



# DATA MINING IN E-COMMERCE: PROTECTING PERSONAL DATA IN INDIA

<sup>1</sup>Neha Bhandari

<sup>1</sup>Student

<sup>1</sup>Christ (Deemed to be University), Bangalore, India

<sup>2</sup>Dr. Shampa I. Dev

<sup>2</sup>Professor

<sup>2</sup>Christ (Deemed to be University), Bangalore, India

**Abstract:** The past decade has witnessed a rise in the e-commerce platforms in India which has led to an enormous influx of data generated through customer interactions, transactions, and browsing behaviors. In this context, data mining acts as a powerful tool in extracting valuable data from large raw data sets, enabling e-commerce platforms to enhance user experiences, optimize platform operations and formulate strategic decisions. However, the use of data mining techniques in e-commerce raises legal concerns that necessitate careful consideration.

This paper highlights the methods used in data mining to understand patterns and relationships between the customer and the platform, thus, increasing the business value, and explores the legal implications of data mining in e-commerce platforms in India. It examines how data mining practices intersect with the existing regulations, primarily focusing on the Information Technology Act, 2000 and the recently enacted Digital Personal Data Protection Act, 2023. Further, it analyses landmark judgments and regulatory guidelines set forth by the Judiciary on the evolving jurisprudence around personal data protection, the concept of consent and liability, in essence, the consumer rights in e-commerce platforms.

One of the primary concerns that data mining raises is, user consent and data ownership. E-commerce platforms collect and store a multitude of personal as well as behavioral data, often without taking explicit consent from the users. This further raise privacy concerns of the users with no right to data portability and the right to be forgotten under the Digital Personal Data Protection Act, 2023, which further brings about a gap between the legislation and precedents.

Further, the cross-border transfer of data to other countries might pose a threat to the privacy of users interacting on such e-commerce platforms in India. The transfer of personal data from India to other countries, unless they are explicitly blacklisted by the Central Government, may be subject to unauthorized use and breaches by private and governmental entities if there is no robust data protection legislation in the transferee country. This further raises the issue of inadequate protection of personal data in the process of data mining by e-commerce platforms that are providing goods and services to users within the territory of India.

This paper delves into concepts of data privacy, data localization, and user consent enshrined in the Information Technology Act, 2000 and the Digital Personal Data Protection Act, 2023 with an attempt to answer the fundamental question of pondering: how can we protect our personal data without parting ways with the evolving technology? With the ever-evolving landscape of data protection and e-commerce, the stakeholders must ensure compliance with the existing and forthcoming regulations. The e-commerce platforms must adopt transparent data collection practices, robust mechanisms for consent, and extensive data protection policies to mitigate legal risks. The research methodology adopted for this research paper is doctrinal, descriptive and conceptual. The sources referred to are secondary in nature, such as books, articles, and journals.

**Index Terms** – E-Commerce, Data Mining, Privacy, Consent, Information Technology Act 2000

## I. INTRODUCTION

In the digital era, where the landscape of commerce has shifted drastically towards online platforms, the importance of data mining in e-commerce cannot be emphasized. Data mining, the process of extracting valuable patterns and insights from large datasets, plays a pivotal role in shaping the strategies of e-commerce businesses. However, with great power comes greater responsibility, especially when it concerns the privacy of personal data. In the context of India, a country with a burgeoning e-commerce industry, safeguarding personal data has become a paramount concern.

Data in the twenty-first century are the equivalent of what was oil in the twentieth century and as an oil refinery, also data centres and fulfils the same role: produce raw materials that fuel the entire economy.<sup>1</sup> Undoubtedly, refined data is the driving force behind all e-commerce services. Similar to oil, data is extracted, refined, stored (in this case, in a data warehouse), valued, and finally sold in various ways. This process of refining is referred to as “data mining,” which involves finding patterns and relationships within datasets. It is an essential procedure since the dissemination of data depends more on its quality than on its quantity. Although the data economy is a relatively new phenomena, it is already causing such disruptions to all industries and sectors that a data revolution is possible.

The goal of data mining is to extract economic value from massive databases by finding patterns and relationships that has not been discovered previously. Its main objective is to derive insights from datasets by identifying correlations between two or more variables. These insights involve translating the data into knowledge that is specifically pertinent to a given use case, such dictating course of action or forecasting results from subsequent occurrences. After examining the data for patterns, certain parameters are used to identify the most prevalent and significant correlations. To do this, a variety of data sorting strategies, including sequence analysis, clustering, and classification, can be applied. Data mining usually employs batch data to provide new insight at a specific moment and it can be used, for instance, to forecast buying patterns, sale trends and to predict adoption of new product.

Data mining is extensively use in E-commerce platforms. E-commerce platform is an interface where consumers to conduct transactions for services/products and provides a rather cost-effective method alternative for conducting business virtually. The E-Commerce strategy mainly focuses on E-Marketing and E-Commerce recommendation system.<sup>2</sup> E-commerce recommendation assists in recommending products to customers. The product searched by the customer through E-Commerce is automatically saved in to the database, from where the vendor focuses on the customer who searched for a particular product and giving suggestions to the customer based on their search<sup>3</sup>. E-commerce recommendation system is fully based on a) customer profile b) history and c) other customer's previous transactions.<sup>4</sup>

E-commerce applications of data mining can be deduced as follows – analyzing customer profiles, selecting a partner organization, assessing the similarity of user browsing patterns, determining which products to market online categorization of web pages based on content, assessment of the similarity of web pages

<sup>1</sup> Pierfrancesco Marano, *Data Economy: Moving from Digital Feudalism toward a Digital Capitalistic Model*, LUISS (2019).

<sup>2</sup> H. Mohana, *An Overview Study on Web Mining in E-Commerce*, 6 IJSR (2017).

<sup>3</sup> Id.

<sup>4</sup> *Supra* note at 2

contents and identifying web pages that are viewed together. Further, the website also assists users in decision-making by listing for instance other books that have been bought by people who also bought the particular book that the online user is currently looking at.<sup>5</sup> One of the most useful applications of data mining in an E-Commerce environment can be found on the Amazon.com website which uses data mining to bundle books that are often bought in pairs, together and to sell or market these at a reduced price if they are bought together.<sup>6</sup> Further, Amazon.com uses data mining to profile its customers, and online users that return to the website after having purchased a book, for instance, will be provided with lists of other book titles that they might be interested in based on the category of book that they just purchased.<sup>7</sup>

One of the primary concerns regarding data mining is the collection of sensitive information without individuals' explicit consent. Companies often gather data from various sources, including social media, online transactions, and mobile apps, without consumers fully understanding the extent of information being collected. This lack of transparency can lead to a breach of trust between businesses and consumers, eroding the privacy boundaries individuals expect in their online interactions.

Furthermore, data mining techniques can create detailed consumer profiles, encompassing preferences, behaviors, and even predicting future actions. Although, this information is valuable for targeted marketing strategies, it can also be misused. Inaccurate profiling or data breaches can lead to identity theft, financial fraud, and other forms of cybercrimes, causing significant harm to individuals.

While legislations such as the Information Technology Act, 2000 and the Digital Data Protection Act, 2023 have been enacted by the Government of India, these do not aim particularly at regulating data mining. Although, data mining offers immense opportunities for businesses and technological advancements, its effects on consumer privacy are significant and complex. Thus, striking the right balance between innovation and privacy protection is crucial to building a digital landscape where individuals' rights are respected, and their personal information is secure from unwarranted intrusion and misuse.

## II. DATA MINING: METHODS AND TECHNIQUES

Data mining techniques play a pivotal role in enhancing the efficiency and competitiveness of e-commerce platforms.<sup>8</sup> By analyzing vast amounts of data generated by online interactions, businesses can gain valuable insights into customer behaviors, preferences, and trends.<sup>9</sup> Here are some essential methods and techniques of data mining in e-commerce platforms:<sup>10</sup>

### 2.1. Association Rule Mining

Association rule mining identifies patterns in customer purchase behaviour. It helps e-commerce platforms understand which products are frequently bought together. For instance, if customers often buy smartphones

---

<sup>5</sup> Dr. Sarika Agarwal, *Use of Data Mining in E-Commerce Platforms in India*, 2 IJIRA (2022).

<sup>6</sup> Id.

<sup>7</sup> *Supra* note at 5.

<sup>8</sup> Hirapara, Jignesh, and Pratik A. Vanjara., *A Comparative Study of Data Mining Techniques for Agriculture Crop Price Prediction*, (2022), <https://doi.org/10.1109/i2ct54291.2022.9824533>.

<sup>9</sup> Irmina Durluk, *Navigating the Sea of Data: A Comprehensive Review on Data Analysis in Maritime IoT Applications* 13 APPLIED SCIENCES (2023).

<sup>10</sup> Alexander, *The Role of Big Data in Business: Opportunities and Challenges*, M (Oct. 25, 2023 9:00 PM), <https://masstamilan.tv/navigating-the-challenges-of-big-data-in-business-tips-and-strategies-for-success/>.

with phone cases, the platform can use this information for targeted marketing, such as offering discounts on phone cases when a customer buys a smartphone.

## 2.2. Clustering

Clustering algorithms group similar customers based on their behaviour, preferences, or demographics. E-commerce platforms can use clustering to segment their customer base. For example, customers who frequently buy electronics might form one cluster, while customers interested in fashion products might form another. By understanding these segments, platforms can tailor their marketing strategies and product offerings for specific customer groups.

## 2.3. Predictive Analysis

Predictive analytics involves using historical data to predict future events or trends. E-commerce platforms can employ techniques like machine learning algorithms and regression analysis to predict customer behaviour, such as forecasting demand for specific products during different seasons. Predictive analytics helps optimize inventory management, pricing strategies, and marketing campaigns.

## 2.4. Recommendation Systems

Recommendation systems use algorithms like collaborative filtering and content-based filtering to suggest products to customers based on their past interactions and preferences. By analyzing a customer's purchase history and behaviour, e-commerce platforms can provide personalized product recommendations. This enhances the user experience, increases sales, and encourages customer loyalty.

## 2.5. Customer Segmentation

Customer segmentation involves dividing customers into distinct groups based on characteristics like demographics, purchase history, or preferences. E-commerce platforms can tailor marketing messages, promotions, and product recommendations to specific segments, increasing the likelihood of conversion and customer satisfaction.

Data mining methods and techniques empower e-commerce platforms to make data-driven decisions, enhance customer experiences, and gain a competitive edge in the digital marketplace. By leveraging these tools effectively, businesses can optimize their operations, improve customer satisfaction, and drive sustainable growth in the dynamic world of online commerce.

### III. DATA MINING AND CONSUMER PRIVACY: A LEGAL CONUNDRUM

In this era of constant digital evolution, absolute privacy is nothing but a myth. Major data sources include server and cookie logs, customer information, intelligent Internet agents and centralized demographic and other official records.<sup>11</sup> Powerful data processing, storage and communication technologies allow these data to be manipulated and used by the other people and agencies freely and in most cases indiscreetly.<sup>12</sup>

## 2.6. Role of Judiciary

The Right to Privacy is a fundamental aspect of an individual's freedom and dignity. It is recognized within the ambit of Article 21 under the Constitution of India. The landmark judgment which broadened the scope

<sup>11</sup> Mehmet Cudi, *On Ethical and Legal Aspects of Data Mining*, JoY 1455, 1455-1456 (2008).

<sup>12</sup> Id.

of Article 21 of the Constitution of India is *Justice K.S. Puttaswamy v. Union of India*<sup>13</sup>. The Supreme Court declared privacy as a fundamental right, stating that it is intrinsic to the right to life and liberty.<sup>14</sup> In this case, commonly referred to as the Aadhaar judgment, the Supreme Court of India upheld the constitutional validity of Aadhaar, the 12-digit unique identification number issued to Indian residents. However, the court imposed several restrictions to safeguard the right to privacy. It ruled that Aadhaar cannot be made mandatory for various services, including bank accounts and mobile connections, thereby protecting citizens from undue intrusion into their privacy. The Supreme Court, however, did not explicitly address the right to be forgotten, and discussions surrounding this right continue in the context of data protection and online privacy, especially considering the vast amount of personal data available on the internet.

### 2.7. Information Technology Act, 2000

The *Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011* under the Information Technology Act, 2000, also provide guidelines for the collection and use of sensitive personal data. These rules aim to protect the privacy of individuals' data in the digital space.

### 2.8. Digital Personal Data Protection Act, 2023

While data mining offers substantial benefits to businesses, it raises ethical concerns regarding consumer privacy. The Digital Personal Data Protection Act, 2023, addresses these challenges by imposing stringent regulations on how e-commerce platforms collect, store, and utilize consumer data. The Act enforces transparency, ensuring that consumers are aware of the data being collected and how it will be utilized. The key provisions of the Act are:

- a. *Explicit Consent*: E-commerce platforms are required to obtain the explicit consent of customers before collecting their personal data. This ensures that consumers have control over the information they share and how it is used.
- b. *Data Minimization*: The Act emphasizes the principle of data minimization, requiring businesses to collect only the data necessary for the intended purpose. Unnecessary data collection is discouraged, reducing the risk of misuse.
- c. *Purpose Limitation*: E-commerce platforms can only use consumer data for the specific purposes for which it was collected. Any deviation from the stated purpose requires re-obtaining consent from the consumer.
- d. *Data Security*: The Act enforces strict security measures to safeguard consumer data from breaches and unauthorized access. E-commerce platforms are required to implement encryption, regular security audits, and other measures to protect consumer information.
- e. *Right to be Forgotten*: Consumers have the right to request the deletion of their personal data once the purpose for which it was collected is fulfilled. E-commerce platforms must comply with such requests in a timely manner.

<sup>13</sup> (2019) 1 SCC 1.

<sup>14</sup> Id.



- f. *Accountability and Transparency:* Businesses are held accountable for their data processing activities. They are required to maintain records of data processing activities, ensuring transparency and compliance with the law.
- g. *Data localization:* The Digital Personal Data Protection Act, 2023, also addresses the critical issue of data localization in the context of E-commerce. Data localization can be understood as the practice of storing digital data within the borders of a specific country, ensuring that sensitive information is subject to local laws and regulations. Under this Act, e-commerce businesses are required to store and process customer data locally, enhancing data security and ensuring compliance with national privacy laws.

This provision serves as a fundamental safeguard, particularly in the realm of e-commerce, where vast amounts of sensitive customer data are generated and processed daily. By mandating data localization, the Act protects the consumers' personal information from unauthorized access as well as cyber threats but also strengthens national cybersecurity frameworks. Furthermore, localizing data fosters a sense of trust among consumers, assuring them that their data is subject to stringent domestic privacy regulations, thus encouraging greater participation in the digital marketplace.

The data localization requirement plays a pivotal role in shaping a secure and trustworthy environment for e-commerce activities, bolstering consumer confidence and ensuring the responsible handling of personal data in the digital era.

Further, the *Digital Personal Data Protection Act, 2023*, significantly impacts data mining practices in e-commerce. While businesses must adhere to stringent regulations, the Act also fosters responsible data mining. By ensuring that data collection is based on explicit consent and used for legitimate purposes, the Act promotes ethical data mining practices. Companies are compelled to invest in technologies that facilitate responsible data analytics, leading to more accurate targeting and improved customer satisfaction.

The Act also empowers consumers by giving them control over their personal information. With clear information about how their data is being used and the ability to opt-out, consumers can make informed decisions. This transparency enhances trust between consumers and e-commerce platforms, encouraging more secure and meaningful online interactions.

Moreover, the Act enables consumers to exercise their rights, such as the right to access their data, rectify inaccuracies, and even move their data from one service provider to another. These rights ensure that consumers are not only aware of their digital footprint but also have the means to manage it effectively.

#### **IV. CONCLUSION AND RECOMMENDATIONS**

In the ever-expanding digital landscape, the synergy between data mining and consumer privacy in e-commerce stands at the forefront of ethical and legal discussions. The implementation of the *Digital Personal Data Protection Act, 2023*, signifies a monumental step towards harmonizing these seemingly opposing forces. By promoting responsible data mining practices and safeguarding consumer privacy, the Act establishes a framework that not only protects individuals but also fosters a more trustworthy and innovative E-commerce ecosystem.

This Act brings a balance to the equation. E-commerce platforms can still harness the power of data mining to understand consumer behaviour and enhance their offerings, but now, they do so under the watchful eye of regulations designed to ensure ethicality and transparency. Consumers, on the other hand, benefit from enhanced privacy protections, empowering them with control over their personal information. This newfound sense of agency encourages active participation in the E-commerce platforms, knowing their data is handled responsibly.

However, the Government can further initiate certain steps in order to influence consumers to become more careful on E-commerce platforms.

1. It is imperative to educate both businesses and consumers about the implications of data mining and the rights conferred by the Digital Personal Data Protection Act, 2023. Awareness campaigns can bridge the knowledge gap, ensuring that individuals understand their rights and how their data is being used.
2. E-commerce platforms should invest in advanced technologies like blockchain and differential privacy techniques. These technologies enable secure data transactions and protect sensitive information, ensuring compliance with the Act while facilitating meaningful data mining processes and promote technological integration.
3. E-commerce platforms should establish internal ethical guidelines governing data mining practices. These guidelines should align with the Act's provisions, emphasizing responsible data usage, customer consent, and transparency.
4. Regular internal audits and compliance checks are essential. E-commerce platforms must ensure that their data mining practices adhere to the Act's Guidelines. This proactive approach not only prevents legal repercussions but also fosters a culture of ethical data use.
5. Whenever possible, businesses should anonymize customer data before processing it. Anonymization techniques strip data of personally identifiable information, which reduces the risk of any data breaches and ensures compliance with the Act.
6. The digital landscape is ever-changing, thus, E-commerce platforms must stay abreast of emerging technologies and evolving consumer expectations by adapting their data mining strategies in response to these changes, businesses can continue to innovate while respecting privacy norms.

The intersection of data mining and consumer privacy in E-commerce platforms is rather complex, yet crucial for the digital economy's growth. The Information Technology Act, 2000 along with the newly laid down *Digital Personal Data Protection Act, 2023*, provides a robust framework, ensuring that businesses can leverage the power of data while safeguarding individual privacy rights. Through education, technological integration, ethical guidelines, regular audits, data anonymization, and continuous adaptation, e-commerce entities can navigate this landscape responsibly, fostering trust, innovation, and mutual benefit between businesses and consumers in the digital realm.

The *Digital Personal Data Protection Act, 2023*, marks a significant milestone in the realm of data mining and consumer privacy in E-commerce, by establishing clear guidelines and stringent regulations which aims to strikes a balance between fostering innovation and protecting consumer interests. E-commerce platforms are now compelled to adopt responsible data mining practices, ensuring that the vast potential of data analytics

is harnessed ethically and securely. As E-commerce platforms navigate the evolving landscape of digital commerce, adherence to the Act not only ensures legal compliance but also fosters trust and credibility among consumers. The Act, therefore, serves as a cornerstone in shaping the future of data mining in e-commerce, promoting a harmonious relationship between businesses and consumers in the digital age.

## REFERENCES

- [1] Pierfrancesco Marano, *Data Economy: Moving from Digital Feudalism toward a Digital Capitalistic Model*, LUISS (2019).
- [2] H. Mohana, *An Overview Study on Web Mining in E-Commerce*, 6 IJSR (2017).
- [3] Id.
- [4] *Supra* note at 2.
- [5] Dr. Sarika Agarwal, *Use of Data Mining in E-Commerce Platforms in India*, 2 IJIRA (2022).
- [6] Id.
- [7] *Supra* note at 5.
- [8] Hirapara, Jignesh, and Pratik A. Vanjara., *A Comparative Study of Data Mining Techniques for Agriculture Crop Price Prediction*, (2022), <https://doi.org/10.1109/i2ct54291.2022.9824533>.
- [9] Irmina Durlik, *Navigating the Sea of Data: A Comprehensive Review on Data Analysis in Maritime IoT Applications* 13 APPLIED SCIENCES (2023).
- [10] Alexander, *The Role of Big Data in Business: Opportunities and Challenges*, M (Oct. 25, 2023 9:00 PM), <https://masstamilan.tv/navigating-the-challenges-of-big-data-in-business-tips-and-strategies-for-success/>.
- [11] Mehmet Cudi, *On Ethical and Legal Aspects of Data Mining*, JoY 1455, 1455-1456 (2008).
- [12] Id.
- [13] (2019) 1 SCC 1.
- [14] Id.