Academic Journal of Research and Scientific Publishing

International peer-reviewed scientific journal

The Fifty-Fifth Issue

Publication date: 05-11-2023

ISSN: 2706-6495

Doi: doi.org/10.52132/Ajrsp.e.2023.55

Email: editor@ajrsp.com

Dedication

It is our pleasure and great privilege to present the fifty-fifth issue of the Academic Journal of Research and Scientific Publishing to all researchers and doctors who published their research in the issue, and we thanks and appreciate to all contributors and supporters of the academic journal and those involved in the production of this scientific knowledge edifice.

Academic Journal of Research and Scientific Publishing

Editorial Board

Table of Content:

Digital Forensic Investigation Tools for Cases Related to Social Media and Cybersecurity

Asma Ibrahim Hijan

Bachelor of Computer Science, King Khalid University, Saudi Arabia

Email: asma.hijan@gmil.com

Abstract:

The entire world has moved into the digital realm as a result of the quickening pace of technological development. However, this change has also led to an increase in cybercrimes and security breach occurrences, which endanger user security and privacy. As a result, this study sought to examine how digital forensics, a significant advancement in cybersecurity, is used to combat cybercrime. The most recent developments in digital forensics, such as cloud forensics, social media forensics, and IoT forensics, have been examined in this study. With the use of these technologies, cybersecurity experts can protect data while identifying fraudsters by using the digital footprints that data processing and storage leave behind.

Technical, operational, and personnel-related issues are among the specific dangers to digital forensics that have been identified by the research. These systems' high level of complexity, volume of data, chain of custody, personnel integrity, and the validity and accuracy of digital forensics are all significant barriers to their widespread adoption. However, the research has also noted the use of artificial intelligence, intrusion detection, and USB forensics as significant prospects for digital forensics that can make the processes simpler, more effective, and safe.

Keywords: Digital forensics, Data security, Cybercrime, Data theft, Security attack

1. Introduction

The emergence of Web 2.0 technologies and the rapid advancement in the digital arena have significantly altered the global paradigm. These days, more and more people participate in internet exchanges, give to open projects, and post information about their Chapter online. However, the anonymity and ease with which any of these can be carried out cause concern over veracity and trust (Gollub, 2013). Particularly, the development of digital technologies has given rise to new strategies for committing computer crimes. In addition, the accessibility of networks and highly optimized data transfer have prompted security worries. Every day, malicious approaches, tools, and software are developed and executed to threaten both public and private networks while also exploiting data storage to obtain valuable information (Aminnezhad & Dehghantanha, 2014).

Digital forensics has received a lot of attention in resolving cybersecurity concerns to counter this new threat. According to, the science of displaying, documenting, analyzing, storing, and identifying information and evidence from electronic and digital devices while protecting user privacy is known as digital forensics. In addition, it recreates and explains the sequence of events using scientific methods. Digital forensics tries to use such illicit artifacts as evidence by analyzing, examining, and recording these sequences (Dezfouli & Dehghantanha, 2014).

Social networks undeniably power the modern world, and as digital technologies have advanced, cybercrime has also advanced, considerably influencing the creation of new strategies, tools, and attacks that allow attackers to breach even well-controlled environments (Sharma et al., 2019).

To combat the growing number of cyber anomalies, security professionals, researchers, and law enforcement organizations use digital forensics. To gather digital evidence, these professionals use scientific techniques including identification, validation, interpretation, and documentation on digital devices like RAM, phones, memory cards, floppy disks, and flash drives. However, as digital forensics tools progress, hackers are also taking use of anti-forensics technologies to either delay or entirely destroy digital evidence (Wazid et al., 2013).

1.1. Study problem

The rapid growth of social media platforms and their integration into various aspects of our lives has given rise to new challenges in the realm of cybersecurity. As social media becomes increasingly prevalent, it becomes a prime target for cybercriminals seeking to exploit

vulnerabilities, commit fraud, engage in harassment, or perpetrate other malicious activities. Digital forensic investigations play a crucial role in uncovering evidence, identifying perpetrators, and providing legal support in social media-related cybersecurity cases.

However, the field of digital forensics faces challenges in effectively investigating and analyzing social media-related incidents due to the dynamic nature of social media platforms, their vast amount of user-generated content, and the complexity of extracting and preserving digital evidence. Therefore, there is a need to evaluate and assess the existing digital forensic investigation tools specifically designed for social media-related cybersecurity cases.

1.2. Study question

1- What are the current challenges faced by digital forensic investigators when conducting investigations related to social media in the context of cybersecurity incidents؟

2- What are the existing digital forensic investigation tools available for analyzing social media-related evidence in cybersecurity cases, and what are their strengths and limitations؟

3- How effective are the existing digital forensic investigation tools in handling social media-related evidence in cybersecurity cases? What are their capabilities, features, and limitations؟

4- Are there any specific requirements or improvements needed in digital forensic investigation tools to enhance their effectiveness in social media-related cybersecurity cases?

1.3. Study Objectives:

1- Assess the current challenges and limitations faced by digital forensic investigators when conducting investigations related to social media in the context of cybersecurity incidents.

2- Identify and evaluate existing digital forensic investigation tools specifically designed for analyzing social media-related evidence in cybersecurity cases.

3- Analyze the capabilities, features, and limitations of the identified digital forensic investigation tools in handling social media-related evidence.

4- Determine the effectiveness of the existing tools in addressing the challenges posed by social media in cybersecurity investigations.

5- Identify specific requirements and improvements needed in digital forensic investigation tools to enhance their effectiveness in social media-related cybersecurity cases.

1.4. Study Importance:

1- Enhancing Investigation Capabilities: The study of digital forensic investigation tools for social media-related cybersecurity cases is crucial for enhancing the capabilities of investigators. By understanding the strengths and limitations of existing tools, researchers can identify areas for improvement and develop more effective methodologies and techniques to investigate social media-related incidents.

2- Addressing Emerging Threats: Social media platforms are constantly evolving, and cybercriminals adapt their tactics accordingly. By focusing on digital forensic investigation tools for social media-related cases, this study addresses the emerging threats and challenges posed by cybercrime on these platforms. It helps in staying ahead of cybercriminals and developing strategies to effectively investigate and mitigate social media-related cybersecurity incidents.

3- Preserving Digital Evidence: Social media platforms generate vast amounts of user-generated content, making it challenging to extract and preserve digital evidence. This study emphasizes the importance of digital forensic investigation tools in accurately collecting, analyzing, and preserving social media-related evidence. It contributes to the development of best practices for maintaining the integrity and admissibility of digital evidence in legal proceedings.

4- Informing Decision-Making: By evaluating the capabilities of existing digital forensic investigation tools, this study provides insights that can inform decision-making processes. Investigators, forensic analysts, and cybersecurity professionals can make informed choices about the tools they employ in social media-related cybersecurity cases, ensuring efficient and effective investigations.

2. Methodology

Approach and Relationship to the Study:

The approach followed in this study is a combination of literature review, tool evaluation, data analysis, and requirement analysis. The study begins with a comprehensive review of existing literature to establish a foundation of knowledge and identify key challenges, tools, and methodologies in the field of digital forensic investigation for social media-related cybersecurity cases. This literature review helps inform the research questions and objectives of the study.

Following the literature review, the study identifies and selects digital forensic investigation tools specifically designed for analyzing social media-related evidence in cybersecurity cases. These tools are then applied to relevant datasets or case studies that involve various social media-related cybersecurity incidents. The evaluation process assesses the tools' effectiveness in extracting, analyzing, and preserving social media-related evidence.

The results obtained from the tool evaluation are analyzed to identify patterns, trends, and gaps in the capabilities of the tools. This analysis forms the basis for the requirement analysis, where specific requirements and improvements needed in digital forensic investigation tools for social media-related cases are identified.

The approach followed in this study is closely aligned with the research questions and objectives. It allows for a systematic evaluation of existing tools, an analysis of their strengths and limitations, and the identification of areas for improvement. By combining literature review, tool evaluation, data analysis, and requirement analysis, the study provides a comprehensive assessment of digital forensic investigation tools for social media-related cybersecurity cases.

3. Literature Review

The literature review conducted for this study aimed to explore existing research, publications, and articles related to digital forensic investigation tools for cases involving social media and cybersecurity. The review sought to establish a foundation of knowledge and identify key challenges, tools, and methodologies in the field. The following sources were consulted:

1- Casey, E. (2018). Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet. Academic Press.

Casey's book provides a comprehensive overview of digital evidence and computer crime, including chapters dedicated to social media forensics. It covers various aspects of digital forensic investigations, including evidence acquisition, analysis, and preservation, with a focus on the challenges and complexities associated with social media platforms.

2- Marrington, A., & Clark, D. (2019). Digital forensics: Challenges and future research directions. Digital Investigation, 29, S88-S96.

This research article discusses the challenges faced in digital forensics, including those specific to social media investigations.

It explores the need for advanced tools and techniques to handle the vast amounts of data generated by social media platforms and identifies areas for future research and development.

4. Digital Forensics

Digital forensics is the practice of collecting, analyzing, and preserving digital evidence to investigate and respond to cybersecurity incidents. It involves the application of forensic techniques and tools to examine digital systems, networks, and data in order to identify the source of an attack, assess the extent of the compromise, and gather evidence for legal or disciplinary actions, Examples of tools used (Forensic Toolkit, Cellebrite UFED, DEFT, OpenPuff).

4.1. Digital forensics and its role in cybersecurity

The role of digital forensics in cybersecurity is multifaceted and vital to effectively respond to and mitigate cyber threats. Here are some key aspects of its role:

1- Incident response: Digital forensics is an integral part of incident response procedures. When a cybersecurity incident occurs, such as a data breach or a network intrusion, digital forensic techniques are used to identify the attack vector, determine the scope of the compromise, and gather evidence to understand the attacker's methods and motives. This information helps organizations contain the incident, remediate vulnerabilities, and prevent future attacks (Casey, 2011).

2- Attribution and threat intelligence: Digital forensics can aid in attributing cyber-attacks to specific individuals, groups, or nation-states. By analyzing the digital evidence left behind, such as malware, network logs, or communication traces, forensic experts can piece together the tactics, techniques, and procedures (TTPs) employed by threat actors. This intelligence can be shared with relevant authorities, security organizations, or the cybersecurity community to enhance overall defense capabilities (Rogers, 2017).

3- Malware analysis: Digital forensics plays a crucial role in analyzing malware, which is often used in cyber-attacks. Forensic investigators examine the behavior, code, and impact of malicious software to understand its functionality, origins, and potential vulnerabilities it exploits. This knowledge helps in developing countermeasures, detecting similar malware variants, and improving overall cybersecurity defenses (Nelson, 2018).

4- Data breach investigations: In the event of a data breach, digital forensics helps uncover how the breach occurred, what data was compromised, and the extent of the damage. Forensic experts analyze system logs, network traffic, and compromised devices to identify the attack vector and determine whether any sensitive data was accessed, exfiltrated, or altered. This information is crucial for organizations to comply with data breach notification requirements, assess the impact on affected individuals, and implement appropriate remediation measures.

5- Legal and regulatory compliance: Digital forensics plays a crucial role in legal proceedings related to cybercrimes. The evidence collected and analyzed through digital forensic techniques can be presented in court to support legal actions against cybercriminals. Additionally, organizations may use digital forensics to demonstrate compliance with industry regulations and legal obligations related to cybersecurity, data protection, and incident response (SANS Institute, n.d).

4.2. Social media digital forensics tools and its mechanism

The process of forensic analysis of social media is a complex process and requires a lot of understanding and tracking. There are auxiliary software tools based on special software such as Python. The forensic investigation process on social media goes through several stages, the most important of which are:

- Data Collection: The process of collecting data through activity log on the digital communication platform and collecting IP addresses, social media platform APIs (e.g., Facebook Graph API, Twitter API, Instagram API) to retrieve user data, posts, comments, and other relevant information. And Implement web scraping techniques to collect information when APIs are not available or insufficient such as (Scrapy, Selenium, Apify) this platform used for web scraping or Implement new specfic web scraping tools using AI programming language.

- Data Preservation: Ensure that collected data remains unchanged and tamper-proof.

- Data Analysis: 1- Text analysis using natural language processing (NLP) to analyze text data, including sentiment analysis, topic modeling, and keyword extraction, 2- Image and video analysis Extract metadata from multimedia files, such as geolocation, timestamps, and camera information, and using Tools to verify the reliability of images or video, not fake or tampered or made with artificial intelligence. 3- User profiling: Create detailed profiles of social media users,

including their connections, interests, and behavior patterns. 4- Network analysis to monitor network devices, servers, and cloud resources, there are many tools using network analysis such as (Cacti: uses SNMP to collect and display network data in graphs, Snort is an open-source intrusion detection system (IDS) and intrusion prevention system (IPS), Angry IP Scanner: Angry IP Scanner is a lightweight, cross-platform network scanner that quickly scans IP addresses and ports to discover devices on a network). 5- Analyze Hashtag the use and popularity of hashtags, and first user who create hashtag and its activity and social manner.

- Geolocation and Timestamp Analysis: Geotagging: Determine the geographical origin of posts and track the movement of users, Timestamp analysis: Establish the timeline of events based on post timestamps.

- Metadata Extraction**:** Extract metadata from images, videos, and posts to understand the context and authenticity of the content. b. Verify image authenticity by checking for digital manipulation, Identify unusual behavior patterns, such as bot accounts, fake profiles, or abnormal posting activity.

- Creating interactive visualizations information map to track a digital forensic investigation.

- Adhere to relevant legal and ethical standards in data collection and analysis, and maintain user's data privacy.

These are some of the mechanism steps for digital investigation on social media, with many tools and software used that can be obtained or developed.

4.3. The importance of digital forensics in investigating cybercrimes

Digital forensics plays a crucial role in investigating cybercrimes by providing valuable insights into the nature of the crime, identifying perpetrators, collecting evidence, and supporting legal proceedings. Here are some key points highlighting the importance of digital forensics in investigating cybercrimes (Smith, 2019):

1. Evidence Collection: Digital forensics enables the collection and preservation of digital evidence from various sources such as computers, mobile devices, networks, and cloud storage. This evidence can include log files, emails, chat conversations, deleted files, and system artifacts. By employing specialized tools and techniques, digital forensic investigators can

extract and analyze this evidence, which is vital for identifying the methods and motives of cybercriminals.

2. Attribution: Digital forensics aids in attributing cybercrimes to specific individuals, groups, or organizations. By analyzing digital evidence, investigators can trace the origin of an attack, track the activities of the perpetrators, and uncover their identities. This attribution is critical for holding cybercriminals accountable and deterring future attacks.

3. Incident Response: Digital forensics is an integral part of incident response efforts. When a cybercrime occurs, such as a data breach or network intrusion, digital forensic techniques help identify the point of entry, the actions taken by the attacker, and the extent of the compromise. This information guides the response team in containing the incident, mitigating the damage, and implementing measures to prevent similar incidents in the future.

4. Legal Proceedings: Digital forensic evidence is admissible in legal proceedings, providing solid proof of cybercrimes. It supports law enforcement agencies, prosecutors, and legal professionals in building a strong case against cybercriminals. Digital evidence, backed by forensic analysis, can establish a clear chain of custody, demonstrate intent, and prove the actions taken by the perpetrator, enhancing the chances of successful prosecution.

5. Prevention and Future Protection: Digital forensics not only investigates cybercrimes but also contributes to preventing future incidents. By analyzing the methods and vulnerabilities exploited by cybercriminals, forensic experts provide valuable insights to organizations and cybersecurity professionals. This knowledge helps in strengthening security measures, implementing effective controls, and developing proactive strategies to mitigate future threats.

4.4. Social media forensics

The development of Web 2.0 and Industry 4.0 technologies has greatly boosted social media platform acceptance, making it a main source of social interaction. Through these websites, users voluntarily exchange their information, set up accounts, and participate in social activities. As a result, hackers are presented with numerous chances to abuse user accounts (Wazid et al., 2013).

In addition, different social media applications like LinkedIn, Instagram, Facebook, and Twitter have been exposed to multiple cyber threats and malware. Attacks on social media platforms can take place outside the system/network or within the network. Outside systems attack usually

include DDoS, or DoS, while attacks within the network include retrieving cookies data (Sharma et al., 2019).

Furthermore, it is known that these social media programs' databases are most susceptible to such attacks. Due to this circumstance, digital investigators are now more interested in social media forensics. Investigators can use social media posts as excellent evidence in criminal investigations because to social media forensics (See Figure 1). Social media networks, which are the best for profiling, are also an ideal source of information about possible offenders, suspects, and witnesses (Rocha, 2016).

In addition, by combining social media with digital forensics, investigators can gain access to a modern and diverse subset of sources of data, including demographic location, photographs, contact lists, geo-location, and text messages. This network data, combined with the metadata, has the potential to assist digital forensics investigations. Furthermore, the metadata can also be used to authenticate online social networking facts. Thus, it can be contended that social media forensics is a rising trend in the digital forensics' domain due to its ability to efficiently providing adequate digital evidence. The advent of social media apps on a mass of platforms has enabled these networking domains to leave digital forensic trace or artifacts that can be of a valuable asset in an investigation. For instance, research like discovered that the chat logs could be extracted from social media applications like Facebook and a huge amount of digital forensic artifacts, such as pictures, location data, friends, posts, passwords, and usernames are left behind as potential evidence. These artifacts are essential evidence, which makes social media forensics as one of the most prominent digital forensic trends (Baggili & Breitinger, 2015).

Studies like these forensically analyzed social media apps on Android, iPhone, and Blackberry devices, including MySpace, Twitter, and Facebook. According to the study, it was successful to recover digital forensic artifacts such user data in text format, timestamps, passwords, URLs, and written comments. This shows that social media forensics is a highly effective method for assessing, verifying, and obtaining digital evidence. It is also a potent tool for tracking down digital evidence dispersed over social media (Al Mutawa, 2012).

Additionally, the three functionalities that social media forensics offer are reverse search integration, tempering localization analysis, and metadata visualization and extraction (Zampoglou et al., 2016).

The first advantage of Google Image Search is that it opens a new tab in your web browser to see the results. Second, it has six separate tampering localization maps designed to locate various signs of social media manipulation. These maps were produced utilizing forensic algorithms. Thirdly, it fully supports metadata listing and shows any potential embedded thumbnails. Professional forensic investigators can go deeper into the data and retrieve pertinent evidence thanks to these characteristics. As a result, the area of digital forensics is seeing a rise in the use of social media forensics.

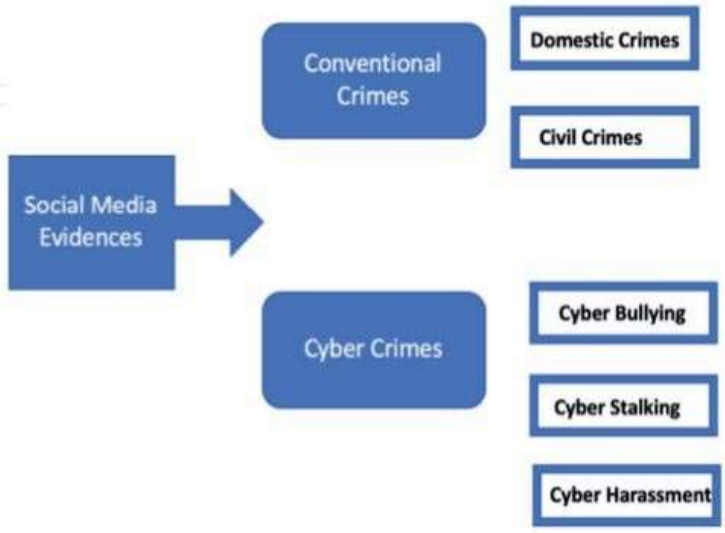

Figure 1

5. Case Studies

Russian Interference in the 2016 US Presidential Election[1]

The investigation into Russian interference in the 2016 US Presidential Election involved extensive digital forensics analysis. The investigation, headed by Special Counsel Robert S. Mueller III, aimed to uncover and document attempts by Russian entities to influence the election through social media manipulation and cyberattacks.

Digital forensic investigators utilized various techniques to identify and analyze online activities related to the Russian interference campaign. They collected and analyzed vast amounts of digital evidence, including social media posts, advertisements, email communications, and internet traffic data.

[1] United States Department of Justice (https://www.justice.gov/sco)

Through the examination of metadata, user profiles, and network connections, investigators traced the origin of deceptive social media campaigns and identified the individuals and organizations responsible for orchestrating the interference. Sophisticated forensic tools were employed to link the digital activities to specific Russian entities and track the dissemination of misinformation and propaganda.

The digital forensic investigation provided critical evidence of Russian involvement in the election interference. The findings were documented in the Mueller Report, which served as a foundation for legal and policy actions to safeguard future elections and protect against foreign influence campaigns.

Methodologies (Mueller, 2019):

1. Social Media Analytics: Investigators utilized social media analytics techniques to analyze large volumes of social media posts, advertisements, and user interactions. This involved mining data from various social media platforms and employing natural language processing and sentiment analysis to identify patterns and trends in the content shared by Russian entities.

2. Metadata Analysis: Digital forensic investigators examined metadata associated with emails, social media posts, and other digital artifacts to establish the origins, timestamps, and user identities. Metadata analysis helps in corroborating the authenticity and integrity of digital evidence.

3. Network Traffic Monitoring: Investigators monitored network traffic, including internet traffic from suspicious IP addresses and connections, to identify patterns and trace the origins of cyberattacks and intrusion attempts. Network traffic monitoring helps in understanding the techniques and infrastructure used by threat actors.

Tools:

1. Social Media Monitoring Tools: Investigators utilized specialized tools for social media monitoring and analytics, such as sentiment analysis platforms, social media listening tools, and content analysis software. These tools helped in identifying coordinated campaigns, tracking influential posts, and analyzing the reach and impact of social media content.

2. Email Forensic Tools: Investigators employed email forensic tools to analyze email headers, extract metadata, and trace the path of emails. These tools assisted in identifying suspicious

email communications and linking them to specific individuals or organizations involved in the interference campaign.

3. Network Analysis Tools: Investigators utilized network analysis tools to monitor and analyze network traffic, identify suspicious connections, and track the flow of information between various entities. These tools helped in tracing the origin of cyberattacks and mapping the infrastructure used by threat actors.

In both case studies, a combination of specialized methodologies and tools was employed to gather digital evidence, analyze data, and establish links between individuals, activities, and criminal operations. These methodologies and tools played a crucial role in the successful digital forensic investigations of social media and cybersecurity incidents

7. Conclusion

In conclusion, digital forensic investigation tools play a vital role in cases related to social media and cybersecurity incidents. These tools enable investigators to collect, analyze, and interpret digital evidence, helping uncover the truth, identify perpetrators, and support legal actions .

For social media investigations, specialized tools for social media analytics and monitoring are crucial. These tools assist in mining large volumes of social media data, detecting patterns, and identifying coordinated campaigns or malicious activities. Social media monitoring tools enable investigators to track user interactions, sentiment analysis, and content dissemination, providing insights into the motives and tactics of individuals or groups involved in social media-based crimes.

In cybersecurity investigations, digital forensic tools are essential for analyzing network traffic, identifying malicious activities, and tracing the origins of cyberattacks. Network analysis tools help investigators understand the techniques used by threat actors, map their infrastructure, and identify vulnerabilities exploited for illegal activities. Additionally, email forensic tools and metadata analysis play a significant role in tracing the origins of communications, verifying authenticity, and establishing the chain of custody for digital evidence.

Blockchain analysis tools are crucial when investigating cases involving cryptocurrencies, such as money laundering or illegal marketplaces. These tools help trace transactions, analyze

blockchain data, and identify the flow of funds, assisting in linking financial activities to individuals or organizations involved in illicit operations.

Collaboration between law enforcement agencies, cybersecurity experts, and social media platforms is essential in leveraging these tools effectively. Sharing expertise, accessing specialized intelligence platforms, and collaborating with platform administrators enhances the investigation process and strengthens the evidence collected.

It is important to note that the field of digital forensics is constantly evolving, and new tools and techniques continue to emerge. Investigators must stay updated with the latest advancements and adapt their methodologies and tools accordingly to effectively tackle the ever-changing landscape of social media and cybersecurity incidents.

In conclusion, the successful utilization of digital forensic investigation tools is crucial in social media and cybersecurity cases. These tools enable investigators to navigate complex digital environments, collect evidence, and uncover the truth, ultimately contributing to the prevention and prosecution of social media-related crimes and cybersecurity incidents.

8. Results:

The study on digital forensic investigation tools for cases related to social media and cybersecurity has yielded several key results:

1. Specialized tools for social media analytics and monitoring play a crucial role in mining and analyzing large volumes of social media data, identifying patterns, and detecting coordinated campaigns or malicious activities.

2. Network analysis tools are essential in cybersecurity investigations, enabling investigators to analyze network traffic, trace the origins of cyberattacks, and map the infrastructure used by threat actors.

3. Blockchain analysis tools are valuable in cases involving cryptocurrencies, assisting in tracing transactions, analyzing blockchain data, and linking financial activities to individuals or organizations involved in illicit operations.

4. Social media plays an important role in influencing public opinion and raising social problems, which requires the presence of official or reliable bodies specialized in analyzing cyber behavior on these platforms.

9. Recommendations:

Based on the study's findings, the following recommendations are proposed:

- Investing in continuous research and development of new tools and techniques in digital forensics.

- Enabling professionals and investigators to access the most advanced and effective tools for social media and cybersecurity investigations.

- Train and develop the skills of cybersecurity professionals, enhance their investigative capabilities, and improve the overall effectiveness of social media and cybersecurity investigations.

- Strengthening mechanisms for cooperation and information exchange, enhancing the investigation process and improving the outcomes of social media and cybersecurity cases

- Educating the public about the risks associated with social media and cybersecurity incidents and providing guidance on safe online practices

10. References:

1. E. A. Gollub, (2013). "Recent trends in digital text forensics and its evaluation," In International Conference of the Cross Language Evaluation Forum for European Languages, pp. 282-302, September.
2. A. Aminnezhad and A. Dehghantanha, (2014). "A survey on privacy issues in digital forensics," nternational Journal of Cyber-Security and Digital Forensics (IJCSDF), vol. 3, no. 4, pp. 183-199.
3. F. Dezfouli and A. Dehghantanha, (2014). "Digital forensics trends and future," International Journal of Cyber-Security and Digital Forensics (IJCSDF), vol. 3, no. 4, pp. 183-199.
4. B. K. Sharma, M. A. Joseph, B. Jacob and L. C. B. Miranda, (2019). "Emerging trends in Digital Forensic and Cyber security-An Overview," In 2019 Sixth HCT Information Technology Trends (ITT), pp. 309-313, November.
5. M. Wazid, A. Katal, R. H. Goudar and S. Rao, (2013). "Hacktivism trends, digital forensic tools and challenges: A survey," n 2013 IEEE Conference on Information & Communication Technologies, pp. 138- 144, April.
6. Casey, E. (2011). Digital evidence and computer crime: Forensic science, computers, and the internet. Academic Press.

7. Rogers, M. K., Goldman, J., & Mislan, R. (2017). Investigating digital crime. Academic Press
8. Nelson, B., Phillips, A., & Steuart, C. (2018). Guide to computer forensics and investigations. Cengage Learning.
9. SANS Institute. (n.d.). Computer Forensics, Investigation Techniques, and Incident Response. Retrieved from https://www.sans.org/reading-room/whitepapers/incident/computer-forensics-investigation-techniques-incident-response-86
10. Smith, R. (2019). The Importance of Digital Forensics in Investigating Cybercrimes. International Journal of Cybersecurity and Digital Forensics, 8(2), 45-58.
11. M. Wazid, A. Katal, R. H. Goudar and S. Rao, (2013). "Hacktivism trends, digital forensic tools and challenges: A survey," n 2013 IEEE Conference on Information & Communication Technologies, pp. 138- 144, April.
12. A. E. A. Rocha, (2016). "Authorship attribution for social media forensics," IEEE Transactions on Information Forensics and Security, vol. 12, no. 1, pp. 5-33.
13. I. Baggili and F. Breitinger, (2015). "Data sources for advancing cyber forensics: what the social world has to offer.," n 2015 AAAI Spring Symposium Series., March
14. N. Al Mutawa, I. Baggili and A. Marrington, (2012). "Forensic analysis of social networking applications on mobile devices," Digital Investigation, vol. 9, pp. S24-S33.
15. M. Zampoglou, S. Papadopoulos, Y. Kompatsiaris, R. Bouwmeester and J. Spangenberg, (2016). "Web and social media image forensics for news professionals.," In Tenth international AAAI conference on web and social media, April
16. United States Department of Justice (https://www.justice.gov/sco)
17. Mueller, R. S., III. (2019). Report on the investigation into Russian interference in the 2016 presidential election. United States Department of Justice. Retrieved from
18. Casey, E. (2018). Digital Evidence and Computer Crime: Forensic Science, Computers, and the Internet. Academic Press
19. Marrington, A., & Clark, D. (2019). Digital forensics: Challenges and future research directions. Digital Investigation, 29, S88-S96

Doi: https://doi.org/10.52132/Ajrsp.e.2023.55.1

Academic Journal of Research and Scientific Publishing | Vol 5 | Issue 55
Publication Date: 05-11-2023 ISSN: 2706-6495

Diabetes Detection without Visiting the Clinic (A Machine Learning Approach)

Bahatheq Tariq Ahmed S
Undergraduate, Computer Science, National University of Singapore, Singapore
Email: tariqbahatheq@outlook.com

Venkat Vishwanth Sreenivasan
Undergraduate, Computer Science, National University of Singapore, Singapore
Email: vishwanth.s22@u.nus.edu

https://colab.research.google.com/drive/1KBDNEQZNKeiISultOQSOcTMTJVq50C65

Abstract:

The research presented in this paper focuses on the application of machine learning techniques for early detection of diabetes, without the need for clinic-dependent data. Utilizing a dataset of 253,680 examples sourced from the Behavioral Risk Factor Surveillance System (BRFSS), the study employs a variety of machine learning models, including Decision Tree, Random Forest, XGBoost, Neural Networks, SVM, and Naive Bayes. The paper highlights the significance of early diabetes detection and the potential of machine learning in making this process more accessible and efficient. The dataset underwent extensive preprocessing, including under-sampling to address imbalance and feature engineering to enhance model performance. The paper meticulously discusses the employed preprocessing techniques, providing insights into the importance of handling data imbalance and feature selection in machine learning applications for healthcare. The neural network model emerged as the top-performing model, achieving an accuracy of 88.76%. This result underscores the potential of machine learning in diabetes detection. We believe that this is fruitful as most people will avoid visiting the clinic to check for diabetes because of costs and loss of time. In conclusion, whilst we believe that this approach is beneficial, we suggest that this model only to be used as a possible indicator with the need to visit the doctor to fully confirm the presence of diabetes.

Keywords: Diabetes Detection, Machine Learning, Neural Networks, Feature Engineering, Data Preprocessing, Responsible AI, Transparency, Model Evaluation, Early Detection.

1. Introduction

Diabetes is a chronic health condition that affects how the body converts food into energy. It is a significant public health concern, with nearly half a billion people worldwide suffering from it. In Singapore, one in three individuals is at risk of developing diabetes, and it is estimated that about one million Singaporeans will be living with the condition by 2050. (1)

For those affected, the implications can be severe, including a lifetime of daily medication or injections, and, in some cases, blindness, amputation, kidney dialysis, and premature death. (1)

Early detection of diabetes can help save patients' health by enabling them to take medications and adopt routines that mitigate the condition's effects. Therefore, we have decided to employ machine learning to detect diabetes as the target variable in individuals based on commonly known factors, eliminating the need for a clinic visit.

2. The Dataset

Selected Dataset: The chosen dataset is sourced from the Kaggle website (2) and classifies individuals as having diabetes or not based on 21 features. It comprises 253,680 examples, with 218,334 of the examples representing people without diabetes and 35,346 examples representing those with the condition. The dataset was collected from the Behavioral Risk Factor Surveillance System (BRFSS), an annual health-related telephone survey conducted by the CDC. The data is anonymized for privacy reasons, and the 2015 dataset was used for this project. Table 1 below provides descriptions of the 21 features.

Alternative Datasets: We evaluated three potential datasets. The second dataset included pre-diabetes as a third classification, while the third dataset was under-sampled based on the target variable. We selected our dataset because the second one was highly imbalanced, and we preferred to perform under-sampling that better suited our dataset instead of the third option.

HighBP	HighChol	CholCheck	BMI	Smoke	Stroke	HeartDisease
High Blood Pressure	High Cholesterol	cholesterol check in 5 years	Body Mass Index	Smoked at least 100 cigarettes in your entire life?	had a stroke.	had/have (CHD) or (MI)

PhysActivity	Fruits	Veggies	Alcohol	Healthcare	NoDocbcCost	GenHlth
physical activity in the past 30 days - not including job	Consume Fruit 1 or more times per day	Consume Vegetables 1 or more times per day	Heavy drinkers	Any kind of health care coverage	Couldn't see a doctor in the past 12 months due to cost	General health: scale of 1-5 1 = excellent 5 = poor

MentHlth	PhysHlth	DiffWalk	Sex	Age	Education	Income
Bad mental health during the past 30 days	Bad physical health during the past 30 days	difficulty walking or climbing stairs	0 = female 1 = male	13-level age category	Education level scale 1-6	Income scale 1-8

Table 1: Dataset features with descriptions

Feature Selection: We removed the high blood pressure and high cholesterol features from the dataset to make the model accessible for anyone from their home, without the need for clinic visits, as these features require lab test confirmation. While this decision might affect the model's accuracy, we believe it is worthwhile as it increases the model's accessibility and usefulness for individuals. People can then use the model as an indicator of potential diabetes risk, prompting them to seek treatment before the condition worsens.

3. Data Preprocessing

3.1. Data Cleaning

The dataset selected has been pre-cleaned and has many features removed. Therefore, there are no missing data such as null values to consider and fix. However, the dataset is highly imbalanced where the minority class is diabetes label.

To overcome this imbalance, under-sampling was used. Under-sampling reduces the number of data points in the majority class till the majority class is approximately the same size as the minority class.

To be precise, NearMiss Algorithm was used to carry out the under-sampling by randomly removing data points in the majority class that are close to each other. Moreover, through this method, information loss is minimized as there still exists a data point that is quite similar to the one removed.

This resulted in a dataset of 70,692 entries with 35346 entries with diabetes label and 35346 entries with No Diabetes label.

3.2. Bias

The provided dataset was pre-reduced from 441,455 entries with 330 features, from the BRFSS, to 253,680 entries with 21 features. The selection criteria used to reduce the features might have been subjected to human bias. For example, in Representative Heuristic, one might incorrectly assume that BMI might be an important feature to judge diabetes based on stereotypes. However, it is not always true. As some might have obtained diabetes through genetics.

4. Data Visualization

Feature-Target Relationship: We visualized the relationship between each feature and the target variable using bar plots, as shown in Figure 1. Some correlations are readily apparent, such as the connection between BMI and diabetes, where both extremely low and high values are linked to the condition. General, physical, and mental health also show a strong association with diabetes, indicating that an unhealthy lifestyle is generally correlated with the disease. Furthermore, health conditions such as strokes, heart disease, or difficulty walking are also related to diabetes. We use these correlations to engineer features that will enhance the accuracy of our models.

Mutual Information Score (MI): A well-established metric for identifying relational information is the Mutual Information score. The advantages of MI include its ability to detect any relationship type (whereas correlation only detects linear relationships), ease of use and interpretation, computational efficiency, robust theoretical foundation, and resistance to overfitting[3]. The mutual information scores for the features are presented in Figure 2.

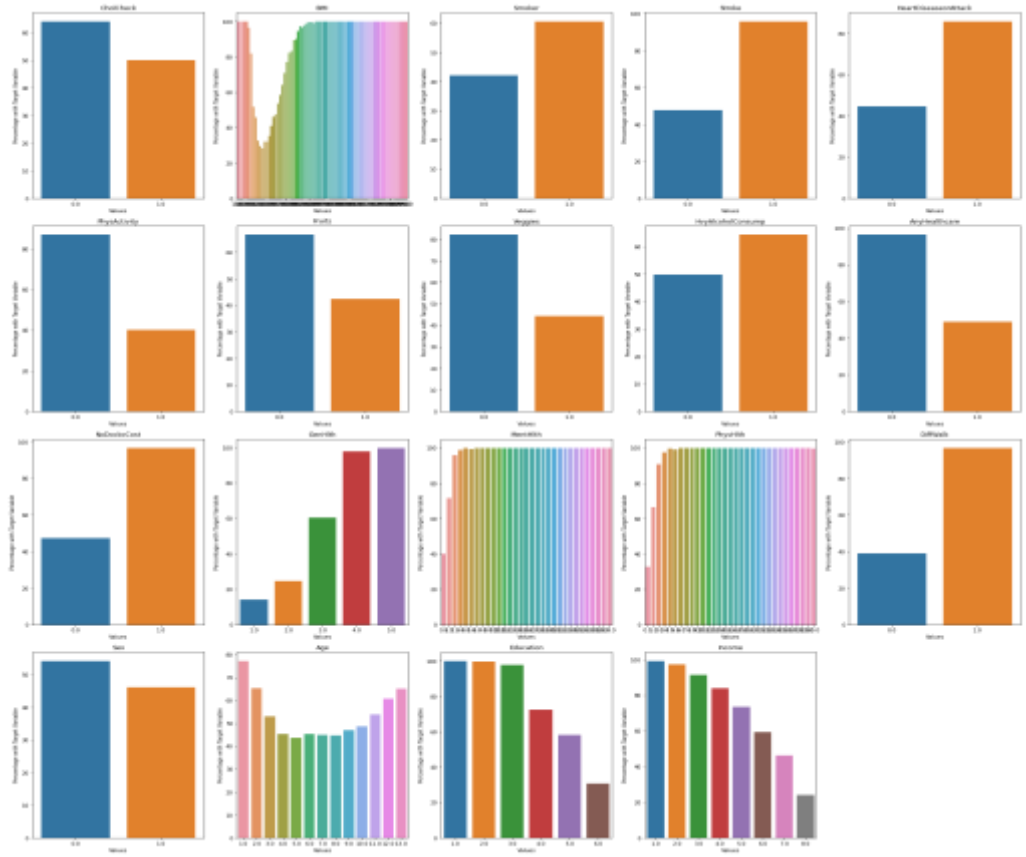

Figure 1: The bar plots show the percentages of people that have the target variable based on each value of the feature.

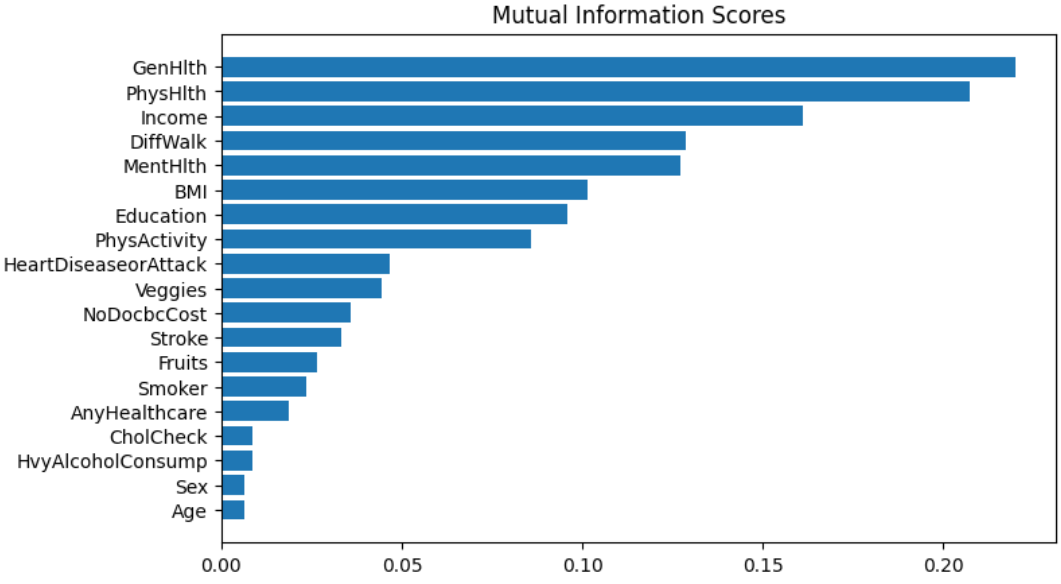

Figure 2: Mutual Information scores

5. Feature Engineering

Based on the relational information gathered from Figures 1 and 2, we experimented with combining new features that fall under categories such as physical health and physical activity, healthy lifestyle, diet, etc. We also explored exponentiating some features to examine potential non-linear relationships. The features we added are shown in Table 2:

PhysHA	PhysHlth + PhysActivity
GenAge	GenHlth * Age
class	Income * Education
BMIwalk	BMI * DiffWalk

Table 2: Feature engineering

Additionally, we experimented with removing some features based on their MI scores. However, this approach yielded slightly worse results. Therefore, we retained the original set of features along with the engineered ones to improve our model's performance. By incorporating these engineered features, we aimed to enhance the accuracy and predictive power of our machine-learning models.

Moreover, we split the dataset into 3 segments, Training, cross-validation, and Testing. This helps to overcome overfitting issues.

6. Model Selection and Performance Metrics

6.1. Model Selection

The target variable used for prediction is categorical. Hence, classification algorithms are required for prediction. We selected the most popular algorithms to determine the best algorithm that has the highest prediction accuracy.

6.2. Performance Metrics

To judge the performance of the classification algorithms, we opted to use Accuracy. Accuracy indicates the ratio of labels correctly predicted. Due to the nature of the dataset selected being balanced, Accuracy performs exceptionally well in this instance. Furthermore, this metric allowed us to better fine-tune the algorithm. A high accuracy would indicate that the model has a higher chance of predicting the label correctly and hence have good performance.

7. Models

7.1 Decision trees

The Model: Decision trees are reliable models for classification tasks such as this one. Furthermore, they are considered white-box models due to their interpretability.

Performance: Utilizing the default parameters for the decision tree from the sklearn library, we achieved an accuracy score of 83.60% on the validation dataset. This serves as a solid foundation for our model. However, further tuning is possible. The most critical parameters affecting decision trees are the error function and maximum leaf nodes. Automatically tuning the model on these attributes using the validation set yields an accuracy score of 86.28%. After feature engineering, the model delivers a score of 86.62%. The testing or generalization accuracy is 86.81%.

Limitations and Advantages: Unfortunately, decision trees are sensitive to changes in data, as slight alterations in the data can modify the tree's structure. On the other hand, decision trees are interpretable and computationally efficient.

7.2. Random Forest

The Model: The Random Forest model effectively compensates for the sensitivity of decision trees. As a tree ensemble, it uses sampling with replacement when constructing each tree, making it more robust.

Performance: Using the default parameters for the Random Forest from the sklearn library, we achieved an accuracy score of 86.00% on the validation dataset, which is already better than the default decision tree. Nonetheless, further tuning is possible. The most critical parameters affecting random forests are the error function, maximum leaf nodes, tree depth, and tree count. Automatically tuning the model on these attributes using the validation set yields an accuracy score of 88.98%. After feature engineering, the model delivers a score of 88.86%. The testing or generalization accuracy is 88.59% using the random forest without feature engineering as the validation accuracy was better.

Limitations and Advantages: Although random forests are more robust, there is still a chance of misrepresenting the dataset during random sampling with replacement, as some data might never be selected.

7.3. XGBoost

The model: XGBoost stands for "Extreme Gradient Boosting".(3) XGBoost is a decision tree ensemble that works together to compute the final prediction by summing up the predictions of multiple trees. Moreover, through the gradient boosting ensemble technique, weak prediction models, such as decision trees, provide a good overall prediction model. (4)

Performance: Using the default parameter settings, the model was able to give an accuracy of 89.00%. Hyperparameter tuning improved this accuracy to 89.16% on the validation data. Feature engineering further improved this to 89.19%. The generalization accuracy is 88.55%. The tuned parameters are shown in Table 3. (5)

Name	Best Value	Description
Colsample_bytree	0.3256299	Specify the fraction of columns to subsample when constructing each tree
Eta	0.8240329	Shrinks the feature weights to make boosting more conservative and prevent overfitting
gamma	8.3293339	The minimum split loss required to make a partition on a leaf node of the tree. A lower value makes the model less conservative (More likely to have False Positive and identify more True Positive predictions)
Max_depth	3789.0	Maximum depth of the tree. Increased depth makes the model more complex and likely to overfit
Min_child_weight	0.0	Determines the partitioning of the leaf node by comparing its minimum sum of instance weight. A lower value makes the model less conservative.
Reg_alpha	0.6574576	L1 regularization term on weights
Reg_lambda	7.2628099	L2 regularization term on weights

Table 3: XGBoost parameters

Limitations and Advantages: XGBoost excels in classification problems, however, it is not good at regression problems due to the use of decision trees that underperform with continuous inputs. An advantage is the ensemble technique used which allows individual models to help correct each other as opposed to all models training in isolation which might result in them making the same errors giving a better overall prediction.

7.4. Neural Networks

The Model: Neural networks are powerful and versatile models that can be used for a variety of tasks, including classification problems such as this one. They consist of interconnected layers of nodes or neurons that can learn complex patterns and relationships within the data.

Performance: Using a basic feedforward neural network with default parameters, with 3 layers and 16 neurons we trained the model for 100 epochs and obtained an accuracy score of 88.31% on the validation dataset. This provides a strong baseline for our model. However, we can further optimize the architecture and parameters. Key factors that affect neural networks include the number of layers, number of neurons per layer, activation functions, learning rate, and optimization algorithm. Tuning the model on these attributes using the validation set yields an accuracy score of 89.07%. The network architecture is shown in Table 4. The final layer is a linear layer because it avoids computational error. (6) After feature engineering, the model delivers a score of 89.34%. The testing or generalization accuracy is 88.76%.

Optimizer Function	Learning rate	Loss Function	Epochs
Adam	0.0001	Sparse Categorical Cross entropy	100
Layer	Neurons	Activation function	Dropout
L1	32	ReLU	0
L2	64	ReLU	0
L3	128	ReLU	0
L4	50	ReLU	0
L5	25	ReLU	0
L6	12	ReLU	0
L7	7	ReLU	0
L8	2	Linear	0

Table 4: Final neural network architecture

Limitations and Advantages: One major drawback of neural networks is their black-box nature, which makes them difficult to interpret compared to decision trees or random forests. Additionally, they can be more computationally intensive and time-consuming to train, especially for large datasets and complex architectures.

On the other hand, neural networks are capable of modelling complex relationships and can often achieve higher performance than other algorithms when properly tuned and trained. Furthermore, their flexibility allows for various architectures and activation functions to be tailored to the specific problem at hand. Moreover, a significant advantage is that they can be used for transfer learning meaning we can keep adding data to the model continuously without having to retrain the model.

7.5. SVM

The Model: Support Vector Machine (SVM) is well suited for classification problems where it finds a hyperplane in the (Num of features – 1) plane to aid in categorizing the data. (7) Additionally, it is using the Radial Bias Function (RBF) that uses the distance between the data points.

Performance: This model was able to achieve an accuracy of 88.00%. This was further improved to 89.20% by tuning the parameters. With c = 44.1658 and gamma = 74.3120 where c is the regularization parameter and gamma is the kernel coefficient (8) and feature engineering this slightly reduced to 89.19%. The generalization accuracy is 79.26%.

Limitations and Advantages: A huge drawback is the performance of SVM in large datasets. When trained on an initial large dataset, SVM took 50 minutes whereas other models only took a few minutes. Additionally, SVM doesn't perform well if imbalanced data is used as this causes the hyperplane to be closer to the minority class resulting in wrong classification. Some advantages of SVM are its ability to handle high dimensional data as well as its robustness to noise in the data.

7.6. Naive Bayes

The Model: Naive Bayes applies the Bayes theorem while assuming that all features are conditionally independent of each other. Hence, the name naive. Due to feature independence, using the Bayes theorem allows the model to use lesser parameters to calculate and make classifications. (9)

Performance: This model was able to achieve an accuracy of 87.41%. By tuning the hyperparameter, alpha, an additive smoothing parameter to aid in the zero-probability problem. With alpha = 0.033446, this model has an accuracy of 87.48% on validation data. This was slightly reduced to 87.18% using feature engineering. The generalization accuracy of this model is 86.78%.

Limitation: Due to the conditional independence assumption, this model is not suitable to make probability estimations. Moreover, in a dataset that has features that are dependent on each other, this model's assumption would make it a poor choice as a classifier.

Moreover, Naïve Bayes has a zero-probability problem, where any feature not present in the training set will be automatically assigned a probability of zero, which prevents this model from making any predictions regarding that feature. (10)

One advantage is the ability of the model to train and predict fast due to the independence assumption made as well as the simple computations done to make predictions.

8. Conclusion

8.1. Best Model

By comparing the generalized accuracy of the models, Neural Network has the best-generalised accuracy of 88.76%, followed by Random Forest with 88.59%, then XGBoost with 88.55%, then Decision Tree with 86.81%, then Naive Bayes with 86.78% and finally SVM with 79.56%.

8.2. Responsible AI

Privacy: This model ensures privacy through the use of a de-identified dataset for training as well as not collecting personal particulars from the users for prediction. Hence, individual rights to privacy are well preserved.

Transparency: Despite the dataset used to train the Neural Network as well as the code for the neural network being released to the public, Neural Network is a Black Box Algorithm that doesn't provide the user with transparency of the decision that leads to the predictions. However, if transparency is necessary, the Decision Tree can be used as it can provide the actual tree used to make the decision.

Regardless of the transparency provided by the Decision tree, the general public would not be able to understand and use the decision tree to their advantage and hence, we would be trading off transparency for higher accuracy as wrong classification can affect the user more than the user having more transparency.

8.3. Limitations

There are three limitations present in this model. Firstly, due to the use of accuracy as an evaluation metric, the dataset has to be balanced. In the real world, some medical datasets have large

imbalances, such as diabetes. Therefore, training this model to classify such datasets would result in a poor evaluation of the performance.

Secondly, the dataset provided consists of details of Americans only. This results in the model wrongly anchoring on select features to make predictions specific to Americans. Applying this model to the population might result in a higher probability of wrong predictions. For example, BMI would be a good indicator of diabetes for Americans however, in Singapore, sugar intake would be a good indicator of diabetes. Therefore, it is not recommended to generalize the predictions for the sample to the population that consists of people around the world.

Finally, our lack of experience in the medical field prevents us from truly ensuring that this model is trained and predicting diabetes correctly. Having sufficient knowledge can allow us to choose features more accurately and perform better feature engineering. Therefore, this model might not be truly accurate in predicting diabetes and would require the consultation of actual doctors to test the presence of diabetes.

Despite these limitations, we would still use this model to help predict diabetes in addition to visiting a doctor for a second opinion when the model predicts the presence of diabetes.

9. Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

10. Ethical Consideration:

This research prioritizes the ethical standards integral to scientific endeavors, particularly in the realm of medical research. Herein, we outline the key ethical considerations that were adhered to:

1. **No Human Experiments**: At no point did this study involve direct experiments on humans by the authors.
2. **Data Sources**: The data used in this research is sourced from a Kaggle challenge. It's crucial to understand that all data used was anonymized and void of any personal identifiers, ensuring the privacy and confidentiality of the responders.
3. **Patient Consent & Approvals**: While the authors did not directly conduct experiments or collect data, it's implicit that the original data collectors sought necessary consent from responders or their guardians.

4. **Protocols Followed**: The research strictly followed data handling and analysis protocols to ensure the integrity of the results. Furthermore, while the models and findings show promise, it's crucial to emphasize their supplementary role in medical diagnosis. Decisions based on these findings should be made with caution, in tandem with expert judgment.
5. **Transparency & Openness**: The research aims to contribute to the broader scientific community. As such, efforts have been made to ensure transparency in methodology, findings, and potential limitations. This open approach facilitates peer review and collective advancements in the field.

By adhering to these principles, this research aims to be both scientifically rigorous and ethically responsible, ensuring that advancements made contribute positively to patient care and the broader medical community.

11. References

1. MOH | News Highlights [Internet]. [cited 2023 Apr 7]. Available from: https://www.moh.gov.sg/news-highlights/details/speech-by-mr-ong-ye-kung-minister-for-health-at-world-diabetes-day-2021#:~:text=Locally%2C%20one%20in%20three%20individuals,will%20be%20living%20with%20diabetes
2. Diabetes Health Indicators Dataset | Kaggle [Internet]. [cited 2023 Apr 7]. Available from: https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset
3. Introduction to Boosted Trees — xgboost 1.7.5 documentation [Internet]. [cited 2023 Apr 6]. Available from: https://xgboost.readthedocs.io/en/stable/tutorials/model.html
4. Gradient boosting. In: Wikipedia [Internet]. 2023 [cited 2023 Apr 6]. Available from: https://en.wikipedia.org/w/index.php?title=Gradient_boosting&oldid=1146274757
5. XGBoost Parameters — xgboost 2.0.0-dev documentation [Internet]. [cited 2023 Apr 6]. Available from: https://xgboost.readthedocs.io/en/latest/parameter.html#parameters-for-tree-booster
6. Google Colaboratory [Internet]. [cited 2023 Apr 7]. Available from: https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/tutorials/quickstart/beginner.ipynb#scrollTo=he5u_okAYS4a&line=1&uniqifier=1
7. 1.4. Support Vector Machines [Internet]. scikit-learn. [cited 2023 Apr 7]. Available from: https://scikit-learn/stable/modules/svm.html

8. sklearn.svm.SVC [Internet]. scikit-learn. [cited 2023 Apr 7]. Available from: https://scikit-learn/stable/modules/generated/sklearn.svm.SVC.html
9. 1.9. Naive Bayes [Internet]. scikit-learn. [cited 2023 Apr 7]. Available from: https://scikit-learn/stable/modules/naive_bayes.html
10. Jayaswal V. Laplace smoothing in Naïve Bayes algorithm [Internet]. Medium. 2020 [cited 2023 Apr 7]. Available from: https://towardsdatascience.com/laplace-smoothing-in-na%C3%AFve-bayes-algorithm-9c237a8bdece

Doi: https://doi.org/10.52132/Ajrsp.e.2023.55.2

Academic Journal of Research and Scientific Publishing | Vol 5 | Issue 55
Publication Date: 05-11-2023 ISSN: 2706-6495

The Effects of listed Price on the Consumer's Online Search and The Optimal Pricing (Online Shoppers in the Saudi Community)

By: Amal Dolem Alsohaimi

Teaching assistant, Department of Economics and Finance, College of Business and Economics in Qassim University, Saudi Arabia

Email: ad.alsuhaimi@qu.edu.sa

Abstract:

Various factors such as disclosure of the item prices on company websites influence consumer responses and activity. This study is a detailed descriptive investigation of the effect that a selected pricing model has on the consumers as they search online for items to purchase. The study is limited to Saudi online shoppers. The research employs a case study approach, where the pricing models of firms such as Uber and Airbnb among others are evaluated. In every particular case study, the maximum price on the prices that appeal to the Saudi community online shoppers was determined. A survey study was executed to collect quantitative and qualitative information from 57 randomly selected Saudi online shoppers using questionnaires. In this case, the independent factor was listed price, whose proxies included the commitment and the non-commitment of the seller and search costs. The dependent elements included consumer online search, optimal pricing, profits, and volume of trade, as influenced by the consumers' chance of visiting. The results of the study indicate that there is a significant positive effect of the listed price on consumer's online search among online shoppers in the Saudi community. The findings also proved that the there is a positive effect of the listed price on optimal price among online shoppers in the Saudi community.

Keywords: Listed Prices, Consumer Search, Optimal Pricing, and Online Shoppers in Saudi Community

1. Introduction

Deciding on a price for a product or service is one of the most important decisions for any organization. Because companies need to cover their costs, it is vital that the price of an item is high enough to cover expenses, but not so high that customers won't be willing to pay for the product.

In many markets, including devices, PCs, and fashions, consumers ordinarily need to visit stores to discover which item they like most. In spite of the fact that elementary data about items sold in these business sectors is generally simple to get either from TV, the Internet, daily papers, specific magazines, or just from neighbors, family, and friends, consumers expression since some applicable item properties are hard to evaluate, print, or promote. In other words, most shoppers take part in a planned action of inquiry for items they want to buy it, since going by stores includes noteworthy search prices (De Los Santos, Hortaçsu, &Wildenbeest, 2017; Honka, 2014; De Los Santos, Hortaçsu, &Wildenbeest, 2012).

In general, consumers prefer the Web, due to the Web offers a place, it offers a more extensive search than what is accessible in physical stores or different channels, and it offers esteem. But, Forrester information indicated that consumers additionally get themselves disappointed from online shopping when its costs are too high. Truth be told, one of the key reasons that customers submission physical shopping because of the cost of buying an item is that transportation expenses are suddenly high (Forrester, 2011).

Moreover, publicized prices are frequently different from conclusive prices in many markets. For instance, web-based shopping, as a rule, includes delivering and shipping, which might be watched simply subsequent to adding an item to a shopping basket or searching for all the important delivery and payment data for items (Dai, 2016). As per Ellison & Ellison (2009), on Pricewatch.com, shipping charges developed to the point that it was normal for firms to list a cost of $1 for a memory module and illuminate consumers of a $40 transporting and shipping at look at. Likewise, a report in the Washington Post archives a case in which one customer expected a $25 ride from Uber, however, a highpoint extra charge prompted a $120 charge.

In a different case, Airbnb postings included $45 benefit charges and $25 cleaning expenses that were not uncovered until the point that well into the booking procedure (Diakopoulos, 2015). These cases pointed the real reasons why most consumers surrender physical shopping for

acquiring an item. A Forrester study found that 44% of Web customers said that they did not finish an online shopping since transportation and its cost was too high (Forrester, 2011).

Hence, in many markets, shoppers cause costs to look/visit a firm, so they seek just in the event that it is justified regardless of the effort. Specifically, customers think about, firstly: what value they pay and secondly: do they get an item. An optimal cost does no use for a thing that is out of stock or an administration that cannot be offered (e.g., no arrangements or seats accessible). What's more, accessibility is not valuable if the cost is too high. So, in these conditions, the firm needs to draw in purchasers with a decent arrangement (cost and accessibility). The firm can do this with two levers: a pricing methodology and a limit decision (Cachon, & Feldman, 2015).

Consumer search models with detectable price have been attracting developing consideration the past studies. The Internet has essentially brought down the cost of gathering valuable data. Presently it is regular to check costs on the web and visit stores just to know all the data of the items and additionally finish a buy. Meanwhile, the model catches some notable highlights of online commercial centers and value correlation sites. A purchaser regularly starts with an outline site page showing different things. She/he clicks a specific arrangement of things, gathers more point-by-point data, and afterward settles on the last buy search (Dai, 2017).

A few consumer search models have been considered in three late papers, Armstrong and Zhou (2011), Shen (2015), and Haan, Moraga-Gonz'alez, &Petrikaite (2015). Each of the three papers investigates asymmetric duopoly condition, however, consider diverse connection structures for purchasers' earlier (known) and coordinate (hidden) values. Both earlier and match values are excellently adversely connected between the items in Armstrong and Zhou (2011), though both are free in Haan, Moraga-Gonz'alez, &Petrikaite (2015). Shen (2015) reviewed a middle of the road situation where every customer earlier values are perfectly poorly agreed, while her match values are free, between the two items.

It is very much perceived that such consumer search models do not accept manageable portrayal. There are two primary troubles. To begin with, the buyer seeks conduct is convoluted and difficult to abridge. Every purchaser experiences successive pursuit, whose multifaceted nature develops quickly as the number of sellers increases or new highlights are brought into the model. This is probably going to be the motivation behind why every past examination has limited thoughtfulness regarding the duopoly case.

Second, the wholesalers' best reaction functions do not carry on well when all is said in done. There may not exist a pure technique balance, and the model once in a while delivers precisely similar statics comes about (Dai, 2017).

1.1. Research statement:

Determining the price of a product or service is one of the most important decisions for any organization, especially if the product is promoted electronically and sold online. Since companies need to cover their costs, it is essential that the price of the commodity is high enough to cover expenses, but not so high that customers do not want to pay for the product. The best strategies for pricing a product or project, and using competitive intelligence from joint market research reports to determine the success of competitors and their failure to help determine pricing strategies that attract or simply discourage customers as a result of their own research should be identified. The problem of the study was the result of the existence of a number of retailers who earn marginal profit every time until the product reaches the consumer, and the lack of clarity of the price fully increase significantly compared to peers who sell the same product and the same specifications.

1.2. Research objectives:

This study aims to achieve the following objectives:

1. Reveal the maximum price to consumers who are looking for online shoppers in the Saudi community.
2. Detection of the maximum price on the best prices among online shoppers in Saudi society.
3. Disclosure of the seller's commitment / non-commitment to the opportunity of consumers to visit online shoppers in Saudi society.
4. Identify the impact of commitment / non-commitment on the seller's search cost among online shoppers in Saudi society.
5. Identify the impact of commitment / non-commitment on the volume of trade between shoppers online in Saudi society
6. Recognize the effect of commitment / non-commitment on the seller's profits to the seller among online shoppers in Saudi society.

1.3. Research questions:

Thus, this study will determine two foremost relationships: the effect of the listed price on consumer's online search and the effect of the listed price on optimal pricing among online shoppers in the Saudi community. Thus, the chief questions of this study are, as follows:

1- What is the effect of the listed price on consumer's online search among online shoppers in the Saudi community?
2- What is the effect of the listed price on optimal pricing among online shoppers in the Saudi community?

Sub-questions of this study are, as follows:

a. What is the effect of the commitment/non-commitment seller on consumers' chance of visiting among online shoppers in the Saudi community?
b. What is the effect of the commitment/non-commitment seller on search cost among online shoppers in the Saudi community?
c. What is the effect of the commitment/non-commitment seller on the volume of trade among online shoppers in the Saudi community?
d. What is the effect of the commitment/non-commitment seller on profits for the seller among online shoppers in the Saudi community?

1.4. The scope of the study

The present study has the following limitations:

1. The study is limited to Saudi community, so the results are not to be generalized to other communities at different countries.
2. This study describes the effect of the listed price (commitment/non-commitment seller) on consumers search (consumers' chance of visiting and search cost) and optimal pricing (volume of trade and profits). Therefore, other factors will not be investigated.
3. The study sample is limited to online Saudi shoppers.

1.5. The Relevance of the Study

We are living in a time described by a wealth of data. Firms are in a race to make utilization of the huge information accessible on their consumers and items. In a comparable manner, consumers have numerous items/administrations to look over and have simple access to an abundance of data sources that can help in their search procedures.

In principle, all the data that is accessible on the web, web-based social networking, item lists, magazines and other distributed media, data communicated on the radio, TV, data got from valued ones are at the transfer of consumers. Notwithstanding, practically, of course, consumers have controlled time and consideration, as well as limited ability to process the data that is obtained (Boyacı&Akçay, 2017).

In this manner, data procurement and handling are an expensive attempt. Thus, consumers need to search how much and what kind of data to focus on (and what to overlook) and settle on buys searches on the premise of this small data. Seeing such limits and how they change into decision conduct is of critical worry to the offering firm (e.g., a retailer) since there is a relationship between the pricing and consumer search techniques (Sims 2006).

With a specific end goal to look at the ideal pricing estimating methodologies of the store, it is basic to catch the prominent highlights of listed price consideration and consumer search in a decision display. Perceptive obliviousness hypothesis (Sims 2006) offers an undoubted methodology for this reason. Differentiation to the perceptive desires hypothesis, which expects that consumers can completely process all openly accessible data about the item, balanced inaccuracy hypothesis accept that they do not have the ability to comprehend the accessible data thoroughly and make an interpretation of it into choices (Akçay, Natarajan, & Xu, 2010).

By the side of the fundamental of rational inattention is thoughtful that consideration is a rare means and consequently must be to be paid intelligently. Especially, the original works of Sims (2006) suggested an outline that is founded on a flow of works on data philosophy, which procedures doubt through entropy and measures data as a discount in doubt. This method does not kind specific expectations on in what way decision-makers obtain data and what they acquire knowledge of.

It expands on utility-boosting consumers who procure data ideally, exchanging off the normal advantage of better data against the cost related to obtain it. In like manner, the consumers ideally select the sort and amount of data they require and overlook the data that does not quality acquiring and not easy to handle with (Boyacı&Akçay, 2017).

Truth be told, in a current paper, Matejka& McKay (2014) demonstrated that when looked with detached decisions with stochastic (pay-off) values, a normally forgetful chief's ideal data handling system endogenously prompts a decision conduct that can be portrayed as summed up

Multinomial Logit (MNL). Specifically, the decision probabilities depend not just on the genuine acknowledge of the decisions, yet additionally on the consumer search and the cost of items.

Consequently, this study is looking to know the relation between the listed price and the customer search and the optimal price, in trying to understand the continuance behavior of online shoppers inside Saudi Arabia.

1.6. Terminology of study:

- **Listed price:** is the price at which the manufacturer recommends that the retailer sell the product.

- **Consumer search:** is the foundation of many marketing departments. The information it provides gives you feedback on products, marketing campaigns and future products or services

- **Optimal pricing:** is the price point at which the seller's total profit is maximized

2. The Review of the Literature

Online shopping conduct and encounters are generally unique to the physical shopping experience. Nelmapius & others (2005) recommend that the idea of the web (where an individual sit alone, in a commonplace situation, before a between associated arrange) imply that a large portion of the basic leadership in regards to Internet shopping is done in disconnection with practically zero connection with others. They consider that the online shopping condition is generally new and complex and that the feeling of novelty and unpredictability is to make worse by the nonattendance of the reminders of touch, taste, and notice, which are accessible in the physical shopping condition. When utilizing the web, on the grounds that the shopping happens in a virtual domain, the buyer is free either to finish the buy or to reject it anytime, if not by any means satisfied, with no social impact from different consumers.

The web has made it simple for consumers to think about prices and get the best prices by means of data cooperation (Punj, 2012). Truth be told, consumers have a few choices through media, which incorporate physical shopping, home-shopping, mail arrange shopping and the web (Card et al., 2003). The WWW has turned into an essential hotspot for information creation, utilization through online groups (Seraj, 2012). The consumers can undoubtedly stream the data through different channels. Past examinations in attire items additionally demonstrate that shopper shopping behavioral goal from the online clothing e-retailers is definitely identified with the data honestly and accessibility from the online trader (Park & Kim, 2007).

Be that as it may, one of the greatest contrasts amongst on the web and physical shopping conditions is how much customers think about prices. In web-based shopping situations, value correlation locales are boundless (Pan, Ratchford, & Shankar, 2004; Iyer&Pazgal, 2003; Häubl&Trifts, 2000). The nearness of value examination destinations brings down buyers' inquiry prices (Brynjolfsson& Smith, 2000). While web-based shopping has turned into a general pattern, online shops have a significantly harder time than at any other time finding a grand slam procedure to protect themselves from cruel competition including data on competitors' prices from value examination destinations, which work as outer reference prices (Kang & Jung, 2015).

Trust has been broadly perceived as a key factor in online buy (Ba &Pavlou, 2002). It decides purchasers' goal to buy and their selection of sellers to visit. Since online purchasers cannot completely recognize either wholesalers or items before buy, they utilize an assortment of accessible signals (e.g., value, notoriety, audits) to help them to decide which items and online shops are the best (Hsieh & Tsao, 2014; Roest&Rindfleisch, 2010). Among these prompts, cost is thought to be essential in assessing future item and the sellers (Han & Ryu, 2009), in light of the fact that individuals as often as possible expect that cost and quality are exceptionally related (Kim et al., 2012; Jin and Kato, 2006; Kardes et al., 2004). Given the way that notoriety and audits can be controlled, it is sensible to expect that cost related data to be essential in consumer's search.

As Wu & others (2015) proposed a hypothetical model to clarify how price scattering cooperates with different factors in Chinese Online Consumer-To-Consumer (C2C) buy, for example, primary trust, supposed price, buying intention and supposed risk. Item sort is considered as a mediator. A total of 261 students were welcomed in a questionnaire-based test. The outcomes from Partial Least Squares (PLS) investigation demonstrate that price scattering contrarily influences consumer search, while, emphatically influences supposed risk, which additionally impacts consumer search adversely. Price scattering additionally adversely impacts primary trust through supposed risk. Besides, the negative impacts of price scattering are more limited when purchasers buy high-touch items.

Escobar-Rodríguez & Carvajal-Trujillo (2014) analyzed determinants of buying flights from Low-Cost Carrier (LCC) sites. In doing as such an increased Unified Theory of Acceptance and Use of Technology (UTAUT) demonstrate is proposed expanding on before work by Venkatesh,

Thong, & Xu (2012). The outcomes, got from test of 1096 Spanish shoppers of LCC flights, demonstrated that key determinants of buying are: propensity, limited price, opportuneness, consumer search, completed buying process, six-factor hedonic shopping motivation (minimalists, the gatherers, the providers, the enthusiasts, and the traditionalists) as well as social variables. Of these factors, online buy goals, tendency, and opportuneness are the most essential.

Besides, Dai (2016) examined the impacts of limited price on the consumer search and the optimal price. It considered a situation in which purchasers are indeterminate about a seller's sense of duty regarding the shown cost. This investigation described the arrangement of pure methodology balances and find that a higher level of seller commitment prompts to decrease the costs. It demonstrated that the effect of the search costs on optimal prices is non-monotone and relies upon the level of seller commitment, in addition, the extent of the search cost. It likewise measured the impacts of regulation that limits the degree of a seller's deviation from the promoted cost and show that limited regulation may not be usefulness improving shopping process. At long last, it considered the situation where sellers have heterogeneous levels of commitment control and explore how the difference in commitment control impacts showcase results of items for consumers. It found that full commitment enables a consumer to continuation visit sellers since match sellers have limited price, while a higher level of limited price does not allow the consumer to decide the request.

Thus, online shoppers are still considered as a difficult issue since they do not have full data about their search for the items. In this condition, prices influence every seller's request not just through their consequences for customers' last buy choices, yet in addition to their impacts on shopper search (Choi, Dai, and Kim, 2016). According to the above past studies, it can be concluded that most of the past studies focus on the consumer and price in general, but there are little past studies that study the relationship between the listed price and the consumer search and the optimal price, in addition, there is no study determine these relationships among the online shopper in Saudi community.

3. The Research Design

The research design of this study followed a quantitative approach since the research design of this study followed a quantitative approach since it is utilized to pick up a comprehension of fundamental reasons, theories, and Perceptions of the study issue (Padgett, 2016). It additionally used to measure the study issue by a method for creating numerical information or information

that can be changed into usable insights. It is utilized to measure mentalities, suppositions, practices, and selected factors, as well as generalize results because of a bigger sample populace. It also utilizes quantifiable information to figure actualities and reveal designs in look into (Wincup, 2017).

3.1. Type of Study

The information gathered by primary source which is a questionnaire that the researcher designed concerning the study issue; which includes three main section: the personal information of the sample and the effect of commitment/non-commitment seller on consumers' chance of visiting and search cost, as well as the effect of commitment/non-commitment seller on volume of trade and profits.

3.2. Hypothesis

Main hypothesis:

H_1: There is a positive effect of the listed price on consumer's online search among online shoppers in the Saudi community

H_2: There is a positive effect of the listed price on optimal price among online shoppers in the Saudi community

Sub-hypothesis:

H_{1a}: There is a positive effect of the non-commitment seller on consumers' chance of visiting among online shoppers in the Saudi community

H_{1b}: There is a positive effect of the commitment seller on consumers' chance of visiting among online shoppers in the Saudi community

H_{1c}: There is a positive effect of the commitment seller on search cost among online shoppers in the Saudi community

H_{1d}: There is a positive effect of the non-commitment seller on search cost among online shoppers in the Saudi community

H_{2a}: There is a positive effect of the commitment seller on the volume of trade among online shoppers in the Saudi community

H_{2b}: There is a positive effect of the non-commitment seller on the volume of trade among online shoppers in the Saudi community

H_{2c}: There is a positive effect of the commitment seller on profits for the seller among online shoppers in the Saudi community

H_{2d}: There is a positive effect of the non-commitment seller on profits for the seller among online shoppers in the Saudi community

3.3. The Sampling Design

This descriptive study aims to study the relationships between the listed price and consumer search and optimal price among the online Saudi shoppers. Thus, the population of this study is the Saudi community. Then, the sample of this study is randomly selected from the online shoppers in Saudi Arabia.

3.4. Statistical Analysis Technique

The collected data has analyzed through the Statistical Package for the Social Sciences (SPSS) program, using the Analysis of Variance (ANOVA) and Multivariate Analysis of Variance (MANOVA) techniques. SPSS is a valuable package for determining variable influences for complex ideas (Kline, 2014). In this study, these statistical procedures were used in the determination of the links between the independent factor; listed price, whose proxies included the commitment and the non-commitment seller, and the dependent elements; consumers' search and optimal pricing. The consumers' chance of visiting and search cost represented the former variable, whereas the volume of trade and the profits denoted the latter.

ANOVA is a statistical technique that is used in assessing the existence or absence of significant differences between the means of two or more groups. It tests the influence of one or more elements by comparing the averages of distinct samples (Loerts, 2008). In this case, the repeated measure ANOVA was used. This procedure entails the comparison of means across one or more variables whose bases are repeated observations (Rayner, 2017). This analysis examined the relationships between the non-commitment seller and the consumers' chance of visiting, the commitment seller and the consumers' chance of visiting, the non-commitment seller and the search cost, the commitment seller and the search cost, the commitment seller and the volume of trade, the non-commitment seller and the volume of trade, the commitment seller and the profits, and the non-commitment seller and the profits.

Conversely, MANOVA compares three or more categories, where two or more dependent variables are involved. Moreover, this technique compares the differences between categories

with variation within groups (Loerts, 2008). Its primary assumptions include the independence of observations, the multivariate normality for explained variables, and the equality of covariance matrices. In this study, the mixed between-within subject MANOVA, which examines the influence of two factors on a group of dependent variables, was performed (Caruth, 2014). This procedure involved the determination of the main effect and the interaction effect among variables, such as the listed price and consumers' online search, the non-commitment seller and consumers' chance of visiting, the commitment seller and the consumers' chance of visiting, the commitment seller and the search cost, and the non-commitment seller and the search cost. Further, the significant value for Box's Test of Equality of Covariance Matrices was evaluated.

Besides the exhibition of the various associations between different variables, the SPSS program displays Cronbach's alpha test which measures indicate the reliability of the study instruments (Rayner, 2017). This survey will be conveyed to a panel of scholarly teachers and specialists for the validation of its contents.

3.5. The Research Model

The research model separated into two foremost portions: depended and independent variables, independent variable is listed price that includes according to (Choi, Dai, and Kim, 2016; Dai, 2016; Wu, et al., 2015; Escobar-Rodríguez & Carvajal-Trujillo, 2014; Venkatesh, Thong, & Xu, 2012): non-commitment and commitment seller, while dependent variables are consumer search and optimal price, which includes: consumers' chance of visiting, search cost, the volume of trade and profits for the seller. Figure 1 below shows the research-developed model.

$Y_1 = \beta_0 + \beta_1 X_1$ $Y_2 = \beta_0 + \beta_1 X_1$ $Y_3 = \beta_0 + \beta_1 X_1$ $Y_4 = \beta_0 + \beta_1 X_1$

$Y_1 = \beta_0 + \beta_2 X_2$ $Y_2 = \beta_0 + \beta_2 X_2$ $Y_3 = \beta_0 + \beta_2 X_2$ $Y_4 = \beta_0 + \beta_2 X_2$

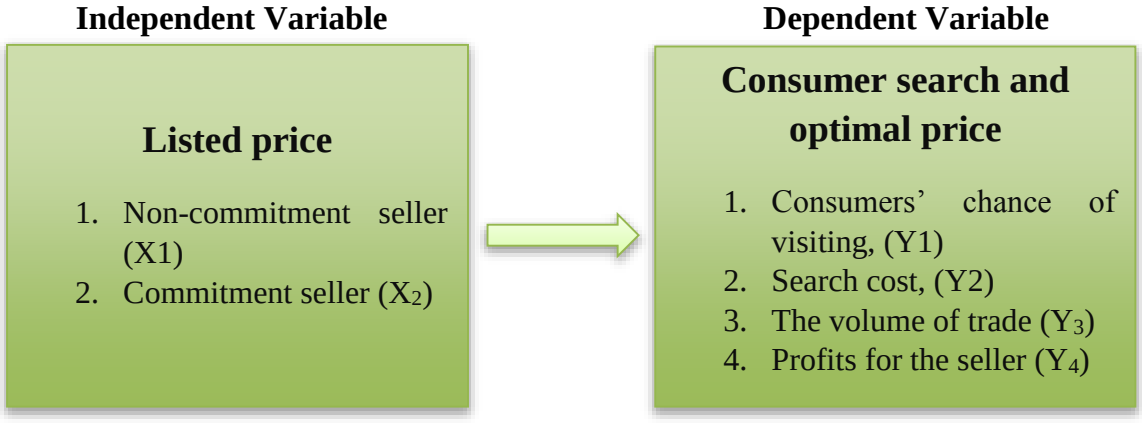

Figure 1: The Research developed Model

4. Analysis and Results

The focus of this study is to the effects of listed price on the consumer's online search and the optimal pricing: online shoppers in the Saudi community. Thus, this study uses a quantitative approach that is the questionnaire. Furthermore, this chapter describes in detail the quantitative data that was obtained throughout this study.

4.1. Reliability Analysis

The researcher has distributed the questionnaire on sample pilot of study (57 respondents) and computes extents questionnaire reliability by calculation of internal consistency using Cronbach' alpha values, table (1) shows that:

Table 1. The result of reliability

No	Variables	Cronbach's Alpha	Item No
1	Listed price	0.502	9
2	Consumer Search	0.814	12
3	Optimal Price	0.771	7
The effects of listed price on the consumer's online search and the optimal pricing		0.819	28

Table (1) shows that the reliability of the Consumer Search is equal to 0.814, the reliability of Optimal Price is equal to 0.771; and the reliability of listed price is equal to 0.502. The highest Cronbach' alpha value reached (0.814) for the total alpha values of **the effects of listed price on the consumer's online search and the optimal pricing** reached (0.819). This indicates to accept reliability.

4.2. The study sample and sampling

A questionnaire was designed to elicit responses on the main constructs investigated in this study (see Appendix A) to gather primary data. The questionnaires were then distributed to (57) the online shoppers in Saudi Arabia. They were selected using the random sampling method. Table 4.1 displays the distribution of respondents.

Table 4.1. Demographic profile of respondents (Source: SPSS results of the field work)

Demographic variable		Frequency	Percent	Valid Percent	Cumulative Percent
Gender	Male	12	21.1	21.1	21.1
	Female	45	78.9	78.9	100.0
	Total	**57**	**100.0**	**100.0**	-
Live	Qassim	36	63.2	63.2	63.2
	Badays	1	1.8	1.8	63.2
	Riyadh	6	10.5	10.5	73.7
	Medina	1	1.8	1.8	75.4
	Najran	8	14.0	14.0	89.5
	Jeddah	1	1.8	1.8	91.2
	Taif	1	1.8	1.8	93.0
	Dammam	1	1.8	1.8	94.7
	Jubail	2	3.5	3.5	100.0
	Total	**57**	**100.0**	**100.0**	-
Age (in years)	19-24	12	21.1	21.1	21.1
	25-30	15	26.3	26.3	47.4
	31-36	17	29.8	29.8	77.2
	More than 36	13	22.8	22.8	100.0
	Total	**57**	**100.0**	**100.0**	-
Monthly income	Less than 10000	32	56.1	56.1	56.1
	10000-20000	21	36.8	36.8	93.0
	More than 20000	4	7.0	7.0	100.0
	Total	**57**	**100.0**	**100.0**	-
Presently	In university	11	19.3	19.3	19.3
	Working	33	57.9	57.9	77.2
	Workless	12	21.1	21.1	98.2
	Superannuated	1	1.8	1.8	100.0

	Total	**57**	**100.0**	**100.0**	-
shop online	Monthly	34	59.6	59.6	59.6
	Around once a year	23	40.4	40.4	100.0
	Total	**57**	**100.0**	**100.0**	-
Educational level	Secondary educational level	5	8.8	8.8	8.8
	Diploma	4	7.0	7.0	15.8
	Bachelor	39	68.4	68.4	84.2
	Master	8	14.0	14.0	98.2
	High Diploma	1	1.8	1.8	100.0
	Total	**57**	**100.0**	**100.0**	-
Ever shopped online	Yes	56	98.2	98.2	98.2
	No	1	1.8	1.8	100.0
	Total	**57**	**100.0**	**100.0**	-

The demographic profile of the respondents in the research is presented in Table 4.1 above and graphically depicted in Appendix A. The results reveal that out of the 57 sampled online shoppers in Saudi Arabia, about 21.1% are males while females represented about 78.9% of the respondents; with an approximate live distribution of Qassim (63.2%), Najran (14%), Riyadh (10.5%), Jubail (3.5%), and other categories such as (Badays, Medina, Jeddah, Taif, and Dammam) accounted in 1.8% of the respondents. Within the age segment, 29.8% of the respondents lie between the age limit of 31 to 36 years, 26.3% lie between age limit of 25 to 30 years, 21.1% lie between age limit of 19 to 24 years, and the remaining 22.8% of the respondents are 36 years and above. Analysis of the presently segment indicate that about 57.9% of the respondents are Working (35%) and Workless (21.1%), with the remaining 19.3% as in university, Superannuated (1.8%). The educational segment of the respondents indicates predominance of educational level (68.4%) from Bachelor, Master (14%) and Secondary educational level (8.8%), Diploma (7%), and High Diploma (1.8%). However, about 56.1% of the respondents relatively earn monthly incomes below 10000, incomes between 10000 to 20000 (36.8%), and incomes more than 20000 (7%); explained shop online of the respondents' are

monthly (59.6%) and the Around once a year (40.4%), about ever shopped online 98.2%, while not ever shopped online represented about 1.8% of the respondents.

4.2.1 Descriptive Statistics Analysis of the Mean Scores in Listed price

Table 4.2. Mean, Standard Deviation, Minimum and Maximum Values

	Listed price
N	57
Mean	3.18
Standard Deviation	0.433
Minimum	2
Maximum	4.11

Table 4.3. Mean and Standard Deviation in Listed price variable

Variable	N	Mean	Standard Deviation
Commitment seller	57	3.08	0.666
Non-commitment seller	57	3.26	0.659

Table 4.2. It shows the mean, standard deviation, minimum and maximum value for the listed price. Whilst in Table 4.3, the mean and standard deviation value is displayed for all listed price variables. It was found that the Non-commitment seller had the highest mean from the Commitment seller.

4.2.2 Descriptive Statistics Analysis of the Mean Scores in Consumer's Online Search

Table 4.4. Mean, Standard Deviation, Minimum and Maximum Values

	Consumer Search	Optimal Price
N	57	57
Mean	3.79	3.85
Standard Deviation	0.479	0.529
Minimum	2.17	1.43
Maximum	4.58	4.86

Table 4.5. Mean and Standard Deviation in Consumer's Online Search variables

Type	N	Mean	Standard Deviation
Consumers' Chance of Visiting	57	3.75	0.469
Search Cost	57	3.89	0.668
The Volume of Trade	57	3.89	0.598
Profits for The Seller	57	3.81	0.433

Table 4.4. It shows the mean, standard deviation, minimum and maximum value for the Consumer Search and Optimal Price. Meanwhile in Table 4.5, the mean and standard deviation value is displayed for all Consumers' Online Search variables. It was found that the Search Cost and The Volume of Trade had the highest mean from the Profits for The Seller. The Consumers' Chance of Visiting obtained the lowest mean.

4.2.3 Inferential Statistics Analysis on the Effect of Between and Within Subjects for listed price and consumer's online search

This section covers the analysis of main effect the listed price on consumer's online search among online shoppers in the Saudi community, mixed between-within subject MANOVA used. The hypothesis measured is as stated:

H1: There is a positive effect of the listed price on consumer's online search among online shoppers in the Saudi community.

H1a: There is a positive effect of the non-commitment seller on consumers' chance of visiting among online shoppers in the Saudi community.

H1b: There is a positive effect of the commitment seller on consumers' chance of visiting among online shoppers in the Saudi community.

H1c: There is a positive effect of the commitment seller on search cost among online shoppers in the Saudi community.

H1d: There is a positive effect of the non-commitment seller on search cost among online shoppers in the Saudi community

A- MANOVA Assumptions and Univariate Analysis

Preliminary assumption testing was conducted for Mixed Between-Within Subject MANOVA. The significant value for Box's Test of Equality of Covariance Matrices was checked.

B- MANOVA Results

The analysis of main effect and the interaction was conducted using MANOVA. Table 4.6 displays the MANOVA results, the analysis showed that the main effect for Non-commitment seller was significant, Wilks'=0.282, F (26, 28)= 0.950, p=0.550, partial eta squared=0.469; main effect for commitment seller was significant, Wilks'=0.319, F (24, 28)= 0.897, p=0.603, partial eta squared=0.435; interaction effect for commitment seller and non-commitment seller was significant, Wilks'=0.318, F (32, 28)= 0.677, p=0.857, partial eta squared=0.436.

Table 4.6. MANOVA Results

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Non-commitment seller	Pillai's Trace	.907	.958	26.000	30.000	.541	.454
	Wilks' Lambda	.282	.950	26.000	28.000	.550	.469
	Hotelling's Trace	1.872	.936	26.000	26.000	.566	.483
	Roy's Largest Root	1.388	1.601	13.000	15.000	.190	.581
Commitment seller	Pillai's Trace	.846	.916	24.000	30.000	.583	.423
	Wilks' Lambda	.319	.897	24.000	28.000	.603	.435
	Hotelling's Trace	1.613	.874	24.000	26.000	.629	.446
	Roy's Largest Root	1.171	1.464	12.000	15.000	.240	.539
Non-commitment seller * Commitment seller	Pillai's Trace	.841	.681	32.000	30.000	.857	.421
	Wilks' Lambda	.318	.677	32.000	28.000	.857	.436
	Hotelling's Trace	1.645	.668	32.000	26.000	.861	.451
	Roy's Largest Root	1.242	1.165	16.000	15.000	.386	.554

C- ANOVA Results

The analysis for ANOVA is displayed in Table 4.7. All the results that were significant are as stated below:

(i) The main effect for Non-commitment seller on Consumers' Chance of Visiting was no significant, F= 0.571, p=0.842, partial eta squared = 0.331. In addition, the main effect for Non-commitment seller on Search Cost was no significant, F= 1.109, p=0.420, partial eta squared = 0.490.

(ii) The interaction effect for Non-commitment seller and commitment seller on Consumers' Chance of Visiting was no significant, F=0.425, p=0.950, partial eta squared =0.312. In addition, the main effect for Non-commitment seller and commitment seller on Search Cost was no significant, F= 0.530, p=0.890, partial eta squared = 0.361.

(iii) The main effect for commitment seller on Consumers' Chance of Visiting was no significant, F=0.624, p=0.792, partial eta squared =0.333. In addition, the main effect for commitment seller on Search Cost was no significant, F= 0.703, p=0.727, partial eta squared = 0.360.

Table 4.7. Tests of Within-Subjects Effects

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Non-commitment seller	Consumers' Chance of Visiting	1.812	13	.139	.571	.842	.331
	Search Cost	6.497	13	.500	1.109	.420	.490
Commitment seller	Consumers' Chance of Visiting	1.829	12	.152	.624	.792	.333
	Search Cost	3.804	12	.317	.703	.727	.360
Non-commitment seller * commitment seller	Consumers' Chance of Visiting	1.662	16	.104	.425	.950	.312

	Search Cost	3.822	16	.239	.530	.890	.361
Error	Consumers' Chance of Visiting	3.663	15	.244			
	Search Cost	6.759	15	.451			

i. Analysis of the effect for non-commitment seller on consumers' chance of visiting

An ANOVA repeated measure was conducted to analyze the between non-commitment seller and consumers' chance of visiting. The analysis showed that the main effect for non-commitment seller on consumers' chance of visiting was significant, F (1) = 13.660, p=0.001, partial eta squared = 0.199. Table 4.8 shows the results.

Table 4.8. Tests of Within-Subjects Effects Results

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	16.234	1	16.234	90.598	.000	.622
Non-commitment seller	2.448	1	2.448	13.660	.001	.199
Error	9.856	55	.179			
Corrected total	12.303	56				

ii. Analysis of the effect for commitment seller on consumers' chance of visiting

An ANOVA repeated measure was conducted to analyze the between commitment seller and consumers' chance of visiting. The analysis showed that the main effect for commitment seller on consumers' chance of visiting was significant, F (1) = 4.708, p=0.034, partial eta squared = 0.070. Table 4.9 shows the results.

Table 4.9. Tests of Within-Subjects Effects Results

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	47.757	1	47.757	231.765	.000	.808
Commitment seller	.970	1	.970	4.708	.034	.079
Error	11.333	55	.206			
Corrected total	12.303	56				

iii. Analysis of the effect for non-commitment seller on search cost

An ANOVA repeated measure was conducted to analyze the between non-commitment seller and search cost. The main effect for non-commitment seller on search cost was significant, F (1) = 15.587, p=0.000, partial eta squared = 0.221. Table 4.10 shows the multivariate results.

Table 4.10. Tests of Within-Subjects Effects

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	12.016	1	12.016	33.978	.000	.382
Non-commitment seller	5.512	1	5.512	15.587	.000	.221
Error	19.451	55	.354			
Corrected total	24.963	56				

iv. Analysis of the effect for commitment seller on search cost

An ANOVA repeated measure was conducted to analyze the between commitment seller and search cost. The main effect for commitment seller on search cost was no significant, F (1) = 1.072, p=0.305, partial eta squared = 0.019. Table 4.11 shows the multivariate results.

Table 4.11. Tests of Within-Subjects Effects

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	46.819	1	46.819	105.165	.000	.657
Commitment seller	.477	1	.477	1.072	.305	.019
Error	24.486	55	.445			
Corrected total	24.963	56				

4.2.4 Inferential Statistics Analysis on the Effect of Between and Within Subjects for listed price and optimal price

This section covers the analysis of main effect the listed price on optimal price among online shoppers in the Saudi community, mixed between-within subject MANOVA used. The hypothesis measured is as stated:

H2: There is a positive effect of the listed price on optimal price among online shoppers in the Saudi community

Sub-hypothesis:

H1d: There is a positive effect of the non-commitment seller on search cost among online shoppers in the Saudi community.

H2a: There is a positive effect of the commitment seller on the volume of trade among online shoppers in the Saudi community.

H2b: There is a positive effect of the non-commitment seller on the volume of trade among online shoppers in the Saudi community.

H2c: There is a positive effect of the commitment seller on profits for the seller among online shoppers in the Saudi community.

H2d: There is a positive effect of the non-commitment seller on profits for the seller among online shoppers in the Saudi community.

A. MANOVA Assumptions and Univariate Analysis for listed price and optimal price

Preliminary assumption testing was conducted for Mixed Between-Within Subject MANOVA. The significant value for Box's Test of Equality of Covariance Matrices was checked.

B. MANOVA Results for listed price and optimal price

The analysis of main effect and the interaction was conducted using MANOVA. Table 4.12 displays the MANOVA results, the analysis showed that the main effect for Non-commitment seller on optimal price was significant, Wilks'=0.243, F (26, 28)= 1.110, p=0.393, partial eta squared=0.508; main effect for commitment seller on optimal price was significant, Wilks'=0.276, F (24, 28)= 1.054, p=0.444, partial eta squared=0.475; interaction effect for commitment seller and non-commitment seller on optimal price was significant, Wilks'=0.467, F (32, 28)= 0.406, p=0.993, partial eta squared=0.317.

Table 4.12. MANOVA Results

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Non-commitment seller	Pillai's Trace	.910	.963	26.000	30.000	.535	.455
	Wilks'	.243	1.110	26.000	28.000	.393	.508

	Lambda						
	Hotelling's Trace	2.493	1.247	26.000	26.000	.289	.555
	Roy's Largest Root	2.209	2.548	13.000	15.000	.043	.688
Commitment seller	Pillai's Trace	.874	.971	24.000	30.000	.524	.437
	Wilks' Lambda	.276	1.054	24.000	28.000	.444	.475
	Hotelling's Trace	2.077	1.125	24.000	26.000	.383	.509
	Roy's Largest Root	1.769	2.211	12.000	15.000	.074	.639
Non-commitment seller * Commitment seller	Pillai's Trace	.623	.425	32.000	30.000	.990	.312
	Wilks' Lambda	.467	.406	32.000	28.000	.993	.317
	Hotelling's Trace	.950	.386	32.000	26.000	.994	.322
	Roy's Largest Root	.655	.614	16.000	15.000	.828	.396

C. ANOVA Results

The analysis for ANOVA is displayed in Table 4.13. All the results that were significant are as stated below:

(i) The main effect for Non-commitment seller on **The Volume of Trade** was no significant, F= 0.762, p=0.685, partial eta squared = 0.389. In addition, the main effect for Non-commitment seller on Profits for The Seller was no significant, F= 0.475, p=0.907, partial eta squared = 0.292.

(ii) The interaction effect for Non-commitment seller and commitment seller on The Volume of Trade was no significant, F=0.289, p=0.991, partial eta squared =0.235. In addition, the main effect for Non-commitment seller and commitment seller on Profits for The Seller was no significant, F= 0.360, p=0.975, partial eta squared = 0.278.

(iii) The main effect for commitment seller on The Volume of Trade was no significant, F=0.430, p=0.926, partial eta squared =0.256. In addition, the main effect for commitment seller on Profits for The Seller was no significant, F= 0.900, p=0.567, partial eta squared = 0.419.

Table 4.13. Tests of Within-Subjects Effects

Source		Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Non-commitment seller	Volume of Trade	5.675	13	.437	.762	.685	.398
	Profits for The Seller	4.104	13	.316	.475	.907	.292
Commitment seller	Volume of Trade	2.956	12	.246	.430	.926	.256
	Profits for The Seller	7.172	12	.598	.900	.567	.419
Non-commitment seller * commitment seller	Volume of Trade	2.646	16	.165	.289	.991	.235
	Profits for The Seller	3.828	16	.239	.360	.975	.278
Error	Volume of Trade	8.594	15	.573			
	Profits for The Seller	9.963	15	.664			

i. Analysis of the effect for non-commitment seller on Volume of Trade

An ANOVA repeated measure was conducted to analyze the between non-commitment seller and Volume of Trade. The analysis showed that the main effect for non-commitment seller on Volume of Trade was significant, F (1) = 5.660, p=0.021, partial eta squared = 0.093. Table 4.14 shows the results.

Table 4.14. Tests of Within-Subjects Effects Results

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	19.576	1	19.576	59.347	.000	.519

Non-commitment seller	1.867	1	1.867	5.660	.021	.093
Error	18.142	55	.330			
Corrected total	20.009	56				

ii. Analysis of the effect for commitment seller on Volume of Trade

An ANOVA repeated measure was conducted to analyze the between commitment seller and Volume of Trade. The analysis showed that the main effect for commitment seller on Volume of Trade was no significant, F (1) = 0.542, p=0.465, partial eta squared = 0.010. Table 4.15 shows the results.

Table 4.15. Tests of Within-Subjects Effects Results

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	43.486	1	43.486	120.712	.000	.687
Commitment seller	.195	1	.195	.542	.465	.010
Error	19.814	55	.360			
Corrected total	20.009	56				

iii. Analysis of the effect for non-commitment seller on Profits for The Seller

An ANOVA repeated measure was conducted to analyze the between non-commitment seller and Profits for The Seller. The main effect for non-commitment seller on Profits for The Seller was no significant, F (1) = 0.011, p=0.915, partial eta squared = 0.000. Table 4.16 shows the multivariate results.

Table 4.16. Tests of Within-Subjects Effects

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	32.666	1	32.666	77.054	.000	.584
Non-commitment seller	.005	1	.005	.011	.915	.000
Error	23.317	55	.424			
Corrected total	23.322	56				

iv. Analysis of the effect for commitment seller on Profits for The Seller

An ANOVA repeated measure was conducted to analyze the between commitment seller and Profits for The Seller. The main effect for commitment seller on Profits for The Seller was no significant, F (1) = 1.029, p=0.315, partial eta squared = 0.018. Table 4.17 shows the multivariate results.

Table 4.17. Tests of Within-Subjects Effects

Source	Type II Sum of square	Df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	44.583	1	44.583	107.108	.000	.661
Commitment seller	.428	1	.428	1.029	.315	.018
Error	22.893	55	.416			
Corrected total	23.322	56				

5. Conclusion

There were two research hypotheses addressed in this chapter. The areas that were measured are a Listed price (Commitment seller, Non-commitment seller), Consumer Search (Consumers' Chance of Visiting, Search Cost), and Optimal Price (The Volume of Trade, Profits for The Seller).

The analyses of the results of the present study revealed that the there is a positive effect of the listed price on consumer's online search among online shoppers in the Saudi community. The results proved that the there is a positive effect of the listed price on optimal price among online shoppers in the Saudi community.

Many previous studies have confirmed the importance of examining the effects of listed price on the consumer's online search and the optimal pricing: Online Shoppers in the Saudi Community, Choi, Dai, Besides and Kim (2016) found that prices influence every seller's request not just through their consequences for customers' last buy choices.

On the other hand, the outcomes by Venkatesh, Thong, & Xu (2012) that key determinants of buying are: propensity, limited price, opportuneness, consumer search, completed buying process, six-factor hedonic shopping motivation (minimalists, the gatherers, the providers, the enthusiasts, and the traditionalists) as well as social variables. Of these factors, online buy goals, tendency, and opportuneness are the most essential.

6. Acknowledgment

It is a great pleasure to have an opportunity like this whereby I am writing about a subject that has immense relevance in today's society. At the time of preparing this research, I have had the chance of reading and reviewing literature, especially online sources, books, journals, and magazines, which give insightful information about the topic.

I acknowledge with appreciation the input of my professor "Muhammad Khawaja" for the continued support, more so, in making me understand the different ways of synthesizing information and how to solve conceptual problems encountered when conducting research. I have no doubt that this paper will be a critical contribution to the existing literature, and furthermore, the findings will be helpful to companies with presence in online platforms.

7. References

Akçay, Y., Natarajan, H. P., & Xu, S. H. (2010). Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science*, *56*(8), 1345-1361.

Armstrong, M., & Zhou, J. (2011). Paying for prominence. *The Economic Journal*, *121*(556), F368–F395.

Ba, S., &Pavlou, P. A. (2002). Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior. *MIS quarterly*, 243-268.

Boyacı, T., &Akçay, Y. (2017). *Pricing when customers have limited attention*. Management Science. ESMT European School of Management and Technology Working Paper, No. 16-01.

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless commerce? A comparison of Internet and conventional retailers. *Management science*, *46*(4), 563-585.

Cachon, G. P., & Feldman, P. (2015). Price commitments with strategic consumers: Why it can be optimal to discount more frequently… than optimal. *Manufacturing & Service Operations Management*, 17(3), 399-410.

Card, J. A., Chen, C. Y., & Cole, S. T. (2003). Online travel products shopping: Differences between shoppers and nonshoppers. *Journal of Travel Research*, *42*(2), 133-139.

Caruth, G. D. (2014). A multivariate analysis (MANOVA) of where adult learners are in higher education. *International Journal of Learning, Teaching and Educational Research*, *1*(1).

Choi, M., Dai, A. Y., & Kim, K. (2016). Consumer search and price competition. *Journal of Economic Literature (JEL)*, D43, D83, L13, 1-40.

Dai, A. Y. (2016). Consumer Search and Optimal Pricing under Limited Commitment. Job Market Paper, University of Iowa, Iowa.

Dai, Y. (2017). *Essays on information, search and pricing*. A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Economics in the Graduate College of The University of Iowa, Iowa.

De los Santos, B., Hortaçsu, A., & Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *The American Economic Review*, *102*(6), 2955-2980.

De Los Santos, B., Hortaçsu, A., & Wildenbeest, M. R. (2017). Search with learning for differentiated products: Evidence from e-commerce. *Journal of Business & Economic Statistics*, 1-16.

Diakopoulos, N. (2015). How Uber surge pricing really works. *Washington Post Wonkblog*. Available at: https://www.washingtonpost.com/news/wonk/wp/2015/04/17/how-uber-surge-pricing-really-works/.

Ellison, G., & Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica*, *77*(2), 427-452.

Escobar-Rodríguez, T., & Carvajal-Trujillo, E. (2014). Online purchasing tickets for low cost carriers: An application of the unified theory of acceptance and use of technology (UTAUT) model. *Tourism Management*, *43*, 70-88.

Forrester, (2011). Smarter Strategies for free Shipping: understanding the true costs and benefits to retailers. A forresterConsulting Thought Leadership Paper commissioned by UPS, Cambridge, USA.

Haan, M., Moraga-González, J. L., &Petrikaite, V. (2015). *Price and match-value advertising with directed consumer search* (No. 15012-EEF). University of Groningen, Research Institute SOM (Systems, Organisations and Management).

Han, H., & Ryu, K. (2009). The roles of the physical environment, price perception, and customer satisfaction in determining customer loyalty in the restaurant industry. *Journal of Hospitality & Tourism Research*, *33*(4), 487-510.

Häubl, G., &Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing science*, *19*(1), 4-21.

Honka, E. (2014). Quantifying search and switching costs in the US auto insurance industry. *The RAND Journal of Economics*, *45*(4), 847-884.

Hsieh, M. T., & Tsao, W. C. (2014). Reducing perceived online shopping risk to enhance loyalty: a website quality perspective. *Journal of Risk Research*, *17*(2), 241-261.

Iyer, G., &Pazgal, A. (2003). Internet shopping agents: Virtual co-location and competition. *Marketing Science*, *22*(1), 85-106.

Jin, G. Z., & Kato, A. (2006). Price, quality, and reputation: Evidence from an online field experiment. *The RAND Journal of Economics*, 37(4), 983-1005.

Kang, J., & Park-Poaps, H. (2011). Motivational antecedents of social shopping for fashion and its contribution to shopping satisfaction. *Clothing and Textiles Research Journal*, *29*(4), 331-347.

Kang, M. Y., & Jung, K. (2015). The Effect of Online External Reference Price on Perceived Price, Store Image, and Risk. *Journal of Business Inquiry: Research, Education & Application*, *14*(1).

Kardes, F. R., Cronley, M. L., Kellaris, J. J., &Posavac, S. S. (2004). The role of selective information processing in price-quality inference. *Journal of Consumer Research*, *31*(2), 368-374.

Kim, H. W., Xu, Y., & Gupta, S. (2012). Which is more important in Internet shopping, perceived price or trust?. *Electronic Commerce Research and Applications*, *11*(3), 241-252.

Kline, P. (2014). *An easy guide to factor analysis*. London: Routledge.

Loerts, H. (2008). Multivariate ANOVA &ANOVA & repeated measures. 1-62. Retrieved March 30, 2019, from http://www.let.rug.nl/~nerbonne/teach/rema-stats-meth-seminar/presentations/Loerts-2008-MANOVA-Repeated-Measures.pdf

Matejka, F., & McKay, A. (2014). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review*, *105*(1), 272-298.

Nelmapius, A. H., Boshoff, C., Calitz, A. P., &Klemz, B. R. (2005). The impact of the information search variables, time pressure and involvement, on buying behaviour in a

three-dimensional hypermedia computer-mediated environment. *South African Journal of Business Management*, *36*(3), 1-13.

Padgett, D. K. (2016). *Qualitative methods in social work research* (Vol. 36). Sage Publications.

Pan, X., Ratchford, B. T., & Shankar, V. (2004). Price dispersion on the Internet: A review and directions for future research. *Journal of Interactive Marketing*, *18*(4), 116-135.

Park, C.H. & Kim, Y.G. (2003). Identifying key factors affecting consumer purchase behavior in an online shopping context. International Journal of Retail & Distribution Management, 31 (1), 16-29.

Punj, G. (2012). Consumer decision making on the web: a theoretical analysis and research guidelines. *Psychology & Marketing*, 29(10), 791–803.

Rayner, J. C. W. (2017). Extended ANOVA. *Journal of Statistical Theory and Practice*, *11*(1), 208-219. https://doi.org/10.1080/15598608.2016.1268550

Roest, H., &Rindfleisch, A. (2010). The influence of quality cues and typicality cues on restaurant purchase intention. *Journal of Retailing and Consumer Services*, *17*(1), 10-18.

Seraj, M. (2012). We create, we connect, we respect, therefore we are: intellectual, social, and cultural value in online communities. *Journal of Interactive Marketing*, *26*(4), 209-222.

Shen, J. (2015). *Ex-ante preference in a consumer search market*. Doctor of Philosophy, Ohio State University, Ohio.

Sims, C. A. (2006). Rational inattention: Beyond the linear-quadratic case. *The American economic review*, *96*(2), 158-163.

Venkatesh V, Thong JYL, Xu X (2012) Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 36: 157-178.

Wincup, E. (2017). *Criminological research: Understanding qualitative methods*. Sage.

Wu, K., Vassileva, J., Noorian, Z., & Zhao, Y. (2015). How do you feel when you see a list of prices? The interplay among price dispersion, perceived risk and initial trust in Chinese C2C market. *Journal of Retailing and Consumer Services*, *25*, 36-46.

Doi: https://doi.org/10.52132/Ajrsp.e.2023.55.3

Academic Journal of Research and Scientific Publishing | Vol 5 | Issue 55
Publication Date: 05-11-2023 ISSN: 2706-6495

A Concise Review on Time of Flight (TOF) Mass Spectrum

By: Dr. Khamis Ali Atayalla

Specialist in instrumental analysis (analytical chemistry), Chemistry Department, Faculty of Science, Bani Walid University, Libya

Corresponding author E-mail: khamis.ali.atayalla@gmail.com

Abstract:

Time of Flight Mass Spectrum is a method of mass spectrometer depending on the time which has been taking to reach the detector (usually located between 1 to 2 min from the source to separate ions) where the Sample is taken and mixed with matrix {the matrix is made up by chemical compound (2,4 Benzoic acid and cinamic acid) and they act as chemical cross link to produce matrix}. The energy source of this procedure is laser. Never the less, the ratio between sample and the matrix is 1:10,000 in order to prevent hampering the. However, TOF has some advantages such as unlimited upper (mass /charge) values to be detect, high resolution, good definition of narrow chromatographic peaks , good sensitivity and high quality spectra, High accurate mass measurement to the nearest 0.1 millimass unit for determining the elemental composition for ions less than 500 Da and It can perform tandem MS. Never the less, it has some drawbacks as requiring higher vacuum, Resolution changes with m/z value and high cost equipment. We conclude from this paper that: TOF is considered as one of the best mass filters because of unlimited upper (mass / charge) values to be detected as well as the capability of acquiring rapidly for averaging and good definition of narrow chromatographic peaks.

Keywords: Instrumental analysis, TOF, Detectors and acceleration

1. Introduction

The concept of time of flight analyser was appearing in American physical society program by Stephens in 1946, while the design of a linear TOF was published in 1955 by Wiley and McLaren which later became the first commercial instrument. There has been renewed in these instruments in 1980s. Time of Flight Mass Spectrum is a method of mass spectrometer depending on the time which it took to reach the detector usually located 1 to 2 m from the source to separate ions. Ions are accelerated by using a known strength electric field and due to this acceleration, each ion has the same charge will has the same kinetic energy.

The speed of the ion is depending on the ratio of mass to charge. After pulsing ions from the ion source, they will travel the same distance through the flight tube but with different velocity depending on their (mass / charge) ratio, which mean, the ions with the smaller ratio of (mass / charge) will take the shorter time to reach the detector, while ions with larger ratio of (mass / charge) will take a longer time to reach the detector. Flight tube indicates positive and negative electrodes in order to enforce ions to move in toward direction. The ions are pulsed from the source at the same time and have the same kinetic energy. (Hoffmann and Stroobant, 2003).

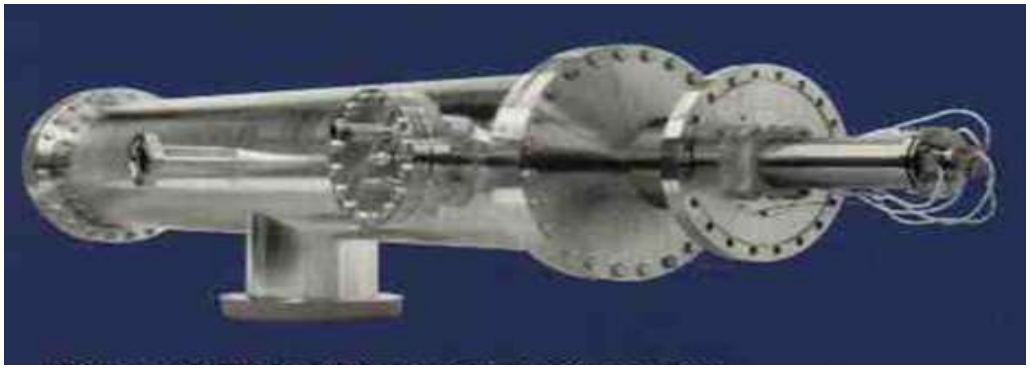

Figure (1): shows TOF mass spectrometer

Theory

The potential energy of ion in an electric field is related to the strength of the field and the charge on the ion $Ep = z\,U$ (1)

Where Ep is the potential energy, z is the charge and U is the voltage.

When ion is accelerated into the flight tube by the electric field (U , the potential energy is converted to kinetic energy ,which mean potential energy (Ep) is equal to kinetic energy (Ek)

Ep = Ek that is lead to Ek = z U (2)

It is known that Ek =1/2 mv2, replacing the value of Ek in equation (2) to be 1/2 mv2 = z U (3)

Where m is the mass, v is velocity (the amount of distance travel over time).

The velocity of the acceleration ion will not change because ion moves in a field-free time of flight tube.

The distance which the ion will cross is known (the length of the tube).

The time of the flight ion can be measured

v = d/t (4)

Where v = velocity, d = distance and t = time

z U = 1/2 m (d/t)2 (5)

by substituting the value of v in equation (3)).

By rearranging equation (5), the flight time will expressed as

t2 = d2 m /2Uz = (d2/2U) x (m/z) (6)

We can consider d2/2U as a constant, the equation (6) will be t2= K (m/z) → t2 α m/z

The ratio of mass to charge varies with the time of flight square which mean the separation of ions depending on time.

As can be seen from the above explanation of TOF, there is no theoretical limit to the upper (m/z)

value that can be detect which make MALDI as the most suitable ion source to use with TOF.

Matrix assisted laser desorption ionisation (MALDI)

Principle \ generate molecular ions

Sample is taken and mixed with matrix {the matrix is made up by chemical compound (2,4 Benzoic acid and cinamic acid) they act as chemical cross link to produce matrix}.

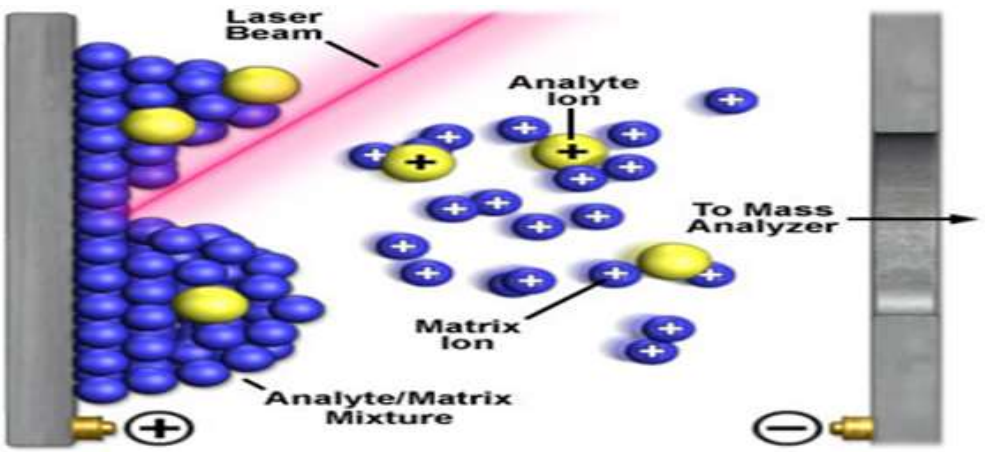

Figure (2): demonstrates MALDI's procedure

The matrix at the beginning is liquid and after adding the sample it should be allowed to be solidify to produce sample with matrix (sample trap in matrix).

Benzoic acid and ceramic acid have aromatic ring which absorb energy (this step is sample preparation step). The energy source of this procedure is laser.

The ratio between sample and the matrix is 1:10,000 which mean that the matrix will take 10,000 times than the sample in order to prevent hampering the sample because of laser.A sample plate is used to hold the matrix and the analyte molecules.The matrix is got the energy from the laser and transmit it to the sample (the sample is not directly hutted by laser in order to protect it from distraction). The sample is kicked out after taking the energy from the matrix because they are charged (they are disrobed from the matrix). (Neville, et al, 2011)

After that, the sample become excited then it generate molecular ions. Next, the sample evaporate from the plate, then the ionised molecules are forced travelling down the flight tube by voltage potential energy. The ionised molecules will being separated according to their molecular mass. Ionised molecules with smaller molecular weight will travel down the tube faster than larger molecules. The mass /charge ratio of each molecule is determined by a detector and used to identify the different components of a sample by measuring the time which ion is taking to reach the detector (the ion is accelerated by an electric field). The result is an output of a spectrum that will give the average molecular weight for the sample analyzed. After pulsing ions into time of flight tube ions with the same energy but having different mass will travel with different velocity.

Delayed extraction

To improve the mass resolution in axial MALDI-TOF, it has to allow the initial burst of ion and the neutrals which have been introduced laser pulse to equilibrate and travel the same distance vertically to the same plate before acceleration ions inside the flight tube. During the ionisation, most of ions start moving from the surface with same velocity. To improve the resolution and to compensate the spread velocity, delay extraction of ions from ion source toward the tube by a few nano second was introduced. Delayed extraction is based on compensation for the initial momentum of the ions. (Rousu, Herttuainen, Tolonen, 2010)

By other words, it provide the same arrival time to the detector for ions which have the same ratio (m/z) but with different initial velocity. The lower momentum ions in the direction of extraction are accelerated at higher potential energy because they are further from the extraction plate when the extraction field turned on.

On the other hand, the ions that have greater momentum start accelerating at lower potential energy due to being closer to the extraction plate. At the exit of acceleration region, the faster ions at the front of the plume will be accelerated to lower velocity than the slower ions at the back of the plume. (Pozo, et al., 2011)

After delaying extraction, group of ions that leaves the ion source later have greater velocity in the direction of acceleration than the other group of ions which have left the ion source earlier have lower velocity. At some distance from the ion source, the faster group catches up the slower group.

So, detector plate which has placed at this distance detects in the same time the arrival groups of these ions. (Kind and Fiehn, 2007)

Reflectron time of flight

Reflectron is used to correct the distribution of kinetic energy in the direction of the ion flight

The principle \ reflecting the ion beam toward the detector by using a constant electric field.

Ions with higher kinetic energy penetrate deeper into the reflectron and take a slightly longer path to the detector, while ions of the same ratio (m/z) but with lower kinetic energy penetrate a shorter distance into the reflectron, as a result, they will take a shorter path to the detector.

Micro channel plate MCP (the flat surface of the ion detector) is placed at the point where ions of the same mass and charge but with different energies reflected by the reflectron hitting the surface of the detector at the same time.

It is providing the high mass accuracy with wide dynamic range.

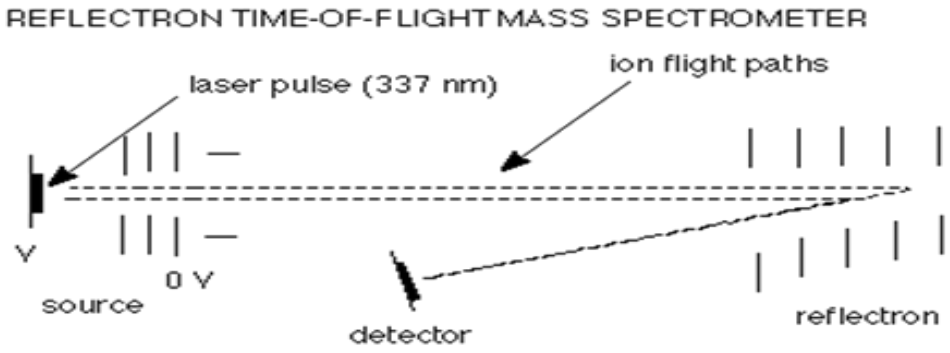

Figure (3) illustrates the aim of reflectron TOF

Detectors

In modern time of flight instrument, the detector is micro channel plate (MCP) which used to detect charge particles and radiation.It is made from high sensitive material of array of thousands of micro diameter channel each one of the latter is effectively a continuous dynode take a place under a strong electric field. Conversion dynode is used to convert ions to electrons and the latter colliding with phosphorescent material in order to convert electrons into photons.

Since these photons enter one of the micro channels throughout a small orifice, they start hitting the wall of the micro channel because the channel is at a special angel to the plate this angel known as (the angel of impact). Consequently, electrons are produced because of this impact and spreading through the micro channel, which amplifies the original signal. The electrons which went out from the opposite side of the channel were detected by a photomultiplier detector. (Watson and Sparkman, 2007)

1-Time-Slice Detection

TSD which typically acquire spectra by sampling only a single time -slice in each transient corresponding to only one value of m/z position in spectrum. A complete mass spectrum can be reconstructed by collecting other time-slice at different time delays after source pulses.

2-Time-Array Detection

With TAD all ions at all value of m/z ratio are detected during each TOF cycle, 10,000 times more data are available with TAD than TSD.

3-TAD with Transient Recorders

Recording all available information in transient signal at the detector facing a problem because of the amount of data and the rate at which they are produced.

Nevertheless, using high speed transient records have allowed many arrival -time window to be sampled for each individual ion source pulse. (Watson and Sparkman, 2007)

4-Hadamard Transform TOF-MS

It is a developed instrument of MS that can analyse many ion packets travelling in the flight tube at the same time, while the traditional TOF was waiting one packet of ions to reach the detector before introducing another one.

5-Tandom TOF-TOF

The aim of this instrument is recording a full spectrum of parent ions by using two time-of-flight spectrometers consecutively. The precursor ions of choice are isolated by a velocity filter in the first TOF mass spectrum, while the second TOF analyses the fragment ions. In order to reduce the instant current load on the ion detector, the second TOF provide with precursor signal quenchers.

TOF is a high resolution mass analyser and the increasing in flight distance and using new type of ion gun are useful to improve resolution. As it known, each mass analyser has a negative and positive aspects depending on the ion source, the procedure, sensitivity, resolution, mass accuracy or the cost of the instrument. (Herttuainen and Tolonen, 2010)

Conclusion

TOF consider as one of the best mass filters because of unlimited upper (mass /charge) values to be detect as well as the capability of acquiring rapidly for averaging and good definition of narrow chromatographic peaks. In addition to that, a good sensitivity and high quality spectra because of capability of detecting all ions of all mass / charge values without the needing for SIM. Although, Resolution changes with m/z value and higher vacuum is required. One the other hand, Pulsed mode of operation make MALDI suitable as ion source to use with TOF and can perform tandem MS.

References

Hoffmann, E. de, & Stroobant, V. (2003). Mass spectrometry: Principles and applications. John Wiley & Sons. https://www.usp.br/massa/2014/qfl2144/pdf/MassSpectrometry.pdf

Watson, J. T., & Sparkman, O. D. (2007). Introduction to mass spectrometry instrumentation, applications and strategies for Data Interpretation. Wiley-VCH.

https://spectrino.com/resources/mass_spec/Intro%20to%20Mass%20Spectrometry.pdf

OKAZAKI, M., HIRATA, I., MATSUMOTO, T., & TAKAHASHI, J. (2005). Advantages of TOF-sims analysis of hydroxyapatite and fluorapatite in comparison with XRD, HR-tem and FT-ir. Dental Materials Journal, 24(4), 508–514. https://doi.org/10.4012/dmj.24.508

https://www.jstage.jst.go.jp/article/dmj1982/24/4/24_4_508/_article/-char/en

Neville, S. A., LeCordier, A., Ziochos, H., Chater, M. J., Gosbell, I. B., Maley, M. W., & van Hal, S. J. (2011). Utility of matrix-assisted laser desorption ionization–time of flight mass spectrometry following introduction for routine laboratory bacterial identification. Journal of Clinical Microbiology, 49(8), 2980–2984.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3147730/

Kind, T., & Fiehn, O. (2007a). Seven golden rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics, 8(1). https://doi.org/10.1186/1471-2105-8-105

https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-105

Rousu, T., Herttuainen, J., & Tolonen, A. (2010a). Comparison of triple quadrupole, hybrid linear ion trap triple quadrupole, Time-of-flight and ltq-orbitrap mass spectrometers in drug discovery phase metabolite screening and identification in vitro – amitriptyline and verapamil as model compounds. Rapid Communications in Mass Spectrometry, 24(7), 939–957. https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/rcm.4465

Doi: https://doi.org/10.52132/Ajrsp.e.2023.55.4