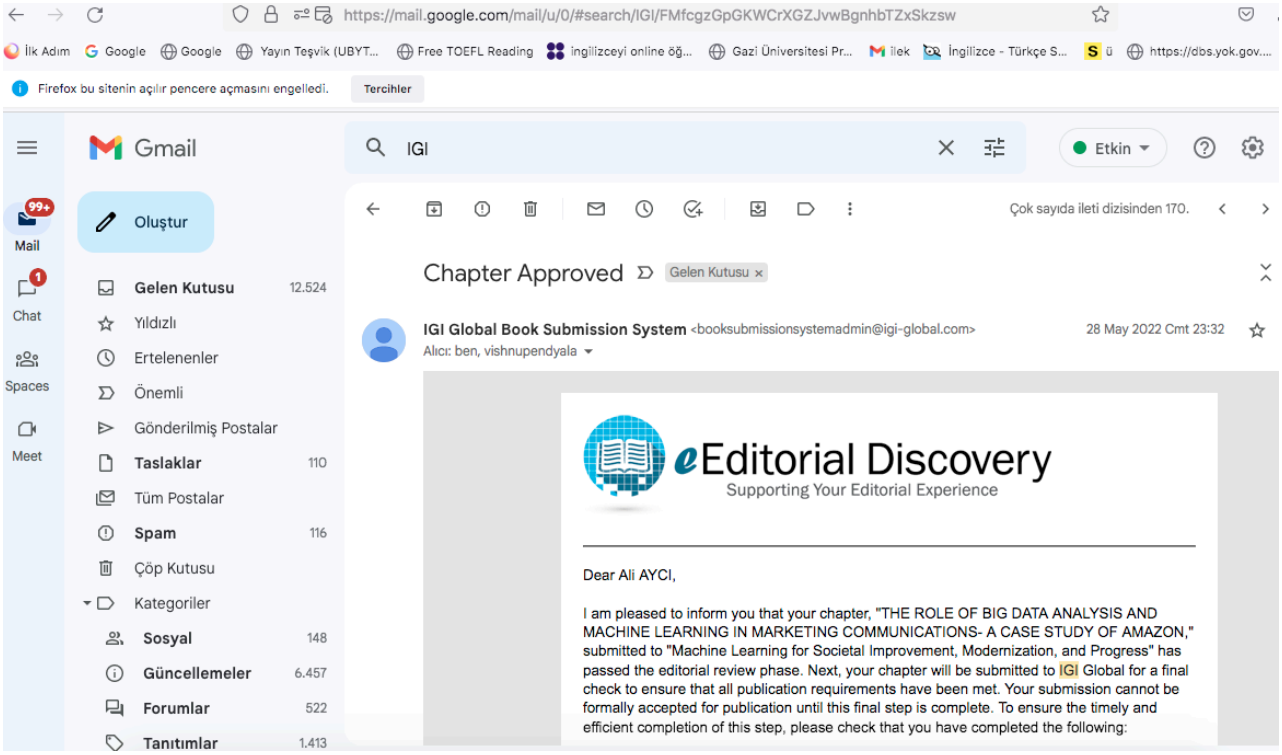
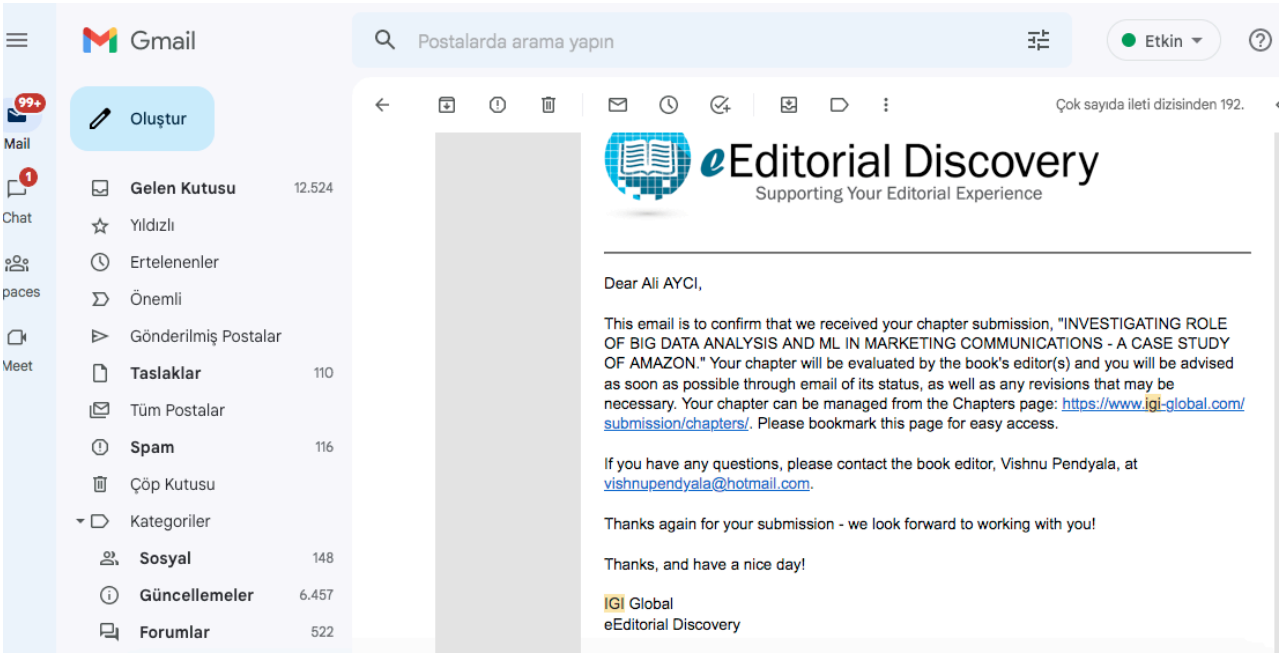
Erişim Linki Editör, İçindekiler

https://www.igi-global.com/book/machine-learning-societal-improvement-modernization/287540#table-of-contents





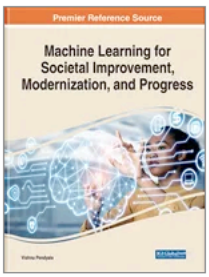
İlk Adım Google Google Yayın Teşvik (UBYT... Free TOEFL Reading İngilizceyi online öğ... Gazi Üniversitesi Pr... İlek İngilizce - Türkçe S...

IGI Global
PUBLISHER of TIMELY KNOWLEDGE

Login/Create Account Language: English

Search title, author, ISBN...

Books Journals e-Collections Open Access Publish with Us Resources Catalogs About Us



The Role of Big Data Analysis and Machine Learning in Marketing Communications: A Case Study of Amazon

Ali Ayci (Ankara Yıldırım Beyazıt Universtiy, Turkey) and Shivam Tyagi (Amazon, USA)

Source Title: Machine Learning for Societal Improvement, Modernization, and Progress
Copyright: © 2022 | Pages: 23
DOI: 10.4018/978-1-6684-4045-2.ch011

OnDemand PDF Download: $29.50
Available
Current Special Offers

Abstract

Technological and commercial developments in the 21st century have increased the demands and expectations of consumers. Big data analysis and machine learning can play a critical role in establishing healthy communication. The insights obtained from big data analysis and ML models can help in making marketing communication more personalized and useful to customers. In this chapter, the authors present an exploratory analysis of the role of big data and ML in the marketing communication of enterprises. For this, Amazon, which provides big data analysis infrastructures to other businesses which are big data users, is being researched as a case study. This study is one of the first studies in the literature to examine the significance of big data and ML in marketing communication and investigate the obstacles encountered when using them. The insights presented in this chapter will help professionals in the marketing communication domain to better understand how they can utilize big data and ML models to make a significant impact on their customers and sales.

Machine Learning for Societal Improvement, Modernization, and Progress

Vishnu S. Pendyala
San Jose State University, USA

A volume in the Advances in Human and Social Aspects of Technology (AHSAT) Book Series

IGI Global
PUBLISHER of TIMELY KNOWLEDGE

Published in the United States of America by
IGI Global
Engineering Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Copyright © 2022 by IGI Global. All rights reserved. No part of this publication may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher. Product or company names used in this set are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI Global of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Names: Pendyala, Vishnu, 1968- editor.
Title: Machine learning for societal improvement, modernization, and progress / Vishnu Pendyala, editor.
Description: Hershey, PA : Engineering Science Reference, [2023] | Includes bibliographical references and index. | Summary: "'Machine learning for societal improvement, modernization, and progress' portrays the application of Machine Learning that is playing a prominent role in improving the quality of life and the progress of civilization. It focuses on not just Machine Learning, but its application to specific domains that are resulting in substantial global progress"-- Provided by publisher.
Identifiers: LCCN 2022008798 (print) | LCCN 2022008799 (ebook) | ISBN 9781668440452 (h/c) | ISBN 9781668440469 (s/c) | ISBN 9781668440476 (ebook)
Subjects: LCSH: Machine learning.
Classification: LCC Q325.5 .M32225 2023 (print) | LCC Q325.5 (ebook) | DDC 006.3/1--dc23/eng20220503
LC record available at https://lccn.loc.gov/2022008798
LC ebook record available at https://lccn.loc.gov/2022008799

This book is published in the IGI Global book series Advances in Human and Social Aspects of Technology (AHSAT) (ISSN: 2328-1316; eISSN: 2328-1324)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

For electronic access to this publication, please contact: eresources@igi-global.com.

Chapter 11
The Role of Big Data Analysis and Machine Learning in Marketing Communications: A Case Study of Amazon

Ali Ayci
Ankara Yıldırım Beyazıt Universtiy, Turkey

Shivam Tyagi
Amazon, USA

ABSTRACT

Technological and commercial developments in the 21st century have increased the demands and expectations of consumers. Big data analysis and machine learning can play a critical role in establishing healthy communication. The insights obtained from big data analysis and ML models can help in making marketing communication more personalized and useful to customers. In this chapter, the authors present an exploratory analysis of the role of big data and ML in the marketing communication of enterprises. For this, Amazon, which provides big data analysis infrastructures to other businesses which are big data users, is being researched as a case study. This study is one of the first studies in the literature to examine the significance of big data and ML in marketing communication and investigate the obstacles encountered when using them. The insights presented in this chapter will help professionals in the marketing communication domain to better understand how they can utilize big data and ML models to make a significant impact on their customers and sales.

INTRODUCTION

Experts estimate that consumers receive an average of between 4,000 and 10,000 marketing messages per day (Simpson, 2017). While on one hand consumers are exposed to thousands of marketing messages every day, each consumer expects a personal approach from the companies. At the same time,

DOI: 10.4018/978-1-6684-4045-2.ch011

Copyright © 2022, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

considering that the target audience of the enterprises is large, how is it possible for the enterprises to communicate with their target audiences in a healthy and effective manner in line with their goals, with personal judgments, past estimates, and unfounded assumptions? In the big data era we live in, computer scientists, physicists, economists, mathematicians, political scientists, informatics, sociologists and other scientists strive to access the enormous amount of information produced about people, objects and their interactions.

The interaction of billions of people using computers, GPS devices, cell phones and medical devices creates an ever-increasing flood of data, and many of these interactions occur through the use of mobile devices in the developing world, used by people whose needs and habits have not been adequately understood until now (WorldEconomicForum, 2012). Scientists across disciplines, businesses, public and private institutions are exploring the potential benefits of analyzing genetic sequences, social media interactions, health records, phone records, government records, and other digital traces left by humans (Boyd and Crawford, 2012, p. 662; Saggi and Jain, 2018, p. 758). Although there is no official definition (De Mauro, Greco and Grimaldi, 2016, p. 122), Big data is a resource that generates new ideas by quickly obtaining and analyzing different sources such as humans and machines (EU, 2021) and ultimately big data is the information asset used to be turned into value (De Mauro at all. 2016, p. 122). Big data is data management that cannot be solved with traditional databases due to the increasing volume, velocity and variety (Amazon, 2021a). Big data can be divided into structured and unstructured big data. Structured data comes in a defined format and offers clear answers. In user-generated content, structured data typically has ratings and data from responses such as totally agree, strongly disagree continues to be useful to marketers.

However, there are also negative aspects of structured big data. For example, the target audience may be uncomfortable with being forced to answer, or people may not be able to accurately and fully express the answer they would like to give (Bendle and Wang, 2016, p. 118). Unlike structured data, unstructured data is not ready and comes in a form that must be processed to be usable. Such unstructured data is often rich and exciting but can be extremely difficult to use. Managers trying to guess what consumers are thinking by browsing these reviews will often create overly biased estimates. When the volume of reviews is very high, reading all reviews can present significant practical difficulties (Bendle and Wang, 2016, p. 118). With the use of big data analysis in businesses, the benefits of creating value for the company, reducing costs, producing innovative solutions, launching new products or services, creating customer value, helping marketing decisions, personalization in the marketing mix can be achieved (Benoit, Lessmann and Verbeke, 2020, p. 235). Understanding the wishes and needs of the target audience is one of the areas where businesses can benefit from big data analysis. On the one hand, businesses use communication tools (Social media, websites, consumer forums, etc.) as large data collection areas, on the other hand, they give their messages regarding their marketing strategies with these marketing communication tools. In other words, big data is both a field where data is collected and a field where data is created. With this feature, big data analysis is becoming an increasingly important channel in terms of marketing communication tools for enterprises.

The findings show that research in big data applications for marketing is at the embryonic stage and therefore more direct efforts towards business need to be developed for big data to evolve in the field of marketing (Erevelles, Fukawa and Swayne, 2016, p. 897). Research is needed to clearly identify the pros and cons of enterprises' big data investments (Amado, at all, 2018, p. 6). When evaluating big data investments, first of all, the benefits of big data for businesses should be revealed. In this study, first of all, a general perspective is presented in order to reveal how and for what purpose big data can be

used in marketing communications. Until recently, big data investments had to be met with business resources, and academic studies were based on high investment values and limited opportunities for businesses. However, today, some leading companies providing big data analysis services are bringing their services and capabilities with other businesses at affordable costs. In this context, Amazon and Amazon Personalize (AP) are researched as a case study in order to investigate the service scopes and usability of outsourcing organizations for the use of big data in marketing communications. technically and practically. The reason why Amazon is included in the research is that it is one of the businesses that use big data and big data analysis most effectively, and more importantly, it provides big data analysis infrastructure services to other businesses, especially SMEs. It makes big data applications more accessible for companies. Because Amazon is a business that markets its services online, the business' website is used as the data source for the exploratory research within the scope of this article. Finding detailed technical information about the services offered on the Amazon web page also provides convenience in accessing information. This study is one of the few studies in the international literature that deals with marketing communication and big data analysis. New and up-to-date studies are needed because consumer expectations, which is one of the most important areas of use of big data, are constantly changing and technology and actors in big data analysis are constantly developing.

BIGDATA, MACHINE LEARNING AND MARKETING COMMUNICATION FROM MARKETING ASPECTS

Marketing Communications and Big Data Needs

Traditionally, marketing has always been a field in which data plays an important role. Since the 1980s, with the development and cheapening of computing technologies, more business processes have been digitized and the data produced and captured by the devices have been made available to marketing researchers and practitioners. Customer choice modelling, market basket analysis, estimation of price (cross) elasticities, etc. have emerged and evolved with the increasing availability of various types of data (Benoit, et al., 2020, p. 236). While deciding on the target customer profile in marketing, defining information sources, and ensuring customer participation in product design, businesses benefit from big data (Zhan, et al, 2018, p. 584). Big data analytics is a powerful analytical tool to understand market and customer needs for products or services and enable them to make better decisions (Saggi and Jain, 2018, p. 782) and serves as a tool to identify customer value creation strategies, define an overall marketing strategy, develop product concept features, identify value creation opportunities, and develop customer value metrics (Smith and Colgate, 2007, p. 15). The main research topic of marketing is to understand the target audience correctly. Based on the results of field research; while 88 per cent of IT managers believe their organization truly understands its customers, only 61 per cent of consumers believe that companies understand their needs (Talend, 2021). Big data analysis offers important opportunities for businesses to understand and respond to consumers' needs and expectations in a timely and accurate manner. (Buganza, Trabucchi and Pellizzoni, 2020, p. 58). Big data is the biggest opportunity that changes the rules of the marketing and sales ecosystem and allows us to understand the target audience in more detail since the internet became mainstream nearly 25 years ago (McKinsey, 2015, p. 4). As machine learning evolves and analyzes different types of data faster; Big data is playing a more central role in marketing than ever before, making the pace of development of marketing messages a more important

revenue item for the business (Luenendonk, 2019). Prior to the recent interest of businesses in analyzing marketing data at a large scale, rapid advances in network and database technology have led to the development of a variety of software applications. This technology not only allowed business processes to be further integrated by interconnected computers but also stimulated interest in analyzing data where the volume and diversity are large, such as the growth of the internet, interacting with potential and existing customers, gathering information about competitors (Benoit, at al, 2020, p. 236). While marketers are trying to use big data analysis commercially, it is seen that the interest of academicians on the subject has increased in the literature. According to the results of Amado et al.'s (2018: 6) study of 1560 articles published between 2010 and 2015 on the use of big data in marketing, big data in marketing has received increasing attention over the years and the number of publications doubles every year compared to the previous year. As a result of the use of big data in marketing activities by researchers and practitioners, businesses have made significant gains. The results of the research conducted by Suoniemi et al. (2020) on 301 marketing managers reveal that the use of big data improves firm performance by improving the firm's market-oriented capabilities.

The main way to make businesses competitive and differentiate them in an internationally competitive environment is healthy communication with the target audience. Marketing communications can provide a sustainable competitive advantage to any organization that taps into its potential (Mihaela, 2015, p. 1447). Marketing communication is the coordinated use of promotional messages and related media to communicate with a market, where marketing communication messages are communicated through one or more channels such as digital media, printed materials, radio, television, direct mail, and personal selling (AMA, 2021). Integrated marketing communication, which expresses the harmony of communication tools in the business, is an activity that allows creating profitable customer relations and creates value for the product/brand or company. It aims to ensure the consistency of the message conveyed through various communication tools such as advertising, sales promotion, direct marketing, public relations, online communication (Mihaela, 2015, p. 1949).

Customer Oriented Marketing

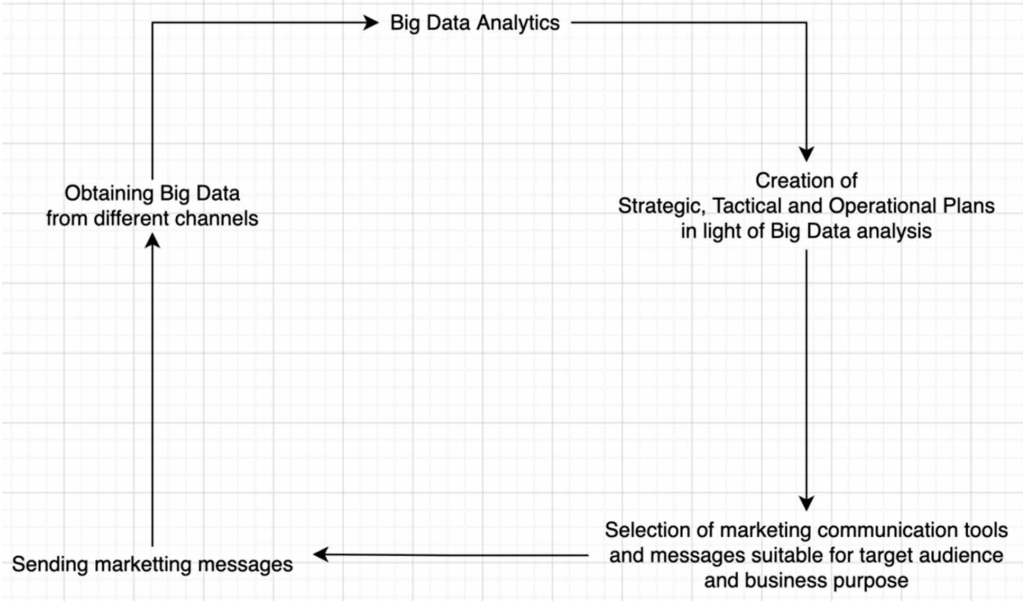
In customer-oriented marketing, it is aimed to collect consumer information, to make effective inferences as a result of detailed analysis of this information, to predict the future from the analysis of past customer interactions, and finally to respond to customer needs in real-time (Gulati and Oldroyd, 2005, p. 97). For this purpose, customer-oriented marketing and customer-oriented communication become more effective with big data analysis, which is one of the most effective data collection and analysis methods. The increased focus on customer experience arises due to the complexity of the communication process as customers interact with companies through multiple channels and numerous touchpoints in the media, due to the redundancy of tools and messages in marketing communication. (Lemon and Verhoef, 2016, p. 69). The integration of technology and big data into business strategy increases the complexity of marketing communications and fosters the need for advanced marketing performance analytics (Buhalis and Volchek, 2021, p. 1). Effective management of the complexity in the communication process is possible with big data analysis tools. Big data enables strategic communicators to analyze the needs, views, attitudes and behaviors of their stakeholders in greater detail. From planning to evaluation, big data analytics make enterprise communication more analytical and potentially more strategic.

Communication professionals use advanced analytics tools to capture and process the needs, opinions, attitudes and user behavior of target groups from all over the world. Marketers by connecting this data to market data can monitor the changing perceptions of their companies, brands and products in different countries and this in more or less real-time. Insights from big data analytics allow companies to conduct more personalized communications in marketing, public relations and internal communications than in the past (Wiencierz and Röttger, 2018, p. 43). A study was conducted by Wiencierz and Röttger (2018, p, 42) with 35 big data and social media analytics experts on the use of big data in the field of communication, and the results of the research show that big data offers benefits to marketing communication practitioners in four main areas. a) providing communicators with a wider knowledge base for situation analysis b) formulating database-based strategies c) automated communication applications d) providing evaluation based on larger data stocks. Marketers benefit from big data analytics in individual product and brand communication, online touchpoint analytics, predictive analytics of product and label message spread, topic or content marketing identification, real-time processing of customer data for online marketing, market share determination through share analytics, and effective communication. In order for big data analysis to be successful in marketing communication, it should be clear which of the goals of the business serves the big data, it should be repeated several times until the results of the efforts such as data generation, extraction and analysis are compatible with the determined goal, and interdepartmental cooperation should be supported (Wiencierz and Röttger, 2018, p. 42).

Behavior-oriented Marketing

The analysis of consumer behavior is of great importance in the effectiveness of advertising decisions, which is one of the leading marketing communication tools. While advertisements in traditional media can present common content regardless of differences, advertisements can be personalized with big data technology, allowing the delivery of advertisements suitable for the region, culture and person (Yavuz and Erdoğan-Gençyürek, 2019, p. 214). Behavioral advertising is one way to better target the most valuable consumers. For example, if past purchasing behavior in one product category is positively correlated with product preference in another category, consumers who purchased the previous category may be more heavily advertised (Shen and Villas-Boas, 2018). Netflix, one of the most valuable media companies in the world, owes its success to its impressive customer retention rates, and the secret of Netflix's customer retention is that it is more successful in determining what they want by communicating effectively with its viewers compared to its competitors. Netflix collects data from its 151 million subscribers and applies data analytics models to explore customer behavior and purchasing patterns, then uses this information to recommend movies and TV shows based on their subscribers' preferences and more than 75% of audience activity is based on personalized recommendations with offers delivered through different marketing communication tools (Dixon, 2019). At the macro level, the two main components of interactive integrated marketing communication are big data marketing information and its technical system and interactive marketing communication plans (Tang, 2020). Social media is the medium where big data is most active with marketing communication. After the analysis of the big data obtained from social media, communication with the target audience is established with different marketing communication tools in line with the strategies compatible with the business objectives and this is a continuous cycle. In other words, while marketing communication tools are used to provide big data, it is also a basic tool used for effective communication after the analysis of big data (Figure 1).

Figure 1. Big Data Usage Cycle in Marketing Communications. Source: (Authors, 2022)



It is seen that different people, institutions and organizations have made big data analysis in social media the building block of their communication strategies and have achieved significant success. Coca-Cola is able to respond to its customers in real-time by analyzing and evaluating the information of Twitter users around the world, using services from Sysomos, a social media analytics business. Former US President Obama predicted the intentions and tendencies of undecided voters by using the Boatbuilder system, which can manage various databases in the 2008 elections, closely analyzed the voter movement and responded to them with a direct communication strategy in real-time and won the US presidential election (Bitnine, 2016). The critical point in the effective design and implementation of these steps and the development of strategies is to collect and analyze as much qualified information as possible from the target audience. With big data analysis, businesses can fully understand consumers' concerns, reach opinions about competitors and the industry, and find the opportunity to build their marketing communications on these findings.

Active vs Passive Recipient's Impact on Marketing

In the era of social commerce, the first part of the 21st century, the basic idea is that customers do not become passive recipients of information from marketers, but interact with each other and with the business to influence consumer purchasing and business decisions. These interactions can take the form of product reviews, product usage descriptions, "homemade ads", blogs, and other consumer-initiated contributions (Fader and Winer, 2012, p. 369). Businesses need to develop timely and effective strategic responses to contributions such as consumer review information, as silence to incoming messages is often not a solution (Chen and Xie, 2008, p. 488). It now has access not only to purchasing and marketing mix data but also to rich information about how consumers interact with each other and the products/services they consume. This provides a much more complete picture of the effects on buyers and has

the potential to significantly improve modelling efforts and ultimately marketing decisions (Fader and Winer, 2012, p. 369). Because analysis of social media data provides valuable information for marketing research purposes, businesses should use this data as a form of exploratory research (Malthouse, et al, 2013, p. 270). By observing what consumers write about products in a category, firms can better understand the online discussion and marketing opportunities, market structure, competitive landscape, and characteristics of their own and competitors' products (Netzer, et al, 2012, p. 521). Businesses soon realized the potential of using internet-based social networks to influence customers and incorporated social media marketing communications into their strategies to strengthen their businesses (Moro, Rita and Vala, 2016, p. 3341). A system that can predict the impact of individual posts would be a valuable advantage in promoting products and services (Chen and Xie, 2008, p. 477). 12 factors can be used to measure the performance of social media posts (Table 1).

Table 1. Social Media Posts Performance Aspects. Source: (Moro et al. 2016, p. 3343)

Characteristic	Indicator
Full access to the post	Number of people who saw the post on the page
	The number of times a post is viewed regardless of whether it is clicked or not. People can see more than one.
Total views	They are the impressions of the same article. For example, someone might see a page update in their news feed once, and then see it a second time if a friend shares it.
Users interacting	The number of people (unique users) who clicked anywhere on a post.
Consumers	The number of people who clicked anywhere on a post
Post consumption	Number of clicks anywhere on a post
Impressions of people who like a page	Total impressions from people who only liked a page
Reach of people who like a page	The number of people who saw a page post because they liked that page (unique users).
People who like a page and are interested in a post	The number of people who liked a page and clicked anywhere on a post (Unique users).
Comments	Number of comments on the post
Likes	Number of likes on the post
Shares	Number of posts shared
Total interactions	Total of likes, "comments and shares"

This dataset is also linked to all phases of branding (cognitive, affective and cognitive phases) and shows that predictions about social media interaction can also be used to predict brand building. Managers will be able to use this information to make informed decisions about the posts they publish, increasing their impact and thus contributing to successful brand communication and brand building (Moro, et al., 2016, p. 3350).

METHODOLOGY

Within the scope of this study, the technical dimensions of big data analysis and machine learning in marketing communication, the problems experienced and the future perspective will be investigated. As a research method, exploratory analysis will be conducted to obtain data by in-depth interview method from experts who work in Amazon. Afterwards, as a case study, Amazon AP, which is the product Amazon offers to other businesses, will be analyzed in accordance with the information provided by the experts, the information obtained from the product and its web page. In face-to-face interviews, the participants were asked to answer in-depth 15 questions spanning over different paradigms including marketing problems and big data, ML solutions, services used and offered by Amazon itself, problems they encounter and the future of machine learning and big data in marketing communication.

Technical Aspects Of Marketing Communication With Bigdata And Machine Learning

While the rapid growth in user content on the Internet has a huge impact on marketing practice at the same time, brand managers are no longer able to fully control the messages they use (Fader and Winer, 2012, p. 369; Malthouse, at al., 2013, p. 270). The huge amount of data makes it difficult to track and measure information, and this rich but unstructured set of consumer data is qualitative in nature. A combination of a text mining device and a network analysis framework can be used to overcome these challenges (Netzer, et al, 2012, p. 521). Unstructured big data from online product reviews are generated in real-time, easily accessible, and contain the messages consumers want managers to hear. Analyzing such data has the potential to revolutionize market research and competitive analysis, but how can the messages be deduced? How can large amounts of data be concentrated on insights to help drive business strategy? How can large amounts of data be concentrated on insights to help drive business strategy?

Data Collection

Different types of data can shed light on different aspects of CX. Depending on the requirement, clients can decide what type of data to be collected on their websites. As in this study, we are focussing on marketing communication, customer interaction and behavioral data is of importance and thus we will focus on strategies to collect that data. For big data applications, customer data is collected in three ways, asking customers directly, by tracking data indirectly and by acquiring data from other companies. There are tools which can be enabled in a website to track what item the customer is looking at, which buttons and links he/she clicked and how much time the customer spent over a specific section or webpage. All these data points when fed into a big data analysis tool or ML algorithm can generate insights that can help in understanding and tracking user interests and behavior.

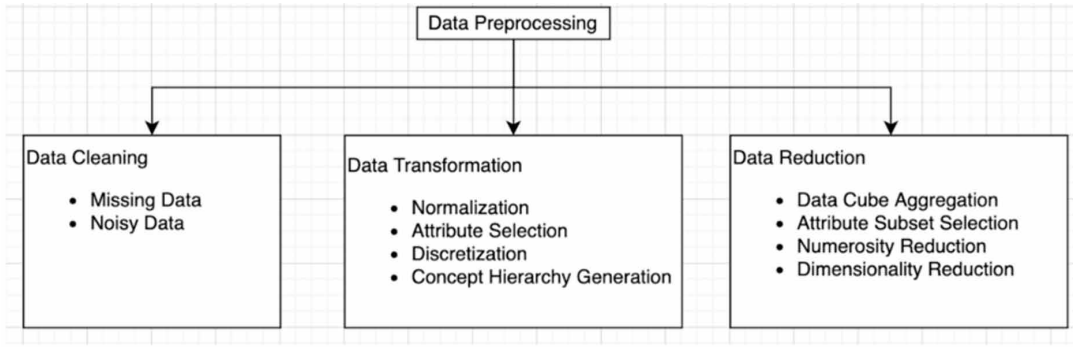
But there are some hurdles that companies face during the data collection phase. With the latest development in privacy controls and new regulations being placed by governments, data collection is becoming a more and more challenging task. In research, it has been observed that after Apple provided an option to customers to block any type of data tracking, almost 25% of customers have opted out of it. This is a huge impact as it means 25% less data for training models. Government regulations like CCPA and GDPR give power to customers for the right to be forgotten. This brings an obligation on companies to delete a customer's data when opted. Due to these new controls and regulations, data collection needs

to be very transparent. Indirect tracking of data and selling of data is not encouraged to practice. True that data collection is a matter of concern but also it has a big role to play in improving CX by helping in identifying patterns in customer behavior and hence allowing companies to create targeted offers and advertisements. Companies like Facebook, Snapchat and Twitter which earn most of their revenue by advertisements are on the receiving end of the push towards privacy controls. They are working on changing their underlying architecture to minimize the impact. In this direction data collection methods like interviews, surveys, GPS based location tracking and data collection through IoT devices are being used more and more.

Data Preprocessing

Most of the time raw data available is not suitable to be used in machine learning models as is. Data preprocessing is a very essential part of the whole process. The quality of inferences directly depends on the quality of the data. If the data is incomplete or incoherent, the results will be unreliable. To ensure that proper inferences are derived, a lot of effort is dedicated to data preprocessing. There are certain techniques like sampling and imputation to clean and prepare data before being used as input to ML models. Once the data is validated to be in good form, the ML model is trained and tested using this data for analysis.

Figure 2. Data Preprocessing. Source: (Authors, 2022)



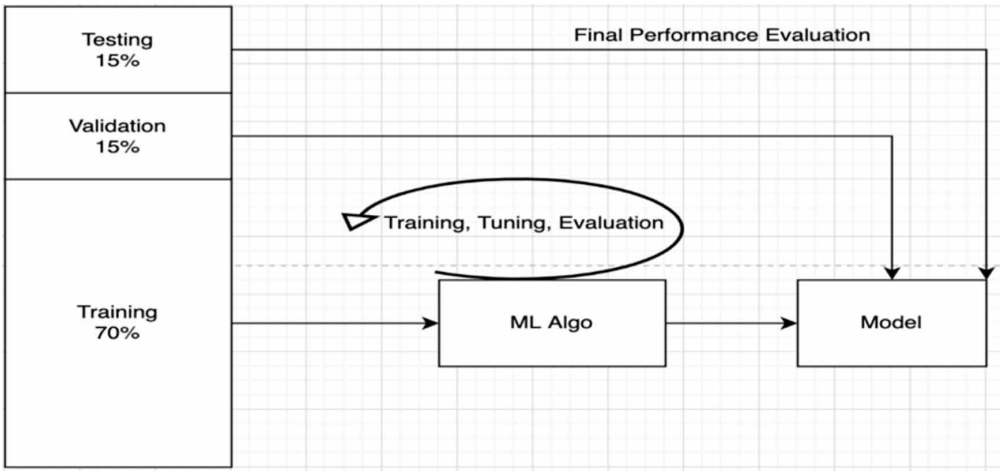
Source: (Authors, 2022)

For training machine learning models, the data is divided into train, test and validation data. There are many ways to do this. Like data can be divided into 70:15:15 shares where 70% data will be used for training, 15% data for testing and 15% for validation. In another approach, there can be overlapping sets that will have some data in common. In some approaches the data set is divided into N number of sets and training is performed using a certain number of sets each time.

There are three main challenges of any data pre-processing stage which can be defined by three Vs, Volume, Variety and Velocity. As their literal meanings, volume means amount of data, variety means different types of data and velocity signifies the speed with which data is processed. All three of these directly impact the data preprocessing step and the quality of results hence obtained. In today's world where IoT and sensors are all around us, millions of data points are obtained every second. To save,

identify, filter and process these data points requires a lot of computing resources. Today hardware designers are working on creating more and more efficient processors to speed up the calculations and hence provide results as quickly as possible. With these advancements in computing power, processing large amounts of data has become possible. But even then as the data increases with every day, new approaches are being researched. Distributed processing is one such field where companies like Google and Amazon have excelled a lot.

Figure 3. Training, Validation and Testing Dataset
Source: (Authors, 2022)



Distributed Data Processing

There are a lot of softwares for big data analysis. Spark is an open-source distributed computing framework used to analyze big data quickly. Its ability to divide the task among workers to process chunks of data in parallel allows it to analyze large amounts of data quickly, hence making it apt for big data applications. For marketing, communication spark can be used to analyze the data of millions of customers in real-time and thus provide inferences that can be integrated with marketing messages instantly. Engineers and scientists can use Spark APIs to analyze user data from different sources to generate insights (Alzahrani, 2021). Hadoop is another distributed data processing framework that uses maps and reduces algorithms to distribute tasks and combine results to get final inferences. The main difference between Hadoop and Spark is the place where data is stored. Spark saves data in RAM while Hadoop uses a hard disk to save data. This makes Spark faster than Hadoop. Thus Spark is good for applications that need real-time inferences while Hadoop is used for applications that require batch processing. Amazon EMR or Elastic Map Reduce service is another big data platform based on Hadoop that provides the capability to run distributed data processing jobs using open source frameworks like Spark. Amazon EMR distributes tasks among virtual servers Amazon EC2 instances and saves data on Amazon S3. Due to being on the cloud, this service has the capability to scale up and down dynamically depending on the requirement. Although these programs are not very expensive to acquire, professionals are needed to use them.

Data Analysis Models for Marketing Communication

Artificial intelligence tools driven by machine learning algorithms are rapidly transforming the business world and this transformation is increasing the interest of researchers. Machine learning methods have flexible model structures that can handle large-scale and unstructured data and provide strong predictive performance (Ma and Sun, 2020, p. 481). Although it is argued that big data will remove the human factor (Akçan, 2016, p. 311) or not (Altunışık, 2015, p. 72), it is a fact that it will enrich the human factor qualitatively. Marketing communication consists of three main types of content, text, images and videos. There are different approaches to analyze each of these types effectively using AI. In this section, we will discuss some of the models for each data type and how they can impact marketing communication.

Text Data

Text analysis of user-generated content provides access to information provided by active, content-producing consumers, providing useful information from the social structure of a brand's follower base, regardless of whether it creates or consumes content. This is particularly useful for predicting perceptual traits that consumers may be less likely to mention in brand conversations rather than key product features (Culotta and Cutler, 2016, p. 347). In order to make these analyzes, Twitter is a very important tool in big data analysis, as it is popular, relevant, social connections are public, and organized as it is organized by users into topic-based lists (Culotta and Cutler, 2016, p. 348). Unlike traditional transaction data, social media is a difficult medium to research due to data volume, diversity and speed (Arrigo, Liberati and Mariani, 2021). Big data analysis and suggestions can be made by people as well as artificial intelligence tools can be used. There are many AI approaches to analyze text data today. Topic-based analysis is one such methodology where an AI model tries to find a topic for each section of the text and then derives an overall topic. This is also known as topic modeling. Latent Dirichlet Allocation (LDA) is a good example of this approach.

LDA is a statistical model that works on the problem of topic discovery. It works by changing each word to its root lexical forms and then checking which topic they relate to. For example Fly, Flying, Flies, Flew will all be converted to Fly and then the system can correlate that word to most probable topics like a bird or plane. So whenever there is an ambiguous situation like this, LDA will try to predict probabilities of all probable topics and present them to us. Plate annotations can be used to represent such probabilistic graphical models (Blei et al. 2003).

In the representation of LDA, their is only one observable variable in the whole process, all others are latent variables which are inferred. There are three levels in LDA, corpus, document (outer box) and words(inner box). LDA can analyze large amounts of text and identify content by focusing on attributes that do not appear with a certain weight. For example, in examining a graphic text, it can be analyzed by focusing 70% on the story and 30% on the graphics. Collecting content from a large number of consumers allows us to understand what is on the minds of consumers collectively, from which it can be revealed what consumers care about and which features are viewed positively or negatively. Thus, for example, if advertisements will be given in newspapers, healthy communication will be established by providing the target audience to concentrate on the areas that the business wants to concentrate on while saving on advertising expenses by including the features that the target audience sees positive. LDA is able to map the relative strategic positions of competitors in the minds of consumers where they matter most (Bendle and Wang, 2016, p. 115). Any method that tries to understand the message must extract the

common themes between the interpretations and thus bring together the unseen qualities expressed at various levels of abstraction. As one of the effective methods of this grouping, LDA has similarities with the principal component methods performed in customer surveys in terms of extracting common core characteristics. The advantage of the method is that it works with a large amount of unstructured data where not every customer review has most attributes. While LDA is somewhat technical, it is relatively easy to implement for statistically trained analysts and the data required for analysis can be accessed from social media. Without in-house expertise, data can be easily and relatively inexpensively purchased from third parties. LDA reveals the hidden nature of the invisible attributes behind comments. This provides more insight into what consumers think, telling management which areas of the customer experience need attention (Bendle and Wang, 2016, p. 123). Thus, very critical information is obtained in terms of strategic communication decisions.

Image Data

Advanced machine learning algorithms integrated with big data power recommendation systems on e-commerce sites and content platforms like Amazon and Netflix. Learning engines analyze and tag billions of images on social media sites like Facebook. Automatic bidding algorithms analyze a user's profile on a millisecond timescale to determine the most appropriate bid for ad serving; chatbots engage in human-like conversations with customers to maintain relationships and loyalty (Ma and Sun, 2020, p. 482). Despite significant advances and growing interest, the use of machine learning methods for big data in marketing is still at an early stage and current work is somewhat scattered. For example, there is no consensus on how machine learning methods should be incorporated into marketing research (Ma and Sun, 2020, p. 482). One challenge with using big data in marketing communications is that while automation can be more efficient for businesses, people still want to talk to people (Ariker and Manuel, 2015, p. 99).

Video Data

The ability of machine learning methods to handle unstructured big data is quite high and unstructured media data (text, audio, video recordings), consumer monitoring data, network data, hybrid data created by the combination of different data are transformed into unique information for marketing managers by machine learning tools (Ma and Sun, 2020, p. 497). Video-based analysis derives real-time inferences by analyzing the customer videos for behavior patterns that can then be materialized for enhancing marketing communication. With the increase in computation power now it is possible to analyze video data with more speed and in more volume. Thus it is apt time to use video-based analysis to derive useful insights. One tool that can be used to benefit from video data in marketing communication is the Video-Based Automatic Suggestion (VAR) System. The system uses a camera that captures the customer's behavior in front of the mirror to make inferences about their preferences based on facial expressions and clothing viewed by consumers at points in time. Second, the system identifies customers with preferences similar to the focus customer from a database of shoppers whose preferences, purchasing and/or evaluation decisions are known. Finally, recommendations are made to the focal customer based on the preferences of these like-minded customers and their purchasing and/or evaluation decisions. Each of the three steps can be implemented in various variations, and a retail chain can choose the very specific configuration that best serves its purpose (Lu, Xiao, and Ding, 2016, p. 484).

General design of the VAR system can be understood by example where a video feed from the changing room can analyze the behavior of the customers and their likings. Then the insights can be used to recommend better choices in marketing communication. If the VAR system is to be implemented in Walmart which has 10526 branches and 2.3 million employees in 24 countries (Walmart, 2021), the size of the data and the scope of its analysis will be understood. Unknown to the consumer, a significant part of the purchasing journey is guided by automated systems.

Search results are generated by an advanced Google ranking system, determined in part by advertisers' bids automatically generated using bidding machines (Ma and Sun, 2020, p. 481). Content on websites is customized according to user profile and questions are answered by chatbots. Reviews read by consumers stand out because they are deemed useful by a rating algorithm, and ads they see repeatedly are delivered through retargeting algorithms with real-time bidding. Coupons offering personalized prices to a consumer are generated just in time by the pricing engine. Finally, their posts on social media are collected by social listening engines and analyzed for emotion and feedback. These automated systems that make instant and context-dependent decisions are artificial intelligence tools implemented using state-of-the-art machine learning algorithms.

CASE STUDY: AMAZON AS A MARKETING COMMUNICATION SYSTEM PROVIDER

Founded in 1994 as a bookselling site, Amazon sold books to people in the US and 45 different countries in its first month (Hartmans, 2021). With 1.3 million employees today, Amazon generated $21.3 billion in annual sales of more than $386 billion in 2020 (Fortune, 2021). Amazon ranks second in the Fortune top 500 list with a market capitalization approaching $2 Trillion (Fortune, 2021). As a leader in customer experience innovation and defining many aspects of customer experience, Amazon is reorganizing its marketing activities around AI and machine learning efforts, not only taking things to the next level but showing how other businesses can do the same. Amazon provides big data analysis services to other large, medium and small businesses as well as serving the American Space and Aviation Administration (NASA) with its solutions based on big data analysis (Morgan, 2018). One of the most effective uses of big data and analytics, Amazon collects more than 2,000 historical and real-time data points on every order, using machine learning algorithms to find transactions with high fraud probability and stopping millions of dollars of fraudulent transactions every year (InsideBigData, 2019). Sharing this information alone is part of Amazon's effective marketing communication for the long-term relationship, not the instantaneous one.

Amazon Personalize (AP) As A Marketing Communication Tool

Amazon provides infrastructure for the use and analysis of big data to other businesses through its AP service. AP makes it easy for businesses to create applications that can offer a wide variety of personalization experiences, such as specific product recommendations, personalized product rankings, and customized direct marketing. AP is a fully managed machine learning service by training, tuning and deploying custom machine learning (ML) models beyond strict static rules-based recommendation systems to provide customized recommendations to customers in industries such as retail, media, entertainment (Amazon, 2021b).

Deep Dive Into AP's Client Space

Amazon Personalize is a service that Amazon provides to its clients who want to obtain recommendations based on their user data. Though AP is being used by companies in many fields most of the AP clients come from two main industries, media entertainment and retail. Amazon pioneered personalization in e-commerce more than 20 years ago so it is a natural fit, bringing the Amazon Personalized service to retail clients, with all the historical learning and experience. Due to Amazon's vast experience in retail, AP has been developed in such a way that it resonates very well with the requirements and expectations of its retail clients(Jory, 2022).

In media entertainment, the scope of AP's use cases is very broad. AP's clients can be media companies that are streaming video on demand or streaming sports or streaming short videos. In addition to these, there are use cases where clients are streaming music or textual data like news and articles. An example of an AP client is Discovery, which is using AP to personalize all aspects of their user experience including recommendations on the content rails. Down under, AP doesn't really care if it's a product client is trying to sell, a video client is trying to stream or a song client is trying to get people to listen to. It just looks at how users are interacting with those items within the client's environment including its application, emails, web app, mobile app and so on (Jory, 2022). The results obtained from AP depend a lot on the quality of data that is fed into the models. The data is owned by the client. AP provides algorithms in which clients can ingest their own data to obtain insights. AP is not like other services which provide already pre-trained models in the form of APIs which clients can call. AP takes the data from the client, uses the recipes to create a trained model. For this clients can save data in AWS S3 buckets or other storage services. The information obtained from Amazon's cloud storage service and the Personalize interface is analyzed and presented to the company, with the models optimized with the right parameters, and presented to the company simultaneously with the models. With AP service clubbed with other AWS services, Amazon has provided an easy-to-use and highly scalable platform that helps clients to generate recommendations for users. By providing the necessary infrastructure, Amazon manages the entire pipeline, including processing data, recognizing features, using the best algorithms, training, optimizing and hosting models (Amazon, 2021b). So in the end, every client has their own dedicated trained model. The focus of AP is on providing the clients with a quality ML model which can be used by them to generate useful insights(Expert, 2022).

The easy integration of the system with different marketing communication tools is a very important feature and it ensures that the people who use the system are not directed to only one communication tool. The most important question in mind is the confidentiality of information, and businesses may rightly be concerned about the confidentiality of shared information. (Altunışık, 2015, p. 61). Amazon's offering for businesses states that the data will be privately and securely encrypted and will only be used to generate recommendations to customers and will not be shared with customers or Amazon.com. (Amazon, 2021b). The benefits of applying big data analytics for the retail industry are listed as follows: Personalizing user home pages with product recommendations based on their shopping history, suggesting similar items on product details pages to help users find what they are looking for easily, helping users quickly find relevant new products, deals and promotions, improving marketing communication through personalization of push notifications and marketing emails with individualized product recommendations, improving sales with relevant product rankings and high-quality up-sell/cross-sell suggestions in the cart (Amazon, 2021b).

Data

One important aspect of the data is temporal details. AP can analyze the temporal details in real-time to help the clients understand the evolving nature of customer interaction. This temporality aspect is really powerful because it allows AP to understand how customer interests are changing over time in contrast to the collaborative filtering approach that's just using a matrix of items. "When did the customer interact" becomes really important when the client is looking to capture "what the customer is interested in now" as opposed to two weeks ago. Sophisticated combination of data and various ML models which are called recipes are used by AP under the hood to provide clients with the best results. These recipes are structured targeting different personalization use cases and hence provide a wide choice of selection to clients(Expert, 2022). Temporal aspect is a groundbreaking discovery as it allows the model to understand the evolving intent of a customer and can predict what he may buy next or alongwith an item already in his cart. This knowledge helps AP models to provide solid recommendations to clients.

There's another aspect to the data that clients can provide to AP and it's very powerful. It is what is called contextual data. This data is very specific to an interaction. Usually clients need to provide a user ID and item ID and a timestamp when a user interacts with an item at a particular time. But the client can also provide additional data about that interaction like when did the interaction occur? Like seasonality, is it winter, fall or spring. So if a client is selling clothing, user interest varies based on the weather conditions for that season. Another example is day of the week, where consumers consume video content differently on weekends. Maybe they watch long movies with their family. But on weekday mornings, maybe they're looking for short 30-minute news clips or sitcoms, whether commuting to work. This type of data. Device type is another really common contextual data. So are customers consuming content on their mobile phones, or on their desktops, or maybe on their set-top boxes. All of this contextual information can help AP learn the nuances of how the same user interacts differently based on different contexts. This allows AP to make even more relevant recommendations based on the current context when you're providing those recommendations. The dynamic filters included in the system create better user experiences by allowing clients to tailor the business rules for each user when recommendations are created, providing the opportunity for clients to control the recommendations in real-time while responding to the individual needs and preferences of the users (Chwastek, Pooniwala and Ashman, 2020).

Examining the customers, it is seen that AP users are international companies (Subway, Yamaha, Dominos etc.). However, the presence of promotions, free apps, training, free trials, minimum fees and no commitments shows that the AP target audience includes other businesses. The potential for SMEs to access these services is very exciting. These developments show that easy and relatively inexpensive applications started to exist shortly after Altunışık (2015) found that "it is extremely difficult to obtain and benefiting from its benefits is extremely difficult, but also very troublesome". AP offers personalized impressions, similar products, personalized redirection, personalized promotions/notifications for businesses using its own infrastructure (Amazon, 2021b).

LotteMart, one of the users of the AP system, is a supermarket chain in South Korea with 189 stores and more than 600,000 users, including an online store (Lottemart, 2021). AP enabled LotteMart to improve the customer experience with tailored recommendations. By focusing only on datasets for predefined interactions, users and items helped Lottemart in time savings tune to almost 50% of what they used to spend earlier in getting insights from the data. The use of coupons, a tool of marketing communication, has more than doubled since the introduction of personalized recommendations. In

addition, there was a 1.7-fold increase in the frequency of new product sales. (Park and Hwang, 2020). By using the AP system at Subway, the fast-food restaurant with branches in 100 countries, can offer personalized recommendations to its customers without the need for machine learning expertise of company personnel(Amazon, 2021b). It is a fact that subjects such as big data, artificial intelligence, data mining, and machine learning are complex and require expertise. It is seen that there are simple package customization services that include all these concepts in the AP system and will be sufficient for the use of the enterprise in practice. AP is now being used in a variety of fields spanning blogs in which the users share their information, training videos, visuals, frequently asked questions and answers sections, certified training programs, customer service support consisting of professional experts, and many others(Amazon, 2021b).

With the effective use of AP, it is seen that the communication with the target audience of the enterprise can be improved in a more effective, fast and healthy way, which can be relatively cost-effective. With the AP product in the literature, as a result of big data analysis, the subject of determining marketing activities: limited access to big data, trust problem, costs, complexity, lack of necessary skills (Boyd and Crawford, 2012, p. 662; Fosso et al. 2015, p. 244; Johnson, 2012; Altunışık, 2015, p. 61; Ariker, Breuer, and McGuire, 2015, p. 91) are partly eliminated.

In the growing space of media and retail which sees most of the AP clients, competition is very fierce. With time it is getting extremely easy for customers to switch from one provider to another. Customer experience was never as important as it is in today's market. It is very essential for providers that they give the customer the best quality and experience. For this using APs recommendations, clients can segment customers that have an affinity for the promoted items. This gives a much better chance of engagement of customers, increasing the probability of getting an open click and then conversion to an order. It's like every customer have their own dedicated, trained model to them (Expert 2, 2022). This is also what differentiates AP from other competitors who provide recommendations as an out of the box SAAS which can be used by calling an API. Since these AP alternates have to focus on end to end products, they somewhere lose the focus on quality of big data analysis and that is where AP excels. These services will provide clients with already built-in libraries that clients have no control to modify as per their use case. So having this one model fits all use-case approach is not correct as every client might have different demographics that will be impacting the sales. AP's approach to personalize the model to individual client and not group them as part of segmented solution brings out the quality recommendations(Expert 1: 2022).

Amazon Personalize (AP) Technical Operating Process

Process Start with first conversation. Every customer that is planning to use Amazon personalized service, Amazon does a touch point meeting to set and decide the things. Amazon starts engaging with a client that's actually looking to use personalized advertising to improve their marketing. So using that user segmentation capability of Amazon mentioned earlier AP enables clients to more intelligently target users that would have an affinity for products that they are trying to promote.

The goal of different AP teams as a whole, including operations, developers, and solution architects is to make the service so easy that clients don't need them at all. Because of this tenet AP have a lot of customers that self adopt the service which means they organically read about the service usings documentation, blog posts and sample codes. AP is so simple to use that most of the time clients never have to contact AP teams at all. Being simple is APs core mission.

Moving to the data being consumed in AP services, clients bring their own data. AP is an industry agnostic tool. Any client can bring their own data and start playing with the recipes provided by AP. AP requires interaction and behavioral data to generate quality recommendations. These data points can be as simple as time a user stayed on an item page. AP takes these datasets and trains the model behind the scene. Different use cases that can benefit from this design are item recommendations based on similarity, popularity and other parameters. Clients have the option to either use already available recipes for some of these common use cases or create their own. AP provides the ability to clients to create multiple versions for testing and then productionalize the best model by using it as an API endpoint from their website to get the inferences(Expert 2, 2022).

The challenges that clients face while working with AP are similar to what they will face with any ML service. The quality of ML service is usually measured by the quality of its recommendations. Quality of recommendations depends a lot on the quality of data used to train and test the models. are usually related to integration issues on either client side or Amazon side. To achieve data quality, clients usually require to do some feature engineering, data cleaning and other data preparation activities. AP is very smart and takes care of feature engineering for clients. Clients can bring their raw data and AP will feature engineer the parameters, perform data splitting for training and evaluation providing performance metrics. Clients however need to develop capability on their side to cache the results, to handle the AP recommendation flow, scaling of application, integrating the application with AP API, and how to productionalize the whole flow. All these tasks need to be taken care of by the client itself.

Selection of Metadata

Selecting the right meta data for the use-case while using AP service can be a game changer in targeted marketing communication. This allows AP clients to optimize results and generate more customer relevant recommendations. This section discusses how to select the right meta data depending on different aspects of a use-case. Right metadata plays a crucial role in training and testing of models. High quality of data ensures high quality results and optimized models. Right metadata allows models to identify patterns of user behavior and preferences. Due to availability of the same data from multiple sources, one needs to compare and determine which data set to use and which to ignore. When a customer visits a website, different aspects of his behavior are captured. This includes data points like how long he was on a web page, how long he hovered over a certain image, which items he added to his wish list etc. All these data points are collected but not all will be useful for every use-case. A proper selection can help in identifying the customer preferences and hence will help in personalized recommendations.

AP has different combinations of models and datasets which are together called recipes. Different recipes lead to different recommendations. Irrespective of the recipe an AP client chooses, there are three main type of data sets that AP deals with. These are, user data, item data and user-item interactions. User data can have metadata like, age, gender, location, ethnicity etc. Similarly, items will have metadata like, size, color, originated from, price etc. The third and most revealing type of dataset is user-item interaction. This dataset will have metadata like, time spent on the item page, item added to wish list etc. The minimum number of interactions Amazon Personalize expects in order to start making recommendations is 1,000 interactions from a minimum of 25 users(Hood and Cleopas, 2020).

Table 2. Different AP Data Types. Source: (Hood & Cleopas, 2020)

Dataset Type	Required Fileds	Reserved Keywords
User	USER_ID (string)	
Item	ITEM_ID (string)	CREATION_TIMESTAMP (long)
Interactions	USER_ID (string) ITEM_ID (string) TIMESTAMP (long)	EVENT_TYPE (string) IMPRESSION (string) EVENT_VALUE (float, null)

Usually there are two types of users that visit a website, guest users and member users. It is always more optimal to use data from members as they are authenticated as well as easy to track then guest or anonymous user. In AP, each data set has set specific requirements and a schema needs to be defined. This ensures a proper expected structure of data and parameters which AP can then use.

One of the latest feature that has been launched is user segmentation (AWS re:Invent, 2021) . Clients can train the models to segment users liking a particular item. So rather than asking what a particular customer will buy, the problem statement becomes, who are all the prospective customers of a specific item. This changes the whole game because now rather than sending email of the prospective items to the customer, clients can send out marketing emails to all the prospective clients of an item. So from a client's perspective these will look like generalized communication to a user segment, but from a customer's perspective this will look like a personalized email of the items he/she likes.

Recommendation Models

In the previous section we discuss how careful and well researched data selection can lead to quality recommendations and optimized models. There are two recommendation models that are used in AP, collaborative filtering and content-based filtering. Collaborative filtering focuses on the user similarities. If two users show similar patterns and interests, then the recommendations to one user depends on the products that other similar users have bought. On the other hand, in content based filtering, focus is one finding similarities between items. The recommendations are the items having similar metadata as the one that the customer likes. Due to the intrinsic differences in the approach, both these models used different data. Collaborative filtering uses user data while content based filtering uses item data. A hybrid model will use both of these data sets.

While a quality data set results in optimized model, quantity of data also makes a big difference too. Too much data will result in creating noise in model while too less data will create an underperforming model. If meta data that has no relation to user is included in training data set, it makes it harder for the learning algorithm to find a relevant pattern in the data. So this displays how a wrong approach in metadata selection can make it hard for algorithms to find patterns and learn anything from data. Recommendations generated by such a model which is developed using skewed, non-relevant data will not be as personalized as expected.

Findings and Suggestions

Marketing communication plays a very critical part in sales of any business. Sales in the soul of any company. A good marketing campaign can make or break the yearly sales revenue. Thus it is of utmost

importance that any insight that can be used to make marketing messages as effective as possible must be utilized. Big data analysis holds this capability to provide useful insights into customer behavior and thus helping create personalized marketing communication. From our analysis of different aspects of using big data analysis to drive marketing campaigns we present some important findings in this section.

One of the most important findings is the impact of metadata on the quality of insights. Specific use cases require specific types of data and thus it is very important to carefully choose the metadata. This is challenging as it is not always easy to find the required parameters. In such cases work arounds like interpolation, mutation and other such techniques can be used to get inferential data which is close to the actual data. Another important finding is the selection of models to process the data and get insights. A single model cannot cater to all use cases. So deciding which will be the best model is a challenge. With this the challenge to train the model in such a way that we can minimize any noise comes into picture. This is closely related with the quantity and quality of data. To decide how much data will be sufficient to properly train a model needs a lot of iterations and validations. Last but not the least the requirement of high speed processing of big data to get quick real-time results is another frontier on which a lot of advancement is required. All these problems bring up the question of cost-benefit analysis of using big data insights in making targeted marketing campaigns more efficient and personalized. No doubt it will bring a lot of value in terms of increased sales and high rate of customer acquisition for the companies.

CONCLUSION

Big data analysis is a very important tool for marketing communication at the point of timely and accurate understanding and the response of consumers whose needs and expectations are constantly increasing. With the development of communication technologies and increasing competition, the messages sent to the consumers and the feedback of the consumers are getting more complex day by day and businesses have to give more importance to healthy communication. Online commerce is increasing very rapidly, and many aspects of communication and product recommendations in online stores are becoming more important. The comments made by the users are read by millions of people within a few minutes, and millions of replies are received to these comments, and if there is no strategic and instant intervention, processes leading to the bankruptcy of the enterprises can be experienced. Big data has a very critical function for effective marketing communication in this environment. Identifying consumer requests, needs, observing comments, complaints and satisfaction, and as a result, determining communication strategies will increase the effectiveness of marketing communication for the business while reducing costs on the other hand. The business will be able to make effective interventions for the right target audience with the right tools.

In addition to this, attention should be paid to the fact that people prefer to talk to people, especially when making decisions about digital communication strategies after big data analysis. Until recently, serious IT personnel, technical equipment, additional personnel, etc. were required by businesses to enable them to collect, analyze and interpret big data in millions of channels within their own structure. Today, outsourced services make big data analytics services relatively accessible to businesses. In the AP example, free usage rights, discounts, training, protocols to ensure security/privacy are indicators that infrastructure services will be offered to a much wider audience at affordable costs.

In this study, big data analysis and machine learning in marketing communication are discussed with different dimensions. In future studies, it is evaluated that big data analysis, which will be carried out

by using primary data, will be beneficial to the literature and practitioners in research on the use cases in the marketing communications of the enterprises, the obstacles encountered while using it, and the solution proposals.

REFERENCES

Akçan, B. (2016). Big Data: Producers and Consumers. *New Media Studies*, (2).

Altunışık, R. (2015). Big Data: My Source of Opportunities or New Bundle of Problems? *Yildiz Social Science Review*, *1*(1), 45–76.

Alzahrani, S. M. (2021). Big Data Analytics Tools: Twitter API and Spark. *2021 International Conference of Women in Data Science at Taif University (WiDSTaif),* 1-6. 10.1109/WiDSTaif52235.2021.9430205

AMA. (2021). *American Marketing Association*. https://www.ama.org/topics/marcom/

Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research Trends on Big Data in Marketing: A Text Mining and Topic Modeling Based Literature Analysis. *European Research on Management and Business Economics*, *24*(1), 1–7. doi:10.1016/j.iedeen.2017.06.002

Amazon. (2021a). *What is Big Data*. https://aws.amazon.com/tr/big-data/what-is-big-data/

Amazon. (2021b). *Personalize*. https://aws.amazon.com/tr/personalize/

Ariker, M., Breuer, P., & McGuire, T. (2015). *How to Get the Most from Big Data? McKinsey. In Marketing & Sales Big Data, Analytics, and the Future of Marketing & Sales*. McKinsey.

Ariker, M., & Manuel, N. (2015). Want Big Data Sales Programs to Work? Get emotional. In Marketing & Sales Big Data, Analytics, and the Future of Marketing & Sales. McKinsey.

Arrigo, E., Liberati, C., & Mariani, P. (2021). Social Media Data and Users' Preferences: A Statistical Analysis to Support Marketing Communication. *Big Data Research,* 24.

Bendle, N., & Wang, X. (2016). Uncovering the Message from the Mess of Big Data. *Business Horizons*, *59*(1), 115–124. doi:10.1016/j.bushor.2015.10.001

Benoit, D. F., Lessmann, S., & Verbeke, W. (2020). On Realizing the Utopian Potential of Big Data Analytics for Maximizing Return On Marketing İnvestments. *Journal of Marketing Management*, *36*(3-4), 233–247. doi:10.1080/0267257X.2020.1739446

Bitnine. (2016). *Introduction About Big Data Marketing*. https://bitnine.net/blog-useful-information/introduction-about-big-data-marketing

Blei, D. M., & Ng, Y. A., & Jordan M. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(4–5), 993–1022. doi:10.1162/jmlr.2003.3.4-5.993

Boyd, D., & Crawford, K. (2012). Critical Questıons for Big Data. *Information Communication and Society*, *15*(5), 662–679. doi:10.1080/1369118X.2012.678878

Buganza, T., Trabucchi, D., & Pellizzoni, E. (2020). Limitless Personalization: The Role of Big Data İn Unveiling Service Opportunities. *Technology Analysis and Strategic Management*, *32*(1), 58–70. do i:10.1080/09537325.2019.1634252

Buhalis, D., & Volchek, K. (2021). Bridging Marketing Theory and Big Data Analytics: The Taxonomy of Marketing Attribution. *International Journal of Information Management*, *56*, 1–14. doi:10.1016/j.ijinfomgt.2020.102253

Chen, Y., & Xie, J. (2008). Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix. *Management Science*, *54*(3), 477–491. doi:10.1287/mnsc.1070.0810

Chwastek, M., Pooniwala, P., & Ashman, S. (2020). *Amazon Personalize Now Supports Dynamic Filters for Applying Business Rules to Your Recommendations on the Fly*. https://aws.amazon.com/tr/blogs/machine-learning/amazon-personalize-now-supports-dynamic-filters-for-applying-business-rules-to-your-recommendations-on-the-fly/

Culotta, A., & Cutler, J. (2016). Mining Brand Perceptions from Twitter. *Marketing Science*, *35*(3), 343–362. doi:10.1287/mksc.2015.0968

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A Formal Definition of Big Data Based on İts Essential Features. *Library Review*, *65*(3), 122–135. doi:10.1108/LR-06-2015-0061

Dixon, M. (2019). *How Netflix Used Big Data and Analytics to Generate Billions*. https://seleritysas.com/blog/2019/04/05/how-netflix-used-big-data-and-analytics-to-generate-billions/

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data Consumer Analytics and The Transformation of Marketing. *Journal of Business Research*, *69*(2), 897–904. doi:10.1016/j.jbusres.2015.07.001

EU. (2021). *Big Data: Definition, Benefits, Challenges (infographics)*. https://www.europarl.europa.eu/news/en/headlines/society/20210211STO97614/big-data-definition-benefits-challenges-infographics

Fader, P., & Winer, R. (2012). Introduction to The Special Issue on The Emergence and Impact of User-Generated Content. *Marketing Science*, *31*(3), 369–371. doi:10.1287/mksc.1120.0715

Fortune. (2021). *Fortune 500*. https://fortune.com/company/amazon-com/fortune500/

Fosso, W. S., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'Big Data' Can Make Big Impact: Findings from A Systematic Review and A Longitudinal Case Study. *International Journal of Production Economics*, *165*, 234–246. doi:10.1016/j.ijpe.2014.12.031

Gulati, R., & Oldroyd, J. B. (2005). The Quest for Customer Focus. *Harvard Business Review*, 92–101. PMID:15807042

Hartmans, A. (2021). *Jeff Bezos Originally Wanted To Name Amazon 'Cadabra,' And 14 Other Little-Known Facts About The Early Days Of The E-Commerce Giant*. Businessinsider: https://www.businessinsider.com/jeff-bezos-amazon-history-facts-2017-4

Hood, A., & Kleopas, I. (2020). *Selecting the right metadata to build high-performing recommendation models with Amazon Personalize*. https://aws.amazon.com/blogs/machine-learning/selecting-the-right-metadata-to-build-high-performing-recommendation-models-with-amazon-personalize/

InsideBigData. (2019). *How Amazon Used Big Data to Rule E-Commerce*. https://insidebigdata.com/2019/11/30/how-amazon-used-big-data-to-rule-e-commerce/

Johnson, E. J. (2012). Big Data + Big Analytics = Big Opportunity: Big Data is Dominating the Strategy Discussion for Many Financial Executives. As These Market Dynamics Continue to Evolve, Expectations Will Continue to Shift About What Should Be Disclosed, When and to Whom. *Financial Executive, 28*(6).

Lemon, N., & Verhoef, C. (2016). Understanding Customer Experience Throughout the Customer Journey. *Journal of Marketing,* (80), 69-96. doi:10.1509/jm.15.0420

Lottemart. (2021). https://www.lotteon.com/p/display/main/lottemart?mall_no=4&chno=100195&ch_dtl_no=1000617

Lu, S., Xiao, L., & Ding, M. (2016). A Video-Based Automated Recommender (VAR) System for Garments. *Marketing Science*, *35*(3), 484–510. doi:10.1287/mksc.2016.0984

Luenendonk, M. (2019). *Uses of Big Data in Marketing*. https://www.cleverism.com/best-uses-big-data-marketing/

Ma, L., & Sun, B. (2020). Machine Learning and AI in Marketing – Connecting Computing Power to Human Insights. *International Journal of Research in Marketing*, *37*(3), 481–504. doi:10.1016/j.ijresmar.2020.04.005

Malthouse, E. C., Haenlein, M., Skiera, B., & Zhang, E. W. (2013). Managing Customer Relationships in The Social Media Era: Introducing The Social CRM house. *Journal of Interactive Marketing*, *27*(4), 270–280. doi:10.1016/j.intmar.2013.09.008

McKinsey. (2015). *Marketing & Sales Big Data, Analytics and the Future of Marketing & Sales*. https://www.mckinsey.com

Mihaela, E. O. (2015). *The Influence of the Integrated Marketing Communication on The Consumer Buying Behavior. Procedia Economics and Finance*.

Morgan, B. (2018). *How Amazon Has Reorganized Around Artificial Intelligence And Machine Learning*. https://www.forbes.com/sites/blakemorgan/2018/07/16/how-amazon-has-re-organized-around-artificial-intelligence-and-machine-learning/?sh=4f75f1573618

Moro, S., Rita, P., & Vala, P. (2016). Predicting Social Media Performance Metrics and Evaluation of The İmpact On Brand Building: A Data Mining Approach. *Journal of Business Research*, *69*(9), 3341–3351. doi:10.1016/j.jbusres.2016.02.010

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. doi:10.1287/mksc.1120.0713

Park, S., & Hwang, K. (2020). *Increasing Customer Engagement and Loyalty With Personalized Coupon Recommendations Using Amazon Personalize*. https://aws.amazon.com/tr/blogs/machine-learning/increasing-customer-engagement-and-loyalty-with-personalized-coupon-recommendat ions-using-amazon-personalize/

Saggi, G. M., & Jain, S. (2018). A Survey Towards An Integration of Big Data Analytics to Big Insights. *Information Processing & Management*, *54*(5), 758–790. doi:10.1016/j.ipm.2018.01.010

Shen, Q., & Villas-Boas, J. (2018). Behavior-based Advertising. *Management Science*, *64*(5), 2047–2064. doi:10.1287/mnsc.2016.2719

Simpson, J. (2017). *Finding Brand Success In The Digital World*. https://www.forbes.com/sites/forbesagencycouncil/2017/08/25/finding-brand-success-in-the-digital-world/? sh=774a012626e2

Smith, J. B., & Colgate, M. (2007). Customer Value Creation: A Practical Framework. *Journal of Marketing Theory and Practice*, *15*(1), 7–23. doi:10.2753/MTP1069-6679150101

Suoniemi, S., Meyer-Waarden, L., Munzel, A., & Straub, A. R. (2020). Big Data and Firm Performance: The Roles of Market-directed Capabilities and Business Strategy. *Information & Management*, *57*(7), 103365. doi:10.1016/j.im.2020.103365

Talend. (2021). *Big Data in Marketing 101: Why it's Important, Where it's Going, and How to Get Started.* https://www.talend.com/resources/big-data-marketing/

Tang, Q. (2020). Construction of Interactive Integrated Communication Marketing System in Big Data Era. *Journal of Physics: Conference Series*, *1550*(3), 1550. doi:10.1088/1742-6596/1550/3/032142

Walmart. (2021). https://corporate.walmart.com/our-story/our-locations

Wiencierz, C., & Röttger, U. (2018). *Big Data-Based Strategic Communication*. https://www.akademischegesellschaft.com/fileadmin/webcontent/Presse/1803_Comm_Director_CW_UR_2018_BD_based_Strat_Comm.pdf

WorldEconomicForum. (2012). *Big Data, Big Impact: New Possibilities for International Development*. http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_ Briefing_ 2012.pdf

Yavuz, V., & Erdoğan-Gençyürek, M. (2019). *Advertising Applications on Internet News Sites*. Current Marketing Studies. in Image Publishing House.

Zhan, Y., Tan, K. H., Ji, G., & Tseng, L. C. (2018). A Big Data Framework for Facilitating Product Innovation Processes. *Business Process Management Journal*, (270), 577–595.