Stephan Böhm
Sid Suntrayuth (Editors)

Proceedings of the
IWEMB 2019

Third International Workshop on Entrepreneurship
in Electronic and Mobile Business

Vestfold, Norway
September 30–October 1, 2019

Program Committee:
Agnese Augello, National Research Council of Italy, Italy
Vesarach Aumeboonsuke, International College of NIDA, Thailand
Thomas Barton, University of Applied Sciences in Worms, Germany
Lasse Berntzen, University of South-Eastern Norway, Norway
Stephan Böhm, RheinMain University of Applied Sciences, Germany
Tove Bøe, University of South-Eastern Norway, Norway
Peerayuth Charoensukmongkol, International College of NIDA, Thailand
Wolfgang Eixelsberger, University of Applied Sciences Kärnten, Austria
Rania Fahim El-Gazzar, University of South-Eastern Norway, Norway
Francis Fang, International College of NIDA, Thailand
Adrian Florea, Lucian Blaga University of Sibiu, Romania
Manuel Gentile, National Research Council of Italy, Italy
Alexander Godulla, Leipzig University, Germany
Milena Krumova, Technical University of Sofia, Bulgaria
Marisa Laokulrach, International College of NIDA, Thailand
Ela Sibel Bayrak Meydanoglu, Turkish-German University, Istanbul
Ariyaporn Nuntiphatprueksa, Silpakorn University International College
Werner Quint, RheinMain University of Applied Sciences, Germany
Piotr Soja, Cracow University of Economics
Cosmin Stoica, University of Craiova, Romania
Ulrike Stopka, Technische Universität Dresden, Germany
Sid Suntrayuth, International College of NIDA, Thailand
Peter Winzer, RheinMain University of Applied Sciences, Germany
Cornelia Wolf, Leipzig University, Germany
To cite this publication please use this DOI link
https://doi.org/10.22602/IQ.9783745870077

Bibliographic information of the German National Library:
The German National Library lists this publication in the German National Bibliography;
detailed bibliographical data are available on the Internet at http://dnb.dnb.de

publIQation
ACADEMIC PUBLISHING

PubliQation – Wissenschaft veröffentlichen

An Imprint of Books on Demand GmbH, In de Tarpen 42, 22848 Norderstedt

Copyright 2020 Stephan Böhm, Sid Suntrayuth (editors).
The authors of the individual contributions are responsible for the content
and for respecting copyrights of third parties.
The copyrights for the contributions in this proceeding remain with the individual authors.

Coverdesign, produced and published by: BoD – Books on Demand GmbH,
In de Tarpen 42, 22848 Norderstedt

ISBN Print 978-3-7458-7000-8
ISBN ePDF 978-3-7458-7007-7
Preface

The International Workshop on Entrepreneurship in Electronic and Mobile Business (IWEMB) is a joint initiative of the Center of Advanced E-Business Studies (CAEBUS) at the RheinMain University of Applied Sciences in Wiesbaden, Germany, and the International College of the National Institute of Development and Administration (ICO NIDA) in Bangkok, Thailand.

The aim of the initiative is to offer a platform for researchers in the fields of electronic and mobile business in order to generate relevant new insights and international exchange of ideas. The mission of this workshop is to bring together young and experienced researchers from institutions all over the world to discuss current electronic and mobile business research topics as well as innovations and trends in related markets. A particular interest of the initiative is to strengthen cooperation in academia between researchers from Europe and Asia.

The third IWEMB was held on September 30 and October 1, 2019 in Vestfold, Norway. The papers in these proceedings were reviewed and accepted for publication by the program committee and presented at the conference.
Contents

1 Analysis and Visualization of Travel Blogs by Means of NoSQL Techniques: Set-Up and Results
   Thomas Barton, Marco Graf, Fahri Özünlü ........................................... 11

2 Big Data and Data Analytics in E-commerce
   Lasse Berntzen, Milena Krumova, Wolfgang Eixelsberger, Tove Bøe, and
   Kim Normann Andersen .............................................................. 23

3 Factors Affecting Online Consumer’s Intention to Repurchase: A Gender Perspective
   Tove Bøe and Lasse Berntzen .................................................. 49

4 Easy access? The Usability of the Payment Processes for Paywalls on
   Newspaper Websites and the Online Kiosk Blendle: A Comparative Analysis
   Barbara Brandstetter and Michael Fürsich .................................. 71
5 Social Media Marketing Communication and Social Media Business Performance: Do Firm Age, Social Media Business Experience, and Presence of Physical Store Matter?

Peerayuth Charoensukmongkol and Penpattra Tarsakoo .......................... 93

6 How Mobile Do You Go: A Study of Five News Media Start-Up Cases in Portugal

Miguel Crespo and Ana Pinto-Martinho .............................................. 123

7 The Implications of the Espoused National Cultural Dimension of Masculinity on the Technology Acceptance of E-Learning: A Call for an Updated Measurement of Masculinity

Wendy Colleen Farrell and Bastian Eine ............................................. 135

8 Online Shopping Intention of International Students in Thailand: An Empirical Study

Sumaia Farzana and Saif-Ur Rahman .................................................. 159

9 Usage of Customer Reviews in Online Shopping: Analyzing Characteristics of Customer Reviews

Saskia Kräuter, Bastian Eine, and Werner Quint ..................................... 191

10 Chatbots in Applicant Tracking Systems: Preliminary Findings on Application Scenarios and a Functional Prototype

Sebastian Meurer, Judith Drebert, and Stephan Böhm ............................ 211
11 Application of Smart Contracts in Trade Financing Processes
Yvonne Muschler, Erik Massarczyk, and Peter Winzer ........................... 235

12 Role of Parents in Making Children’s Use of Media Screen Time more Worthwhile
Saif-Ur Rahman and Sumaia Farzana ................................................. 283

13 Chatbots for Recruiting: An Empirical Study of Recruiters’ Requirements
Anika Regber, Judith Drebert, and Stephan Böhm ................................ 311

14 Mobility-as-a-Service (MaaS): Intermodal Mobility Bundling and Packages in Urban Areas
Ulrike Stopka, René Pessier, and Christian Günther .............................. 335

15 An Integrated Model for Customer’s E-Loyalty: An Evidence from Online Shopping in Vietnam
Bui Nhat Vuong and Sid Suntrayuth ................................................... 367
Analysis and Visualization of Travel Blogs by Means of NoSQL Techniques: Set-Up and Results

Thomas Barton, Marco Graf, and Fahri Özünlü

Department of Computer Science
University of Applied Sciences Worms
Worms, Germany

Abstract—In order to analyse user-generated content NoSQL technologies have shown to be very helpful. These technologies are used to analyse and visualize data of a public travel-blog platform. Starting with the help of an in-house development in JavaScript and the Google Maps API first results had been achieved which help to understand why people travel to special destinations in the example of New Zealand and Thailand. As a next step, the whole set up for analysis and visualization is evolved by using the R software for statistical computing and graphics in order to use a standard software environment which allows to apply more sophisticated methods for analysis and visualization. Results are presented for travel reports written in the United States of America and in Australia.

Keywords—NoSQL, travel blogs, analysis, visualization, USA, Australia.
1. Introduction

Travelers today are able to share their travel experiences using mobile information and communication technologies. This lead to a phenomenon, which is known as Smart Tourism. Within Smart Tourism, travel experiences may assist other travelers in their decision making process to plan and carry own travel experiences (Gretzel et al., 2015: 179–188). To understand travelers’ sentiments user-generated content is analysed (Xiang et al., 2017: 2). Travel blogs are sources of information which contain personal travel experiences (Bosangit et al., 2009: 61–71). In order to investigate the behaviour of travel bloggers, the anonymized data of a travel blog platform have been analysed. Opposite to the studies, which have been carried out earlier, the set up for the analysis and the visualization of data has been modified. Instead of a JavaScript program a standard software environment is used, now, which allows to use already implemented algorithms for Artificial Intelligence. Furthermore, the investigations have been extended to travel reports for two countries, which were not analysed before: The United States of America and Australia. The paper is organized as follows: in Section 2 related work is listed. Section 3 specifies the source of the blogging data and the methods to analyse and visualize the data. In Section 4 the achieved results for the USA and Australia are presented. A summary of the results follows in Section 5. Section 6 gives an outlook to future work directions.

2. Related Work

Within the analysis of online diaries, content analysis and narrative analysis have been the main research methods (Banyai et al., 2012: 267–277). In order to analyse contents, methods of Artificial Intelligence have become very popular (Lee et al., 2019: 615–644). In case of travel blogs, Artificial Intelligence allows for example, to visualize the contents of blogs as a network of words (Yuan et al., 2016: 1306–1319) or to classify blog entries into tourism types (Shibata et al., 2020: 27–37). The analysis of travel reports by some of the authors of this paper started on the basis of a JavaScript program, which was developed by one author. This program was also used
to analyse the travel blogs about two countries, New Zealand and Thailand, with the following results: In the case of New Zealand, the geographic coordinates, which were derived from the posted travel reports, could be assigned successfully to the national parks of the visited destinations (Barton and Graf, 2018: 139–142). In case of Thailand, a classification of travelers into two different groups, who both show different interests and who prefer different travel destinations, has been executed successfully (Barton and Graf, 2019: 35–48).

3. Analysis and Visualization of Travel Blogs

Travel reports have been investigated on basis of JSON-documents, which had been imported in the NoSQL database MongoDB. Its aggregation framework allows to set up and to run queries in a SQL-like manner (Banker et al., 2016). The investigation of data was carried out in the cross-platform programming environment R. In order to analyse and visualize data several collections of extension software libraries were installed. In order to execute queries in the R environment the package MongoLite was applied. Thus, the results from the queries running in the MongoDB, are transferred back to the R environment for further analysis. The analysis of contents by means of text mining is carried out using the tm package (Feinerer et al., 2008: 1–54). Within this package, a text is represented by a matrix, the so called document term matrix. In this matrix, documents are represented by rows and terms are represented by columns. And the contents of cells indicate, how often a term occurs in a document (Welbers et al., 2017: 245–265). The document term matrix is used to visualize travel blog entries as a word cloud, where the size of the words represents the occurrence in the text. To visualize the most informative words as a word cloud, the package wordcloud was used. For the spatial visualization of maps using the Google maps API the ggmap package was used (Kahle et al., 2013: 144–161). In addition, the tmap package was applied for visualization of data (Tennekes, 2018: 1–39).

Within the scope of this study the travel blog platform traveloca.com served as data source (Barton et al., 2016: 712–720). At the end of 2018
the platform consisted of 96,521 travel reports with 319,694 photos or videos in fifty-nine languages over all five continents. For text mining and the visualization in word clouds only blogs of German visitors have been used.

4. Results

4.1 Travel Reports across the United States of America

4367 travel reports about The United States of America were analysed. A visualization of all travel reports across the country is shown in Figure 1.1 as a heat map.

Figure 1.1: Geographical Distribution of Travel Reports across the USA (map data © 2019 Google).

Areas with the highest number of travel reports are shown in red colour. Yellow areas represent a medium number of reports. Green areas indicate, that in these areas travel reports are published with minor numbers. Popular destinations can be found at the west coast of the United States of America, in particular, the federal state of California with the cities of Los Angeles
and San Francisco. Also the city of Las Vegas seems to be an interesting destination. At the east coast the area between Washington DC and New York City is very popular for travel bloggers, too. In the south eastern part of the United States the federal state of Florida also seems to be very interesting for travelers.

Figure 1.2 shows a word cloud for the travel reports, which were created at the western coast of the United States of America. San Francisco is the word, which is mostly used one in these travel reports. Additionally, also the cities of Las Vegas, Los Angeles, San Diego, Santa Monica and Santa Barbara are favoured. Besides these cities, the Yosemite National Park, Death Valley, Lake Tahoe and Hollywood are popular destinations. Also the term beach is used frequently in the blog entries about California.

Figure 1.2: Word Cloud for Travel Reports
Created at the West Coast of the USA

According to Figure 1.3 travel bloggers, who visit the east coast of the USA, prefer metropolises like New York, Washington DC, Philadelphia or Boston at one side. On the other side, they are interested in the Niagara Falls or in a trip to San Francisco.
People, who visit Florida, prefer according to Figure 1.4. The beaches of Miami or Daytona, stay at Orlando to visit the Universal Studios or visit the Everglades National Park in the South of Florida.
4.2 Travel Reports across Australia

3,524 travel reports about Australia were investigated. 82.9 percent of the travel reports are written in German, 15.2 percent in English and 1.4 percent in French language. The most popular destinations can be found at the eastern and southern coast as shown in the heat map in Figure 1.5. In addition, three areas are popular: One area around Perth, at the northern coast and in the north eastern coast. Finally, also an area in the middle of Australia seems to be quite popular.

In order to get an impression, which destinations are favored by the travelers, the word cloud for the corresponding areas are shown. According to Figure 1.6 the big cities at the southern and the eastern cost are very popular destinations. The most popular one is Sydney, followed by Melbourne, Adelaide and Brisbane. The Blue Mountains National Park, Grampians National Park and Kangaroo Island are also popular destinations. People seem
also to enjoy the ocean and road trips along beaches and bays. Examples are South Coast of New South Wales, Byron Bay and Gold Coast.

Figure 1.6: Word Cloud for Travel Reports
Created in the South and East Coast

Figure 1.7: Word Cloud for Travel Reports
Created in the Northern Part of Australia
In the northern part of Australia, Darwin is according to Figure 1.7 the most important destination for travelers. Darwin is the capital of the Northern Territory and owns an international airport. The Litchfield National Park with its waterfalls and the Kakadu National Park are also popular destinations. Animals like crocodiles and kangaroos, which can be found in these national parks, are subjects of travel reports. In addition, Mount Archie seems to be popular in this region. According to Figure 1.8, Alice Springs is the most important destination in the centre of Australia. It is the starting point for trips and tours into the Outback to the Uluru (or Ayers Rock) or to Uluru-Kata Tjuta National Park.

Figure 1.8: Word Cloud for Travel Reports
Created in the Central Part of Australia
5. Summary
User-generated content in terms of travel reports of a travel blog platform has been analysed and visualized using the R software environment and NoSQL technologies. In case of travel reports created in the United States of America and Australia heat maps were presented which show the spatial distribution of the reports across the two countries. Thus, regions are identified in the two countries, which are very popular for travelers. For these regions word clouds in German language were derived, which show the preferred destinations and interests of the German speaking visitors within the regions.

6. Further Work
In order to carry forward the analysis and visualization of travel reports there are different directions for future research. One approach is to analyse the destinations in more detail with more sophisticated Artificial Intelligence techniques. In this context, a cooperation with stakeholders could be very helpful. Another approach is to enrich maps with additional data sources. Another approach is to use Bayesian Statistics in order to predict the prospective behaviour of travelers. And, there are still countries, where the analysis and the visualization of travel reports have not been started, yet.

References


Big Data and Data Analytics in E-commerce

Lasse Berntzen¹, Milena Krumova², Wolfgang Eixelsberger³, Tove Bøe¹, and Kim Normann Andersen⁴

¹School of Business, University of South-Eastern Norway
e-mail: lasse.berntzen@usn.no, tove.boe@usn.no
²Faculty of Management, Technical University of Sofia, Bulgaria
e-mail: mkrumova@tu-sofia.bg
³School of Management, Carentia University of Applied Sciences, Austria
e-mail: w.eixelsberger@fh-kaernten.at
⁴Department of Digitalization, Copenhagen Business School, Denmark
e-mail: andersen@cbs.dk

Abstract—This paper presents an overview of applications of big data and analytics in e-commerce. The presentation is based on a review of recent literature, case descriptions from trade magazines, and blogs and news websites. The paper identifies some specific applications: Improved logistics, targeted marketing and sales, and enhanced customer support. It also proposes an e-commerce technologies portfolio 4.0 model, in which technologies generate diverse data for the e-commerce process. The discussion points out the potential of these applications to help a company gain a competitive advantage.

Keywords—E-commerce, logistics, marketing, customer support.
1. Introduction

Digitalization is essential for all kind of businesses and includes digital trans-
formation of the business (Ross, 2019), the emergence of the platform busi-
ness models (Parker et al., 2016), blockchain technology (Werbach, 2018),
and AI-powered business (Davenport, 2019).

This paper focuses on how big data and analytics can support e-com-
merce (Laudon & Traver, 2019). Thus, we revisit the e-commerce concept
and explore how business is supported by information and communication
technology to make operations more efficient, reach larger customer groups,
and handle payments. One of the most profound changes over time is how
big data and the associated capabilities of extracting, analyzing, and capital-
izing on them are increasingly defining the business of e-commerce. The
change is fostered by the emergence of smart products and companies but
also still more smart customers and, in the end, smart competition. In this
paper, we offer an overview of big data and data analytics applications in e-
commerce. Developments in big data and analytics play a crucial role in
how and the magnitude of the impacts on business and have resulted in
bestselling paperbacks (Clegg, 2017) and articles in leading business jour-
nals (Fountaine et al., 2019). However, online retail companies face several
challenges while also demonstrating strengths compared to traditional retail
companies.

One of the most significant challenges is increased global competition.
A few global companies have managed to get large market-shares and are
continually threatening smaller companies. Platform companies such as
Amazon and their online marketplace and planned global roll-out of the
AmazonGo Just Walk Out stores, are by some observers seen as the possible
end for retail business as we know them today. However, we do also see a
digital transformation of the physical stores with, for example, self-scan-
nning of products at the cashier, robotization of service tasks, e.g., CafeX
(Bandoim, 2019) and the Jumbo robots in Incheon airport in Seoul (Lee,
2017). Another challenge is the lack of physical presence where customers
can inspect the goods before buying. The third challenge is the lack of trust
in e-commerce, mostly caused by low-quality products and intellectual
property rights infringements (fake products). The main strength is that customers prefer to shop online from anywhere at any time. One specific kind of company ("click and bricks") combines virtual presence and physical stores.

A review of current literature shows several critical applications where big data and analytics may help a company gain a competitive advantage. These applications are improved logistics, targeted marketing and sales, and enhanced customer support.

1.1 The Value Chain

Michael Porter has, in his recent work, focused on smart products (Porter & Heppelman, 2014, 2015) and earlier on made a significant contribution to understanding how the Internet might impact the dynamics of the competition (Porter, 2001). In this paper, we will bring attention to the concept of the corporate value chain as a general framework for thinking strategically about business activities and assessing their relative cost and role in differentiation (Porter, 1985). The value chain is based on a set of primary activities to handle the transformation of inputs into finished products or services (inbound logistics, operations, and outbound logistics, but also marketing and sales, and service. The model also includes support activities, namely infrastructure, human resource management, technology development, and procurement. Figure 2.1 shows the value chain, as proposed by Porter. In using this model for teaching purposes, it is essential to acknowledge the right-hand side of the figure (margin) and emphasize that the supportive and primary activities are part of a cost-driven chain if the margin plus support costs plus primary cost are lower than the sales price.

The primary activities consist of inbound logistics, operations, outbound logistics, marketing and sales, and service. The primary activities generate the revenues of the company. Table 2.1 explains the different primary activities with focus on retail.
Table 2.1: Primary Activities

<table>
<thead>
<tr>
<th>Primary Activities</th>
<th>Technologies in Retail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inbound Logistics</td>
<td>Inbound logistics is about acquiring individual products to fulfill an order. For retail companies, this is about ordering goods, receiving the goods, and storing the goods for future processing. Big data and analytics will play a role in predicting the demand for each item to reduce the stockpile, but at the same time serve customers without delay. Typically, this is handled through an enterprise resource planning (ERP) system that communicates with the corresponding supplier systems.</td>
</tr>
<tr>
<td>Operations</td>
<td>Operations are the activities required to fulfill an actual order. This includes checking that the individual components are in stock, and if they are, fetching the individual components from the stock and pack them together. Big data and analytics can be used to optimize the localization of the products in the warehouse and the process of fetching the individual items of an order.</td>
</tr>
</tbody>
</table>
Table 2.1: Primary Activities (continued)

<table>
<thead>
<tr>
<th>Primary Activities</th>
<th>Technologies in Retail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outbound Logistics</td>
<td>Outbound logistics handle the final step to get the order to the customer, including ordering transport. Big data and analytics can play a role in selecting the best transport solution and optimize transport when orders are for different addresses.</td>
</tr>
<tr>
<td>Marketing and Sales</td>
<td>Marketing is to make existing and potential customers aware of products. Sales is handling orders and billing. Big data and analytics play an essential role in customer relationship management (CRM). Predicting potential sales may optimize the marketing and sales organization.</td>
</tr>
<tr>
<td>Service</td>
<td>Service includes customer support and maintenance. Big data and analytics may be used to automate customer support, providing users access to relevant solutions at all times. Chatbots can help to make better customer support. Prediction may be used for preventive maintenance.</td>
</tr>
</tbody>
</table>

The secondary activities (shown in Table 2.2) include procurement, human resource management, technological development, and company infrastructure. The secondary activities are essential for the running of the company, but they are generally not generating revenues themselves. Secondary activities may be candidates for outsourcing. Outsourcing is when a business buys services from another company instead of providing the services itself. Outsourcing may be beneficial for economic reasons, but also to obtain competence and robustness.

Table 2.2: Secondary Activities

<table>
<thead>
<tr>
<th>Secondary Activities</th>
<th>Technologies in Retail</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procurement</td>
<td>Procurement is selecting and ordering goods to be retailed. It is closely connected to inbound logistics. Big data and analytics may play a role in deciding on what suppliers to use, the size of orders, and when to order.</td>
</tr>
<tr>
<td>Human resource management</td>
<td>Human resource management is about handling employees. Big data and analytics may be used for predicting the need for human workforce on an hourly, daily, weekly and monthly basis. It can also be used to analyze workforce performance and sick-leaves.</td>
</tr>
</tbody>
</table>
Table 2.2: Secondary Activities (continued)

| Technology Development | Technology development is important for the retail segment. The electronic storefront needs to be updated as new technologies enter the market. Recommendation systems and new payment solutions need to be implemented. For retail companies, technology may be a critical asset. Even if the technology itself is delivered by a third-party, the data obtained from the systems represent value for the company. |
| Firm Infrastructure    | Infrastructure is buildings, computers, networks, etc. Big data and analytics may be used to keep track of infrastructure, including maintenance. |

1.2 Data in Business Applications

Chapman et al. (2000) introduce a model to explain the use of (big) data in business contexts. The model is shown in Figure 2.2, and includes six steps:

1. Business understanding
2. Data understanding
3. Data preparation
4. Modeling
5. Evaluation
6. Deployment

While the value chain model focuses on business activities, this model shows the steps for developing data-centric applications within the company. The model starts with understanding the business and its data. By using a set of iterations, the data are prepared and modeled. The result is evaluated and deployed but also used to improve business understanding.
1.3 Big Data and Analytics

Laney (2001) defined big data as having high volume, high velocity, and/or high variety. High volume refers to large amounts of data, demanding both specialized storage and processing. High velocity refers to streams of real-time data, for example, from sensor networks or large-scale transaction systems. Finally, high variety is about dealing with data from different sources, having different formats.

According to Marr (2015), the real value of big data is not in the large volumes of data itself, but in the ability to analyze vast and complex data sets beyond anything we could ever do before. Due to recent advances in data analysis and cloud computing, the threshold for using big data has diminished.

Big data can be structured or unstructured. Structured data is simple to process since the format is fixed. Transaction records, sensor readings, and GPS (Global Positioning System) locations are examples of structured data. Unstructured data has no fixed format and requires other means of processing. Unstructured data includes text, but also images and video content. Often unstructured data has some meta-data attached that provide some
information about the unstructured data, for example, tags describing the content, date, and location of an image or a video.

Revenue opportunities are closely connected to the value chain described above. Each part of the value chain can be separate units that can either be done in-house or outsourced to external partners.

The next section will examine business opportunities by using big data and analytics. Section 3 will describe e-commerce 4.0 and a technology portfolio. The last section concludes the paper.

2. Digital Business Applications of Big Data and Analytics

Our literature review identified a set of typical applications for big data and analytics in the e-commerce domain: Improved logistics, targeted marketing and sales, and enhanced customer support.

2.1 Improved Logistics

A key issue in retailing is to optimize logistics. Having items in stock is expensive for the company. The company needs to pay for the item within a specific period after it is received. If the item stays in the warehouse beyond that time, the company gets capital tied up. The inbound logistics needs to focus on “just-in-time” delivery. This can be achieved by prediction and good communication within the supply chain. “Just-in-time” delivery has, to a large extent, been established by the automotive industry. In some cases, a supplier has its own store within the production plant, with the responsibility to deliver when demanded.

In retail, internal logistics needs to be efficient. The process of fulfilling an order should be optimized with regard to costs. The most effective way to reduce costs is to reduce the use of workforce. With advances in robotics and Industry 4.0, the storage and fulfillment of orders can be automated.

The last step is the delivery of orders to the customer. It is necessary to reduce the costs of delivery to make e-commerce attractive. One possibility is to form partnerships with delivery companies to ensure fast delivery and low transportation costs.
An EU Horizon 2020 project is currently investigating the role of big data in transforming the current e-commerce logistics practices for bringing value to all the involved stakeholders, including third-party logistics providers, online retailers, and customers (TT, 2018).

More specifically, the project will investigate how big data exploitation can enable collaborative practices that will enhance the efficiency of the distribution process in e-commerce. The project will also try to provide alternative shipping methods to the consumer to respond to the growing e-commerce consumer demand. The ambition is to provide generic guidelines about the role of big data in e-commerce logistics and provide a roadmap for applying big data solutions with the purpose of tackling specific requirements in e-commerce logistics. The project is ongoing and is a clear indication of the importance of big data and analytics in the logistics domain.

Figure 2.3 shows how big data and analytics impact the logistics from suppliers through a warehouse or retailer to the customer.

![Figure 2.3: E-commerce Logistic Concept Model](image)

2.2 Targeted Marketing and Sales

Amazon.com was a pioneer in recommending specific books to customers based on their individual shopping preferences. In 1998 Amazon.com employees applied for a patent on “item-to-item” collaborative filtering using association between the products themselves (Mayer-Schönberger & Cukier, 2013). The method could be used on all kinds of products, not only
books. Other companies, for example, Netflix, have adopted such automatic recommendations. The use of big data has shown the possibility of making valuable correlations without knowing the underlying causes.

The ultimate scenario is explained by Agrawal, Gans, and Goldfarb (2018). They present a vision of moving from shopping-to-shipping to shipping-to-shopping. By using prediction, the seller may know what the customer wants. This can result in higher sales, but also larger number of returns.

A physical store selling goods does also take the risk of ordering goods that will not be sold. The physical store uses prediction. In most cases, unsold items cannot be returned, so they end up in sales and eventually flea markets or as waste.

The ultimate prediction solution is where customers are unlikely to return goods. To analyze customers, and especially groups of customers, the concept of personas may be used. Personas are fictional characters representing a group of users with similar attributes and similar behavioral patterns. Personas are used in software development (e.g., user-centered design) and for online marketing supporting decision processes related to the development and marketing of services and products. Descriptions about personas typically include information about behavioral patterns, skills, and attributes and will change over time. The creation of a persona is often based on intuition and guesses about users. To increase the quality of personas, the development may be based on data analysis resulting in data-based personas (Mijač et al., 2018). For the creation and update of personas, data may be collected via web services while the users are using the service or product. There is even a potential of automatically generating personas based on, for example, social media analytics (An et al., 2016).

A recent trend in marketing is conversational marketing. “Conversational marketing and sales is the process of having real-time, one-to-one conversations in order to capture, qualify, and connect with your best leads” (Tuzovic & Paluch, 2018: 12). Customers see the online world as a real-time on-demand world. If companies do not respond to a new lead or sales inquiry within a short period of time (e.g., five minutes), the likelihood of
starting a sales process or conversion will be significantly reduced. Live chat is a possibility to communicate with visitors instantly. Live chat is, in many ways, a difficult and expensive way of interacting. To automate the conversation with visitors, chatbots are often used to answer questions, route visitors to responsible people, or schedule a meeting. The conversation is stored in chat logs that can be analyzed. Chat logs are not only offering information about the questions and problems of visitors but also about the wording that customers are using (voice-of-customer language). To use the same wording in the chatbot may increase the sales numbers because visitors like to communicate with systems using their own words. The voice-of-customer language may not only be used within chatbots but also in other channels like website, e-mail, and social media postings.

To optimize the handling of customer’s data, data mining approaches are used to identify valuable customer groups and interesting purchasing patterns. The RFM (Recency, Frequency, Monetary) analysis is used to extract customer profiles and for customer segmentation. An extended version of the RFM analysis has been proposed (called R+fm) because the analysis is complicated since user behavior changes all the time, the original models need supervision for providing meaningful segments and the necessity to pay attention to the customer behavior similarity (Tavakoli et al., 2018). The main difference between the two models is that R+fm dynamically builds customer segmentation models. The market basket analysis (i.e., frequent pattern mining) is used to identify interesting purchasing patterns. The analysis is based on association rules, and the idea is to find relationships between purchases. Both approaches, RFM and market basket analysis, can be successfully integrated to discover patterns (denoted as RFM-customer-patterns) (Hu & Yeh, 2014).

Purchasing patterns are also used in recommender systems. Such systems serve two purposes—(1) stimulate uses to do something like buying a product or (2) supporting users in dealing with information overload (guiding users through the complex task of finding things they like). The methods typically used for recommender systems are collaborative, content-based,
knowledge-based, and hybrid methods (Jannach et al., 2011). Recommender systems may include not only customer preferences but also customer reviews to support the customer and also the seller. Consumer review-driven recommender services semantically explore the topics in reviews to derive product features and combine them with customer preferences to develop recommendations (Lin et al., 2017). The process is divided into four different phases: data preprocessing, product profile construction, consumer preference inference, and recommendation generation. The approach can be combined with other recommendation techniques to augment the performance.

2.3. Discovering Customer Satisfaction

Customer satisfaction measure of how products and services supplied by a company meet or surpass customer expectation. Customer satisfaction with the e-commerce system is a crucial issue because of the fact business depends on its customers (Khalid et al., 2018). The studies show that the key performance indicators (KPIs) for measuring customer satisfaction in e-commerce are diverse (Jiradilok et al., 2014; Mockingfish, 2018; Rakuten, 2019; Webfx, 2019). One of the many examples is the monitoring of e-commerce site traffic, which is among the most important KPIs. For instance, the business can use Google Analytics data to monitor traffic and check out the referrals to see if another website or social media network is sending any or a lot of traffic. Another valued source of data for e-commerce KPIs is social media, like Facebook and Twitter. By gaining knowledge through social media data analytics about customer behavior and satisfaction, e-commerce organizations can develop a strategy for making effective sales of their products. Approaches to be used in the process of customer satisfaction discovery are presented in Figure 2.4.
Data mining provides data for natural language processing, social media listening, and sentiment analysis, and the analysis is used to report results (KPIs).

- **Data Mining (DM)** (Facthosts, 2017; Matsatsinis et al., 1999)—technique frequently adopted by large-scale e-commerce businesses to aid with marketing and product development. The raw data can come from the database associated with an e-commerce website that holds all account and payment information, as well as from web analytics tools like Google Analytics and social media. This allows the company to start using data mining techniques to build visitor profiles to analyze the demographic of visitors to their site.

- **Market segmentation**—Using data mining techniques, it is possible to start building profiles for the type of person that visits the site.
Basket analysis—Data is taken at every stage of the user’s journey through the site. This data can be processed to tell the company that perhaps they should consider offering free delivery, or maybe they should send an e-mail to that customer in a few days with a special offer to prompt them to return and finish their purchase.

Sales forecasting—Sales forecasting with data mining is where a company can use data to predict things useful for stock control, pricing, and marketing.

Fraud prevention—Data mining techniques detect abnormal patterns in a data sequence.

Social media listening—Also known as social media monitoring, is the process of identifying and assessing what is being said about a company, individual, product, or brand on the Internet. Social listening software is to have a screenshot of the public sources fast.

Natural Language Processing (NLP)—Businesses that seek to better grasp the sentiments of their customers might have to sift through thousands of messages in order to get a feel for what customers are saying about the products or services. NLP may help businesses with a high volume of customer feedback garner insights from it in the form of quantifiable trends, or in other words, increases and decreases in the frequency of specific customer complaints (Azulay, 2019). NLP tools can scan texts for specific meanings on all kinds of data sources. NLP software is the necessary decision-making tool.

Sentiment Analysis (SA)—Sentiment analysis refers to the process of mining the texts in order to identify the tone of the passage written by the user. There are usually three tones for a passage written by the user: positive, negative, and neutral (Murugavalli et al., 2017).

Reporting (KPIs). Reporting includes sales, marketing, and customer service key performance indicators.
2.4 Enhanced Customer Support

Customer support is an essential part of an online retail company, but the provision of customer support is expensive in salaries. Therefore, many companies are implementing solutions for automated support, for example, by chatbots. Chatbots are software robots doing tasks usually done by employees. The idea is to make customer service more efficient and available at all times. The chatbots can employ machine learning to provide more accurate help over time. Chatbots can simulate a conversation (or a chat) with a user in natural language through messaging applications, websites, mobile apps, or through the telephone. From a technological point of view, a chatbot only represents the natural evolution of a question-answering system leveraging Natural Language Processing (NLP).

In the context of customer experience, three stages of Chatbot maturity can be identified (Jani, 2018):

- **Informational Interactions**—the bots understand natural language and respond to the questions asked. Such bots are used for a less interactive customer experience like FAQ responses.

- **Personalized Interactions**—the bots are connected back end applications and can thereby generate user-specific responses. The bots act similarly to a human and may deliver impressive customer experiences.

- **Transactional interactions**—Intelligent bots that help the user through a series of instructions (or interactions). The integration of customer data is possible through connected back end systems. The bots may deliver remarkable customer experiences.

Bots ask users for user’s preferences and inclinations and collect thereby data from users. The architecture of bot systems may contain data analytics capabilities delivering analytical information. The data may be used to understand the customer’s preferences or choices.
The capabilities of chatbots will be measured through specific metrics (AI Multiple, 2019):

- **User Metrics** describe metrics related to users like active or engaged users.

- **Message Metrics** represent interactions of individuals with the chatbot. Examples are Bot Messages (number of messages sent by the chatbot in an interaction) or Miss Messages (messages that the chatbot is not able to process).

- **Bot Messages** measuring the performance of a chatbot. Examples are the Goal Completion Rate (percentage of successful engagements) or Retention Rate (percentage of users returning to use the chatbot in a defined time frame).

Chatbots are not limited to texts. Three recent student projects at the University of Southeast Norway demonstrated the use of speech in chatbots. The quality of voice recognition and speech generation is rapidly improving and may provide opportunities for companies to provide even better customer support.

3. **E-commerce 4.0 and Technology Portfolio**

The dynamic increase in numbers of online orders, as well as the sales fluctuations to which e-commerce is always exposed, present new challenges for businesses. The logistics departments must be configured for high speed and ever-shorter delivery times. Along with the challenges in the typical applications, the e-commerce business faces a lot of opportunities for value creation based on the technologies used, such as data analytics, dashboards visualizations, and much more e-commerce 3.0 and e-commerce 4.0.

- **E-commerce 1.0**—Traditionally, B2B organizations used sales reps and physical stores to sell products and services to the market.

- **E-commerce 2.0**—By giving customers more information and free rein when using the web store, businesses were able to create and
retain a loyal customer base by providing them with everything they wanted and more.

- E-commerce 3.0—Businesses are looking at third-generation platforms that are able to increase order conversion and revenue using e-commerce. They want a solution that can be fully integrated within their existing infrastructure, and third-generation solutions bring both the internal and external facing systems together to deliver greater personalization and customer experiences.

- E-commerce 4.0—Customer demands continue to evolve with time. Merchants respond by using artificial intelligence (AI), so they know what the customer wants before the customer does. They also respond by immediately fulfilling orders. As e-commerce 4.0 sits on the horizon, so does predictive purchasing, the technology will start to think for you cutting out any human contact (Ham, 2018).

- Industry 4.0 refers to a new phase in the industrial revolution that focuses heavily on interconnectivity, automation, machine learning, and real-time data. As such, the specific segment of the industry—e-commerce 4.0 possesses the same characteristics, where data plays a key role in business advancement, value creation, and competitiveness. The research argues that e-commerce (Totonchi & Kakamanshadi, 2011) technologies are used for each stage of the e-commerce process, starting from the marketing (Hudak et al., 2017) and sales, customer order to the stage of the product, and service delivery, including different cloud models IaaS, PaaS, and SaaS (HCL, 2019; IJCST, 2018; Lackermair, 2010; Zymr, 2019). The knowledge on how to use technology in the web 4.0 era, where e-commerce 4.0 operates, has a high impact on the competitive advantage. The practice shows that the knowledge for the new e-commerce models, such as e-commerce 4.0, impacts the efficiency of the technologies used. The Pareto principle (also known as the 80/20 rule, the law of the vital few, or the principle of factor sparsity) states that, for many events, roughly 80 percent of the effects come from 20 percent of the
causes. In Figure 2.5, the e-commerce 4.0 model and Pareto abstraction are proposed.

E-commerce technologies can be discussed in different contexts and for different purposes, such as the industry sector (IT, textile industry, education, health, etc.), the stage of the e-commerce process, the business maturity or type of the company (SME or large corporation), and more. For example, e-commerce recommendation systems are used to filter choices for a particular user based on their past searches or other customer’s search or purchase data. It gives users a personalized view on the e-commerce website and helps them to select relevant products. Three techniques are widely used: collaborative filtering, content-based filtering, and hybrid recommendation filtering.

**Figure 2.5: E-commerce 4.0 Model and Pareto Abstraction—Technologies and Knowledge**

Big data analytics technology is a combination of several techniques and processing methods. What makes them effective is their collective use to