

IMPLEMENTING ARTIFICIAL INTELLIGENCE IN THE EVALUATION OF PACKAGING
DISTRIBUTION AND LABEL DESIGN MODELING

By

Shiva Esfahanian

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Packaging - Doctor of Philosophy

2024

ABSTRACT

Packaging evaluation is an important process to ensure the safety of packages. The process involves evaluating the packaging of the product to make sure it reaches the customer in a good condition. Packaging engineers evaluate the packaging design by using different methods such as mechanical testing, computer simulations like FEM, and collecting data from participants to simulate the real environment. Despite the value of data collection and mechanical tests in initial design, there remains a need for a fast and inexpensive method to continuously evaluate packaging designs post-implementation. While mechanical testing and computer simulations have been implemented to fulfill this need, these methods have limitations in effectively evaluating packaging performance. By increasing the abundance of data collections, data-driven approaches such as artificial intelligence (AI) have gained more attention. In this work, an attempt was made to implement AI in packaging evaluation. The implementation was examined in two phases:

1. A Novel Packaging Evaluation Method using Sentiment Analysis of Customer Reviews.

As mentioned, current approaches toward evaluating packaging design (test and FEM) have their own limitations. These methods try to simulate the actual environment, but since these simulations cannot yet perfectly replicate real-life scenarios, the question remains of whether they are successful in this pursuit or not. Therefore, it is essential for the packaging engineers to evaluate their packaging design even after it is implemented. To address these shortcomings, a method was proposed to evaluate the packaging performance in an actual environment through customers' reviews using natural language processing (NLP). NLP is a branch of AI that teaches computers to understand and process human language, enabling tasks like translation and sentiment analysis. Based on the results, engineers can identify potential sources of design failure. Moreover, the percentage of failures over various months and years is examined to identify the potential effects of seasonal changes on packaging failure. In our work, we compared three different TVs, labeled as A, B, and C. The percentage of packaging failure for each of them was 5.73%, 9.60%, and 11.19%, respectively, which means TV C had the worst protection through distribution. By using this method, the packaging performance can be evaluated from customer reviews instead of

physical testing which will save time and cost.

2. Machine Learning Modeling of Patients' Attention of Over-the-counter Medication Label Design.

Progress has been made toward creating a model that can be used to study the effects of different parameters on label design and predict the patient's behavior with respect to parameter changes. In this respect, data provided by the Healthcare, Universal Design, Biomechanics (HUB) research group at Michigan State University (MSU) was utilized. They implemented a new design for Over-the-Counter (OTC) medication. They put a small box in front of the package (FOP), highlighting important information for the patients to increase patients' attention, and their likelihood to read the label of OTC medication. The HUB research group gathered their results from ninety-two participants by using a change detection method. The goal of this project was to predict the response time of the participants with respect to changes. In this project, the impact of content change on response trials with three classes (hit, miss, and time out) was studied using three ML approaches: Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN), chosen for their effectiveness with categorical data. The models were trained and tested for accuracy and area under the curve (AUC), which measures their ability to differentiate between response classes. Results showed accuracies of 85% for DT, 88% for RF, and 90% for KNN, with corresponding AUC values of 0.5, 0.54, and 0.5. Despite satisfactory accuracies, the AUC scores indicated that the models' performances were not as effective as expected. Further analysis indicated that the models' limitation to predicting only in the 'hit' class was due to imbalanced data. By applying SMOTE, the dataset was balanced more effectively, resulting in a boost in model performance. Although the accuracy decreases to 78% and 81% for RF and DT models, respectively, the AUC value increased to 0.81 and 0.79 for RF and DT, demonstrating SMOTE's efficacy in managing imbalanced datasets. The model significantly improved its classification of the 'hit,' 'miss,' and 'time out' classes.

Copyright by
SHIVA ESFAHANIAN
2024

"This dissertation is dedicated to my husband, **Hamid**,
for his continued and unfailing love, support and understanding during my pursuit of
PhD degree,
also
to my respectful parents and parents-in-law for always believing in me."

ACKNOWLEDGEMENTS

Writing this acknowledgment made me see how hard it is to properly thank everyone for their help, support, and understanding while I worked on this dissertation. I'm worried I might forget to thank someone or not give enough credit to their help. But it's clear that a lot of people helped me finish this work, and I'm only able to mention a few of them here.

Firstly, I would like to extend my sincere thanks to my advisor, Dr. Euihark Lee. Your guidance has been important in helping me develop as an independent researcher. I am truly grateful for the opportunity and support you have provided me.

Next, I want to deeply thank my committee members, Dr. Laura Bix, Dr. Amin Joodaky, and Dr. Qiben Yan.. Their extensive experience has been a valuable source of learning for me and has significantly contributed to my growth as a graduate student.

I am really thankful to Dr. Abdol-Hossein Esfahanian for always being there when I needed help or advice. Your advice has been so valuable to me, and I'll always be grateful for it.

I would love to list all of my friends here, but I'm grateful to each one of you for making my journey more enjoyable and for being like a second family to me while I was away from home. Wishing all of you success and happiness in your future endeavors.

My dissertation's completion owes much to my best friend and wonderful husband, Hamid Mohammadi. His steady support, patience, and sacrifices during my struggles, times away, and moments of frustration and impatience are truly commendable. Additionally, my heartfelt thanks go out to my dear parents, in-laws, siblings and their families, especially my sister, Elaheh and brother-in-law, Alireza, for their immense support.

Thank God for being with me all the times.

TABLE OF CONTENTS

CHAPTER 1 Introduction	1
1.1 Motivation	1
1.2 Objective of the Study	3
1.3 Structure of the Thesis	3
CHAPTER 2 Background and Literature Review	4
2.1 Packaging Distribution and Evaluation	4
2.2 Over the Counter (OTC) Drug Labeling	15
2.3 Overview of Artificial Intelligence (AI)	19
CHAPTER 3 A novel packaging evaluation method using SA of customer reviews	35
3.1 Introduction	35
3.2 Method	37
3.3 Result	46
3.4 Conclusion	56
CHAPTER 4 ML modeling of patients' attention of OTC medication label design	58
4.1 Introduction	58
4.2 Method	59
4.3 Result	71
4.4 Conclusion	78
CHAPTER 5 Discussion	80
5.1 Summary	80
5.2 Challenges	82
5.3 Suggestion for the Future Study	83
BIBLIOGRAPHY	85
APPENDIX	99

CHAPTER 1

Introduction

1.1 Motivation

Packaging evaluation is an important process to ensure the safety of packages during distribution. It is a systematic process of assessing and analyzing packaging materials, designs, and functions to determine their effectiveness in achieving specific objectives and meeting predefined criteria. The process involves a comprehensive examination of various aspects of packaging, such as its visual appeal, structural integrity, sustainability, usability, and its ability to protect and preserve the product it contains.

Packaging evaluation varies significantly across industries, each with specific focuses. Cosmetics packaging combines product protection with aesthetic appeal [1]. The distribution and logistics sector emphasizes durable and efficient packaging for safe product transit and storage [2]. In the medical field, especially for over-the-counter (OTC) medications, packaging ensures sterility, safety, and efficacy with clear labeling. Finally, food packaging aims to maintain freshness, prevent contamination, and offer consumer convenience [3], ensuring product quality and safety.

This research primarily focuses on two key industries: distribution and medical. It aims to explore and compare the distinct challenges and methodologies in packaging evaluation within these sectors.

Packages experience different hazards during distribution, like shocks, vibration, temperature changes, and pressure change thus, packaging plays an important role in protecting products. Also, every new package should undergo a series of physical tests, to ensure the protection of the package. These tests are designed to replicate real-world conditions, leading to the question: do they accurately emulate actual scenarios? At this stage, it becomes essential for engineers to assess the package design after it has been commercialized, emphasizing the critical nature of packaging

evaluation.

Three prevalent methods for evaluating packaging are field testing, lab testing, and computer simulation, such as finite element modeling (FEM). Each of these evaluation methods comes with its own strengths and weaknesses. Field testing involves assessing packaging in real-world scenarios. This approach gives insights into how packaging holds up during transportation, storage, and actual consumer use. Although field testing offers invaluable real-world insights, it can also be time-consuming and expensive. Lab testing, by contrast, consists of controlled experiments conducted in a laboratory environment. It aims to evaluate specific facets of packaging performance. There are several standardized test methods available, including those from the international organization for standardization (ISO), american society for testing and materials (ASTM), and international safe transit association (ISTA). While lab testing ensures controlled conditions, it might not always emulate real-world situations perfectly. Computer simulations like FEM employ mathematical models to anticipate how packaging will behave under different conditions, negating the need for physical prototypes. Such simulations are not only cost-effective but also aid in design optimization. However, their success largely depends on the precision of the models used.

In the medical field, particularly for OTC medication, packaging evaluation is a critical process that ensures consumer safety and compliance with regulatory standards. This evaluation focuses on the clarity and accuracy of labeling, which is essential for guiding consumers in the correct usage of the medication. Key elements of an OTC label include detailed information on active ingredients, dosage instructions, intended use, warnings, potential side effects, and contraindications. The design of the packaging also plays a role in ensuring the product's integrity and preventing tampering or misuse. Regulatory bodies like the U.S. Food and Drug Administration (FDA) set stringent guidelines for OTC medication labeling to maintain high standards of public health and safety [4].

Consequently, we are exploring methods that are cost-effective, capable of simulating real environments, and easy to implement. Artificial intelligence (AI) can be considered as a solution that encompasses all these factors. In recent years, the capabilities of AI have expanded and now

offer solutions to various challenges. Among its diverse applications are machine learning (ML) and natural language processing (NLP). NLP, in particular, has shown prowess in analyzing and understanding human language. For instance, it can be effectively used to determine sentiments from textual data. In our initial project, we leveraged NLP to probe the sentiment present in customer reviews, especially those related to the packaging of products. On the other hand, ML specializes in pattern recognition among vast sets of data. In a subsequent project, this research employed ML to detect specific patterns within data related to OTC medications. By doing so, the aim is to refine and potentially reduce the amount of data collection required when dealing with new datasets.

1.2 Objective of the Study

The primary objective of this study is to integrate AI into the evaluation processes within the packaging industry. To achieve this, AI methodologies were applied in two distinct projects.

Objective 1: The first project delves into the realm of NLP, aiming to gauge packaging performance by analyzing customer reviews from online platforms. This approach offers a unique perspective by capturing direct consumer feedback, potentially highlighting areas that traditional assessment methods might overlook.

Objective 2: The second project utilizes machine learning techniques, specifically decision trees (DT), random forest (RF), and k-nearest neighbors (KNN), to classify study models. This classification can pave the way for optimized evaluation processes, thereby reducing the need for resource-intensive in-person testing in the future.

1.3 Structure of the Thesis

The structure of this thesis is outlined as follows: The first chapter delves into the motivation and objectives of the study. In the second chapter, the background and a literature review related to the work are presented. The subsequent two chapters detail two projects that integrate packaging with AI. Lastly, the final chapter offers a discussion on the findings and implications of this thesis.

CHAPTER 2

Background and Literature Review

2.1 Packaging Distribution and Evaluation

To gain a clearer understanding of this section, begin with the definition of 'Packaging'. Packaging involves the process of designing, evaluating, and producing packages. It can be described as a coordinated system of preparing goods for transport, warehousing, distribution, logistics, sale, and end use [5]. This part is divided into two sections: 'Packaging Distribution' and 'Packaging Evaluation', each of which is explained in detail in the subsequent text.

2.1.1 Packaging Distribution

Packaging distribution refers to the entire process of storing, transporting, and delivering products in their respective packaging to retailers or directly to end consumers. This process ensures that products remain intact and retain their quality from the point of manufacture to the final destination. During their journey through the distribution chain, packages encounter a variety of hazards [6, 7]. In this context, packaging assumes a critical function, acting as a shield to safeguard the products within. It ensures that goods remain undamaged and intact with minimum cost, maintaining their quality and value as they move through various stages of distribution [2].

Packaging is categorized into three types based on its function and objective: primary packaging, secondary packaging, and tertiary packaging [8].

- **Primary Packaging:** This is the material that first envelops the product and holds it. It is in direct contact with the product itself. Examples include a bottle containing liquid detergent or a carton containing eggs.
- **Secondary Packaging:** This packaging groups several primary-packaged products together. It is used to protect the product during transportation and handling. Examples include the

cardboard box containing individual packs of crisps or the plastic rings holding together a six-pack of soda.

- **Tertiary Packaging:** It is used for bulk handling, warehouse storage, and transport shipping. The primary objective is to protect the product during transportation and storage. An example would be the pallets on which goods are placed for transport or the shrink wrap used to secure them. The figure 2.1 shows each of these types.



Figure 2.1. Primary-Secondary-Tertiary Packaging [9]

Although primary, secondary, and tertiary packaging provide substantial protection against many distribution and handling challenges, some risks remain. These include manual and mechanical handling challenges, as well as distribution risks such as compression, impacts, vibrations, and environmental conditions. Each of these will be briefly explained below:

- Manual handling: It refers to the basic operations like folding, inserting, wrapping, sealing, and labeling, which are part of the packaging process and are performed manually by workers[10].

- Mechanical handling: Refers to the use of machinery, such as forklifts, to handle and transport materials, specifically focusing on how different types of forklifts, pallet designs, entry speeds, and

loads impact the physical forces exerted during the process of moving pallets [11].

- **Compression:** In the context of packaging and distribution, refers to the force exerted on a package or its contents due to external pressures. This can happen during storage (such as when products are stacked on top of each other) or transportation (due to tight packing or shifting cargo). Understanding and testing for compression resistance is crucial to ensure product integrity and safety during handling and shipping [12].
- **Impact:** The impact refers to the effects that handling, storage, and transportation conditions have on the integrity and stability of packaged goods. Throughout the distribution process, packages are subject to various dynamic forces, like impacts resulting from loading, unloading, and vehicle movements. Such forces can jeopardize the structural integrity of the packaging and the safety of the product inside, especially for fragile items or those susceptible to environmental conditions. As such, the impact in packaging distribution is a fundamental consideration, directly influencing product protection, distribution efficiency, and overall customer satisfaction.
- **Vibration:** Typically involves a recurring pattern with a relatively mild intensity. During transportation, the movement of the vehicle continuously produces a constant level of vibration. To analyze this motion more easily, these vibrating systems are often simplified and depicted as systems consisting of springs and masses [13].
- **Environmental Conditions:** Climatic hazards occur as a result of variations in temperature, atmospheric pressure, and humidity. Exposure to low temperatures can lead to the freezing of liquid solutions and the cracking of their containers. On the other hand, high temperatures bring about several negative effects, including faster diffusion rates and the intrusion of water vapor, which can result in contamination. Additionally, high temperatures can lead to the loss of volatile substances in products, as well as chemical reactions like hydrolysis and oxidation [14].

Additionally, there are various types of transportation, such as rail, truck, and ship, each with its own risks. In order to get the protection for these risks, a series of physical tests are being done on new packaging. A few of these tests include shock, vibration, compression, and others [15].

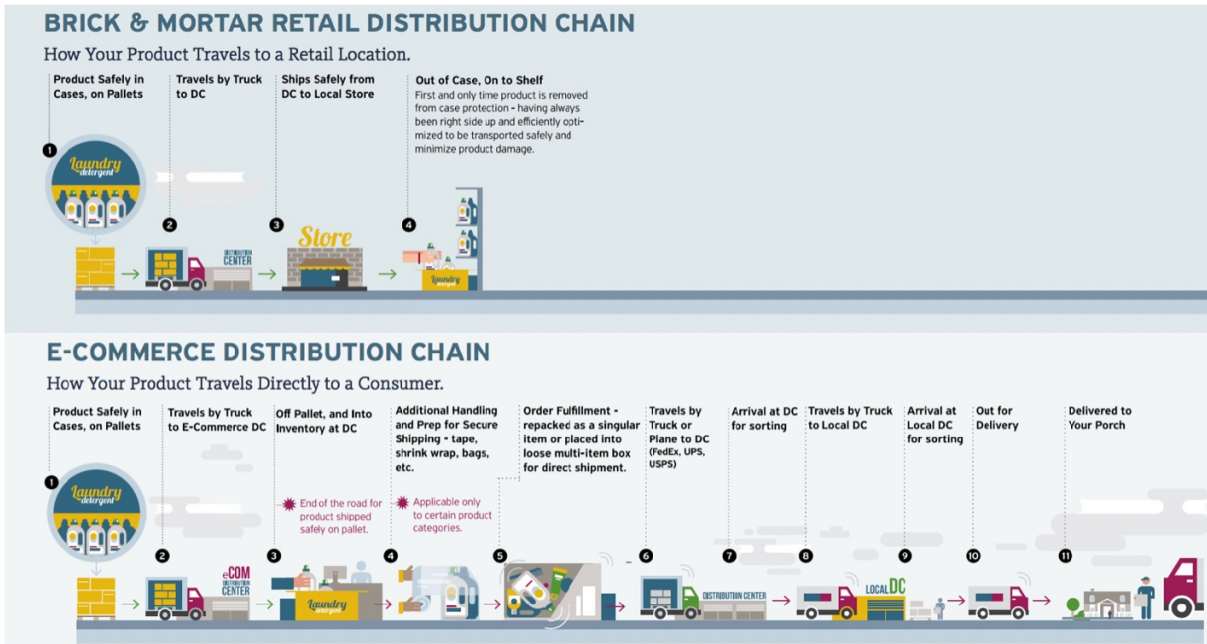


Figure 2.2. Brick and Mortar vs E-Commerce[16]

Also, beyond the various modes of transportation, there are distinct distribution channels such as e-commerce and brick-and-mortar retailing. Therefore, the different type of distribution need different packaging. The packaging requirements for e-commerce differ significantly from those for brick-and-mortar retail due to the increased complexity of the distribution process. Figure 2.2 showed the journey of a package from the manufacturer to the consumer in both brick-and-mortar and e-commerce settings. It is clear that the e-commerce route involves a more extensive sequence of touch points. Let's delve into the detailed journey a package undergoes in both brick-and-mortar retail and e-commerce.

Brick and mortar retail distribution chain is: the process starts with manufacturing the product, which is then packaged and sent in pallets to a distribution center (DC). At the DC, these products are then dispatched in cases to retail stores. It's at the retail store where the products are finally

unpacked from their protective cases, and placed on shelves for customers to buy and use. It's the consumers who browse the store aisles, choose the items they want, scan them at checkout, pack them into bags, and then carry their purchases home.

The e-commerce approach is characterized as selling products directly to the consumer. In e-commerce distribution chain, after the product is transported by truck, it goes to the inventory at the DC. There, it is taken off the pallet, and some items are wrapped and prepared for shipping, which involves additional handling. Orders may consist of single or multiple-item packages for delivery. These then travel by plane or truck to various DCs, such as UPS, FedEx, or USPS. Upon arrival at the DC, they are sorted and sent to local DCs before finally being dispatched for delivery to customers [16]. As Sara [16] has demonstrated, packages in e-commerce are subject to three times more touch points compared to those in traditional retail channels.

2.1.2 Packaging Evaluation

Packaging evaluation involves the comprehensive analysis and testing of packaging components and systems to determine their efficacy in preserving product quality, ensuring user convenience, complying with regulations, minimizing environmental impact, and achieving other desired objectives [17]. According to Robertson[18], to fairly evaluate packaging, it is necessary to recognize the various packaging functionality.

Packaging function based on Robertson [18] can be categorized into six main groups: containment, protection, apportionment, unitisation, convenience, and communication. Each of these will be explained in detail.

- Containment: Packaging's primary role, which many might overlook, is containment. Except for large items, products need packaging to be transported. Whether it's a milk bottle or a cement wagon, the package needs to securely hold the product. If not contained properly, there could be significant environmental pollution, like cement spilling from an open truck or chemicals from a leaking drum. Effective containment through packaging is crucial in modern society to prevent environmental damage as numerous products are moved daily. Poor packaging could lead to substantial environmental pollution.

- Protection: The main job of packaging is to shield its contents from external factors like water, gases, bacteria, and physical damages, and also to safeguard the environment from potentially harmful products like toxic chemicals. For many foods, packaging is vital for preservation. For instance, juices and milk in aseptic cartons stay safe only while the package is intact; similarly, vacuum-packed meat's shelf life relies on the packaging being airtight. If the packaging is compromised, the product's preservation is lost.

- Apportionment: Packaging's role in dividing large industrial quantities into consumer-friendly sizes is often overlooked but crucial. For instance, a large vat of wine is divided into bottles, and bulk butter is packaged into small portions. In essence, modern society's mass production relies on packaging to distribute products into manageable sizes for consumers. The affordability of many products is due to large-scale production and its associated cost savings. As production scales up, so does the importance of packaging to break down products into consumer-friendly amounts.

- Unitisation: Packaging streamlines the process of transporting goods both nationally and internationally. Instead of handling each item separately, primary packages are grouped into secondary ones, like corrugated cases. These secondary packages are then combined into tertiary packages, such as stretch-wrapped pallets. This can even extend to a fourth level, where multiple pallets are placed in a container. Through this layered packaging approach, handling becomes more efficient as fewer individual packages or loads need to be managed.

- Convenience: Modern lifestyles and societal shifts, such as changing family structures and more women in the workforce, have influenced the packaging industry. Increases in single-person households, changing eating habits like snacking on the go, a variety of food and drink needs at outdoor events, and more free time, have spurred a demand for convenience in products. People want pre-made foods that can be quickly prepared, easy-to-use cleaning products, organized medication packs, and mess-free dispensers. Packaging has been crucial in meeting these needs by making products user-friendly.

- Communication: The old saying "a package must protect what it sells and sell what it protects" holds true today. Packaging plays an important role in marketing, making products easily

recognizable through branding and labels, which aids in efficient self-service shopping. Without distinctive packaging, shopping would be tedious and confusing. Additionally, modern checkouts use universal product codes (UPC) on packages for quick scanning. Packaging also communicates essential details in warehouses and distribution centers; without proper labels, operations can become chaotic. In international trade, clear symbols are crucial due to language differences, but many packages still miss out on providing this vital information [18].

Now that there is familiarity with the various functions of packaging, it becomes important to understand the methods of packaging evaluation, especially in the context of distribution. These methods include field testing, lab testing, and computer simulations, each of which is explained in detail.

In the realm of user experience, field testing refers to evaluating a product or system in the user's environment to understand its usability and the user's interactions with it under real-world conditions.

Field testing in distribution testing for packaging refers to evaluating the packaging's performance under real-world conditions. It involves testing the package during actual transportation and handling processes to observe how well it protects its contents. This could include monitoring the package through different transportation modes (like trucks, ships, or planes), handling at warehouses, and exposure to various environmental conditions [13].

An example of field testing in packaging is using Lansmont sensors (Figure 2.3) which involves placing these sensors inside or on a package to monitor real-world conditions during transit. Lansmont sensors are capable of recording data on impacts, vibrations, temperature, and humidity. For instance, a company shipping fragile electronics might use Lansmont sensors in their packages to track the conditions experienced during various transportation modes such as trucks, ships, or planes. The collected data helps in analyzing how well the packaging protects its contents against real-world stresses, enabling the company to make informed decisions on packaging design and materials to enhance product safety during shipping.

One of the significant advantages of field testing is that it offers insights that reflect real-world



Figure 2.3. Lansmore Sensors [19]

use and performance, providing a more genuine picture of how a product or system will function in its intended environment. This real-world testing often facilitates direct interaction with end-users, yielding valuable feedback that might not surface in lab settings. Additionally, field tests can be adapted to a variety of environments, capturing the range of conditions a product might encounter. By accounting for unexpected variables, field testing can illuminate unforeseen challenges or benefits[20].

However, field testing is not without challenges. Due to its real-world nature, there's a distinct lack of control, making it difficult to isolate specific variables and potentially introducing uncertainties in results. Field tests, especially those conducted in diverse or remote locations, might prove to be more time-consuming and costlier than their lab counterparts. Ensuring consistent test conditions across multiple field tests can be a challenge, complicating comparisons between tests. External factors such as weather conditions, human error, or other unpredictable elements can also influence and potentially skew the outcomes [20].

Lab testing refers to the evaluation or examination of a product, material, or system in a con-

trolled environment where conditions can be monitored and manipulated to assess various properties or functionalities [21].

Lab testing in packaging distribution typically involves a series of controlled experiments designed to simulate the stresses that packaging might encounter during shipping and handling. This includes vibration testing, where packaging is placed on a vibration table to mimic the effects of transportation by truck or rail, assessing the structural integrity and the protection of contents. Drop tests from various heights and angles are conducted to simulate potential falls during handling, evaluating the packaging's ability to withstand impact shocks. Additionally, compression testing is performed to determine the maximum load the packaging can endure, simulating the pressures of stacking during transit and storage. Environmental testing in climate-controlled chambers assesses the packaging's resilience to extreme temperatures and humidity changes, mirroring varying weather conditions it might face (Figure 2.4). The results from these tests are crucial for identifying weaknesses in the packaging design and making necessary improvements to ensure product safety and integrity throughout the distribution process [22].

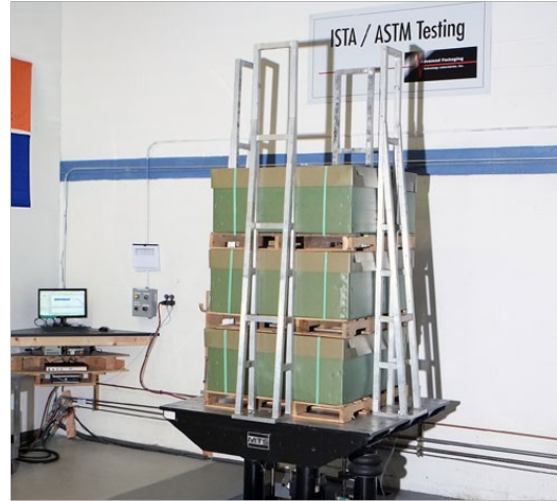
According to Kirk [26], lab testing offers several advantages, foremost being the ability to conduct tests in a controlled environment. This controlled setting ensures the consistency of test conditions, allowing for a higher degree of precision and accuracy in the results. Such an environment is also beneficial for safety, especially when dealing with potential hazards or dangerous materials. The isolation of specific variables in a lab setting makes it easier for researchers to determine causal relationships, and labs typically house advanced equipment, providing the ability to garner detailed insights.

However, lab testing is not without its drawbacks. One significant limitation is that results from a lab might not always be applicable or directly translatable to real-world scenarios due to the highly controlled conditions. Setting up and maintaining a lab, especially with advanced equipment, can be quite costly.

In certain fields, such as medical research, lab testing can raise ethical concerns, particularly when it involves testing on animals. The controlled nature of lab tests might mean that some



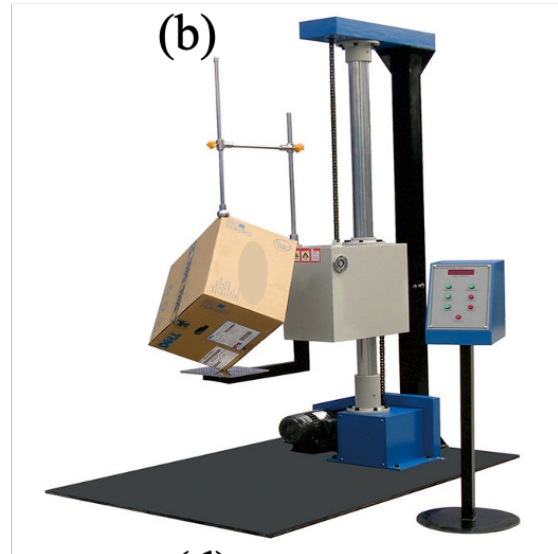
(a)



(b)



(c)



(d)

Figure 2.4. Lab Testing: (a) Compression [23], (b) Vibration [24], (c) Enviromental Chamber, and (d) Drop [25]

real-world variables get overlooked, potentially limiting the scope of the results. Additionally, the entire process of designing, setting up, and conducting lab tests can be time-intensive[26].

A computer simulation is a method of using a computer to mimic a physical experiment. Essentially, a simulation runs a model of the system to which you want to make inferences, in place of performing a test on the real system, which may be hazardous, time-consuming, or expensive [27].

Computer simulation has different models; one of them is finite element method (FEM). FEM is extensively used in structural engineering to predict and analyze the behavior of structures under various loads. By breaking down a larger structure into smaller, simpler parts (finite elements), engineers can simulate stress, strain, and deformation to ensure a structure's safety and efficiency before physical construction begins[28]. Fadiji et al. [29] utilizing finite element analysis (FEA) in packaging distribution focuses on the structural performance of ventilated corrugated paperboard packaging. It emphasizes how FEA can effectively model and predict the behavior of packaging materials under various mechanical loads common in distribution scenarios. Their study particularly examines how design elements such as vent holes and material thickness influence the overall strength and integrity of packaging. It demonstrates the importance of FEM in optimizing packaging design, enhancing durability and efficiency in distribution environments.

Louong et al.[30] employs FEA to model the strength of corrugated board boxes subjected to impact dynamics. The researchers utilize finite element simulations to study the structural behavior and mechanical properties of corrugated board packaging under various impact conditions (Figure 2.5). Through their analysis, they aim to better understand the factors influencing the strength and durability of corrugated board boxes, providing insights into optimizing their design for improved performance and protection of packaged goods during transportation and handling.

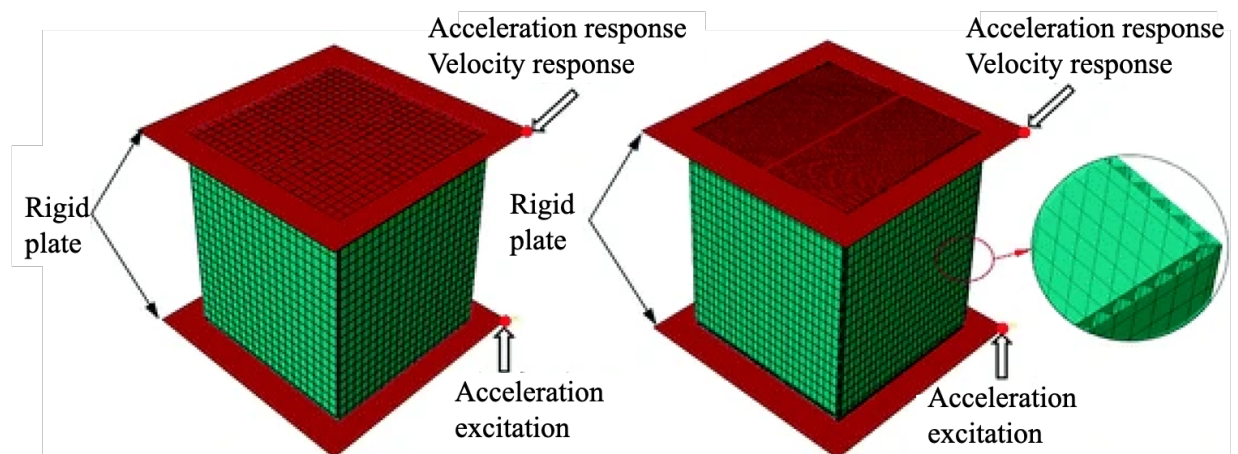


Figure 2.5. Finite element simulation of corrugated board box under impact dynamics [30]

The FEM is a renowned computational approach extensively used in fields like engineering and

physics to solve complex equations on intricate geometries. Its major strength lies in its versatility, aptly handling non-uniform shapes and varying material characteristics. This makes it instrumental in areas like structural analysis, fluid movement, and heat transfer. FEM's ability to detail localized effects is particularly useful when addressing stress points or specific heat sources.

However, FEM does have its shortcomings. Its computational intensity, especially for detailed three-dimensional models, often demands high-end computational equipment. The reliability of the results is also dependent on the mesh's quality and the selection of element types. Incorrect choices can lead to discrepancies in outcomes. Moreover, for those unfamiliar with FEM, understanding its results requires a deep grasp of the method's intricacies [31]. Therefore, a method for packaging that relies on computer simulation and can be time-consuming is being sought, which aims to be streamlined by using artificial intelligence.

2.2 Over the Counter (OTC) Drug Labeling

2.2.1 OTC medicines

Medications in the US in 1951, with the passage of the Durham-Humphrey Amendment to the Federal Food Drug and Cosmetic Act (FEDCA), were legislated into two categories, over-the-counter (OTC) and prescriptions. Prescribed medications are those that supervision by a doctor and are mandated to carry the label, "Caution: Federal law prohibits dispensing without a prescription." This requirement is due to their habit-forming nature or the risk of harm that could arise from improper use [32]. In contrast, OTC medications, also known as non-prescription drugs, are deemed safe and effective for use by individuals without the need for a physician's prescription or supervision [33]. For every dollar expended on OTC, the health care system saves approximately 7.20 dollars (Instead of using the healthcare system to get prescription medications), which is an estimated total of 146 billion dollars savings annually [34]. Furthermore, OTC has other advantages like privacy, convenience, flexibility, and quick access. Consequently, OTC found a vital role in America's health system [35]. A drug must typically undergo a comprehensive, data-focused process called the "RX to OTC switch" to be classified as an OTC medication. In the United States, the US Food and Drug Administration (FDA) regulates this process [36].

In the U.S., the number of OTC drug products on the market is estimated to range from 100,000 to 300,000. A growing quantity of these products is being imported from international manufacturers and distributors. Annually, American consumers spend billions of dollars on these OTC medications. Additionally, as healthcare costs continue to rise in the U.S., an increasing number of consumers are turning to OTC drugs for self-medication [37].

The definition of self-medication is described as the act of using drugs, herbs, or home remedies based on one's own decision or following someone else's suggestion, without seeking a doctor's guidance [38]. Successful self-medication involves accurately identifying symptoms and choosing a suitable medicine or product, along with the right dosage and timing. Additionally, it requires knowledge of the individual's previous medical history, any existing co-morbid conditions, and any other medications being taken. A crucial part of this process is also regularly monitoring how the treatment is working and watching out for any possible side effects [39].

Despite its popularity and advantages, self-medication with OTCs comes with risk. Simple, and routine decisions about OTCs can have negative consequences [40]. Like negative effects or adverse drug reaction (ADR) due to drug misuse. Drug misuse can be a consequence of drug-drug interactions or drug diagnosis interactions. ADR can be defined as "an appreciably harmful and unpleasant reaction resulting from an intervention related to the use of a medicinal product" [41]. A meta-analysis showed annually 106,000 US deaths occur because of ADRs[42].

These effects are more commonly seen in groups at higher risk, such as the elderly, individuals with limited literacy, and those who speak languages other than the native one. Additionally, people following complicated medication schedules are also more susceptible [40].

To reduce the chances of an ADR, various approaches are necessary. However, in the context of OTC drugs, labeling is a key preventive strategy. While there are multiple sources of information available for choosing and using OTCs, research indicates that often the label is the only source consumers refer to [43]. Consequently, proper labeling plays a significant role in preventing ADRs, including issues like overdosing, interactions between different drugs, or conflicts between a drug and a specific medical diagnosis [44].

2.2.2 Labeling

In Section 201(m) of the Federal Food, Drug, and Cosmetic Act (FFDCA), the concept of "labeling" is specifically defined:

"all labels and other written, printed, or graphic matter (1) upon any article of any of its containers or wrappers, or (2) accompanying such article"

The term "label" is included under the wider definition of "labeling," and it is described in section 201(k) of FFDCA as:

"display of written, printed, or graphic matter upon the immediate container of any article..." [45].

Labeling is widely recognized as an effective way to convey important information. Consumers often prefer using OTC product labels as a reliable source of information for making healthcare decisions [46]. For self-medication, labels on OTC medicines are important as they provide essential details for the safe and effective use of the medication. This includes information about active ingredients, how to use the medicine, safety warnings, and dosage instructions, all aimed at helping patients make informed choices and correctly use the medicine [47]. Building on this area of theoretical research, researchers have employed various versions of these models in efforts to categorize and comprehend the way patients interpret and use labeling information on packaging. This understanding assists in decision-making regarding medical products, focusing on how consumers evaluate and utilize them [48, 49, 50].

Regulated information for OTC drugs consists of two main parts (that shows in Figure 2.6): (1) the Principal Display Panel (PDP)(21 CFR 201.66), referred to as "the part of a label that is most likely to be displayed, presented, shown, or examined under customary condition of display for retail sale" and (2) the Drug Facts Labeling (DFL)(21 CFR 201.66) that contain "the active ingredients" and their purpose the product's uses, warnings, directions, other information, and inactive ingredients.

Numerous studies have focused on enhancing patient attention through modifications in the design of labels. These changes include using larger and more prominent font sizes, strategically

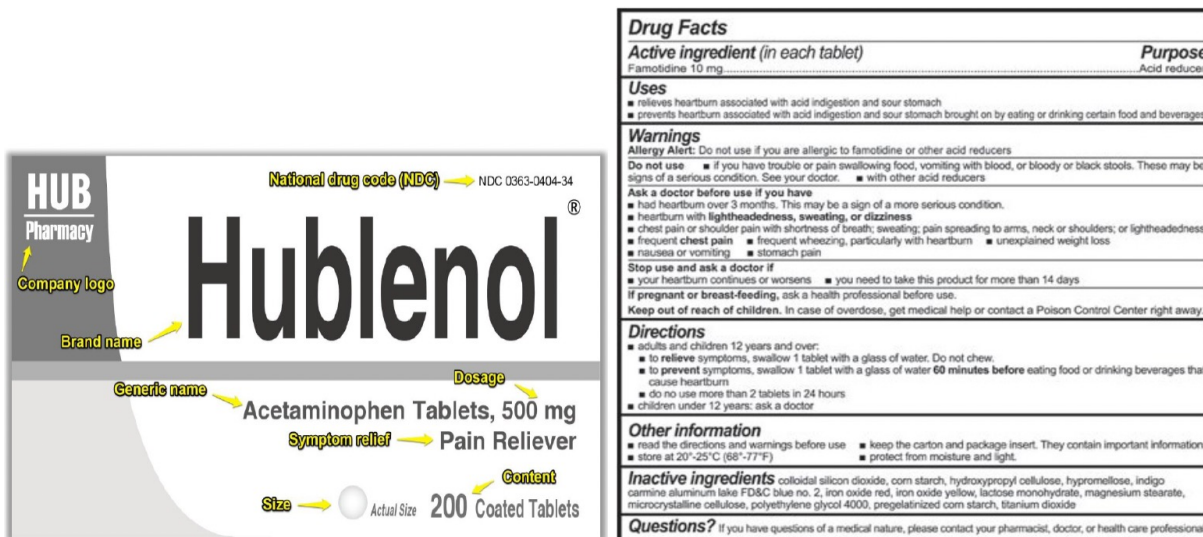


Figure 2.6. Principal Display Panel and Drug Facts [40]

positioning crucial information on the front part of the packaging, and emphasizing warning instructions to ensure they stand out[51, 52]. One of these studies [40], the basis of this work is a study which suggests the implementation of a small box on the Front of the Package (FOP) and highlight important information.

In this work [40], researchers collected information from ninety-two participants through a change detection technique. The study necessitated participants to come to Michigan State University for an initial assessment and the main test, involving a two-hour computer session where they responded to a range of questions and shared their personal information. This method, however, is expensive and time-intensive for both participants and researchers. Moreover, each time a new label design is introduced or different groups of participants are included, new in-person testing is required. To address the challenges associated with in-person testing, this research introduces the concept of employing machine learning models, a subset of artificial intelligence, for modeling patients' attention to label designs. The details of this study will be explained in Chapter 4.

2.3 Overview of Artificial Intelligence (AI)

2.3.1 Introduction of AI

AI was first defined by Stanford Professor John McCarthy in 1955 as "the science and engineering of making intelligent machines." In other words, it is related to understanding human intelligence by using computers [53, 54]. AI has a rich and complex history that spans several decades [55].

The field of AI was officially established in 1956 during the Dartmouth Conference, where researchers gathered to explore the possibility of creating machines that could simulate human intelligence [56]. Early AI research focused on areas such as problem-solving, symbolic reasoning, and natural language processing. Key figures during this period include Alan Turing, John McCarthy, Marvin Minsky, and Allen Newell, who laid the foundations of AI research [57].

However, the field faced significant challenges during the 1970s and 1980s, which became known as the "AI Winter." High expectations for AI capabilities did not match the actual results, leading to a decrease in funding and interest. Progress in AI was slower than anticipated, and symbolic AI and expert systems dominated the field [58]. This period of reduced enthusiasm led to a reassessment of goals and approaches in AI research.

The resurgence of AI came in the 1990s and beyond, driven by advancements in machine learning (ML) and neural networks. ML gained prominence as a subfield of AI (Figure 2.7), emphasizing the development of algorithms that enable computers to learn from data. The availability of large datasets and increased computing power propelled ML, leading to breakthroughs in areas such as computer vision, speech recognition, and data mining [59, 60]. Furthermore, deep learning (DL), a subset of ML (Figure 2.7) that utilizes neural networks with multiple layers, has revolutionized AI applications, achieving remarkable results in image and speech recognition, natural language processing, and autonomous systems [61].

AI systems, particularly those based on ML, work by utilizing algorithms and large datasets to learn patterns and make predictions or take actions. A key aspect of ML is the training process, where algorithms are exposed to labeled data and adjust their internal parameters to minimize the difference between predicted outputs and true labels. This process allows the algorithm to learn

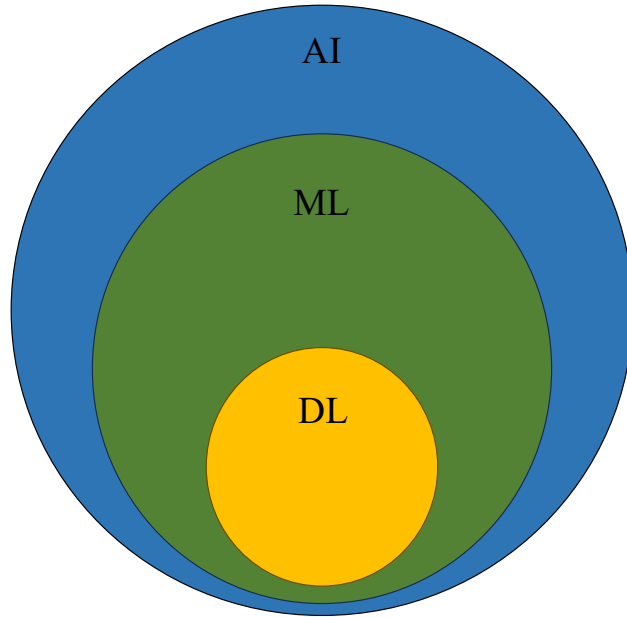


Figure 2.7. AI, ML, and DL

complex representations and generalize from the training data to new, unseen examples. According to Pedro Domingos, a professor of computer science, AI algorithms are designed to "learn from experience and extract knowledge from data to make accurate predictions or take actions" [62].

2.3.2 Introduction of ML

ML a scientific discipline, focuses on developing algorithms and statistical models that allow computer systems to perform tasks without explicit programming. Instead, the system learns to identify patterns and make predictions based on data provided to it [63].

ML has a rich history that dates back several decades. In the 1950s and 1960s, early work on neural networks and perceptrons laid the foundation for ML. Arthur Samuel's development of a program that could learn to play checkers in the 1950s is often cited as one of the first practical applications of ML [64]. However, ML faced challenges in the 1970s as symbolic AI and expert systems gained prominence [65]. The resurgence of ML came in the 1980s and 1990s with the rediscovery of the backpropagation algorithm, enabling efficient training of neural networks [66]. This led to significant advancements in the field, including the development of support vector machines (SVMs) and decision trees. The availability of large datasets and increased computing

power in the 2000s further propelled ML, giving rise to data-driven approaches and the era of big data [67]. Throughout its history, ML has evolved from early neural networks to sophisticated data-driven approaches, thanks to key milestones and contributions from researchers.

ML has various algorithms like, supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, multi-task learning, ensemble learning, neural network, instance based learning [63].

2.3.2.1 Supervised Learning

In supervised learning, ML algorithms learn from labeled training data, where each example is associated with a known output or label [68]. The goal is to develop a model that can accurately predict the output for new, unseen inputs. Supervised learning is commonly used in tasks like classification and regression. For example, in image classification, an algorithm can learn to classify images into different categories based on labeled training data [69]. Famous applications of supervised learning is decision tree (DT), naive bayes, SVM [63], k-nearest neighbour (KNN), and random fores (RF). Each of which is explained in detail in the following.

A DT is a ML approach that recursively partitions the input data into subsets based on the values of input features. It uses a tree-like structure, where each internal node represents a feature or attribute, each branch represents a possible value or outcome of that feature, and each leaf node represents a decision or prediction. The tree is constructed by recursively splitting the data based on feature values to optimize a criterion such as information gain or Gini Impurity [68, 70].

Gini impurity is a measure of impurity or uncertainty to evaluate the quality of a split at a particular node. It quantifies the probability of misclassifying a randomly chosen element from the dataset if it were randomly labeled according to the class distribution at that node. The formula for calculating Gini Impurity is as follows:

$$Gini\ Impurity = 1 - \sum (p_i)^2 \quad (2.1)$$

where p_i represents the probability of an element belonging to class i [71].

Naive Bayes is a supervised ML algorithm based on Bayes' theorem and the assumption of feature independence. It is commonly used for classification tasks, particularly in NLP and document categorization. Naive Bayes calculates the probability of a class label given a set of features by assuming that the features are conditionally independent of each other, given the class label [72, 73]. The Naive Bayes algorithm is efficient, simple, and particularly well-suited for text classification tasks. Despite the assumption of feature independence, Naive Bayes can perform well in practice and often provides fast and accurate predictions [72].

Support Vector Machine (SVM) is used for both classification and regression tasks. It works by finding an optimal hyperplane in a high-dimensional feature space that best separates the data points of different classes or predicts the target values. SVM is particularly effective in handling high-dimensional data and datasets with clear class boundaries [69]. The SVM algorithm operates as follows: given a labeled training dataset, SVM aims to find the best hyperplane that maximally separates the data points of different classes. The hyperplane is defined as a decision boundary that separates the data with the largest possible margin. SVM transforms the input data into a higher-dimensional feature space using a kernel function. The kernel function computes the similarity between data points in the original feature space [74]. In the transformed feature space, SVM searches for the hyperplane that maximizes the margin, which is the distance between the hyperplane and the nearest data points of each class. The data points closest to the hyperplane, known as support vectors, play a crucial role in defining the decision boundary. For classification, SVM assigns new data points to classes based on which side of the decision boundary they fall. For regression, SVM predicts the target values based on their position relative to the hyperplane [75].

K-nearest neighbors (KNN) is used for classification and regression tasks. It operates on the principle of similarity, where the class or value of a new data point is determined by its proximity to the k nearest neighbors in the training dataset [68]. In KNN, the value of K represents the number of neighbors considered and it works as given a labeled training dataset, where each data point consists of a set of features and a corresponding class label (for classification) or value (for

regression) [73]. When a new unlabeled data point needs to be classified or predicted, the algorithm identifies the k nearest neighbors in the training dataset based on a distance metric (such as Euclidean distance or Manhattan distance) that measures the similarity between feature vectors. For classification, the class label of the new data point is determined by a majority vote among its k nearest neighbors. The class label that occurs most frequently among the neighbors is assigned to the new data point. For regression, the predicted value of the new data point is typically calculated as the average or weighted average of the values of its k nearest neighbors [76, 77].

RF is an ensemble ML algorithm that combines multiple DT to create a more robust and accurate model [78]. It is widely used for both classification and regression tasks. RF builds an ensemble of DT by training each tree on a random subset of the training data and using a random subset of features for each split [79]. The RF algorithm operates as follows: given a labeled training dataset, RF creates an ensemble of DT. The number of trees in the ensemble is a user-defined parameter. For each tree in the ensemble, a random subset of the training data is selected with replacement (known as bootstrap aggregating or "bagging"). This random sampling introduces diversity in the training data for each tree [80]. Additionally, for each split in a DT, only a random subset of features is considered. This further enhances the diversity among the trees. As mentioned previously, each DT is trained using the selected data and features, typically using a criterion such as information gain or Gini Impurity to determine the optimal splits at each internal node.

One of the differences between RF and DT is, RF is usually more accurate compared to DT. RF is used in different industries such as banking, healthcare, and marketing. This is because of its versatility, robustness, and ability to handle complex datasets with a mix of categorical and numerical features [81].

2.3.2.2 Unsupervised Learning

Unsupervised learning is a ML paradigm where the goal is to discover underlying patterns, structures, or relationships within a dataset without the use of explicit labels or target outputs. Unlike supervised learning, unsupervised learning algorithms work with unlabeled data, relying on inherent patterns or similarities present in the data to uncover meaningful information [82, 83]. In

unsupervised learning, the algorithm explores the data and seeks to identify inherent structures or clusters, capture dependencies, or reduce the dimensionality of the input space. It does so by leveraging techniques such as clustering, dimensionality reduction, anomaly detection, or generative modeling [69].

Bishop [59] discusses unsupervised learning as a fundamental aspect of ML and pattern recognition. It covers various unsupervised learning techniques, including clustering, dimensionality reduction, density estimation, and generative models. The book explores the theoretical foundations, algorithmic approaches, and practical applications of unsupervised learning, providing valuable insights into the field. In unsupervised learning, clustering algorithms group similar data points together based on their proximity or similarity measures. Dimensionality reduction techniques aim to reduce the dimensionality of the data while preserving its important characteristics [77]. Density estimation methods, estimate the probability distribution of the data, providing insights into its underlying structure. Generative models learn the underlying data distribution and can generate new samples from it. Unsupervised learning is widely used in various domains, including data mining, computer vision, NLP, and anomaly detection. It plays a crucial role in exploratory data analysis, data preprocessing, and feature learning [84]. Example of unsupervised learning is customer segmentation, imagine a retail company that wants to understand its customer base better for targeted marketing campaigns. They have a large dataset containing various customer attributes such as age, income, purchase history, and browsing behavior. By applying unsupervised learning techniques, such as clustering, the company can group similar customers together based on their shared characteristics. Using clustering algorithms like k-means or hierarchical clustering, the company can identify distinct segments within their customer base [85]. These segments may represent different customer profiles, such as young professionals, families, or high-income individuals. The algorithm autonomously discovers these segments based on patterns and similarities found in the data, without any predefined labels [68]. Once the customer segments are identified, the company can tailor their marketing strategies to each segment's specific needs and preferences. For example, they can design targeted promotions or personalized recommendations for each cus-

customer segment, maximizing the effectiveness of their marketing efforts. This real-life example demonstrates how unsupervised learning can help businesses gain insights from unlabeled data. By utilizing unsupervised learning techniques, companies can uncover hidden patterns and structures within their data, enabling them to make informed decisions and develop targeted strategies in various domains, including marketing, customer analysis, and business intelligence [86].

One key distinction between supervised and unsupervised learning is the presence or absence of labeled training data. Supervised learning relies on labeled examples to learn patterns and make predictions, while unsupervised learning explores the data's inherent structure without the need for labels [87]. Supervised learning is suitable for tasks where the desired output is known and requires accurate predictions, while unsupervised learning is beneficial when exploring and understanding the underlying patterns and relationships in the data. Additionally, supervised learning often involves a clear optimization objective (minimizing prediction errors), whereas unsupervised learning tasks may have more varied goals, such as clustering or dimensionality reduction [69, 59].

2.3.2.3 Semi-supervised Learning

Semi-supervised learning is a ML approach that utilizes both labeled and unlabeled data to improve the performance of learning algorithms [88]. It aims to leverage the abundance of unlabeled data in conjunction with a limited amount of labeled data to enhance the learning process and achieve better predictive accuracy. In semi-supervised learning, the labeled data contains instances with known class labels, while the unlabeled data consists of instances without any class labels. The algorithm aims to exploit the underlying structure and patterns in the unlabeled data to make more informed predictions on the labeled data [89]. One of the fundamental assumptions in semi-supervised learning is the "cluster assumption". This assumption states that data points that are close to each other in the input space are likely to share the same class label. By incorporating this assumption, the algorithm can propagate labels from labeled instances to neighboring unlabeled instances, effectively utilizing the unlabeled data to refine the decision boundaries. Semi-supervised learning algorithms often employ techniques such as co-training, self-training, or generative models to leverage the unlabeled data. These methods iteratively update the model using the labeled

data and then utilize the model to make predictions on the unlabeled data, incorporating the confident predictions back into the training process [90].

The example of semi-supervised learning is a social media platform that wants to detect and classify toxic or offensive comments posted by users. The platform has a small labeled dataset where certain comments are labeled as toxic or non-toxic. However, the volume of user-generated content is massive, and it is not feasible to manually label every comment. In this case, semi-supervised learning can be employed to improve the detection of toxic comments. The platform can start by training a classifier on the initial labeled dataset using supervised learning techniques [91]. This model learns from the labeled comments and can make predictions on new, unseen comments. Next, the platform can leverage the unlabeled comments available on the platform. It can use the predictions of the initial model on the unlabeled comments and identify a subset of comments where the model is highly confident in its predictions, either as toxic or non-toxic. To obtain labels for the subset of unlabeled comments, the platform can utilize user feedback mechanisms. For example, it can ask users to report whether a comment is toxic or not. By collecting these user reports, the platform can obtain pseudo-labels for the subset of unlabeled comments. With the newly labeled subset of comments, the platform can retrain the model, incorporating the pseudo-labeled instances alongside the initial labeled data [92]. This iterative process of using the model's predictions, collecting user feedback, and retraining the model can continue, gradually expanding the labeled dataset and refining the model's performance. By leveraging semi-supervised learning, the social media platform can harness the collective intelligence of its user community to effectively detect and classify toxic comments. The iterative process of model improvement ensures that the system becomes more accurate over time, allowing for better moderation and maintaining a healthier online environment. This example demonstrates how semi-supervised learning can be applied in real-life scenarios where the volume of unlabeled data is extensive, and manual labeling is challenging or impractical. By combining the power of labeled data with user feedback and iterative learning, semi-supervised learning enables more effective and scalable solutions for content moderation and user safety [89].

2.3.2.4 Implement ML to the Packaging

ML has seen growth across various scientific fields, and packaging is no exception. ML has been implemented in numerous packaging domains, such as food, delivery systems, medical supplies, beverages, and supply chain management. Below, we explore some of these studies in detail:

The paper [93] titled "Machine Learning for Predicting Chemical Migration from Food Packaging Materials to Foods" likely explores the application of ML techniques to predict the transfer of chemicals from packaging materials into food products. This research is crucial for ensuring food safety and compliance with health regulations. Typically, such a study would involve collecting data on various types of packaging materials, the chemicals they contain, and the conditions under which food is stored. ML models are then trained on this data to identify patterns and predict the likelihood and extent of chemical migration under different scenarios. This predictive capability can help manufacturers and food safety authorities in assessing potential health risks and in making informed decisions about the suitability of packaging materials for different types of food products. The ML models would be particularly useful in simulating various conditions and predicting outcomes without the need for extensive physical testing.

This study [94] addresses the challenge of ensuring fast and reliable delivery in online retailing by developing a data-driven approach to estimate and promise real-time delivery times for new customer orders. Recognizing the importance of accurate delivery time promises in managing customer expectations and enhancing satisfaction, the research utilizes tree-based models to generate distributional forecasts. These models account for the complex interplay between delivery time and various operational factors. A key innovation of the study is the introduction of an asymmetric loss function in quantile regression forests, tailored for cost-sensitive decision-making. The effectiveness of this approach is demonstrated through real-world data from JD.com, showing that the proposed method not only improves forecasting accuracy but also has the potential to increase sales volume by 6.1 percent compared to the existing policy. The study highlights the managerial significance of accurately estimating delivery time distributions, thereby enabling online retailers to strategically set promised times to maximize customer satisfaction and drive sales.

The study by Ting [95], focuses on using DL to identify blister-packaged drugs to prevent medication errors. It utilizes a DL drug identification (DLDI) model, which employs the you only look once (YOLO) DL framework for image processing. The study trained the model with images of drug blister packs, both front and back, to identify drugs accurately. The model significantly outperformed traditional computer vision solutions in identifying drugs, demonstrating over 90 percent accuracy, and has the potential to assist pharmacists in correctly dispensing drugs and reducing errors caused by look-alike packaging. This approach highlights the application of ML in enhancing the safety and efficiency of medical packaging processes.

The paper [96] titled "Deep Learning-Based Bottle Caps Inspection in Beverage Manufacturing and Packaging Process" likely focuses on the application of DL techniques to inspect bottle caps during the manufacturing and packaging process in the beverage industry. In such a study, a DL model, possibly a convolutional neural network (CNN), is trained on a large dataset of images capturing various states of bottle caps - both defective and non-defective. This model would learn to identify and classify these caps accurately, ensuring quality control in real-time during the production line. The DL system automates the inspection process, detecting issues like misalignment, improper sealing, or physical damage, which are critical for maintaining product quality and safety. The use of DL not only enhances the efficiency and accuracy of the inspection process but also significantly reduces the reliance on manual quality checks, leading to cost and time savings in the beverage manufacturing and packaging industry.

Knoll et al. [97] automate packaging planning by using different ML models. The manufacturing industry is significantly influenced by the growing trend of mass customization and the rapidly changing life-cycles of products, leading to an extensive variety of part variants. This trend necessitates an increased effort in logistics and, more specifically, in the planning of packaging. The paper introduces a method to automate packaging selection for each part based on its unique features, employing ML techniques. Historical data of product parts along with their respective packaging details are used to develop and train a two-phase ML model. This model demonstrates the capability to accurately recommend suitable packaging options, achieving an 84 percent accu-

racy rate when benchmarked against actual data from the industry.

This research [98] delves into the application of ML in supply chain management (SCM). It addresses the complexity of supply chains, which consist of various interconnected entities needing to work in collaboration to reduce total costs. A key challenge within SCM is the gap between theoretical and practical aspects, often exacerbated by unpredictable factors and difficulties in accurately forecasting customer demand. This study reviews real-world cases where ML techniques have been employed to optimize supply chain operations. By examining these examples, the research aims to showcase how ML can enhance SCM, particularly in predicting customer demand more accurately and improving overall operational efficiency. The ultimate objective is to leverage ML to narrow the gap between the current and ideal states of supply chain networks.

Evidently, the important role of ML is demonstrated in various areas of packaging, and there are even more sectors where ML can be implemented, such as language processing.

2.3.3 Introduction of NLP

NLP is a branch of AI that focuses on the interaction between computers and human language [99]. It involves the development of computational algorithms and models to understand, analyze, and generate natural language in a way that is meaningful and useful [100]. NLP tools play a crucial role in assisting companies in comprehending how their customers perceive them through various communication channels, including emails, product reviews, social media posts, surveys, and more [101]. These AI-powered tools not only facilitate the understanding of online conversations and customer sentiments towards businesses but also offer the potential to automate repetitive and time-consuming tasks. By leveraging NLP technology, companies can enhance operational efficiency, allowing employees to devote their time and energy to more fulfilling and strategic responsibilities [102].

2.3.3.1 Overview of NLP

NLP has a rich history dating back to the 1950s when researchers began exploring ways to enable computers to understand and process human language. The field has evolved over the years, influenced by advancements in linguistics, ML, and AI [103]. Early approaches focused on rule-based

systems, while later developments incorporated statistical models and neural networks. Here is a timeline highlighting key milestones: in 1950s-1960s the early years of NLP were marked by foundational work. In 1950, Alan Turing proposed the "Turing Test" as a measure of machine intelligence. In the late 1950s and early 1960s, researchers such as John McCarthy, Marvin Minsky, and Allen Newell explored the possibility of using computers to understand and generate natural language. NLP focused on rule-based approaches, during 1970s-1980s. Prominent systems like SHRDLU (developed by Terry Winograd in 1970) showcased the ability to understand and respond to simple English sentences in restricted domains. Statistical approaches gained prominence in NLP in 1990 [104]. Researchers started using ML techniques, such as Hidden Markov Models (HMMs) and Maximum Entropy Models, for tasks like part-of-speech tagging, parsing, and machine translation. The availability of large text corpora and computational resources led to significant progress in NLP in late 1990s-early 2000s. Researchers started exploring probabilistic models, such as the Naive Bayes classifier and conditional random fields (CRFs), for various NLP tasks. In 2010 deep learning revolutionized NLP with the use of neural networks. Recurrent neural networks (RNNs) and especially long short-term memory (LSTM) networks gained popularity for tasks like language modeling, machine translation, sentiment analysis, and more. NLP encompasses a range of tasks, including language understanding, sentiment analysis (SA), machine translation, speech recognition, and text generation. It involves techniques such as parsing, part-of-speech tagging, named entity recognition, and language modeling, among others [105].

NLP has different applications like: SA, text classification, chatbots and virtual assistants, text extraction, machine translation, text summarization, market intelligence, auto-correct, intent classification, Urgency detection, speech recognition [106, 107], and word co-occurrence. This overview will explain about them.

2.3.3.2 Application of NLP

As mentioned earlier, NLP has other applications which, it will explain briefly here. Text classification is a task that involves categorizing or classifying text documents into predefined categories or classes. The goal is to automatically assign relevant labels or tags to new, unseen documents

based on their content and characteristics [108]. Text classification algorithms learn from a labeled dataset, where each document is associated with a known category. The algorithms extract relevant features from the text, such as word frequencies, n-grams, or semantic representations, and use them to build a model capable of predicting the appropriate category for unseen documents [109]. Chatbots and virtual assistants are computer programs designed to simulate human conversation and provide automated assistance to users. They utilize NLP and AI techniques to understand user inputs, interpret their intent, and generate appropriate responses. Chatbots are software applications that interact with users through textual or spoken conversation [110]. They can be rule-based or AI-powered. Virtual assistants, often referred to as voice assistants, are advanced chatbots designed to provide more personalized and interactive experiences. They are typically integrated into devices or platforms and respond to voice commands or text inputs. Virtual assistants leverage speech recognition, NLP, and AI algorithms to perform tasks, answer questions, and assist users with various activities such as setting reminders, playing music, or controlling smart devices [111]. Machine translation involves automatically translating text from one language to another. NLP techniques enable the development of translation systems that facilitate cross-lingual communication [112]. machine translation is defined as a task that entails converting one string of text into another [113]. Text summarization involves condensing large amounts of text into shorter summaries while preserving the main ideas and important information. It is used in news aggregation, document summarization, and information retrieval systems [114]. Market intelligence refers to the use of AI algorithms, NLP technologies, and data mining techniques to extract valuable insights from a wide range of unstructured data sources. This includes customer feedback, social media content, websites, reports, and more. These tools are essential for analyzing large datasets to uncover patterns and trends that inform strategic business decisions. Auto-correct is defined as a type of software program that detects misspelled words, employs algorithms to determine the most likely intended words, and then modifies the text accordingly. Intent classification is a process in AI and ML where the intent of a user is determined by analyzing the language they use [115]. For example, in a customer service context, a message like "How can I find my order status?" would be

classified to understand that the customer is seeking information about their order status. Urgency detection can be defined as the process of identifying urgent communication needs by analyzing the speech acts and communicative intentions expressed in messages [116]. Speech recognition can be defined as the technology that enables the recognition and translation of spoken language into text by computers. It involves the development of algorithms and systems that can understand human speech in various conditions, including noisy environments [117].

2.3.3.3 Sentiment Analysis

SA involves determining the sentiment or emotional tone expressed in text data, such as customer reviews, social media posts, or survey responses. SA has numerous practical uses, including brand monitoring, social media, customer feedback analysis and etc [118]. In brand monitoring, companies can analyze customer sentiment towards their products or services to assess brand perception and make informed decisions for marketing and customer satisfaction improvement [119]. In social media monitoring, NLP techniques can analyze social media posts to understand public opinion on specific topics, track trends, and identify emerging issues [120]. In customer feedback analysis, SA helps businesses to automatically classify customer feedback as positive, negative, or neutral, enabling them to address concerns, improve product quality, or enhance customer support [121]. Machines, face significant challenges in understanding natural language, particularly when it comes to opinions, as humans often employ sarcasm and irony [122]. However, SA possesses the ability to perceive delicate differences in emotions and opinions, precisely determining whether they are positive or negative. SA serves as a valuable tool in the realm of social media, enabling businesses to gain insights into customer opinions about their products [123].

By utilizing SA periodically, businesses can comprehend customer preferences and concerns regarding specific aspects of their operations. For example, they can identify if customers are enthusiastic about a new feature but dissatisfied with customer service. Such valuable insights empower companies to make informed decisions, pinpointing areas for improvement and enabling them to take smarter actions to enhance their offerings [124].

2.3.3.4 Word Co-occurrence Network

A word co-occurrence network visualizes the connections between words in a text collection. In this network, an edge is formed between two words when they appear together within a certain range in a sentence, even if they are not next to each other. This graph illustrates how words are associated within the language data of the corpus [125]. In graph-based methods for NLP, word co-occurrence networks play a important role, particularly in applications such as extracting key words [126]. Figure2.8 shows an example of a word co-occurrence network.

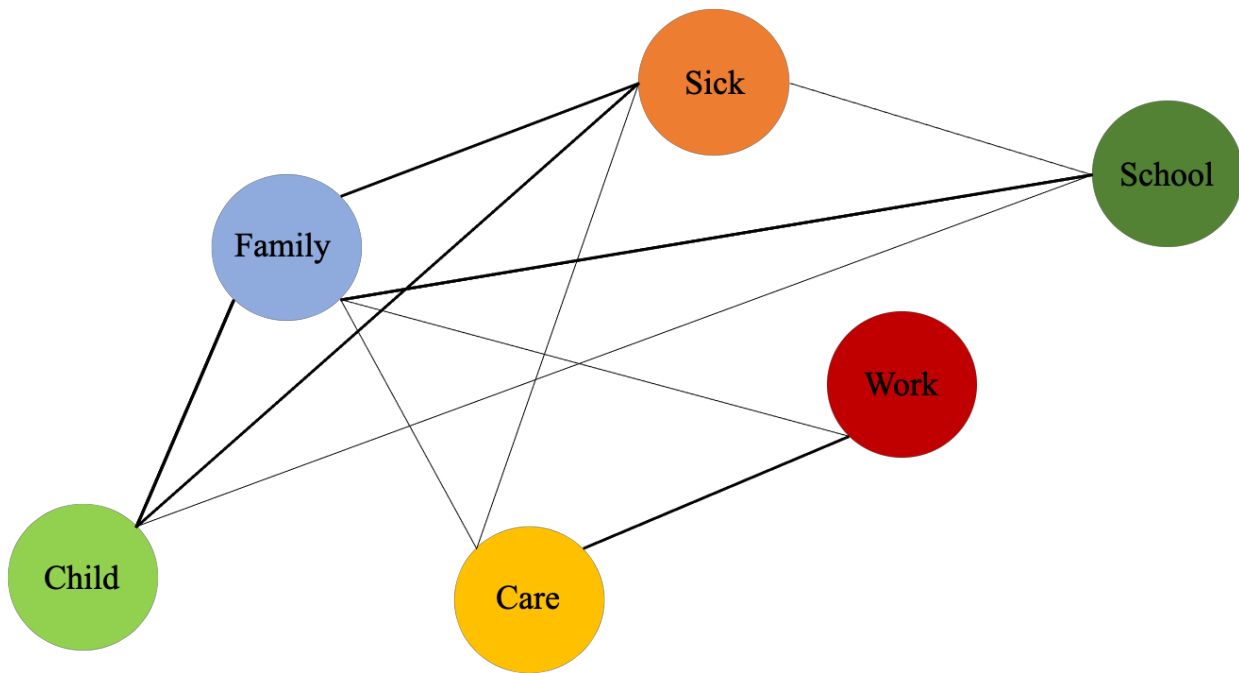


Figure 2.8. Word Co-occurrence Network

There are several studies that have utilized word co-occurrence networks like: The research by Camilo et al. [127] focuses on using word co-occurrence networks, which link words based on their occurrence within close range in a text, to identify text authorship. The study demonstrates that these networks, which do not require extensive linguistic knowledge, can be effective in capturing stylistic features of different authors, providing a robust method for authorship recognition. This approach is particularly useful as it simplifies more complex syntactic networks and utilizes dynamic fluctuations in network metrics to characterize authorship.

Mikaela et al. [128] explores the relationship between word co-occurrence and sentiment in

political tweets. The study uses word co-occurrence networks, where nodes are words and edges represent the frequency of two words appearing together in tweets. It focuses on tweets related to Hillary Clinton's 2016 presidential bid, using hashtags. The research aims to understand the collective sentiment by analyzing how words in these tweets group together and their associated sentiment scores. The study provides insights into the sentiment structure of political discourse on Twitter through network science techniques. Therefore, the first study applied this technique to explore the relationships between various words in packaging reviews, which will be explained in detail in Chapter 3.

The writing demonstrates that AI is a promising solution for packaging evaluation. The remainder of this dissertation will discuss two studies where AI is applied. Chapter 3 explores the use of NLP in evaluating packaging through customer review sentiments. Chapter 4, ML models are utilized to analyze the effectiveness of a newly designed label for OTC medications. Finally, Chapter 5 summarizes the result of these two studies.

CHAPTER 3

A novel packaging evaluation method using SA of customer reviews

3.1 Introduction

The growth of E-commerce is evident in many countries [129], and this growth is predicted to increase up to 25 percents by 2026 [130]. During the Covid-19 pandemic, people clearly preferred online shopping over traditional methods for their safety [131]. Therefore, it is important that packaging designers should focus seriously on tackling any online shopping problems.

Packaging in E-commerce should be different than it is for brick-and-mortar retail stores. In E-commerce, packaging has four major challenges to consider: supply chain, weight of the material, product integrity, and safety and customisation. To clarify, a potentially longer supply chain calls for more robust packaging [132]. Packages during the distribution process in E-commerce have more touch points compared to brick-and-mortar distribution [133]. The weight of material directly affects the cost for the freight. Also, more weight not only increases the distribution cost but also the environmental impacts. Moreover, the package should be able to ensure product integrity and safety. Consumers expect to receive a product damage free, thus the packaging should save the product from impacts, moisture, and excess air. Last but not least, a package needs to convey a sense of customization during online shopping. Customers want an opening experience to remember the product because no salesperson is there to promote the product. All in all, the design of the package should be changed for the E-commerce [134]. All of these challenges show the important role of packaging evolution in E-commerce.

Packaging evaluation is an important process to ensure the safety of packages during distribution. Packages experience different hazards [6, 7] and packaging plays an important role to protect products through distribution. To ensure protection, a series of physical tests, such as shock, vibration, and compression, are done on new packaging designs. For example, Dodds and Plummer

[135] evaluated various laboratory road simulation technologies that have been developed over the last decades. Rouillard [136] indicated that more realistic simulations of road-related package vibrations could be obtained by using statistical models. Böröcz and Singh [137] measured the vibration levels that occur during parcel shipment by small delivery vehicles over ground transportation. Nygard et al. [138] conducted a series of drop and compression tests on gable top packages, and concluded that loading history has a large impact on the compression properties. Although these tests are as close as one can get to real conditions, they are costly and time consuming processes.

In recent decades, with the advancement of computers, researchers have started to use computer models to simulate test conditions to improve them. Many researchers implemented the finite element method (FEM) to simulate different stages that a package will face from producer till it reaches the customer. For instance, the storage condition of packages has been modeled by using a FEM to analyze the loads on pallets, packages and shelf life [139, 140, 141]. Furthermore, shipping conditions have been simulated by various researchers [142, 143, 144]. For example, commercial FEM packages were implemented to create an FEM model of a wheel [143]. Then, experimental tests were used to measure the impact loads that would be applied on the wheel from the road, and those loads were used to carry out FEM analyses [144]. These analyses are important since the resultant load will be moved to the suspension system and after that to the package itself. FEM has been used by other researchers to simulate the handling condition of packages after shipping, such as for drop test simulations as a way to model the drops that might occur during handling [145, 146, 147]. Additionally, the effect of environmental conditions on the mechanical properties of paperboard packaging has been studied by using FEM models [148]. Although FEM is a powerful method, it has its own limitations to simulate actual packaging behavior because of the complexity of materials, large deformations, non-linear behaviors, and other issues.

Machine learning (ML) is an application of artificial intelligence (AI). The importance of ML has become evident in different areas of science and engineering, such as disease discovery [149] and language translation [150]. Packaging design has not been excluded from this growth in ML

usage. Quanz et al. have used multiple ML techniques, to help designers produce creative designs in packaging such as bottles [151]. Zhao et al. used a combination of ML and packaging to demonstrate packaging sentiment analysis (SA) and ML models via the Acumos AI platform [152].

This work focuses on customer reviews to evaluate packaging performance. Customer reviews are one of the available sources to express customers opinion regarding the goods and their packaging [153]. Overall, a review has different components: name of the writer, title, star rating, date of review, and text of review. These components convey invaluable information regarding their respective product. Some researchers [154, 155] have demonstrated the importance of reviews on the purchase rate of a product. Furthermore, by analyzing these reviews, packaging designers can detect any possible issues with the product's package in the beginning stages, which is of great importance.

In this study, customer reviews regarding packaging performance for different items during distribution have been evaluated using a natural language processing (NLP) model. NLP is an application and field of research that determines how computers can manipulate and understand human languages [156]. NLP contains various areas of study, such as spam filtering, speech recognition, and machine translation [157]. This work used SA or opinion mining [158, 159] that is mostly applied to the voice of the customer. By using SA, the sentiment of reviews is categorized into positive or negative. Then, Pack-List, an in-house library containing package-related words, is introduced to identify the packaging-related reviews. By reflecting the opinions of the customers, the model is able to identify packaging failure without physical distribution testing, and it can be used to improve the packaging design.

3.2 Method

A flowchart of the procedure used in this work is shown in Figure 3.1. The procedure starts by selecting a product from the e-commerce platform. The next step is extracting all customer reviews of the product in a text format by using the Scrapy.

Tokenization is then used to split an entire text, paragraph, phrase, or sentence into smaller units, which are called tokens. Then, words are converted to the root form or base form by us-

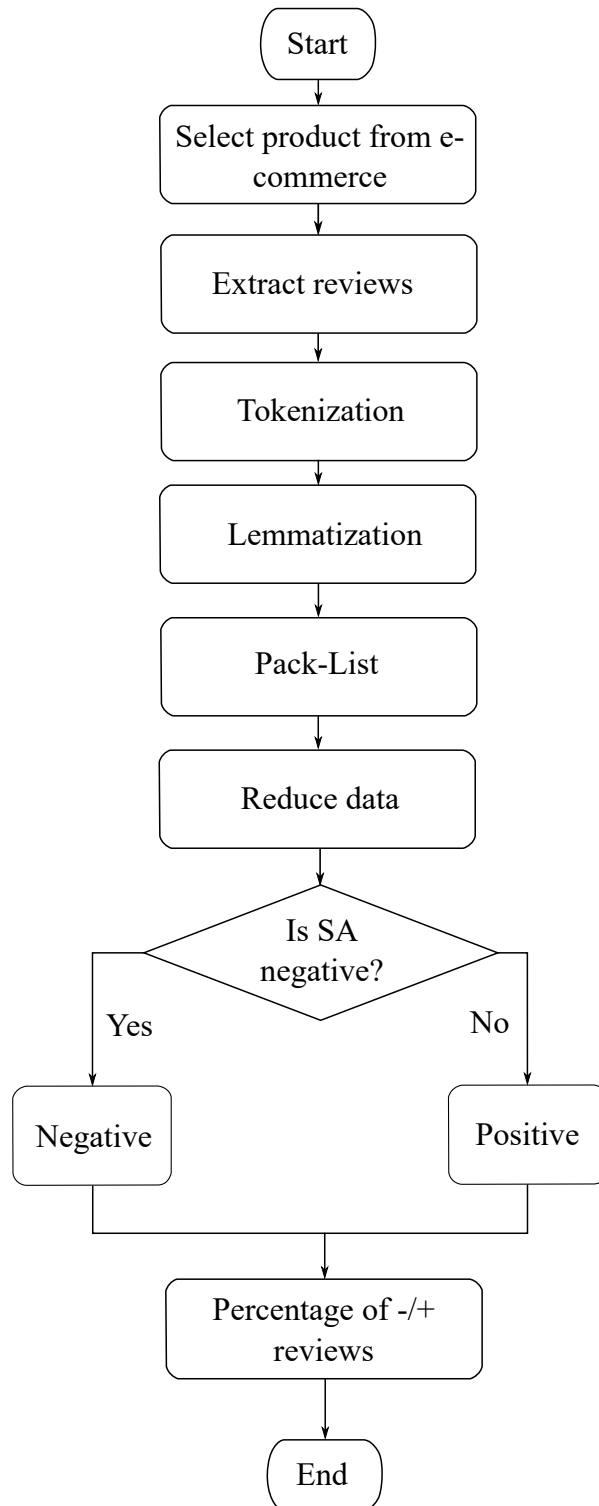


Figure 3.1. Flow Chart of the Procedure

ing lemmatization. To identify the package-related comments, a group of packaging words was created, which is called Pack-List. The number of comments was reduced to the group of packaging words by using Pack-List. Next, negative reviews were separated from positive reviews by using SA. Then, the number of negative and positive reviews was identified. The percentage of negative (failure) and positive reviews were calculated by using Equation 3.1. The percentage of failure can extract valuable data for designers. For example, by comparing the percentage of failure between various brands, one can find which packaging design works better during distribution. Furthermore, the assurance level can be checked in the real-word environment.

$$\begin{aligned} \text{Percentage of Negative or Positive} = & \hspace{15em} (3.1) \\ & \left(\frac{\text{Number of Negative or Positive reviews}}{\text{Total Number of reviews}} \right) \times 100 \end{aligned}$$

Additional details regarding Scrapy, tokenization, lemmatization, Pack-List, and SA are provided in the following sections.

3.2.1 Scrapy

Scrapy is an open source web-crawling framework for extracting data from various sources [160]. Working with Scrapy is very easy and it does not need complete coding. The first step is to define a website, like "amazon.com" which in Scrapy is allowed domains, then should enter start URL. Each product has a specific URL or address, that refers to the page(s) containing data regarding the product. Then, identify the pattern of product or HTML structure. To do this, we should go to inspect by right clicking on the review page. Then, find the pattern of reviews.

For instance:the pattern for review title in this work is

`"/div[@id = "cm_cr - review_list"]//span[@class = "a - profile - name"]/text()",`

so one should find patterns for names, star rating and review's content. Finally, Scrapy will extract all of this information from that specific website. Each product has a specific URL or address, that refers to the page(s) containing data regarding the product. In Scrapy the address of the product's

website should be entered like `https://www.amazon.com/`. Next, by using `inspect`, we can find the code regarding the product website.

3.2.2 Tokenization

Tokenization is the first step of text processing, and involves separating the text into smaller parts or subunits, which are called tokens [161]. Tokens can be words, subwords, or characters. Tokenization can classify text into three different units: words, characters, and subwords. A subword is a unit that can be equal or smaller than a word. For instance smaller can be "small" + "er". The word tokenization algorithm is most common and splits text into individual words, while character tokenization splits data into characters, and subword tokenization splits text into subwords. There are several methods for tokenization like using Regular Expressions (RegEx), natural language Toolkit (NLTK), spacy library and more. NLTK which is library for python language for statistical Natural Language Processing and symbolic is implemented in this work. There is a module in NLTK named `tokenize()` with two categories: sentence tokenize and word tokenize. At the beginning, by using sentence tokenize, split document to the sentence. Usually, sentences end with ".", therefore "." can be use for separating sentence. Then, it breaks sentence to words or terms, by splitting the string after each space. For example, after the tokenization process, the sentence "The package was damaged" is converted to: "The" "package" "was" "damaged".

3.2.3 Lemmatization

Usually, morphological analysis of words is done to reduce their inflectional forms or sometimes derivationally related forms. This process is called lemmatization. In short, lemmatization reduces the complex form of the words to their common base form, which is known as the lemma. The important role of lemmatization is obvious when reviews contain different forms of the same word instead of the base form, such as "go","went" and "gone" rather than the base form: "go". Therefore, lemmatization increases the efficiency of text data management. Two different python packages, Wordnet and Spacy lemmatizer were tried in this study. Although both packages work well, for the data provided in this paper, Wordnet performed better in identifying the number of negative sentence; as shown in Table 3.1. In this study by using wordnet lemmatizer, each word

convert to root form and then we replace the root form of words with their original.

Table 3.1. Number of Negative Sentences

	Wordnet	Spacy
Product A	22	15
Product B	18	15
Product C	0	1

3.2.4 Pack-List

A list or lexicon of package-related words, called Pack-List, was needed to identify the packaging-related feedback. To create a Pack-List, the first step was choosing the lowest rating reviews (1- and 2-star rating) for the targeted products, which can be multiple products in the same category. Next, to find the frequency of each word, these reviews were put in a word count platform. Then, these words have been ranked based on the frequency. The words related to packaging and the possible cause of package damage were selected depending on the products. The reviews related to packaging were collected by using Pack-List as the guide. Then, the sentiment of the reviews was realized by using SA.

3.2.5 Sentiment Analysis (SA)

SA is a subset of NLP which helps in comprehending the sentiment in customer reviews. SA is one of the tools usually used in social media to obtain knowledge regarding the opinion of customers about a product. SA consists of different algorithms that can be used based on the circumstances. The criteria for choosing the best algorithm lies in the accuracy of the work and the number of available data. The automatic algorithm with Naïve Bayes classifier was implemented in this work.

Naïve Bayes uses statistical and probability methods which had been discovered by Thomas Bayes (British scientist) to find the highest probability value to classify test data in its appropriate category [162]. Due to its simplicity, the Naïve Bayes classifier is a famous learning algorithm for data mining applications [163]. Furthermore, its efficiency has been proven in various applications, such as medical diagnosis, system performance management, and text classification [164, 165, 166]. The Naïve Bayes method is rooted in Bayes theorem, which uses past experiences to predict future opportunities [162]. The formula regarding Bayes Theorem can be shown as follows:

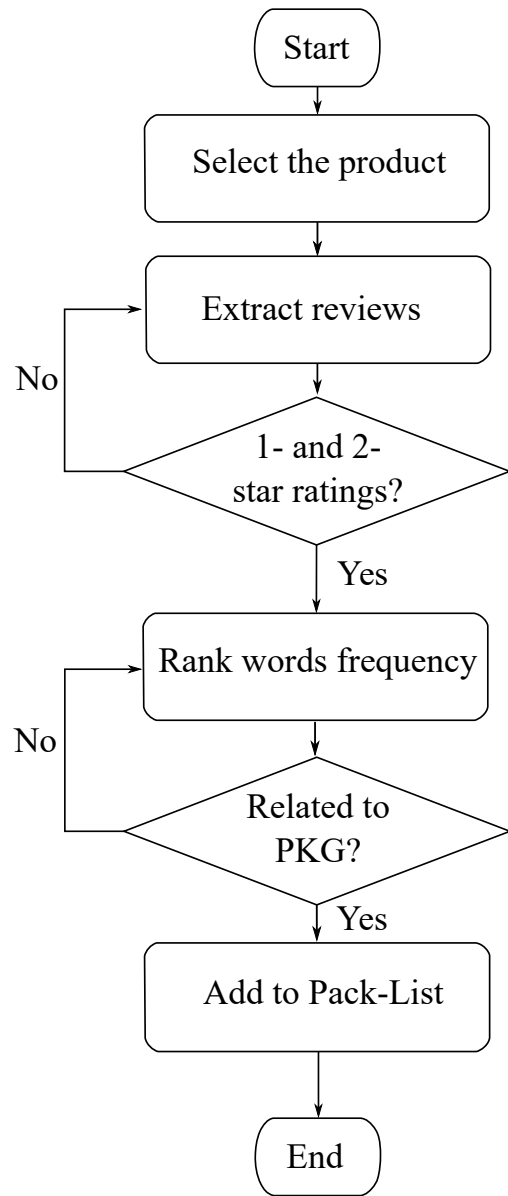


Figure 3.2. Procedure for Pack-List

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.2)$$

This means that the probability of A happening if B has occurred can be calculated based on the probability of occurrence of B and A and occurrence of B if A has occurred. Instead of using

the probability of a single feature like A , the Naïve Bayes algorithm has used a matrix of features (\mathbb{X}). Moreover, it uses a vector of response (\mathbf{y}), rather than using an output (B). There are two important hypotheses in Naïve Bayes: I. independency, and II. equality of the probability of the features. Therefore, Equation 3.3 shows that y is a class variable and \mathbf{X} is its dependent vector of features, which is a row from the features matrix (\mathbb{X}):

$$P(y|\mathbf{X}) = \frac{P(\mathbf{X}|y)P(y)}{P(\mathbf{X})} \quad \text{where } \mathbf{X} = (x_1, x_2, x_3, \dots, x_n) \quad (3.3)$$

Then, through the assumption of independency [72], the probability is calculated:

$$P(\mathbf{X}|y) = \prod_{i=1}^n P(x_i|y) \quad (3.4)$$

For instance, a classifier is created to decide whether a product review is negative or positive. The algorithm performs in the following manner: first, a certain number of tagged reviews are produced by a training data set. The data set is formed from a series of tagged reviews, each one has a positive or negative tag attached to it. Now, the question is in case of a new untagged review, which tag should the classifier attach to it? For example, consider the sentence in a review that says "The package was received broken.". To classify this sentence, the probability of both (positive or negative) should be calculated. Subsequently, the correct tag is the largest probability value. To put it mathematically, $P(\text{Negative} | \text{The package was received broken})$ means the probability that the tag of a sentence is negative, given that the sentence is "The package was received broken."

The next vital step in the algorithm is feature selection. Features are the pieces of information that the algorithm needs to perform. Therefore, it took them out of the text. The feature that was selected by the Naïve Bayes algorithm is word frequencies. Therefore, it treats every document as a set of words that it contains and ignores word order and sentence construction. The features will be the counts of each of these words. Although the process seems simple, it works great.

The final step is shown in Equation 3.5, which calculates the probability of each word in the sentence.

$$P(\text{"The package was received broken"}) = \tag{3.5}$$

$$P(\text{The})P(\text{package})P(\text{was})P(\text{received})P(\text{broken})$$

In order to calculate the probability, the data set containing the list of tagged words was used. Next, by using Equations 3.6 and 3.7, the probability of sentences in positive and negative classifier will be calculated.

$$P(\text{Positive} \mid \text{The package was received broken}) = \tag{3.6}$$

$$P(\text{The} \mid +) \times P(\text{package} \mid +) \times P(\text{was} \mid +)$$

$$\times P(\text{received} \mid +) \times P(\text{broken} \mid +)$$

$$P(\text{Negative} \mid \text{The package was received broken}) = \tag{3.7}$$

$$P(\text{The} \mid -) \times P(\text{package} \mid -) \times P(\text{was} \mid -)$$

$$\times P(\text{received} \mid -) \times P(\text{broken} \mid -)$$

To calculate the first term on the right-hand side of Equation 3.6, the algorithm only counts the frequency of the word "package" in positive reviews divided by the total number of words that are available in positive reviews. This process should be done for all the words in the sentence. Lastly, the probability of the sentence being positive or negative will be calculated. The equation with the higher value shows the sentiment of the sentence.

The training data set has positive and negative reviews. To calculate the terms in Equation

3.6, the first step is to calculate the probability of each word by using training positive reviews. Therefore, to calculate the first term on the right-hand side of Equation 3.6, the algorithm only counts the frequency of the word "the" in positive reviews divided by the total number of words that are available in positive reviews. This process should be done for all the words in the sentence. Lastly, the probability of the sentence being positive or negative will be calculated by Equations 3.6 and 3.7, respectively. The equation with the higher value shows the sentiment of the sentence. A question come into the mind, if we have words that are not in the training data set? then, the probability will be zero and multiply to other probability will be zero. To prevent this problem, we add 1 to all of the words and then add the value to the total. For instance, if the frequency of the word "the" in positive training is 15 and the total number of words in training positive review is 150 the value of $P(\text{The}|\text{positive})$ is $15+1/150+4$ (+4 is for total number of words in the specific sentence). Based on the data from table 3.2 for "package", "was", and "broken" will be $21/154$, $6/154$, and $2/154$, respectively. Therefore, the value of Equation 3.8 will be $45/154$ and should compare to the value of Equation 3.9 which is $44/104$. At the end, the value of Equation 3.9 is higher so the sentence is considered as negative.

Table 3.2. Frequency of the words in positive and negative training

	Positive	Negative
The	15+1	10+1
Package	20+1	15+1
Was	5+1	7+1
Broken	1+1	8+1

$$P(\text{word}|\text{positive}) = P(\text{The}|\text{positive}) * P(\text{Package}|\text{positive}) * \tag{3.8}$$

$$P(\text{Was}|\text{positive}) * P(\text{Broken}|\text{positive}) =$$

$$\frac{16}{154} * \frac{21}{154} * \frac{6}{154} * \frac{2}{154} = \frac{45}{154} = 0.292$$

$$P(\text{word}|\text{negative}) = P(\text{The}|\text{negative}) * P(\text{Package}|\text{negative}) * \quad (3.9)$$

$$P(\text{Was}|\text{negative}) * P(\text{Broken}|\text{negative}) = \frac{11}{104} * \frac{16}{104} * \frac{8}{104} * \frac{9}{104} = \frac{44}{104} = 0.423$$

3.3 Result

A mirror, a TV, and a comparison of three TVs from different brands have been selected in this paper to demonstrate the effectiveness of the proposed method. Mirror is a fragile product and can be damaged easily if the packaging is not suitable. Moreover, TVs are high demand products that gravely suffer from packaging problems like delivery with shattered screens. Hence, three famous TV brands with similar size and price have been chosen to compare packaging properties of the same product.

The result of this study is divided into four categories: percentage of negative and positive reviews, tracking percentage of failure over time, word clouds, and word co-occurrence.

3.3.1 Mirror

A rectangular wall mirror with 319 total number of reviews was selected. The product is 5.55 pounds and its dimensions are 11.5*11.5*0.51 inches. It has free shipping and free return features. Table 3.3 shows the final Pack-List of the mirror. First, words frequency of low star (1-and 2 star) rating of negative sentences related to mirror were collected. Then, some words relevant to packaging of mirrors were added to this list to create the final Pack-List.

Table 3.3. Pack-List for Mirror

Pack-List for Mirror		
Break	Fail	Scratch
Damage	Failure	Protect
Crack	Defect	Protective
Deliver	Defective	Shatter
Delivery	Distort	Fragile
Destroy		

Figure 3.3 shows the percentage of negative and positive reviews. It shows 11.15% of negative reviews were regarding the packaging of the mirror and only 7.0% of customers wrote positive opinions about the packaging. This data can further be used to compare the assurance level of the mirror. Assurance levels use experimental tests' results to give a sense of safety in using a design. However, laboratory conditions seldom match real-world conditions. On the other hand, SA results from the reviews can be representative of the assurance level of the package in real-world condition. Therefore, the designer can compare these two results to make sure that the product meets all the requirements.

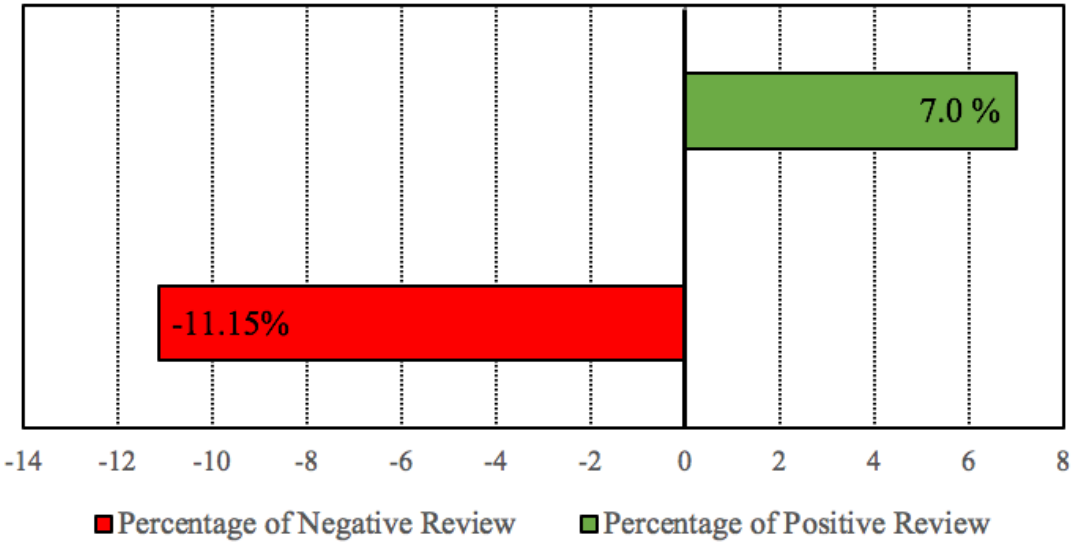


Figure 3.3. Percentage of Negative and Positive for Mirror in 2019, 2020, and mid 2021

Figure 3.4 shows the percentage of negative reviews in various months through the years in the study. This figure clearly shows an increase in failure percentage after May. Therefore, one can check for any changes that might have happened in the design, handling method or etc. after this period. Hence, the designer can find the cause and solve the issue immediately. Next, Figure 3.5 shows negative words from the review as a word cloud form. With word cloud figures, packaging designers can find that most of the negative reviews are regarding the break of a mirror, which happens during shipping.

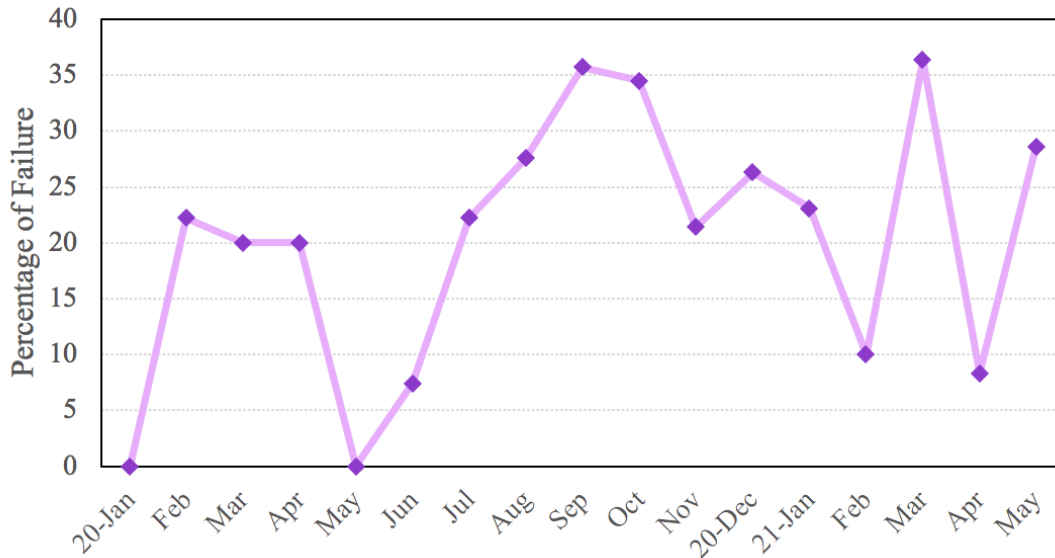


Figure 3.4. Percentage of Failure for Mirror Overtime(2020 and mid 2021)



Figure 3.5. Word Cloud for Mirror

3.3.2 Television

Another example used in this study is a TV, which is a fragile product during shipping and requires careful packaging to protect it during handling. The selected TV, which has 991 reviews, is 75

inches (191 cm) and has free shipping and returns features. Table 3.4 displays the Pack-list for this TV.

Table 3.4. Pack-List of TV

Pack-List of TV		
Break	Damage	Crack
Delivery	Deliver	Destroy
Fail	Failure	Defective
Distort		

Figure 3.6 illustrates the percentages of negative and positive packaging reviews for TV A. It shows that approximately 6% of the packaging reviews are negative, while around 3% are positive. Additionally, Figure 3.7 presents a word cloud for this example. As demonstrated, terms such as 'damage', 'box', 'screen', 'crack', 'fail', and 'break' appear frequently in the packaging reviews, indicating these are common concerns among packaging reviewers.

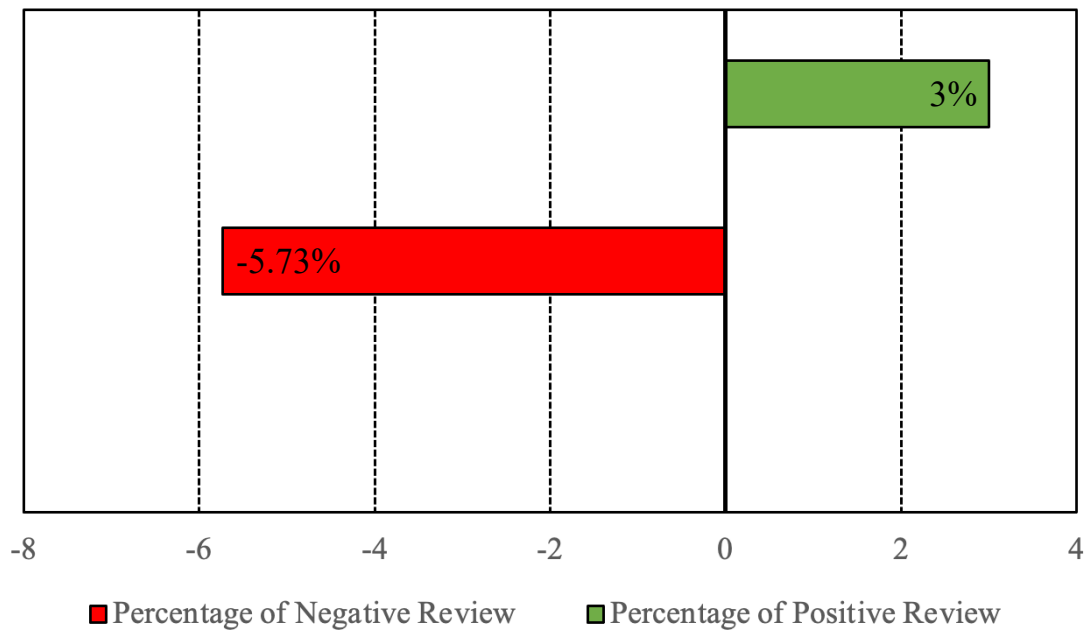


Figure 3.6. Percentage of Negative and Positive for TV A

In next example, the aim was to compare three brands within a single product category. Unlike previous examples that focused on a single product, this case aims to demonstrate the strength



Figure 3.7. Word Cloud for TV A

of the method in comparison, highlighting its effectiveness. Three televisions (TVs)(as shown in Figure 3.8) were selected to demonstrate a comparison of different packaging performances. The three TVs have similar sizes and prices, have a free shipping feature, and have different number of reviews: TV A has 991 reviews, TV B has 1270, and TV C has 820. Once the reviews were collected, Pack-List was constructed, as shown in Figure 3.9. In the packaging evaluation words, words with the highest frequency are collected from 1 and 2 star rating reviews. This step is done for each targeted product. For each new product, new words might be added to the list that are shown in bold format. For these three TVs, these words are collected and are shown in the middle section of Figure 3.9. Therefore, the final Pack-List can be seen in the right side of the Figure 3.9.

Figure 3.10 shows the percentage of failure and percentage of positive reviews with the Pack-List of three different TV brands (A, B, and C). Also, the package related reviews are identified and SA is implemented to identify negative and positive reviews. The results show that TV C has the highest percentage of failure which is 11.19 percent, meaning packaging of TV C provides the worst protection during distribution as compared to TV A and TV B. By comparing the results, TV A shows the best protection performance during the distribution. Furthermore, Figure 3.10 provides a criterion for evaluating assurance level. By comparing the assurance level of the product



Figure 3.8. Images of Three TVs

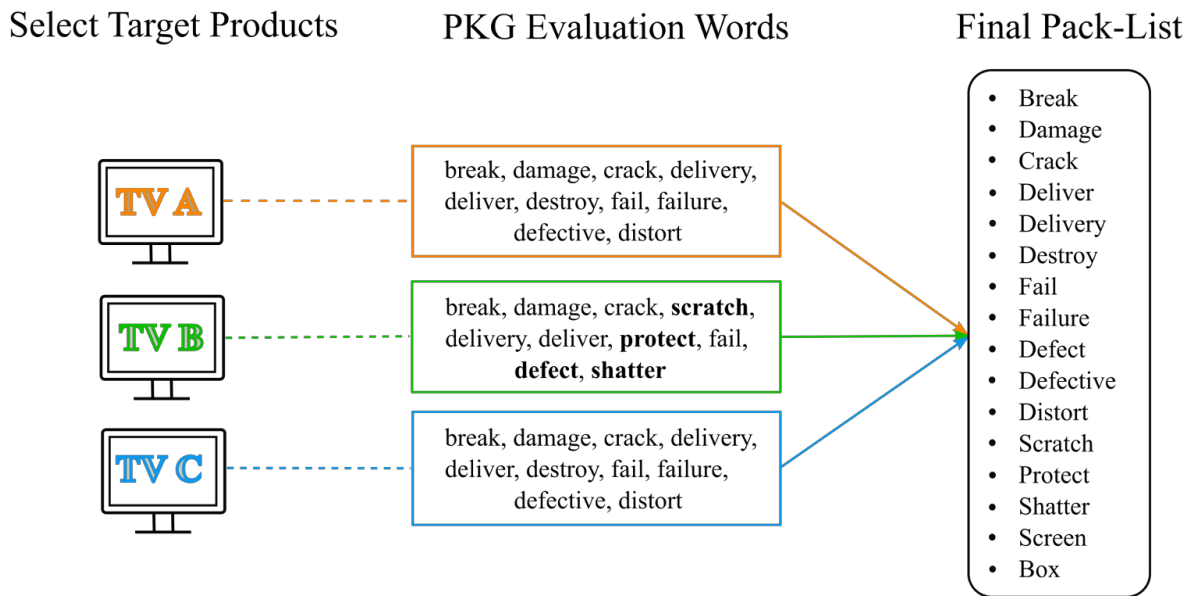


Figure 3.9. Pack-List for Three TVs

and percentage of failure, designers can examine whether the product meets all the requirements. By comparing the percentage of positive reviews with the pack-list, it is evident that TV B has more positive reviews compared to TVs A and C.

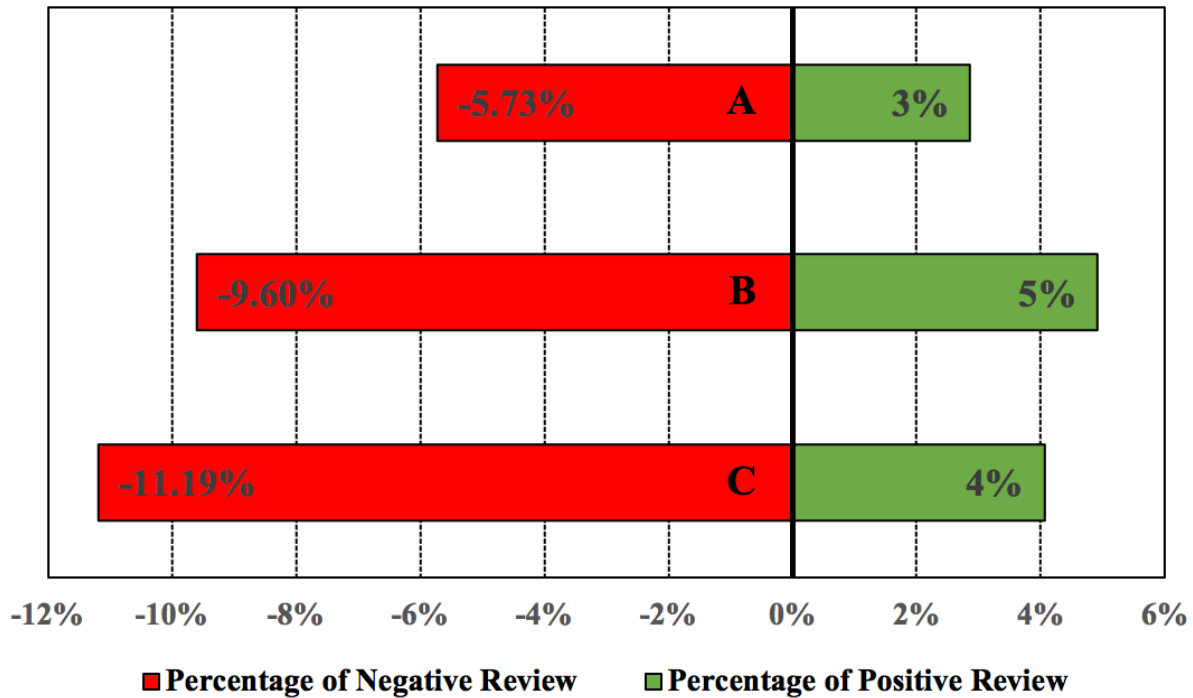


Figure 3.10. Percentage of Negative and Positive Reviews for Three TVs

Next, Figure 3.11 shows the percentage of failure of negative reviews of three TV brands during different months and years. By analyzing this chart for 2 years, the chart shows that the percentage of failure was higher in 2020 as compared to 2019. Furthermore, the highest peak of negative reviews in both years can be seen during May through October. Therefore, packaging designers can use this data to identify the cause of the increasing packaging failure at the moment.

The word cloud is the visualization method to demonstrate data. Figure 3.12 shows word clouds of three TV brands, which identifies the part of the TV's package has the problem during distribution. The word cloud of three TVs indicates that the most frequent problem is damage/crack/break/defect screen, which means the cushioning part of the package is not protecting the product as expected. Furthermore, these results display that the screen part of TV has the highest possibility of problem.

Word cloud results can also be shown in the format of a bar chart for better comparison. Figure 3.13 shows such a chart. As it can be seen from this figure, now the results for these three TVs

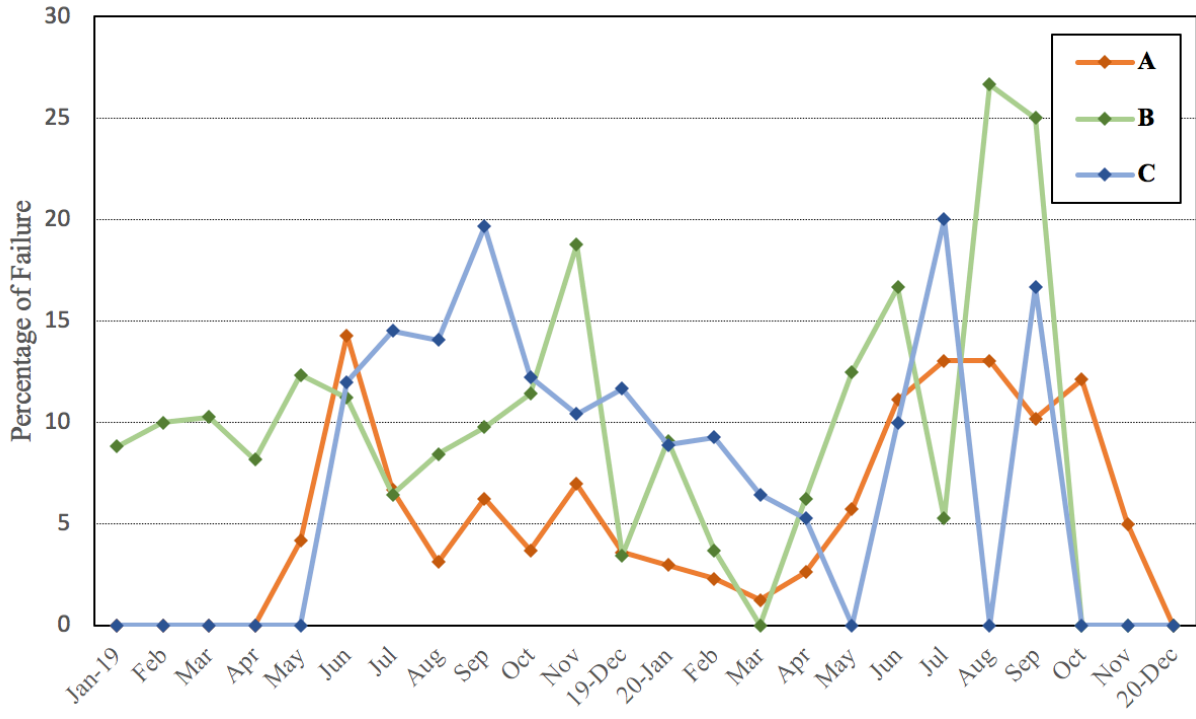


Figure 3.11. Percent of Failure of Negative Reviews for Three TVs Overtime(2019, and 2020)

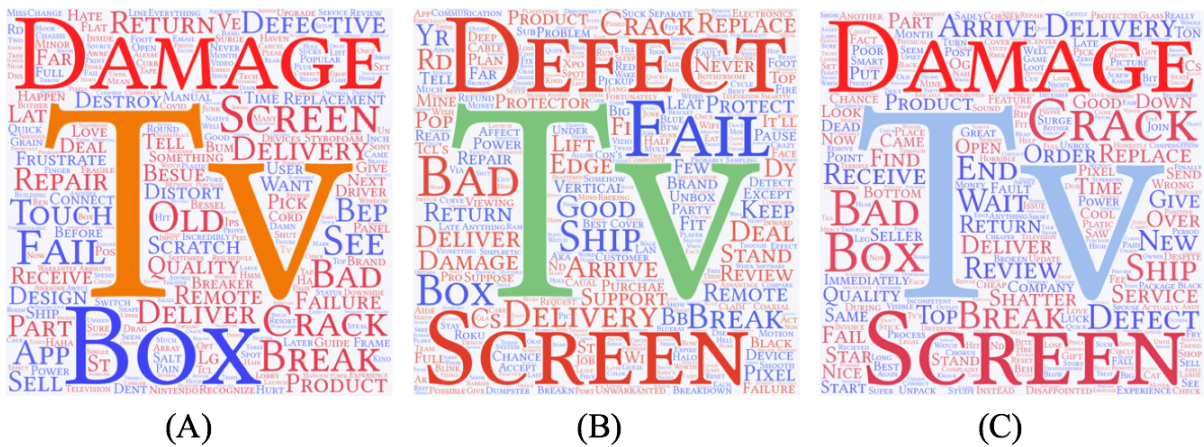


Figure 3.12. Word Cloud for Three TVs

can be easily compared and the TV with the better performance in each category can be detected. In the word cloud, it is possible to identify the words with the highest frequency in the reviews. However, it does not reveal the relationships between these words. To uncover these relationships, we should utilize word co-occurrence analysis.

The analysis of feedback on packaging was represented using a word co-occurrence network,

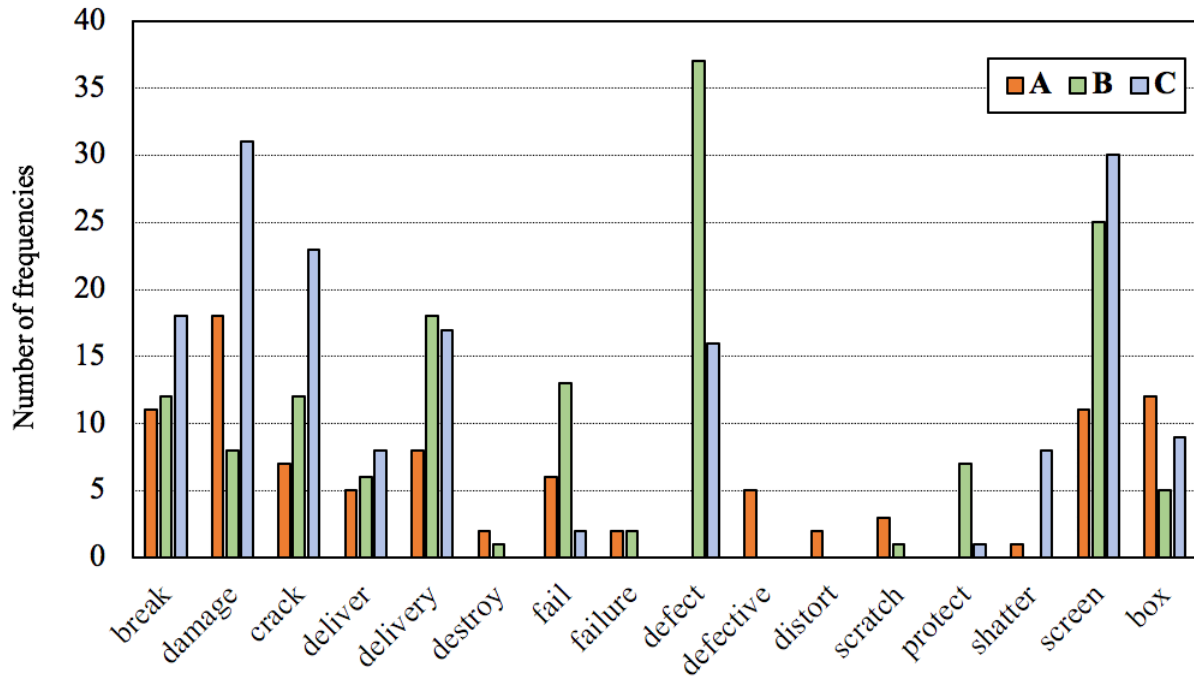
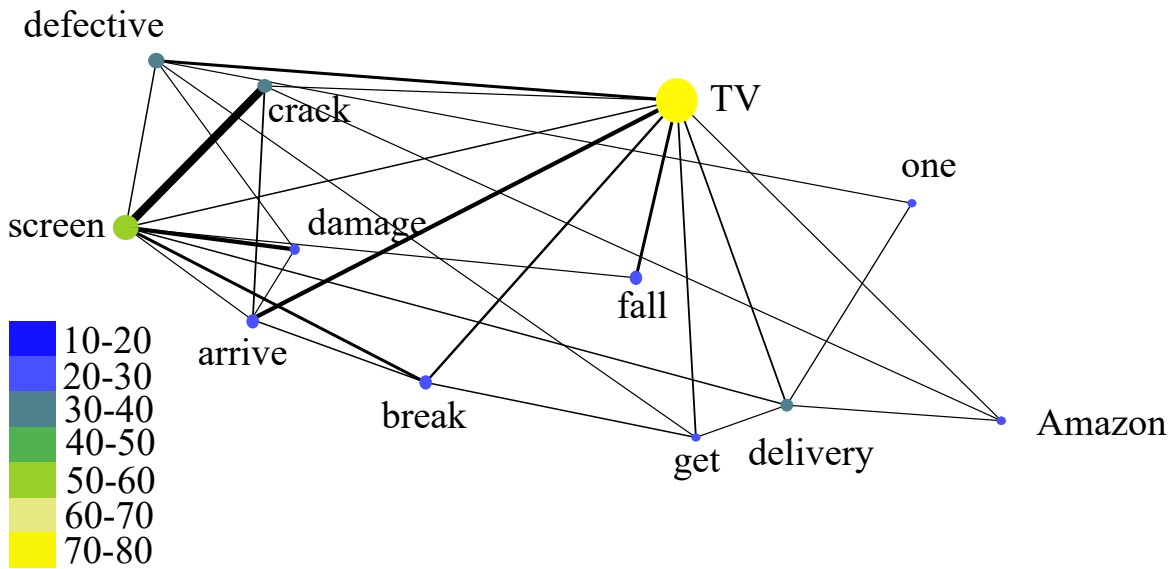
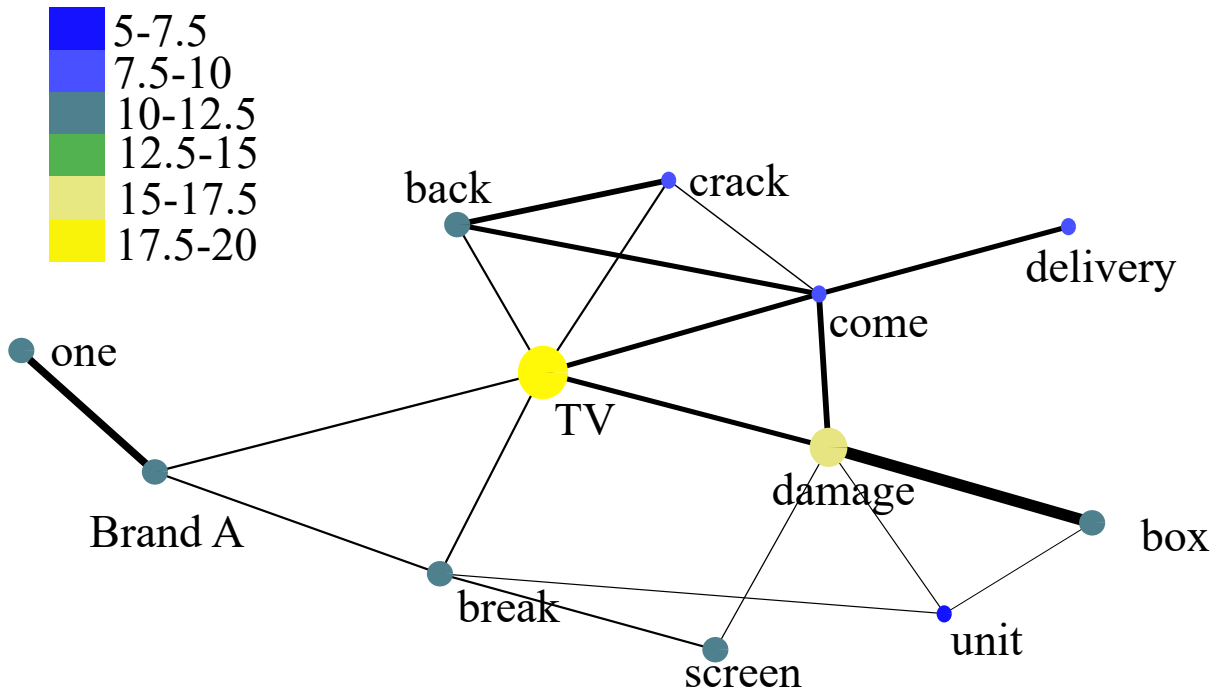


Figure 3.13. Words Frequency of Pack-List for Three TVs

as shown in Figures 3.14,3.15, and 3.16 . This network is essentially an undirected graph where each node symbolizes a unique word from the given vocabulary, and the edges indicate how often these words appear together within a document [167]. In Figures 3.14,3.15, and 3.16 nodes that correspond to words occurring more frequently are depicted as larger and colored in yellow, whereas those representing less common words are smaller and shown in blue. The thickness of the lines between nodes indicates the rate at which two keywords appear together.

For TV A, the term "damage" is significantly connected with "box," "come," "tv," and "screen" in the analysis. This implies, as per the reviews, that damage to the box during delivery is a major issue, often leading to screen damage. In the case of TV B and TV C, the term "screen" shows a strong link with words like "crack," "damage," and "break." While there are no specific remarks about box damage for TV B and TV C, this pattern suggests that screen damage is a common problem in e-commerce distribution.



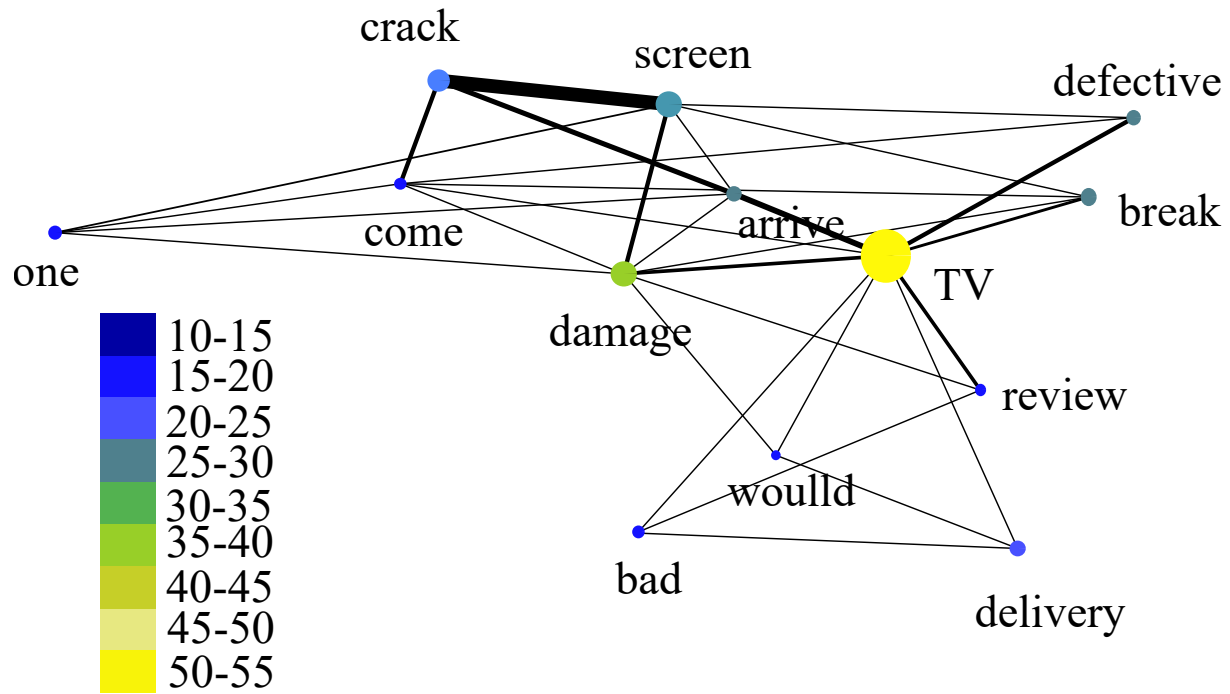


Figure 3.16. Word Co-occurrence for TV C

3.4 Conclusion

This study introduced a packaging evaluation method by analyzing customer reviews in an e-commerce platform. For the purpose of analyzing the reviews, SA, a subset of NLP and Pack-List, a packaging keywords library, were implemented in this paper. With the proposed method, the package performance can be evaluated from the customer reviews instead of physical testing. As a result, the evaluation process can be more efficient in comparison to using laboratory tests. In addition, this analysis can provide meaningful data to the packaging designers. Overall percentage of failure can be used to ensure the assurance level of the current design. The percentage of failure over time helps to identify any seasonal issues. Moreover, word cloud of negative reviews is an indicator of the most problematic packaging areas, which reduces the need for laboratory tests tremendously. Lastly, analyzing the word co-occurrence of negative reviews offers valuable insights into the relationships between specific words used by customers. By examining how frequently certain words appear together in these reviews, we can identify common issues that customers are experiencing.

In conclusion, this paper proposed a novel method for packaging evaluation using customer reviews. It should be noted that the aim of the proposed model is evaluating packaging performance and based on the available data, there is a limitation to the in-depth qualitative analysis that it can provide. However, with more data gathering it is possible to enhance the capabilities of its predictions. As a result, it contributes to enhancing the efficiency of the packaging evaluation process.

In this study, customer reviews regarding packaging performance for different items during distribution have been evaluated using a natural language processing (NLP) model. NLP is an application and field of research that determines how computers can manipulate and understand human languages [156]. NLP contains various areas of study, such as spam filtering, speech recognition, and machine translation [157]. This work used SA or opinion mining [158, 159] that is mostly applied to the voice of the customer. By using SA, the sentiment of reviews is categorized into positive or negative. Then, Pack-List, an in-house library containing package-related words, is introduced to identify the packaging-related reviews. By reflecting the opinions of the customers, the model is able to identify packaging failure without physical distribution testing, and it can be used to improve the packaging design.

CHAPTER 4

ML modeling of patients' attention of OTC medication label design

4.1 Introduction

Medications in the US in 1951, with the passage of the Durham-Humphrey Amendment to the Federal Food Drug and Cosmetic Act (FEDCA), were legislated into two categories, over-the-counter (OTC) and prescriptions [153, 168]. OTC medication, or non-prescription medications, refers to a safe and effective drug that does not need physician's oversight or prescription [33]. Moreover, for every dollar expended on OTC, the health care system saves approximately 7.20 dollars, which is an estimated total of 146 billion dollars savings annually [34]. Furthermore, OTC has other advantages like privacy, convenience, flexibility, and quick access. Consequently, OTC found a vital role in America's health system [35].

Beside all of these advantages, OTC medication has its own risks. Like negative effects or Adverse Drug Reaction (ADR) due to drug misuse. Drug misuse can be a consequence of drug-drug interactions or drug diagnosis interactions. ADR can be defined as "an appreciably harmful and unpleasant reaction resulting from an intervention related to the use of a medicinal product" [41]. A meta-analysis showed annually 106,000 US deaths occur because of ADRs[42]. One way to prevent ADR is to read the label of medication. Obviously, the label of OTC medication has an important role, and it is important for patients to read it before purchasing. Therefore, it is imperative to increase the patients' attention to the label of OTC drugs. There are many studies that have tried to increase patients' attention by changing the label's design. For example, bigger and bolder fonts, placement of vital information on the front of the package, and highlighting warning instructions [51, 52]. One of these studies, which is the basis of our paper, suggests putting a small box on Front of the Package (FOP) and highlighting important information. In that study, the researchers collected data from ninety-two participants by using a change detection method.

Participants should come to Michigan State University and do the pre-test (containing demographic information of the participant), which involves sitting in front of a computer for two hours and answering some questions. This process is very costly and time consuming for both participants and researchers. On top of that, new in-person tests are required whenever new label design is proposed, or new groups of participation are required. To overcome the disadvantages of in-person tests, this paper proposed a Machine Learning (ML) modeling of patients' attention of label design. ML is a branch of artificial intelligence (AI), and is the analysis of computational algorithms [169]. The crucial role of ML is obvious in various areas of engineering and science [150, 151, 170, 171, 172], especially in medical science like disease discovery [149], medical diagnosis [173], and medical imaging [174]. Packaging design is not excluded from this growth. For instance, Knoll et al. automate packaging planning by using different ML models. The researchers use a combination of regression and classification models to find the fill rate based on packaging characteristics [97].

This paper focuses on the effectiveness of different ML model implementation on the data from previous labeling design. The chosen models for this study are random forest (RF), decision tree (DT) and k-nearest neighbors (KNN) as they are most commonly used for classification modeling. These methods will be explained in section 4.2. As a result, the accuracy, and area under the curve (AUC) in section 4.3.1, and confusion matrix of these three models are compared with each other in section 4.3.2.

The remainder of this paper is organized as follows. Section 4.2 explains the method of different models on the data. Sections 4.3 and 4.4 present the results and conclusion, respectively.

4.2 Method

The objective of the proposed method is to evaluate and predict the accuracy of medical packaging for OTC drugs by using different ML models. The procedure of this work is demonstrated in Figure 4.1. At the beginning, the data is filtered to include only those records with the features 'critical/FOP/Highlight/IBU/PDP'. Then, 70% of this data is used for training the model, and the remaining 30% for testing. Three ML models - DT, RF, and KNN - are implemented. To compare the results of these three models, metrics such as accuracy, AUC, and the confusion matrix are

utilized.

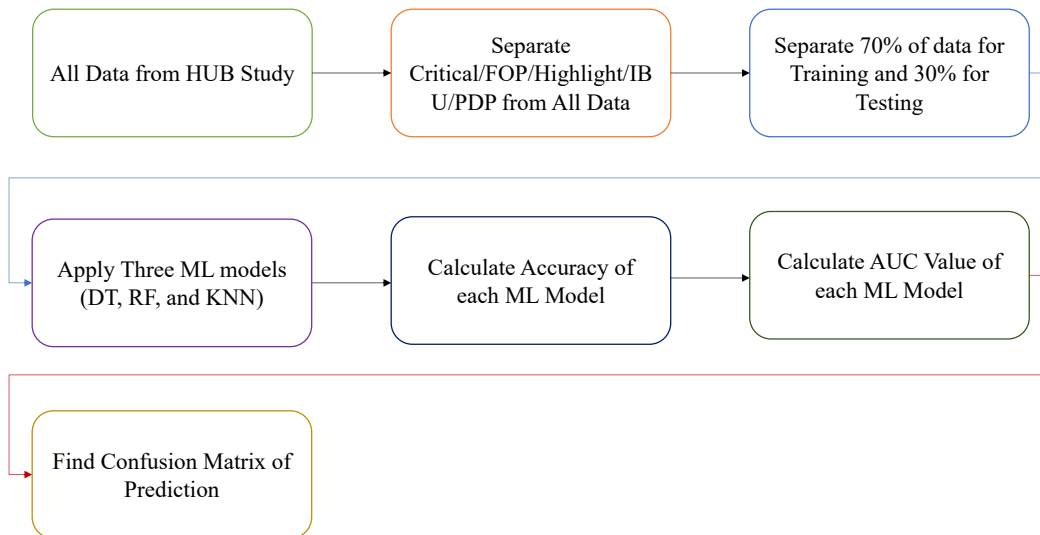


Figure 4.1. Procedure of ML Modeling

4.2.1 Review of label recognition in-person test

In this study, data collected by the Healthcare, Universal Design, Biomechanics (HUB) research group at Michigan State University (MSU) was used for training and testing the models. The data was collected through an in-person test. To be eligible to participate in this study, participants must met the following criteria:

- Be at least 18 years old
- Not be legally blind
- Have used OTC drugs during the past 6 months
- Have No history of seizure
- Are willing to come to the HUB lab at the MSU, where the research was conducted.

Each participant came to MSU and completed the consent form and pre-test. The consent form contains questions regarding the participant’s demographic information; sex, age, ethnicity, education, and language. Participants should then complete a pre-test containing three assessments regarding their visual acuity, literacy, and color differentiation ability. Each of these participants should answer 168 different trials. At the beginning, researchers explain and demonstrate the procedure. A total of 92 participants, participated in this study using the label change detection method. This study has been used E-prim 3.0 version for run and built change detection. Also, the test designed based on Rensink’s change detection timing [175]. In change detection, or the flicker task, two images (original and modified) along with a grey one are shown to participants intermittently, as demonstrated in Figure 4.2. This loop of images will be displayed for eighteen seconds. During this time, if the participants identify the difference between the original and the modified image, they can click on the difference using a mouse. Otherwise, the computer will automatically proceed to the next trial.

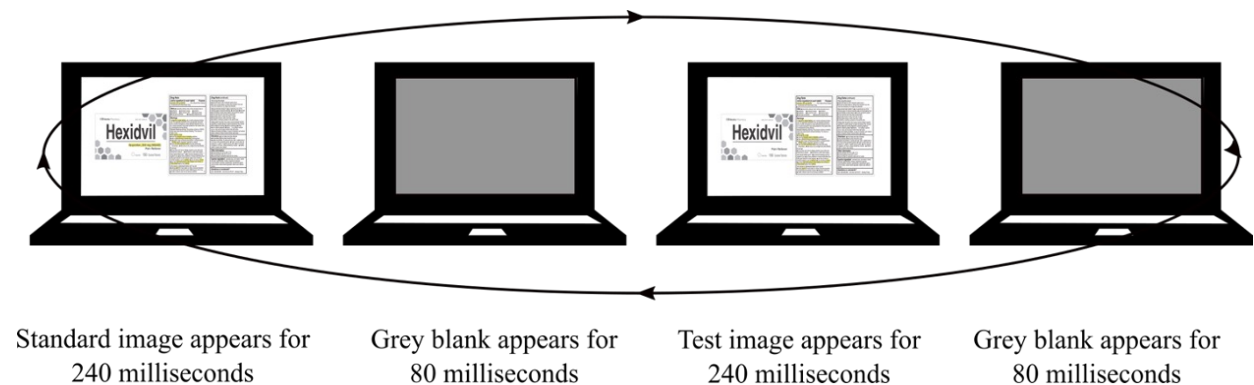


Figure 4.2. Change Detection Process

There are three mock brands used in the designing of this study. Hexidvil (pain relief/fever reducer), Circussin (antitussive), and Recantan (acid reducer), with their three different active ingredients being Ibuprofen (IBU), Dextromethorphan (DEX) and Ranitidine (RAN), respectively (Table 4.1).

Each of these brands have 56 trials in the test which, in total is 168 trials ($3 \times 56 = 168$). Each of these brands are divided into front of package (FOP), which means a small box in front of package

Table 4.1. Active Ingredient and Drug Category and Mock Brand Information for labeling test

Active Ingredient	Drug Category	Mock Brand
Ibuprofen (IBU)	Pain reliever	Hexidvil
Dextromethorphan (DEX)	Cough and cold	Circussin
Ranitidine (RAN)	Anti-diarrhea	Recantac

or standard (STD) without box. Then, depending on the change that occurs in the front or side of the package, these label designs are further divided into principal display panel (PDP) or a drug fact label (DFL), respectively. If the changes in information are related to the safety and effectiveness of the products, then they will be classified as critical. Also, each label can be highlighted or none highlighted. Final classification depends on change in content and are as follows: AI has any change in active ingredient, DD1 means drug-drug interaction and DD2 means drug- diagnosis. Each brand was created in four treatments: FOP present/ highlight, FOP present / non highlight, STD (FOP absent)/ highlight, and STD (FOP absent)/non highlight (Figure 4.3).

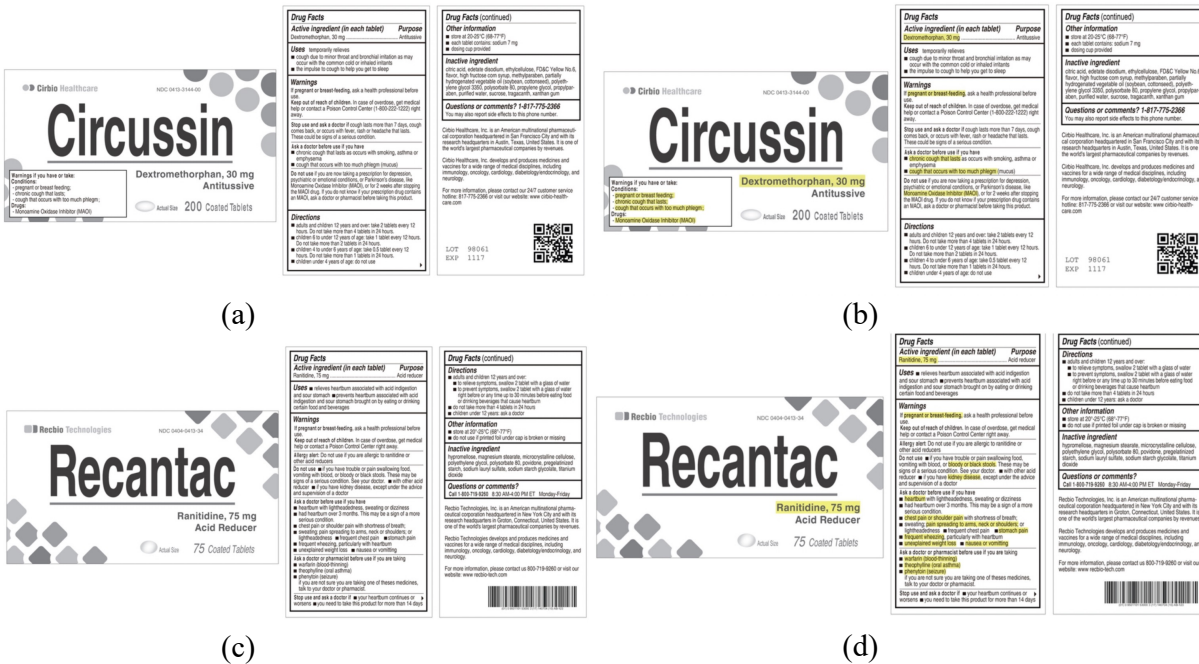


Figure 4.3. Image (a) FOP/non-highlight, (b) FOP/highlight, (c) STD/non-highlight, and (d) STD/highlight [176]

This study focuses on the first treatment, FOP, and presents highlights with these features;

critical, IBU, and PDP. This work only considers change on one of trial features which is change content because it focuses on participants demographic information with response trial as a target. Therefore, the machine predicts the response trial without collecting new data from in person participants. Various ML models such as DT, RF, and KNN were utilized for training, as explained in sections 4.2.2, 4.2.3, and 4.2.4, respectively. To apply these models to the dataset, Python version 3.10 was employed. Python is a versatile programming language, highly favored in data science for its readability and efficiency. Among the numerous libraries available for implementing ML algorithms, 'scikit-learn' and 'pandas' were chosen for this study due to their comprehensive features and user-friendly nature. Scikit-learn is widely acknowledged in the ML community for its extensive set of tools for data mining and analysis. Pandas, renowned for its data manipulation capabilities, was utilized for preprocessing the data, including the important step of splitting it into training and testing sets. The combination of these libraries ensured a streamlined and effective workflow for this multi-class classification problem.

4.2.2 Decision Tree

DT is a supervised ML method for regression and classification. It is a classification algorithm that makes a model in the shape of a tree [169]. The aim of this model is training the machine based on data features to predict the value of the target [177]. The tree contains different parts as root node, branches and leaf node. As you can see in Figure 4.4 root node is the top-level that represents a big decision or objective.

Branches are used to show different options and mostly represented by an arrow. Finally, the leaf node is the result of the decision [178]. DT is dividing a node to different sub nodes and at the end to leaf nodes. To find the root node, the first step is to measure the impurity of each feature. There are different methods for the measurement like entropy and Gini index/ Gini impurity [179]. Entropy is the first step of the decision tree to find randomness in data. The formula of entropy is as follow:

$$Entropy = -\sum_{(i=1)}^n p_i \times \log(p_i) \quad (4.1)$$

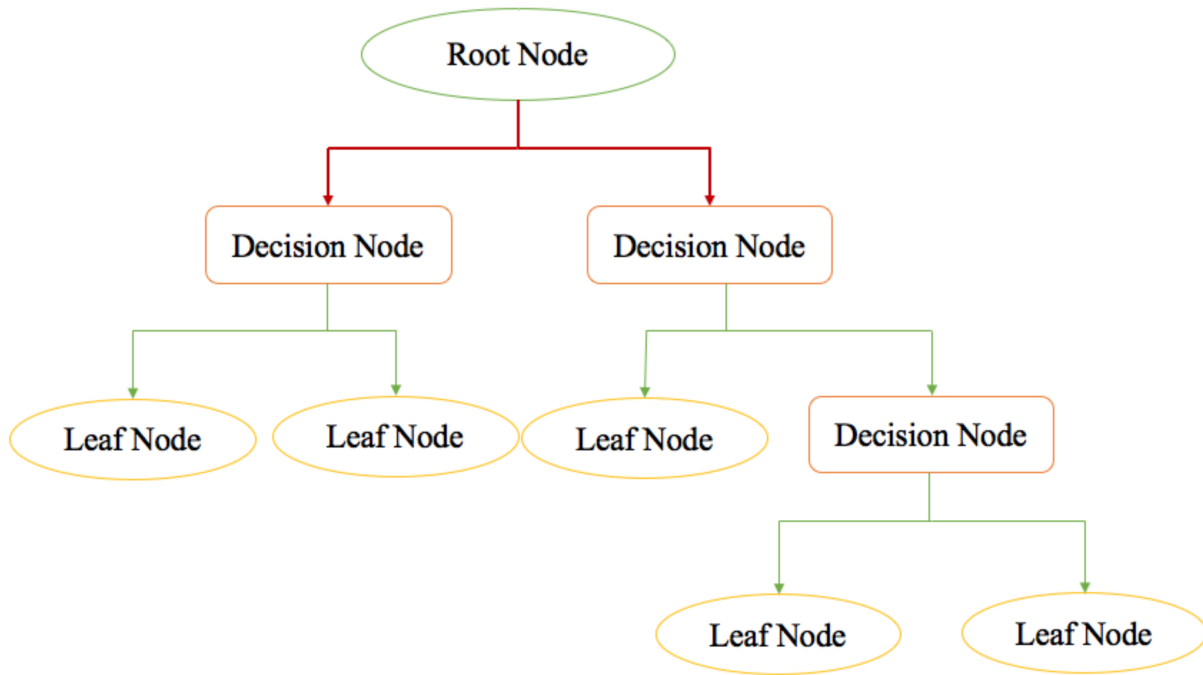


Figure 4.4. Demonstration of DT

Which i is the total number of samples, and p is the probability of i . At the beginning the entropy for the entire data set should be calculated. Then, the information gain should be calculated based on Equation 4.2 [180].

$$InformationGain(T, X) = Entropy(T) - Entropy(T, X) \quad (4.2)$$

T is the target variable and X is the feature and $Entropy(T, X)$ is entropy after the data point is split on feature X . At the end, the feature with highest information gain is the root node [180]. The advantage of DT is that it is suitable for large data sets and works for continuous and categorical data. There are some disadvantages with DT such as, usually it has an overfitting problem, and although it works well in training data, it has a low accuracy on new data leadings [181].

The DT created from this dataset is both detailed and extensive (Figure 4.5). Initially, a dataset

is imported using pandas and divided into features (inputs) and a target (the predicted outcome). For compatibility in a multi-class context, the target labels are binarized, and the dataset is split into training and testing subsets. The resulting DT consists of 79 nodes, indicating the points of decision, and 40 leaves, denoting the possible outcomes or categories the model predicts. The decision tree begins with a root node or primary decision point, where the feature "Change Content" is evaluated against a threshold of 1.5. This criterion divides the dataset into two distinct paths, depending on whether the "Change Content" value is less than or equal to 1.5. The node's impurity is quantified by the Gini index, a metric where 0 indicates a perfectly homogeneous node. Here, a Gini index of 0.18 suggests a moderate level of impurity. The node processes 193 samples, as indicated before the split. The value array, noted as [13, 6, 174], shows the distribution across the classes: 13 samples in class 0, 6 samples in class 1, and a majority of 174 in class 2. Consequently, class 2 is the majority label at this node. The decision tree then continues to branch into two nodes, to the left and right, with similar processes occurring at each node until the end of the tree is reached. After training the model with the training data, its efficacy is evaluated on the test set through pivotal metrics such as accuracy, confusion matrix, and ROC AUC score. These metrics are crucial as they illuminate the model's accuracy and its proficiency in distinguishing between various classes. The results of this prediction will be explained in the section 4.3.

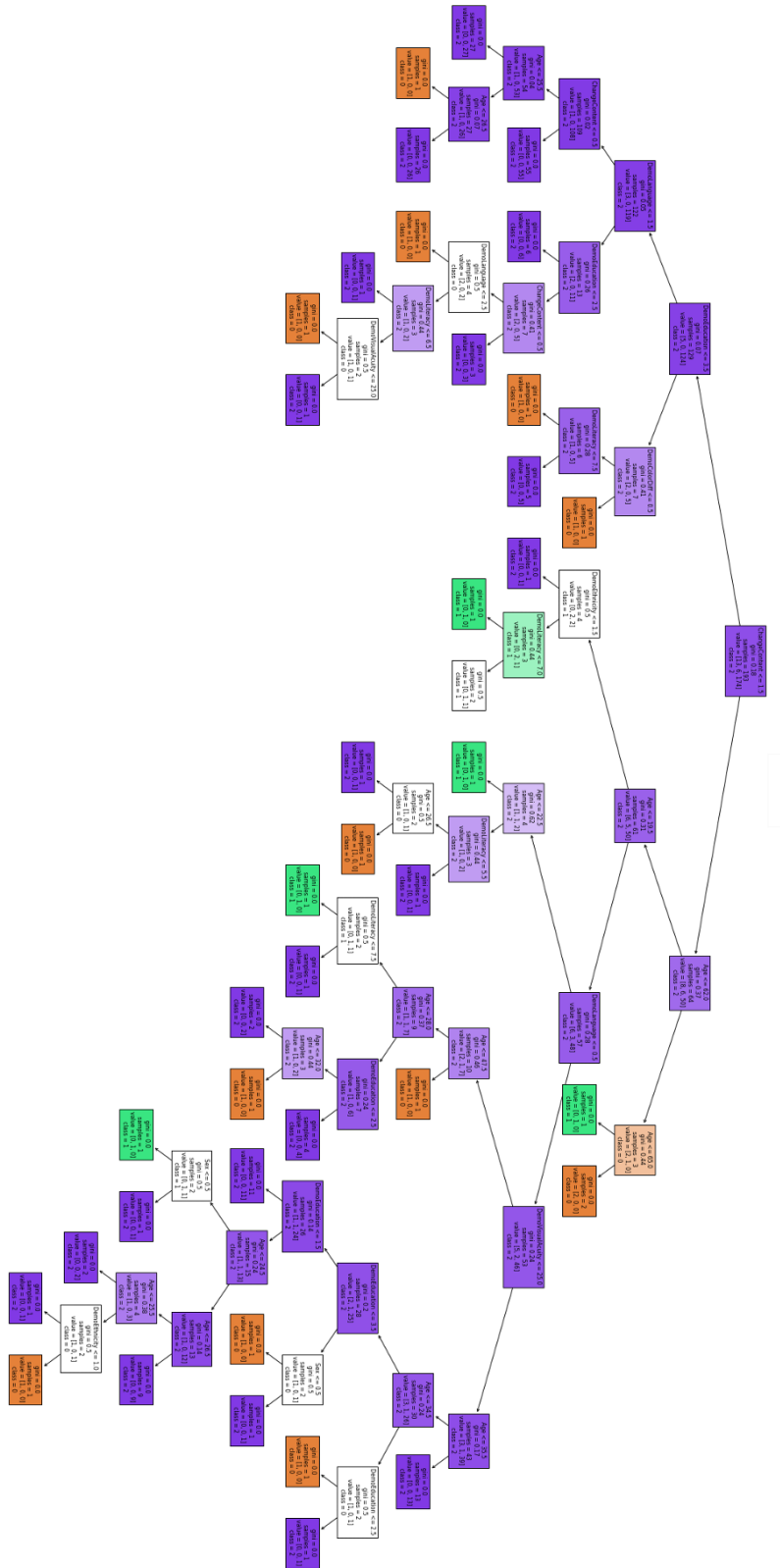


Figure 4.5. Demonstration of the DT Model in the Dataset

4.2.3 Random Forest

RF algorithm is one of the ML algorithms which can be used for different tasks, like classification and regression. RF is a combination of multiple different DTs. One of the differences between RF and DT is, RF is usually more accurate compared to DT. RF is used in different industries such as banking, healthcare, and marketing. As shown in Figure 4.6, it gets prediction results from each tree and the majority of the prediction is the final prediction of RF [182]. RF uses Bagging or Bootstrap Aggregation as ensemble technique [183]. The procedure is: choosing random samples from original data with replacement, then each model (DT) created based on samples (generate DT based on entropy or Gini impurity). At the end, each model has its own result and final output comes from the majority voting of samples [180]. The advantage of RF is having low risk of overfitting and its ability to handle large data sets. The downsides of using RF are that it is very slow in training the data and sometimes working with biased categorical variables [181].

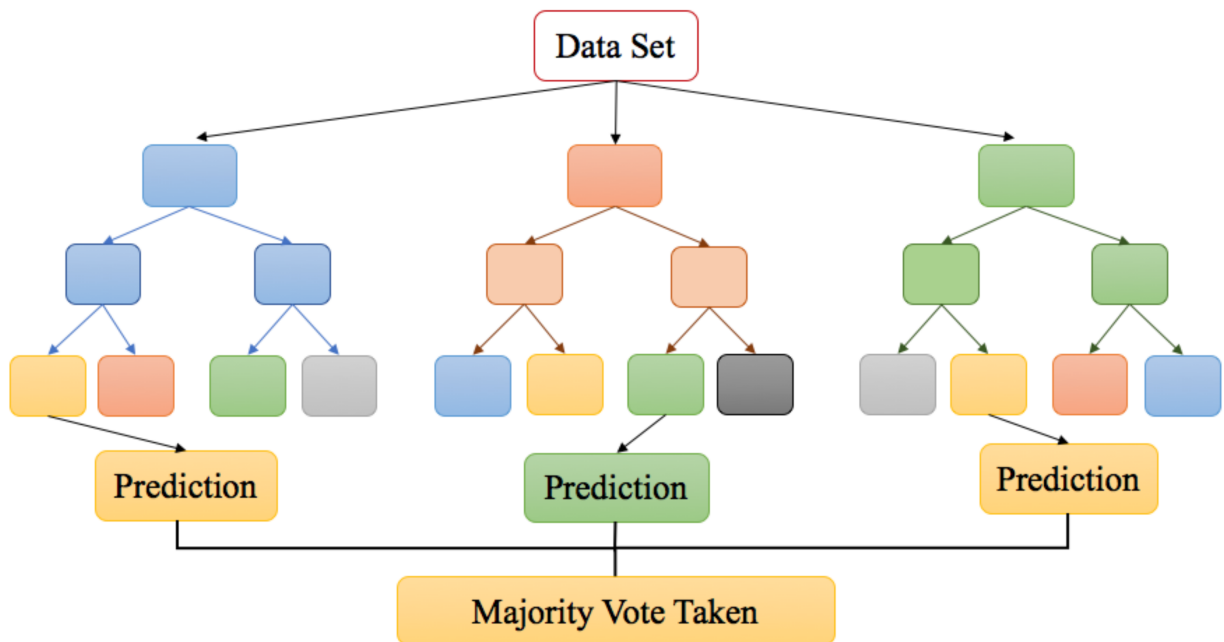


Figure 4.6. Demonstration of RF Model

After applying the RF model to the dataset, the ensemble comprises 200 individual trees (One of these is demonstrated in Figure 4.7), each characterized by a varying number of nodes and leaves (refer to Table 4.2). This diversity among the trees contributes significantly to the robust-

ness of the RF model. Each tree independently makes predictions, and their collective decisions lead to a more comprehensive and reliable final prediction. The workflow commences with the importation of the dataset using pandas, followed by the division into features (inputs) and a target variable (outcome to predict). The target labels are binarized to suit a multi-class context. Subsequently, the dataset undergoes a 70-30 partitioning into training and testing sets. Post-training, the model's performance on the test set is meticulously evaluated using pivotal metrics such as accuracy, confusion matrix, and AUC value, offering profound insights into the model's predictive capabilities and its proficiency in distinguishing between various classes. The detailed results and performance analysis of this model will be presented in section 4.3.

Table 4.2. Random Forest Trees with Number of Nodes and Leaves (Entire table in the Table A1)

Trees	Number of Nodes	Number of Leaves
1	59	30
2	59	30
3	55	28
:	:	:
198	67	34
199	57	29
200	81	41

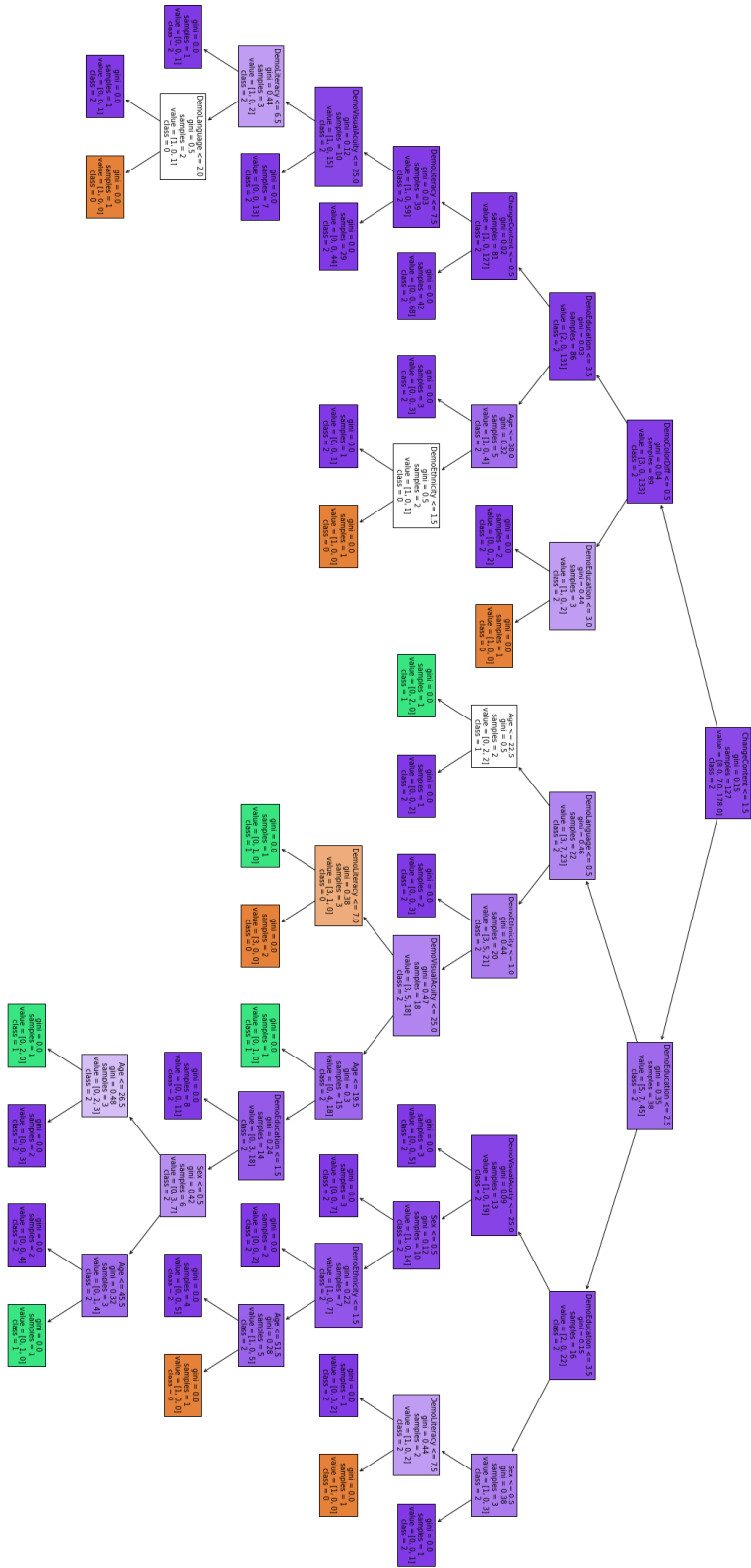


Figure 4.7. Demonstration of One of the Trees in the RF Model within the Dataset

4.2.4 K-Nearest Neighbours

KNN is a supervised ML algorithm used for both regression and classification prediction issues. But it is more used for classification problems. KNN algorithm works by estimating the likelihood that a data point will become a member of one group, or another based on what group the data points is closest to. In essence KNN works on the assumption that similar data points are close to each other [184]. The first step is to find the K value, which does not have any role, mostly chosen as 5. Second, the similarity among all data points and new data should be calculated by using the Euclidean distance (Equation 4.3 [185]). Equation 4.3 measures the distance between two points of p and q. Then, by using this equation the nearest K amount is selected in training data [185]. At the end, the nearest point to the new data is the prediction. If we have two classes of data represented as stars and triangles, and the k nearest points to a given point are illustrated in Figure 4.8, then for k=4, the result is determined by a majority vote. In this case, as shown in Figure 4.8, the majority class among the nearest points is 'stars', hence the prediction for the query point would be 'stars'.

$$\begin{aligned}d(p, q) &= \sqrt{((q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2)} \\ &= \sqrt{(\sum_{i=1}^n (q_i - p_i)^2)}\end{aligned}\tag{4.3}$$

As you can see it is very easy to work with KNN, but this algorithm is slow and does not work very well on imbalanced data [186].

The data, preprocessed with pandas, is split into training and testing sets. A key component of this implementation is *GridSearchCV*, which determines the optimal 'k' (number of neighbors) by conducting a thorough search over a specified parameter range (in this case, '*n_neighbors*' from 1 to 30) and employing cross-validation for robust model validation. The KNN model, once configured with the optimal 'k' determined to be 5, is then trained and evaluated on the test data. Performance metrics such as accuracy, confusion matrix, and AUC score are used to assess the

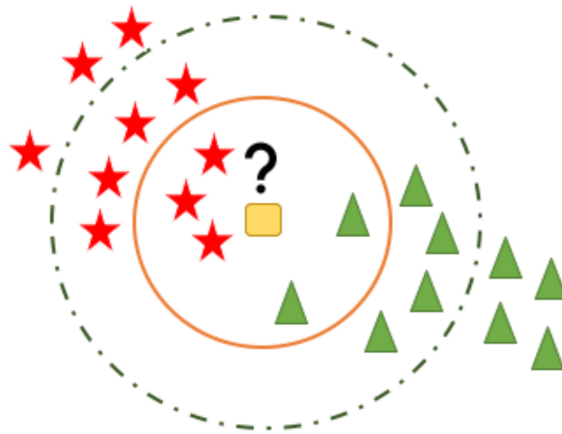


Figure 4.8. Demonstration of KNN

model's predictive accuracy and its ability to distinguish between classes. The results of this analysis will be presented in section 4.3.

In this study, patients' attention has been classified and predicted using the three methods of DT, RF, and KNN. The comparison between these methods is done by using the accuracy, AUC, and the confusion matrix of each of them. The AUC is a measure of the classifier's capability to differentiate between classes [187]. The confusion matrix indicates a summary of results of prediction with their values for each model. The results of this study are presented in the next section.

4.3 Result

The labeling design study has 14 different features. Eight of features refer to the participants' demographic information in terms of, age, color differentiation, education, ethnicity, language, literacy, visual acuity, and sex. Remaining 6 are dependent on the trial features. The way that the model is trained is as follows, 5 features (critical, FOP, highlight, IBU, PDP) are considered as constant and the remaining one (change of content) is able to alternate. Note that change of content by itself consists of AI, DD1, and DD2. So, the number of data points with these features are $92 \times 3 = 276$. The model is trained over 70% of these data points (193 trials) and tested over

the remaining 30% (83 trials). Furthermore, the target of this study is a response trial, which has three answers, hit (correct response), miss (wrong response), and time-out (the time of trial finished without an answer). The results are categorized into three sections: accuracy and AUC values in section 4.3.1, Confusion matrix in section 4.3.2, and addressing the problem of imbalanced data in section 4.3.3.

4.3.1 Accuracy and AUC Value

The accuracy of each model is calculated based on Equation 4.4. Additionally, the area under the receiver operating characteristic (ROC) curve (AUC) of each model is determined by plotting the true positive rate (TPR) versus false positive rate (FPR) at various threshold settings. TPR and FPR are calculated based on Equation 4.5 and 4.6, respectively. In these equations, TP represents true positives, FN represents false negatives, FP represents false positives, and TN represents true negatives [188]. The AUC is a common metric used to evaluate the performance of classification models, particularly in binary classification tasks [189]. The ROC curve illustrates the trade-off between sensitivity (true positive rate) and specificity (1 - false positive rate) for different threshold values. A perfect classifier would have an ROC curve that hugs the top left corner, indicating high sensitivity and low false positive rate across all threshold settings. A high AUC value (closer to 1) indicates good model performance, while a low AUC value (closer to 0.5) suggests poor performance [77].

$$Accuracy = \frac{\textit{Number of Correct Predictions}}{\textit{Total Number of Predictions}} \quad (4.4)$$

$$TPR = \frac{TP}{TP+FN} \quad (4.5)$$

$$FPR = \frac{FP}{FP+TN} \quad (4.6)$$

The first model used is the DT. The tree of this data is very large, as it has 79 nodes and 40 leaves. The classification accuracy of this model is 85% with AUC of 0.5. The second model is RF, which is a combination of different DTs. The number of trees in this model are 200 and each of these trees has a different number of nodes and leaves. The classification accuracy of RF is 88% with AUC of 0.54. The last model that is discussed here is KNN with the number of neighbors of 5. The classification accuracy for this model is 90% with AUC of 0.5. Figure 4.9 shows the accuracy and AUC comparison of three models.

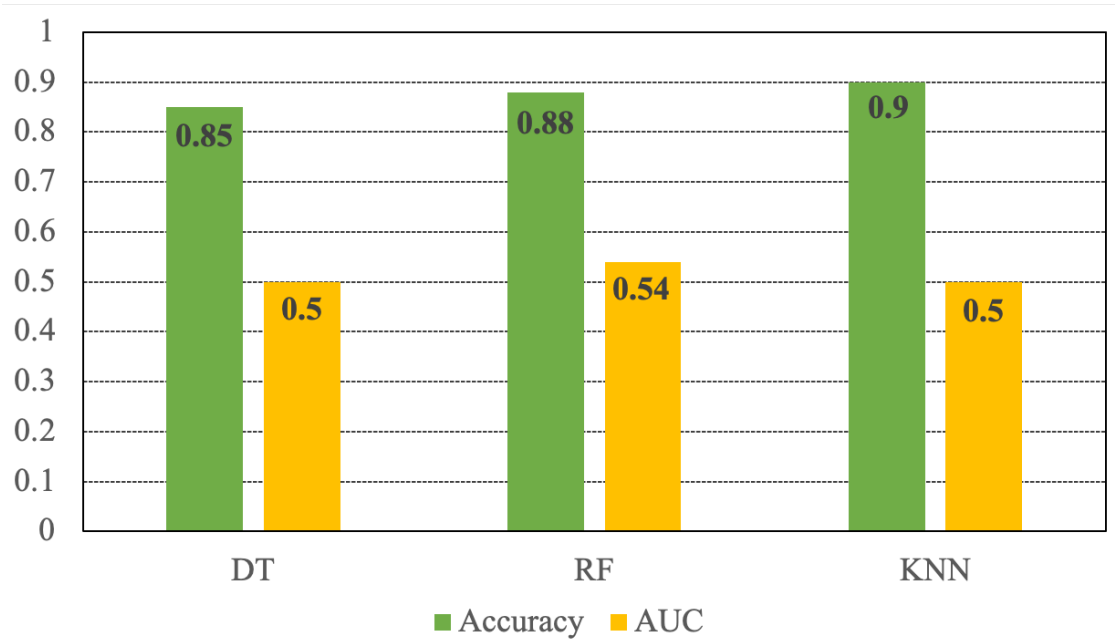


Figure 4.9. Accuracy and AUC of Three Models

It should be mentioned here that although the accuracy of the three models is good, the AUC value is not satisfactory. The lower AUC indicates that there is a problem in the dataset, so it is better to examine the confusion matrix to delve deeper into it.

4.3.2 Confusion Matrix

To better understand the low AUC results, the confusion matrix has been utilized. Figure 4.9 indicates the classification accuracy and AUC value, but it does not indicate the results of the prediction. As mentioned in the previous section, the confusion matrix is used to summarize the prediction results [190]. Figure 4.10 shows the comparison of the confusion matrix between these three models.

As seen in Figure 4.10, all three models accurately predict 'hit' but struggle with 'time out' and 'miss.' Out of 75 'hit' cases, RF and KNN correctly predicted all, while DT predicted 72. For 'time out,' only KNN successfully predicted one instance, and for 'miss,' DT correctly predicted only one out of the total three.

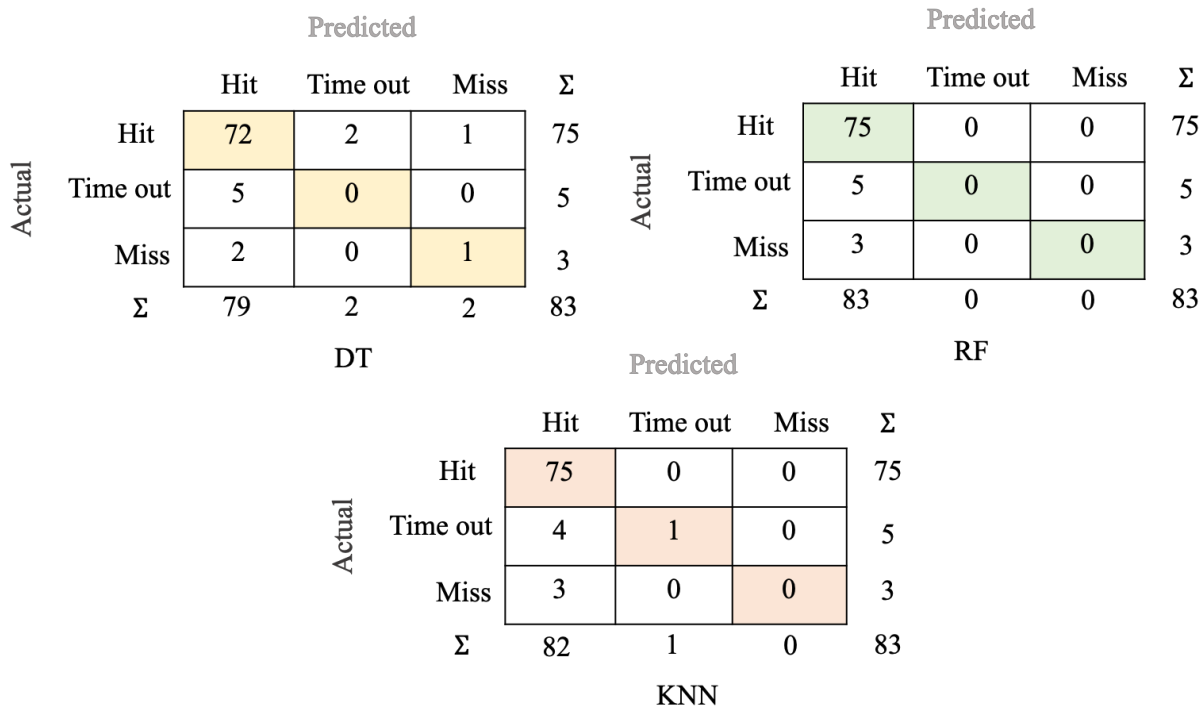


Figure 4.10. Confusion Matrix of Three Models

As can be seen from the confusion matrices of these models, all three models demonstrate high accuracy in predicting hits for the dataset. However, they fail to correctly predict time-outs or misses. This discrepancy arises because accuracy, as defined by Equation 4.4, reflects the pro-

portion of correct predictions out of the total number of predictions made. The predominance of hit data within the overall dataset significantly influences this measure of predictive accuracy, as observed in this case. To illustrate in detail, accurate predictions of hits substantially boost overall accuracy, while predictions for the other two categories—time-outs and misses—remain unreliable. This issue is indicative of an imbalanced dataset, where the number of hits (175 out of 195 trials) far exceeds the instances of the other two outcomes (20 trials), thus heavily influencing the outcome. Consequently, an additional criterion, such as the AUC value, is necessary to provide a more balanced comparison of the models. It is worth noting that a low AUC value further underscores the imbalance issue. The following subsection 4.3.3 addresses the problem of imbalanced data and how to resolve this issue.

4.3.3 Addressing the Imbalanced Data Problem

There are several methods to address imbalanced data, such as resampling, which includes upsampling the minority classes and downsampling the majority class, as well as the synthetic minority over-sampling technique (SMOTE).

Resampling techniques for imbalanced datasets involve adjusting the class distribution either by increasing the instances in the minority class, known as oversampling, or by decreasing the instances in the majority class, referred to as downsampling. Oversampling aims to equalize the minority class size to that of the majority, while downsampling reduces the majority class size to align with the minority [191].

SMOTE is a technique in data analysis that addresses the imbalance in datasets by creating synthetic examples in the minority class. SMOTE works by taking samples from the minority class and generating new, synthetic samples that are similar yet slightly different, thereby increasing the size of the minority class in the dataset. This approach helps in balancing the dataset, improving the performance of ML models on imbalanced datasets [192, 193].

The SMOTE offers a significant advantage over traditional resampling methods for handling imbalanced datasets. Unlike simple duplication of minority class instances in resampling, SMOTE synthesizes new examples from the minority class. This approach is effective as it creates syn-

thetic examples that are close in feature space to existing minority class examples. As a result, it augments the representation of the minority class without simply duplicating existing examples, contributing to a more effective learning of the decision boundary for the model [192, 193]. Therefore, in this study, the SMOTE technique was used to address the problem of imbalanced data. It should be noted that 'hit' is the majority class in this study, while 'miss' and 'time out' are the minority classes.

4.3.3.1 Effect of SMOTE on Outcome

After applying SMOTE, the number of data points in the minority classes increases. SMOTE raises the count of these minority data points to match that of the majority class. In this specific case, the number of data points for both the 'miss' and 'time out' classes has increased to 249, aligning with the 'hit' class, which originally had the most data points.

The process of implementing SMOTE in this code involves several steps. Initially, 30% of the 'hit'(75 trials) data along with all the original data for 'miss' and 'time out' (18+9=27 trials) are selected. Then, SMOTE is applied to the entire dataset. Subsequently, the same 30% portion of 'hit' data and the original 'miss' and 'time out' data are removed to form the testing set. The model is subsequently trained with the remaining data, which comprises both the original and the SMOTE-augmented data. With this methodology, the RF model achieved an accuracy of 78%. Furthermore, it recorded an AUC value of 0.81. Detailed performance metrics of accuracy and AUC are provided in Figure 4.11. DT and RF exhibit similar accuracy values, approximately 81% and 78% respectively, while KNN demonstrates a lower accuracy compared to these two models, at 61%. It's worth mentioning that the AUC values of all three models have increased, with DT, RF, and KNN showing AUC values of 0.79, 0.81, and 0.74 respectively.

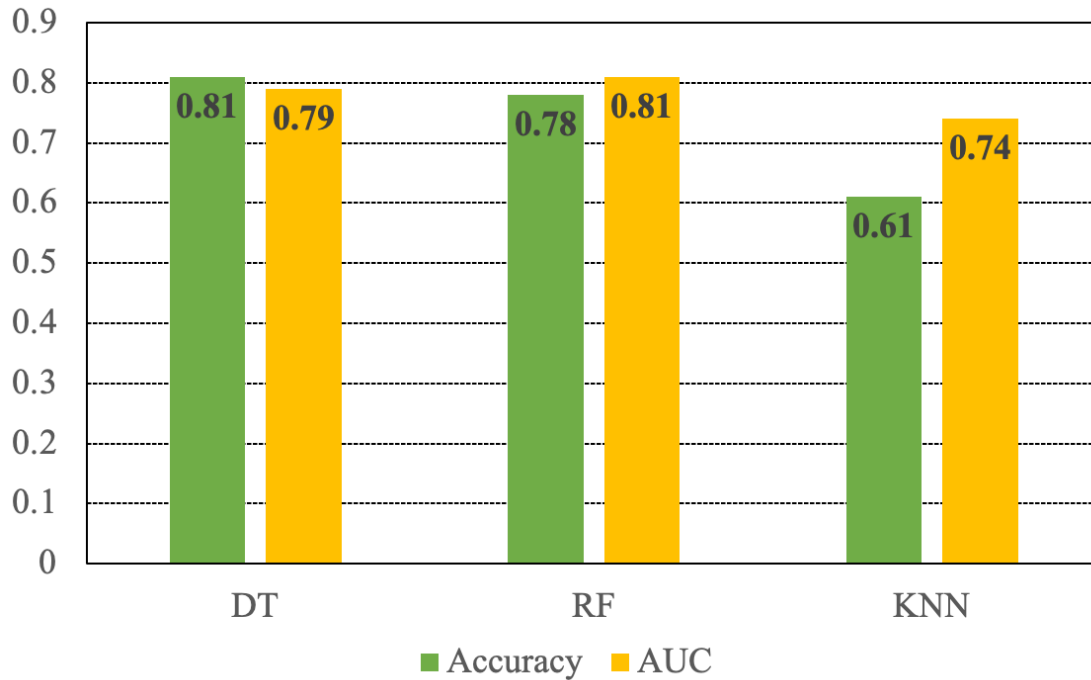


Figure 4.11. Accuracy and AUC of Three Models after SMOTE

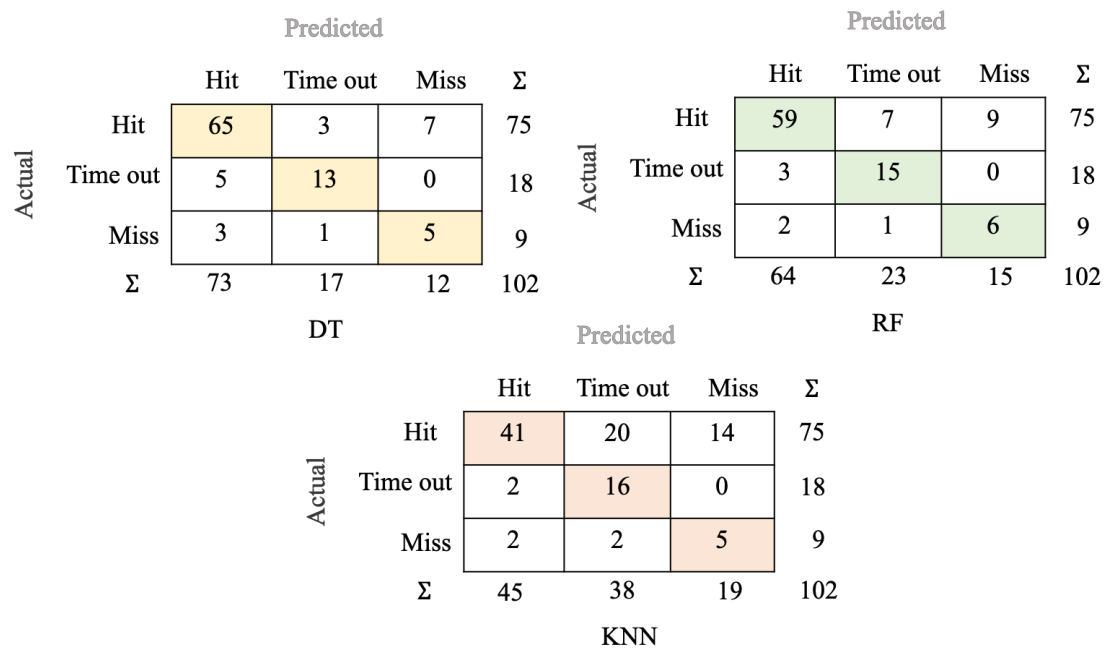


Figure 4.12. Confusion Matrix of Three Models After SMOTE

Although there is a slight decrease in accuracy, the AUC values and confusion matrix in Figure 4.12 indicate that all three models perform better for the three target classes after applying SMOTE.

In the DT model, out of 75 data points classified as 'hit', 65 were predicted correctly, 3 were predicted as 'time out', and 7 were predicted as 'miss'. For the 18 data points classified as 'time out', 13 were predicted correctly, 5 were incorrectly predicted as 'hit', and none were predicted as 'miss'. In the last row of the confusion matrix for DT, out of 9 data points classified as 'miss', 5 were predicted correctly, 3 were incorrectly predicted as 'hit', and one was incorrectly predicted as 'time out'. As observed in the confusion matrices for DT, RF, and KNN, there is a notable decrease in the number of 'hit' predictions compared to the initial predictions. However, despite this reduction in 'hit' predictions, the overall performance of the models has improved across all three classes: 'hit', 'time out', and 'miss'. This improvement can be attributed to the models achieving better balance and accuracy in their predictions for all classes, leading to a more robust and reliable classification outcome.

This result demonstrates that SMOTE effectively resolves the data imbalance issue present in the dataset, allowing all three models to predict the results of all three classes. This marked improvement in performance underscores the efficacy of SMOTE in effectively addressing the data imbalance issue prevalent in the dataset. Prior to the implementation of SMOTE, the models faced challenges in predicting beyond a single class (hit). However, post-SMOTE implementation, they exhibit the capability to predict the results for all three classes, indicative of the successful resolution of the previously encountered limitation.

4.4 Conclusion

This paper introduces three distinct Machine Learning (ML) algorithm models—Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN)—applied to Over-The-Counter (OTC) labeling design data. The implementation of ML in this study is aimed at predicting patterns in newly introduced features without the necessity of collecting additional data. While the models demonstrated satisfactory classification accuracies of 85%, 88%, and 90% for DT, RF, and KNN respectively, their lower AUC values suggest limitations in accurately predicting when one target category ('hit') is predominant over the others ('time-out' and 'miss'). This is a common challenge in scenarios where data is imbalanced, with one class significantly outnumbering the others.

To mitigate this imbalance, various strategies are explored, including resampling, which includes upsampling the minority classes and downsampling the majority class. One effective technique for dealing with such imbalances is synthetic minority over-sampling technique (SMOTE), which this study utilized. By applying SMOTE, the imbalance in the dataset was effectively addressed, enabling the model to be trained on a more balanced mix of original and SMOTE-generated data. This revised training methodology significantly enhanced the model's performance. The improvements indicate a substantially increased capability of the model to effectively classify all three target classes: 'hit,' 'miss,' and 'time out'. The enhancement of the confusion matrix in 'time out' and 'miss' underscores the effectiveness of SMOTE in managing imbalanced datasets, as clearly evidenced by the results.

CHAPTER 5

Discussion

5.1 Summary

The aim of these two studies was to apply AI and ML technologies in the field of packaging. Evaluating packaging is essential as it prevents potential damage and ensures the package meets key functions such as protection, containment, apportionment, unitization, communication, and convenience. With the rise of global e-commerce, distribution risks are increasing, driven by factors such as manual and mechanical handling, transport vehicle impacts and vibrations, and environmental hazards. To assess packaging, various methods including field tests, laboratory evaluations, and numerical solutions have been utilized. However, due to certain limitations in these traditional methods, an alternative approach using AI has been proposed.

In recent years, the use of AI has spread across many areas of science, including engineering fields, medicine, and more, demonstrating its adaptability and transformative impact. The realm of packaging is no exception to this technological growth and innovation. AI's integration into packaging offers novel solutions for design, efficiency, and sustainability challenges. This study implements AI and ML in two areas of packaging: evaluating customer feedback and modeling patient attention towards OTC medications.

5.1.1 A Novel Packaging Evaluation Method Using Sentiment Analysis of Customer Reviews

This research develops a new approach to assess packaging by examining customer feedback on an e-commerce platform. Utilizing Sentiment Analysis (SA), a component of Natural Language Processing (NLP), and Pack-List, a library of terms related to packaging, the study analyzes these reviews. This innovative method allows for the assessment of packaging effectiveness through customer insights, presenting an efficient alternative to conventional physical testing. This technique

not only streamlines the evaluation process compared to lab tests but also yields valuable information for packaging designers. Analyzing the overall failure rate informs the reliability of the current packaging design, and tracking failure rates over time pinpoints potential seasonal problems. Additionally, a word cloud generated from negative feedback highlights critical areas in packaging, significantly reducing the dependence on physical testing. Furthermore, studying word co-occurrence in negative reviews uncovers patterns and common issues highlighted by customers.

In conclusion, this paper introduces an innovative approach for assessing packaging effectiveness by leveraging customer reviews. This method signifies a shift from traditional evaluation techniques, focusing on the analysis of real-world user feedback for a more practical understanding of packaging performance. While the model effectively evaluates packaging quality, it's important to acknowledge that its depth in qualitative analysis is somewhat restricted by the scope of currently available data. This limitation highlights the potential for even more detailed and refined insights with the collection of additional data. Expanding the dataset would not only refine the model's predictive accuracy but also enable a more comprehensive analysis of customer sentiments and preferences. Ultimately, this approach stands to greatly improve the efficiency and effectiveness of the packaging evaluation process, providing valuable insights that can inform future packaging design and enhancements.

5.1.2 Machine Learning Modeling of Patients' Attention of Over-the-Counter Medication Label Design

This research presents the application of three ML models—Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN)—to analyze Over-The-Counter (OTC) labeling design data. The goal of utilizing ML in this context is to identify trends in new features without the need for additional data collection. The models achieved classification accuracies of 85%, 88%, and 90% for DT, RF, and KNN, respectively. However, their lower Area Under the Curve (AUC) values highlight a challenge in predicting accurately when one class ('hit') dominates the others ('time-out' and 'miss'), a typical issue in datasets with an imbalance of classes.

To address this imbalance, the study explores several methods, such as increasing the rep-

resentation of minority classes (upsampling), reducing the representation of the majority class (downsampling), and synthetic minority over-sampling technique (SMOTE). The study specifically employs the SMOTE to counter this issue. The application of SMOTE effectively corrects the imbalance, allowing for the training of the model on a dataset that includes both original and synthetically generated data. This approach markedly improves the model's performance, elevating its accuracy to 78%, 81% for RF and DT and the AUC to 0.81 and 0.79, respectively. The improved results in the 'time out' and 'miss' categories within the confusion matrix highlight SMOTE's success in handling imbalanced datasets, as the outcomes clearly demonstrate.

5.2 Challenges

While this dissertation presents valuable insights, it is important to acknowledge that, as with any research, it comes with its own limitations.

In the evaluation of customer feedback, the analysis is limited to reviews written only in English. This presents a limitation as there are some reviews that are written in languages other than English, which are not included in the analysis. Consequently, important insights from non-English feedback could be missed, potentially affecting the comprehensiveness of the evaluation. Including multiple languages in future analyses could provide a more holistic understanding of customer sentiments and enhance the accuracy of the findings.

This research utilized online data sources to evaluate customer feedback, as these were the only available resources for assessment in this study. However, it should be noted that there are other forms of feedback which were not accessible for inclusion in this analysis, such as feedback received via telephone, email, or in-person interactions. The incorporation of these additional feedback channels could potentially provide a more comprehensive view of customer sentiments and experiences. Including a broader range of feedback sources in future studies might offer more detailed insights and a more thorough understanding of customer perspectives, thereby enriching the data analysis and enhancing the overall findings of the research.

Additionally, in the ML modeling of patient behavior, the dataset size was limited. This study utilized data collected from Lanqing Liu's research [40] in the Healthcare, Universal Design,

Biomechanics lab (HUB) at Michigan State University, where this research was conducted. The relatively small dataset size presents a challenge, particularly in achieving a balanced dataset, as the limited number of data points contributes to the issue of data imbalance. More comprehensive data would be beneficial, as it could provide a broader understanding of patient behavior patterns and lead to more accurate and reliable ML model predictions. Expanding the dataset could also help in addressing the imbalance issue by offering a more representative sample of the various behavioral patterns, thereby enhancing the overall effectiveness of the study.

5.3 Suggestion for the Future Study

For future studies, an important enhancement would be to include multiple languages in the analysis, not just English. Expanding the scope to encompass different languages would allow for a richer, more diverse collection of customer feedback, providing a broader perspective on consumer sentiments across various cultural and linguistic backgrounds. This multilingual approach could uncover unique insights that are specific to different regions and demographics.

Additionally, implementing a spell-checking algorithm for all words in customer reviews before analysis could significantly improve the accuracy of the findings. Customer reviews often feature informal language and may contain various spelling errors or casual expressions. Correcting these errors beforehand would ensure a more precise and reliable text analysis, leading to better quality data.

Another key area for future research is the removal of stop words in text analysis. Stop words, which are commonly used words that carry minimal meaningful information, can clutter and dilute the significance of the analysis. Eliminating these words would sharpen the focus on more relevant terms, enhancing the depth and clarity of the analytical results.

An additional aspect that could be explored is the integration of image processing with text analysis. Often, customer reviews include images that can provide additional context or highlight specific aspects of product packaging. By analyzing these images alongside the textual content, researchers could gain a more comprehensive understanding of customer opinions and sentiments. This dual approach of text and image analysis could lead to a more detailed and comprehensive

evaluation of product packaging, offering a holistic view of consumer feedback and preferences.

In the second project, a range of ML models were considered for classification purposes. Among these, three widely recognized and popular models were selected and implemented. However, it is worth exploring the potential of other ML models in this context. Different models may offer varied strengths and could potentially lead to enhancements in key performance metrics. Exploring alternative models could yield improvements in accuracy, enhance the AUC values, and produce more informative confusion matrices. These alternative models might pick up different details in the data or fit better with the unique features of the dataset being used. Therefore, experimenting with a broader range of models could provide valuable insights and possibly optimize the overall performance of the classification task in this project.

BIBLIOGRAPHY

- [1] Kit L Yam. *The Wiley encyclopedia of packaging technology*. John Wiley & Sons, 2010.
- [2] Frank A Paine and Heather Y Paine. *A handbook of food packaging*. Springer Science & Business Media, 2012.
- [3] Gordon L Robertson. *Food packaging: principles and practice*. CRC press, 2005.
- [4] US Food and Drug Administration. Understanding Over-the-Counter Medicines, howpublished = <https://www.fda.gov/drugs/buying-using-medicine-safely/understanding-over-counter-medicines>, 2018. Online; accessed 16 May 2018.
- [5] Walter Soroka. Fundamentals of packaging technology, institute of packaging professionals, st. Charles, IL, 2002.
- [6] Julien Lepine, Vincent Rouillard, and Michael Sek. Review paper on road vehicle vibration simulation for packaging testing purposes. *Packaging Technology and Science: An International Journal*, 28(8):672–682, 2015.
- [7] Dennis E Young. Testing and evaluation of transport packaging: a view to the future. *Packaging Technology and Science: An International Journal*, 13(1):3–6, 2000.
- [8] Henrik Pålsson. *Packaging Logistics: Understanding and managing the economic and environmental impacts of packaging in supply chains*. Kogan Page Publishers, 2018.
- [9] Cartier Packaging Optimized. PRIMARY, SECONDARY AND TERTIARY PACKAGING: WHAT'S THE DIFFERENCE?, howpublished = <https://www.emballagecartier.com/en/article/primary-secondary-and-tertiary-packaging-whats-the-difference/>, 2019. Online; accessed 9 Aug 2019.
- [10] SG Lee and SW Lye. Design for manual packaging. *International Journal of Physical Distribution & Logistics Management*, 33(2):163–189, 2003.
- [11] Jorge Masis, Laszlo Horvath, and Péter Böröcz. The effect of forklift type, pallet design, entry speed, and top load on the horizontal shock impacts exerted during the interactions between pallet and forklift. *Applied Sciences*, 12(14):7035, 2022.
- [12] ASTM. Standard Test Method for Determining Compressive Resistance of Shipping Containers, Components, and Unit Loads, howpublished = <https://www.astm.org/d0642-20.html>, 2020. Online; accessed 22 OCT 2020.
- [13] Daniel Goodwin and Dennis Young. *Protective packaging for distribution: design and development*. DEStech Publications, Inc, 2011.
- [14] Pallab Mandal, Jasmina Khanam, Sanmoy Karmakar, Tapan Kumar Pal, Sujata Barma, Soumya Chakraborty, Rakesh Bera, and Sourav Poddar. An audit on design of pharmaceutical packaging. *Journal of Packaging Technology and Research*, 6(3):167–185, 2022.
- [15] William I Kipp. Packaging hazard, 206.

- [16] Sara Shumpert Dunn. E-commerce packaging strategy: Design with the end in mind. <https://www.linkedin.com/pulse/e-commerce-packaging-strategy-design-end-mind-sara-shumpert-dunn/>, 2018.
- [17] Hugh Lockhart and Frank A Paine. *Packaging of pharmaceuticals and healthcare products*. Springer Science & Business Media, 1996.
- [18] Gordon L Robertson. Good and bad packaging: who decides? *International Journal of Physical Distribution & Logistics Management*, 20(8):37–40, 1990.
- [19] Lansmont Website. Lansmont product, howpublished = <https://www.lansmont.com/de>.
- [20] Kathy Baxter, Catherine Courage, and Kelly Caine. *Understanding your users: a practical guide to user research methods*. Morgan Kaufmann, 2015.
- [21] Basem El-Haik. *Axiomatic quality: integrating axiomatic design with six-sigma, reliability, and quality engineering*. John Wiley & Sons, 2005.
- [22] Richard Kenneth Brandenburg and Julian June-Ling Lee. Fundamentals of packaging dynamics. (*No Title*), 1985.
- [23] Presto Testing Instruments. Box Compression Tester – A Key Solution For Packaging Industries, howpublished = <https://www.testing-instruments.com/blog/box-compression-tester-a-key-solution-for-packaging-industries/>. Online; accessed 19 Dec.
- [24] Advanced Packaging Technology Laboratories, Inc. VIBRATION TESTING, howpublished = <https://advanced-labs.com/vibration/>.
- [25] Vihaan Nagal. How To Drop Test A Box To Determine If It’s Right For Your Business, howpublished = <https://packagingguruji.com/drop-test-a-box/>, 2022. Online; accessed 23 July 2022.
- [26] Roger E Kirk. *Experimental design: Procedures for the behavioral sciences* (4th), 2013.
- [27] Averill M Law, W David Kelton, and W David Kelton. *Simulation modeling and analysis*, volume 3. Mcgraw-hill New York, 2007.
- [28] Olek C Zienkiewicz, Robert L Taylor, and Jian Z Zhu. *The finite element method: its basis and fundamentals*. Elsevier, 2005.
- [29] Tobi Fadiji, Alemayehu Ambaw, Corné J Coetzee, Tarl M Berry, and Umezuruike Linus Opara. Application of finite element analysis to predict the mechanical strength of ventilated corrugated paperboard packaging for handling fresh produce. *Biosystems Engineering*, 174:260–281, 2018.
- [30] V Dung Luong, Fazilay Abbès, Boussad Abbès, PT Minh Duong, Jean-Baptiste Nolot, Damien Erre, and Ying-Qiao Guo. Finite element simulation of the strength of corrugated board boxes under impact dynamics. In *Proceedings of the International Conference on Advances in Computational Mechanics 2017: ACOME 2017, 2 to 4 August 2017, Phu Quoc Island, Vietnam*, pages 369–380. Springer, 2018.

- [31] Olgierd Cecil Zienkiewicz and Robert Leroy Taylor. *The finite element method: solid mechanics*, volume 2. Butterworth-heinemann, 2000.
- [32] THE DURHAM-HUMPHREY AMENDMENT. *Journal of the American Medical Association*, 149(4):371–371, 05 1952.
- [33] Steven M Albert, Laura Bix, Mary M Bridgeman, Laura L Carstensen, Margaret Dyer-Chamberlain, Patricia J Neafsey, and Michael S Wolf. Promoting safe and effective use of otc medications: Chpa-gsa national summit. *The Gerontologist*, 54(6):909–918, 2014.
- [34] Consumer Healthcare Products Association (CHPA). New Study: Over-the-Counter (OTC) Products Used by Millions of Americans Saves Healthcare System Billions Annually. <https://apnews.com/press-release/business-wire/business-health-58f7099d41a6445382dac7c361455083>, 2019. Online; accessed 18 March 2019.
- [35] Consumer Healthcare Products Association (CHPA). OTC Sales Statistics. <https://www.chpa.org/about-consumer-healthcare/research-data/otc-sales-statistics>. Online.
- [36] US Food and Drug Administration. Prescription-to-Nonprescription (Rx-to-OTC) Switches. <https://www.fda.gov/about-fda/changes-science-law-and-regulatory-authorities/part-iii-drugs-and-foods-under-1938-act-and-its-amendments>.
- [37] US Food and Drug Administration. Over-the-Counter (OTC) Drugs Branch: The OTC Drug Review. <https://www.fda.gov/drugs/enforcement-activities-fda/over-counter-otc-drugs-branch-otc-drug-review>.
- [38] M Hernandez-Juyol and JR Job-Quesada. Dentistry and self-medication: a current challenge. *Medicina oral: organo oficial de la Sociedad Espanola de Medicina Oral y de la Academia Iberoamericana de Patologia y Medicina Bucal*, 7(5):344–347, 2002.
- [39] Tanmay Mahapatra. Self-care and self-medication: A commentary. *Annals of Tropical Medicine and Public Health*, 10(3):505–505, 2017.
- [40] Lanqing Liu. *Improving Interactions Between Self-Medicating Consumers and Over-the-Counter Packaging with Front-of-Pack and Personalized Labeling as Strategies*. Michigan State University, 2022.
- [41] Jeffrey K Aronson and Robin E Ferner. Clarification of terminology in drug safety. *Drug safety*, 28(10):851–870, 2005.
- [42] Jason Lazarou, Bruce H Pomeranz, and Paul N Corey. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *Jama*, 279(15):1200–1205, 1998.
- [43] Eric P Brass and Michael Weintraub. Label development and the label comprehension study for over-the-counter drugs. *Clinical Pharmacology & Therapeutics*, 74(5):406–412, 2003.

- [44] Sven Schmiedl, Marietta Rottenkolber, Joerg Hasford, Dominik Rottenkolber, Katrin Farker, Bernd Drewelow, Marion Hippus, Karen Saljé, and Petra Thürmann. Self-medication with over-the-counter and prescribed drugs causing adverse-drug-reaction-related hospital admissions: results of a prospective, long-term multi-centre study. *Drug safety*, 37:225–235, 2014.
- [45] US Food and Drug Administration. Guidance for Industry: Food Labeling Guide. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-food-labeling-guide>.
- [46] Mohamed Altai, Kristina Westerlund, Justin Vellella, Bogdan Mitran, Hadis Honarvar, and Amelie Eriksson Karlström. Evaluation of affibody molecule-based pna-mediated radionuclide pretargeting: development of an optimized conjugation protocol and 177lu labeling. *Nuclear Medicine and Biology*, 54:1–9, 2017.
- [47] Vivien Tong, David K Raynor, and Parisa Aslani. User testing as a method for identifying how consumers say they would act on information related to over-the-counter medicines. *Research in Social and Administrative Pharmacy*, 13(3):476–484, 2017.
- [48] Laura Bix, Raghav Prashant Sundar, Nora M Bello, Chad Peltier, Lorraine J Weatherspoon, and Mark W Becker. To see or not to see: Do front of pack nutrition labels affect attention to overall nutrition information? *PLoS One*, 10(10):e0139732, 2015.
- [49] Eric F Shaver and Michael S Wogalter. A comparison of older vs. newer over-the-counter (otc) nonprescription drug labels on search time accuracy. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 47, pages 826–830. SAGE Publications Sage CA: Los Angeles, CA, 2003.
- [50] Alyssa Harben, Shiva Esfahanian, and Laura Bix. An assessment of older adults’ selection of over-the-counter medication: What information are they utilizing during the selection process? *Packaging Technology and Science*, 2023.
- [51] Olayinka O Shiyabola, Brittney A Meyer, Michelle R Locke, and Sara Wettergreen. Perceptions of prescription warning labels within an underserved population. *Pharmacy practice*, 12(1), 2014.
- [52] Olayinka O Shiyabola, Paul D Smith, Sonal Ghura Mansukhani, and Yen-Ming Huang. Refining prescription warning labels using patient feedback: a qualitative study. *PLoS One*, 11(6):e0156881, 2016.
- [53] John McCarthy. What is artificial intelligence. URL: <http://www-formal.stanford.edu/jmc/whatisai.html>, 2004.
- [54] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Pearson, 2016.
- [55] Arya Yaghoobzadeh-Bavandpour, Omid Bozorg-Haddad, Babak Zolghadr-Asli, and Vijay P Singh. Computational intelligence: an introduction. In *Computational intelligence for water and environmental sciences*, pages 411–427. Springer, 2022.

- [56] John McCarthy, Marvin L Minsky, Nathaniel Rochester, and Claude E Shannon. A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI magazine*, 27(4):12–12, 2006.
- [57] Peter Norvig Russell. Artificial intelligence: a modern approach by stuart. *Russell and Peter Norvig contributing writers, Ernest Davis...[et al.]*, 2010.
- [58] James Hendler. Avoiding another ai winter. *IEEE Intelligent Systems*, 23(02):2–4, 2008.
- [59] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [60] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.
- [61] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [62] Pedro Domingos. *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Basic Books, 2015.
- [63] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9:381–386, 2020.
- [64] Arthur L Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, 3(3):210–229, 1959.
- [65] Pamela McCorduck and Cli Cfe. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. CRC Press, 2004.
- [66] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [67] Alon Halevy, Peter Norvig, and Fernando Pereira. The unreasonable effectiveness of data. *IEEE intelligent systems*, 24(2):8–12, 2009.
- [68] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.
- [69] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.
- [70] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1:81–106, 1986.
- [71] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.
- [72] Irina Rish et al. An empirical study of the naive *Bayes* classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, pages 41–46, 2001.

- [73] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.
- [74] Ameet V Joshi. Support vector machines. In *Machine learning and artificial intelligence*, pages 89–99. Springer, 2022.
- [75] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20:273–297, 1995.
- [76] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.
- [77] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [78] Mariana Belgiu and Lucian Drăguț. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31, 2016.
- [79] Zhenfeng Shao, Muhammad Nasar Ahmad, and Akib Javed. Comparison of random forest and xgboost classifiers using integrated optical and sar features for mapping urban impervious surface. *Remote Sensing*, 16(4):665, 2024.
- [80] Gérard Biau and Erwan Scornet. A random forest guided tour. *Test*, 25:197–227, 2016.
- [81] A Lokesh Reddy, T Sathish, and N Sangeetha. Prediction of student results using novel random forest in comparison with decision tree to improve accuracy. In *AIP Conference Proceedings*, volume 2853. AIP Publishing, 2024.
- [82] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.
- [83] Jing Wang and Filip Biljecki. Unsupervised machine learning in urban studies: A systematic review of applications. *Cities*, 129:103925, 2022.
- [84] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [85] Oliver Theobald. *Machine learning for absolute beginners: a plain English introduction*, volume 157. Scatterplot press United States, 2017.
- [86] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European conference on computer vision*, pages 69–84. Springer, 2016.
- [87] Kevin P Murphy. *Probabilistic machine learning: an introduction*. MIT press, 2022.
- [88] Xiaojin Zhu, Zoubin Ghahramani, and John D Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pages 912–919, 2003.
- [89] Mohamed Farouk Abdel Hady and Friedhelm Schwenker. Semi-supervised learning. *Handbook on Neural Information Processing*, pages 215–239, 2013.

- [90] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]. *IEEE Transactions on Neural Networks*, 20(3):542–542, 2009.
- [91] Anders Søgaard. *Semi-supervised learning and domain adaptation in natural language processing*. Springer Nature, 2022.
- [92] Jesper E Van Engelen and Holger H Hoos. A survey on semi-supervised learning. *Machine learning*, 109(2):373–440, 2020.
- [93] Shan-Shan Wang, Pinpin Lin, Chia-Chi Wang, Ying-Chi Lin, and Chun-Wei Tung. Machine learning for predicting chemical migration from food packaging materials to foods. *Food and Chemical Toxicology*, 178:113942, 2023.
- [94] Nooshin Salari, Sheng Liu, and Zuo-Jun Max Shen. Real-time delivery time forecasting and promising in online retailing: When will your package arrive? *Manufacturing & Service Operations Management*, 24(3):1421–1436, 2022.
- [95] Hsien-Wei Ting, Sheng-Luen Chung, Chih-Fang Chen, Hsin-Yi Chiu, and Yow-Wen Hsieh. A drug identification model developed using deep learning technologies: experience of a medical center in taiwan. *BMC health services research*, 20(1):1–9, 2020.
- [96] Pinkaew Horputra, Rateepat Phrajonthong, and Phisan Kaewprapha. Deep learning-based bottle caps inspection in beverage manufacturing and packaging process. In *2021 9th International Electrical Engineering Congress (iEECON)*, pages 499–502. IEEE, 2021.
- [97] Dino Knoll, Daniel Neumeier, Marco Prüglmeier, and Gunther Reinhart. An automated packaging planning approach using machine learning. *Procedia Cirp*, 81:576–581, 2019.
- [98] Sandhya Makkar, G Naga Rama Devi, and Vijender Kumar Solanki. Applications of machine learning techniques in supply chain optimization. In *ICICCT 2019–System Reliability, Quality Control, Safety, Maintenance and Management: Applications to Electrical, Electronics and Computer Science and Engineering*, pages 861–869. Springer, 2020.
- [99] Jacob Eisenstein. *Introduction to natural language processing*. MIT press, 2019.
- [100] Ivano Lauriola, Alberto Lavelli, and Fabio Aiolli. An introduction to deep learning in natural language processing: Models, techniques, and tools. *Neurocomputing*, 470:443–456, 2022.
- [101] Sridhar Ramaswamy and Natalie DeClerck. Customer perception analysis using deep learning and nlp. *Procedia Computer Science*, 140:170–178, 2018.
- [102] Rachel Wolff. what Is Natural Language Processing. <https://monkeylearn.com/blog/what-is-natural-language-processing/>.
- [103] Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551, 2011.

- [104] Keith D Foote. A brief history of natural language processing (nlp). *DATAVERSITY, May*, 22, 2019.
- [105] Dan Jurafsky and James H Martin. Speech and language processing (draft). 2021. URL: <https://web.stanford.edu/~jurafsky/slp3>, 2020.
- [106] Rachel Wolff. 11 NLP Applications & Examples in Business. <https://monkeylearn.com/blog/natural-language-processing-applications/>.
- [107] Jacob Eisenstein. Natural language processing. *Jacob Eisenstein*, 2018.
- [108] Shervin Minaee, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. Deep learning–based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, 54(3):1–40, 2021.
- [109] Fabrizio Sebastiani. Machine learning in automated text categorization. *ACM computing surveys (CSUR)*, 34(1):1–47, 2002.
- [110] Mohammad Nuruzzaman and Omar Khadeer Hussain. A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)*, pages 54–61. IEEE, 2018.
- [111] Ewa Luger and Abigail Sellen. "like having a really bad pa" the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5286–5297, 2016.
- [112] Philipp Koehn. *Statistical machine translation*. Cambridge University Press, 2009.
- [113] Amr Hendy, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. How good are gpt models at machine translation? a comprehensive evaluation. *arXiv preprint arXiv:2302.09210*, 2023.
- [114] Ani Nenkova, Kathleen McKeown, et al. Automatic summarization. *Foundations and Trends® in Information Retrieval*, 5(2–3):103–233, 2011.
- [115] Xiaokang Liu, Jianquan Li, Jingjing Mu, Min Yang, Ruifeng Xu, and Benyou Wang. Effective open intent classification with k-center contrastive learning and adjustable decision boundary. *arXiv preprint arXiv:2304.10220*, 2023.
- [116] Laurenti Enzo, Bourgon Nils, Farah Benamara, Mari Alda, Véronique Moriceau, and Courgeon Camille. Speech acts and communicative intentions for urgency detection. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 289–298, 2022.
- [117] Jinyu Li, Li Deng, Reinhold Haeb-Umbach, and Yifan Gong. Robust automatic speech recognition: a bridge to practical applications. 2015.

- [118] Mayur Wankhade, Annavarapu Chandra Sekhara Rao, and Chaitanya Kulkarni. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7):5731–5780, 2022.
- [119] Bo Pang, Lillian Lee, et al. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135, 2008.
- [120] Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow, and Rebecca J Passonneau. Sentiment analysis of twitter data. In *Proceedings of the workshop on language in social media (LSM 2011)*, pages 30–38, 2011.
- [121] Bing Liu. *Sentiment analysis and opinion mining*. Springer Nature, 2022.
- [122] Bleau Moores and Vijay Mago. A survey on automated sarcasm detection on twitter. *arXiv preprint arXiv:2202.02516*, 2022.
- [123] Bing Liu. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge university press, 2020.
- [124] Federico Alberto Pozzi, Elisabetta Fersini, Enza Messina, and Bing Liu. *Sentiment analysis in social networks*. Morgan Kaufmann, 2016.
- [125] Ramon Ferrer I Cancho and Richard V Solé. The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482):2261–2265, 2001.
- [126] Rada Mihalcea and Paul Tarau. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pages 404–411, 2004.
- [127] Camilo Akimushkin, Diego Raphael Amancio, and Osvaldo Novais Oliveira Jr. Text authorship identified using the dynamics of word co-occurrence networks. *PloS one*, 12(1):e0170527, 2017.
- [128] Mikaela Irene Fudolig, Thayer Alshaabi, Michael V Arnold, Christopher M Danforth, and Peter Sheridan Dodds. Sentiment and structure in word co-occurrence networks on twitter. *Applied Network Science*, 7(1):1–27, 2022.
- [129] Sílvia Escursell, Pere Llorach-Massana, and M Blanca Roncero. Sustainability in e-commerce packaging: A review. *Journal of cleaner production*, 280:124314, 2021.
- [130] K Taylor. The retail apocalypse is far from over as analysts predict 75,000 more store closures[www document]. bus. insid, 2019.
- [131] Rae Yule Kim. The impact of covid-19 on consumers: Preparing for digital sales. *IEEE Engineering Management Review*, 48(3):212–218, 2020.
- [132] Al Pizzuti. How to overcome 4 challenges of ecommerce packaging. <https://www.amcor.com/insights/blogs/how-to-overcome-4-challenges-of-e-commerce-packaging>, 2017. Online; accessed 14 December 2017.

- [133] Janay Cooper. E-commerce Packaging Has Different Needs. <https://www.netrush.com/insights/e-commerce-packaging-has-different-needs>, 2020. Online; accessed 7 January 2020.
- [134] Emily Anne Page. CRUCIAL DIFFERENCES IN E-COMMERCE VS. BRICK-AND-MORTAR PACKAGING. <https://www.emilyannepage.com/post/crucial-differences-in-e-commerce-vs-brick-mortar-packaging/>.
- [135] CJ Dodds and AR Plummer. Laboratory road simulation for full vehicle testing: a review. *SAE technical paper*, pages 26–0047, 2001.
- [136] Vincent Rouillard. Generating road vibration test schedules from pavement profiles for packaging optimization. *Packaging Technology and Science: An International Journal*, 21(8):501–514, 2008.
- [137] Péter Böröcz and S Paul Singh. Measurement and analysis of delivery van vibration levels to simulate package testing for parcel delivery in Hungary. *Packaging Technology and Science: An International Journal*, 31(5):342–352, 2018.
- [138] Mikael Nygårds, Stefan Sjökvist, Gustav Marin, and Jonas Sundström. Simulation and experimental verification of a drop test and compression test of a gable top package. *Packaging Technology and Science: An International Journal*, 32(7):325–333, 2019.
- [139] Eduardo Molina, Laszlo Horvath, and Robert L West. Development of a friction-driven finite element model to simulate the load bridging effect of unit loads stored in warehouse racks. *Applied Sciences*, 11(7):3029, 2021.
- [140] Fayi Hao, Lixin Lu, and Jun Wang. Finite element simulation of shelf life prediction of moisture-sensitive crackers in permeable packaging under different storage conditions. *Journal of Food Processing and Preservation*, 40(1):37–47, 2016.
- [141] Chiara Cevoli and Angelo Fabbri. Heat transfer finite element model of fresh fruit salad insulating packages in non-refrigerated conditions. *Biosystems Engineering*, 153:89–98, 2017.
- [142] Chia-Lung Chang and Shao-Huei Yang. Simulation of wheel impact test using finite element method. *Engineering Failure Analysis*, 16(5):1711–1719, 2009.
- [143] F Ballo, R Frizzi, M Gobbi, G Mastinu, G Previati, and C Sorlini. Numerical and experimental study of radial impact test of an aluminum wheel: Towards industry 4.0 virtual process assessment. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 58158, page V003T01A015. American Society of Mechanical Engineers, 2017.
- [144] F Ballo, G Previati, G Mastinu, and F Comolli. Impact tests of wheels of road vehicles: A comprehensive method for numerical simulation. *International Journal of Impact Engineering*, 146:103719, 2020.

- [145] Onder Kabas, H Kursat Celik, Aziz Ozmerzi, and Ibrahim Akinci. Drop test simulation of a sample tomato with finite element method. *Journal of the Science of Food and Agriculture*, 88(9):1537–1541, 2008.
- [146] Somaye Yousefi, Habib Farsi, and Kamran Kheiralipour. Drop test of pear fruit: Experimental measurement and finite element modelling. *Biosystems Engineering*, 147:17–25, 2016.
- [147] Onder Kabas and Valentin Vladut. Determination of drop-test behavior of a sample peach using finite element method. *International Journal of Food Properties*, 18(11):2584–2592, 2015.
- [148] Tobi Fadiji, Tarl Berry, Corne J Coetzee, and Linus Opara. Investigating the mechanical properties of paperboard packaging material for handling fresh produce under different environmental conditions: Experimental analysis and finite element modelling. *Journal of Applied Packaging Research*, 9(2):3, 2017.
- [149] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.
- [150] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google’s neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.
- [151] Brian Quanz, Wei Sun, Ajay Deshpande, Dhruv Shah, and Jae-eun Park. Machine learning based co-creative design framework. *arXiv preprint arXiv:2001.08791*, 2020.
- [152] Shuai Zhao, Manoop Talasila, Guy Jacobson, Cristian Borcea, Syed Anwar Aftab, and John F Murray. Packaging and sharing machine learning models via the Acumos AI open platform. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 841–846. IEEE, 2018.
- [153] M Mudambi Susan and Schoff David. What makes a helpful online review? a study of customer reviews on Amazon. com. *MIS Quarterly*, 34(1):185–200, 2010.
- [154] B Ebehavior and PA Pavlou. Evidence of the effect of trust building technology in electronic markets: price premiums and buyer. *MIS Quartely*, 26(3):243–268, 2002.
- [155] Paul A Pavlou and David Gefen. Building effective online marketplaces with institution-based trust. *Information Systems Research*, 15(1):37–59, 2004.
- [156] Gobinda G Chowdhury. Natural language processing. *Annual Review of Information Science and Technology*, 37(1):51–89, 2003.
- [157] Richard Socher, Yoshua Bengio, and Christopher D Manning. Deep learning for NLP (without magic). In *Tutorial Abstracts of ACL 2012*, pages 5–5. 2012.

- [158] Soo-Min Kim and Eduard Hovy. Determining the sentiment of opinions. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 1367–1373, 2004.
- [159] Jesus Serrano-Guerrero, Jose A Olivas, Francisco P Romero, and Enrique Herrera-Viedma. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*, 311:18–38, 2015.
- [160] Dimitrios Kouzis-Loukas. *Learning scrapy*. Packt Publishing Ltd, 2016.
- [161] Gregory Grefenstette. Tokenization. In *Syntactic Wordclass Tagging*, pages 117–133. Springer, 1999.
- [162] Risky Novendri, Annisa Syafarani Callista, Danny Naufal Pratama, and Chika Enggar Puspita. Sentiment analysis of *YouTube* movie trailer comments using Naïve *Bayes*. *Bulletin of Computer Science and Electrical Engineering*, 1(1):26–32, 2020.
- [163] Mark Hall. A decision tree-based attribute weighting filter for Naive *Bayes*. In *International conference on innovative techniques and applications of artificial intelligence*, pages 59–70. Springer, 2006.
- [164] Pedro Domingos and Michael Pazzani. On the optimality of the simple *Bayesian* classifier under zero-one loss. *Machine Learning*, 29(2):103–130, 1997.
- [165] Joseph L Hellerstein, TS Jayram, Irina Rish, et al. *Recognizing end-user transactions in performance management*. IBM Thomas J. Watson Research Division Hawthorne, NY, 2000.
- [166] Tom M Mitchell et al. *Machine Learning*. 1997. *Burr Ridge, IL: McGraw Hill*, 45(37):870–877, 1997.
- [167] MathWorks. Create Co-occurrence Network. <https://www.mathworks.com/help/textanalytics>.
- [168] US Food and Drug Administration. Part III: Drugs and Foods Under the 1938 Act and Its Amendments. <https://www.fda.gov/about-fda/changes-science-law-and-regulatory-authorities/part-iii-drugs-and-foods-under-1938-act-and-its-amendments>, 2018. Online; accessed 2 January 2018.
- [169] Nasiba Mahdi Abdulkareem, Adnan Mohsin Abdulazeez, et al. Machine learning classification based on random forest algorithm: A review. *International Journal of Science and Business*, 5(2):128–142, 2021.
- [170] Keith T Butler, Daniel W Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh. Machine learning for molecular and materials science. *Nature*, 559(7715):547–555, 2018.
- [171] Jing Wei, Xuan Chu, Xiang-Yu Sun, Kun Xu, Hui-Xiong Deng, Jigen Chen, Zhongming Wei, and Ming Lei. Machine learning in materials science. *InfoMat*, 1(3):338–358, 2019.

- [172] Wassim Ben Chaabene, Majdi Flah, and Moncef L Nehdi. Machine learning prediction of mechanical properties of concrete: Critical review. *Construction and Building Materials*, 260:119889, 2020.
- [173] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1):89–109, 2001.
- [174] Bradley J Erickson, Panagiotis Korfiatis, Zeynettin Akkus, and Timothy L Kline. Machine learning for medical imaging. *Radiographics*, 37(2):505, 2017.
- [175] Ronald A Rensink. Change detection. *Annual review of psychology*, 53(1), 2002.
- [176] Shiva Esfahanian. *A Patient-Centered Approach to Labeling for Over-The-Counter Medications: Using Data to Drive Design Decisions for the Benefit of Older Adults*. Michigan State University, 2020.
- [177] Scikit-learn’s development and maintenance. Decision Trees. <https://scikit-learn.org/stable/modules/tree.html>, 2007-2022. Online.
- [178] Rachel Cravit. What is a Decision Tree and How to Make One [Templates + Examples]. <https://venngage.com/blog/what-is-a-decision-tree/>, 2021. Online, Aug 03 2021.
- [179] Sebastian Raschka and Vahid Mirjalili. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt publishing ltd, 2019.
- [180] Stacey Ronaghan. The Mathematics of Decision Trees, Random Forest and Feature Importance in Scikit-learn and Spark. <https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3>, 2018. Online, May 11 2018.
- [181] Daksh Trehan. Why choose Random Forest and Not Decision Trees. <https://towardsai.net/p/machine-learning/why-choose-random-forest-and-not-decision-trees>, 2020. Online, July 2 2020.
- [182] Onesmus Mbaabu. Introduction to Random Forest in Machine Learning. <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/>, 2020. Online, Dec 11 2020.
- [183] Adele Cutler, D Richard Cutler, and John R Stevens. Random forests. *Ensemble machine learning: Methods and applications*, pages 157–175, 2012.
- [184] Onel Harrison. Machine Learning Basics with the K-Nearest Neighbors Algorithm. <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighbors-algorithm-6a6e71d01761>, 2018. Online, Sep 10 2018.
- [185] Nour Al-Rahman Al-Serw. The maths behind it, how it works and an example. <https://medium.com/analytics-vidhya/k-nearest-neighbor-the-maths-behind-it-how-it-works-and-an-example-f1de1208546c>, 2021. Online, May 17 2021.

- [186] Genesis. Pros and Cons of K-Nearest Neighbor. <https://www.fromthegenesis.com/pros-and-cons-of-k-nearest-neighbors/>, 2018. Online, Sep 25 2018.
- [187] Aniruddha Bhandari. AUC-ROC Curve in Machine Learning Clearly Explained. <https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/>, 2020. Online, Jun 16 2020.
- [188] Mithat Gönen et al. Receiver operating characteristic (roc) curves. *SAS Users Group International (SUGI)*, 31:210–231, 2006.
- [189] John Muschelli III. Roc and auc with a binary predictor: a potentially misleading metric. *Journal of classification*, 37(3):696–708, 2020.
- [190] Chris Albon. *Machine learning with python cookbook: Practical solutions from preprocessing to deep learning*. " O'Reilly Media, Inc.", 2018.
- [191] Wonjae Lee and Kangwon Seo. Downsampling for binary classification with a highly imbalanced dataset using active learning. *Big Data Research*, 28:100314, 2022.
- [192] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
- [193] Alberto Fernández, Salvador García, Mikel Galar, Ronaldo C Prati, Bartosz Krawczyk, and Francisco Herrera. *Learning from imbalanced data sets*, volume 10. Springer, 2018.

APPENDIX

Table A1. Random Forest Trees with Number of Nodes and Leaves

Trees	Number of Nodes	Number of Leaves	Trees	Number of Nodes	Number of Leaves
1	59	30	2	59	30
3	55	28	4	75	38
5	61	31	6	23	12
7	69	35	8	61	31
9	73	37	10	57	29
11	59	30	12	65	33
13	79	40	14	61	31
15	65	33	16	79	40
17	65	33	18	69	35
19	47	24	20	91	46
21	53	27	22	45	23
23	87	44	24	79	40
25	73	37	26	69	35
27	77	39	28	75	38
29	65	45	30	89	45
31	43	22	32	71	36
33	61	31	34	61	31
35	65	33	36	65	33
37	77	39	38	71	36
39	75	38	40	51	26
41	65	33	42	63	32
43	59	30	44	77	39
45	77	39	46	69	35
47	77	39	48	61	31
49	59	30	50	77	39
51	45	23	52	63	32
53	81	41	54	73	37
55	77	39	56	51	26
57	69	35	58	89	45
59	81	41	60	67	34
61	47	24	62	73	37
63	85	43	64	73	37
65	93	47	66	65	33
67	69	35	68	61	31
69	59	30	70	57	29
71	61	31	72	55	28
73	41	21	74	63	32
75	71	36	76	67	34

Table A1 (cont'd)

77	69	35	78	73	37
79	59	30	80	71	36
81	69	35	82	75	38
83	61	31	84	69	35
85	69	35	86	55	28
87	69	35	88	63	32
89	65	33	90	57	29
91	81	41	92	63	32
93	65	33	94	45	23
95	41	21	96	49	25
97	49	25	98	63	32
99	47	24	100	75	38
101	71	36	102	67	34
103	61	31	104	75	38
105	71	36	106	41	21
107	53	27	108	57	29
109	57	29	110	75	38
111	77	39	112	57	29
113	73	37	114	71	36
115	67	34	116	51	26
117	63	32	118	59	30
119	67	34	120	55	28
121	47	24	122	73	37
123	71	36	124	71	36
125	63	32	126	69	35
127	63	32	128	55	28
129	59	30	130	69	35
131	53	27	132	71	36
133	71	36	134	69	35
135	60	35	136	71	36
137	71	36	138	59	30
139	55	28	140	65	33
141	73	37	142	63	32
143	73	37	144	67	34
145	75	38	146	63	32
147	59	30	148	59	25
149	73	37	150	75	38
151	53	27	152	89	45
153	53	27	154	73	37
155	63	32	156	57	29
157	49	25	158	43	22

Table A1 (cont'd)

159	77	39	160	71	36
161	67	34	162	55	28
163	61	31	164	79	40
165	63	32	166	59	30
167	67	34	168	59	30
169	45	23	170	73	37
171	65	33	172	69	35
173	63	32	174	69	35
175	65	33	176	89	45
177	79	40	178	63	32
179	51	26	180	59	30
181	63	32	182	77	39
183	79	40	184	67	34
185	71	36	186	69	35
187	63	32	188	69	35
189	55	28	190	59	30
191	63	32	192	75	38
193	69	35	194	63	32
195	71	36	196	33	17
197	51	26	198	67	34
199	57	29	200	81	41