E-GOVERNMENT, E-COMMERCE, AND DIGITAL SKILLS: AN ANALYSIS OF TURKEY USING TÜİK MICRODATA

AUTHOR
Dr. Deniz TANIR

Karabük Üniversitesi Yayınları

Karabuk University Publications

E-GOVERNMENT, E-COMMERCE, AND DIGITAL SKILLS: AN ANALYSIS OF TURKEY USING TÜİK MICRODATA

Author

Dr. Deniz TANIR

Editor

Dr. Emrullah DEMiRAL

Design

Deniz TANIR

Contact

Karabuk University

https://www.karabuk.edu.tr/tr–https://www.karabuk.edu.tr/en

Karabük University Main Campus

444 0 478

Karabuk University Publications

AUTHOR
Dr. Deniz TANIR
ORCID: 0000-0001-6593-6625

EDITOR
Dr. Emrullah DEMiRAL
ORCID: 0000-0001-7633-5761

Library Record Card

E-GOVERNMENT, E-COMMERCE, AND DIGITAL SKILLS: AN ANALYSIS OF TURKEY USING TÜİK MICRODATA
TANIR, Deniz

45 pages, Bibliography included.

KARABUK UNIVERSITY PUBLICATIONS – 145

E-ISBN: 978-625-93475-3-0

December 2025

Contents

CHAPTER 1
INTRODUCTION

Digital technologies have become a defining force in contemporary societies, reshaping public administration, economic structures, and everyday social practices. The widespread diffusion of the internet, mobile devices, and digital platforms has fundamentally transformed how individuals communicate, access information, interact with public institutions, and participate in economic activities. Digitalisation is therefore no longer understood solely as a technological development but as a multidimensional process with deep social, economic, and institutional consequences (Castells, 2010; OECD, 2020).

One of the most prominent outcomes of this transformation is the expansion of e-government services, which aim to improve efficiency, transparency, and accessibility in public service delivery. Governments increasingly rely on digital platforms to reduce administrative costs, simplify bureaucratic procedures, and strengthen citizen–state interaction (United Nations, 2022). In parallel, e-commerce has emerged as a core component of the digital economy, altering traditional consumption patterns and enabling individuals to access goods and services beyond physical marketplaces (Eurostat, 2020). Together, these developments have redefined the boundaries between public administration, private markets, and everyday digital life.

Despite these advances, participation in digital society remains uneven. A growing body of research emphasises that digital inequality cannot be fully explained by differences in internet access alone. Instead, meaningful digital participation depends on individuals' digital skills, levels of digital literacy, and their capacity to manage issues related to privacy, data protection, and online security (van Dijk, 2020; Helsper, 2021). Individuals who lack these competencies may face barriers in using e-government services or engaging in e-commerce, even when technological infrastructure is available. As a result, digital skills have become a key determinant of digital inclusion and digital citizenship.

In the Turkish context, understanding these dimensions of digital transformation requires comprehensive and methodologically robust data sources. The Household Information and Communication Technology (ICT) Usage Survey, conducted annually by the Turkish Statistical Institute (TÜİK), constitutes the primary official dataset for monitoring ICT access, internet usage, e-government engagement, e-commerce behaviour, and digital skills at both household and individual levels. The survey is harmonised with international standards developed by Eurostat and the OECD, ensuring comparability across countries while capturing Türkiye's specific institutional and social dynamics (TÜİK, 2021). Importantly, the availability of microdata enables in-depth analysis of demographic and socioeconomic disparities in digital participation.

This book provides a systematic empirical analysis of e-government usage, e-commerce participation, and digital skills in Türkiye based on the 2021 TÜİK Household ICT Usage Survey microdata. By adopting a microdata-based approach, the study moves beyond aggregate indicators and examines how digital practices vary across age groups, gender, educational attainment, labour market status, income levels, and regional classifications. Such an approach allows for the identification of structural patterns of digital inequality and sheds light on the mechanisms that shape digital participation among different segments of society.

The central objective of this book is not only to measure the prevalence of digital service use but also to assess the depth and quality of digital engagement. Particular attention is paid to the determinants of e-government usage, the factors influencing e-commerce behaviour, and the distribution of digital skills related to information management, software use, content creation, and online security awareness. By integrating these dimensions within a unified analytical framework, the book conceptualises digital participation as an interconnected process shaped by access, skills, trust, and institutional design.

From a policy perspective, understanding the relationships between digital skills, e-government usage, and e-commerce participation is essential for designing inclusive digital strategies. Identifying which groups are disadvantaged in digital transformation—and the reasons behind these disparities—provides critical input for education policies, public service digitalisation, and digital capacity-building initiatives. In this respect, the findings presented in this book aim to contribute to evidence-based policymaking in the areas of digital inclusion, public sector digital transformation, and the development of digital citizenship in Türkiye.

Overall, this book seeks to bridge rigorous empirical analysis with broader theoretical discussions on digital society. By combining nationally representative microdata with advanced analytical methods, it offers a comprehensive portrait of Türkiye's digital landscape and serves as a reference for researchers, policymakers, and practitioners interested in the social dimensions of digital transformation.

Some of the empirical analyses related to e-commerce participation presented in this book build upon and extend the author's previously published work, in which data mining methods were applied to analyse the determinants of individual e-commerce usage in Türkiye

(Tanır & Ramazanov, 2023). In the present volume, these analyses are substantially expanded by incorporating digital skills, e-government participation, and a comparative multi-model framework.

This book is structured to provide a coherent and progressive analysis of digital participation in Türkiye by integrating conceptual discussion, data description, methodological design, and empirical findings.

Chapter 2 presents the population framework, sampling design, and construction of the analytical sample based on the 2021 TÜİK Household ICT Usage Survey microdata. This chapter details the target population, survey methodology, weighting procedures, data collection process, and the preprocessing steps applied to construct the final analytical datasets used in the empirical analyses.

Chapter 3 focuses on variable selection, operational definitions, and analytical preparation. In this chapter, e-commerce usage frequency and e-government usage are formally defined as separate dependent variables, while digital skills are operationalised through a composite index. The chapter explains the rationale behind variable grouping and index construction and establishes the empirical foundation for the modelling framework.

Chapter 4 introduces the data mining and machine learning methods employed in the study. This chapter outlines the classification algorithms used, including decision tree–based methods and support vector machines, and explains the model design, training–testing procedures, and performance evaluation metrics applied to both analytical models.

Chapter 5 presents the empirical findings of the study. Results are reported separately for the e-commerce usage model and the e-government usage model, with particular emphasis on the role of digital skills and socioeconomic factors in shaping digital participation patterns.

Chapter 6 discusses the findings in relation to the existing literature on digital inclusion, digital divide, and digital citizenship. The implications of the results for public policy, digital service design, and skill development strategies are critically evaluated.

Finally, Chapter 7 concludes the book by summarising the main findings, outlining the methodological and empirical contributions, and suggesting directions for future research on digital transformation and inclusive digital development in Türkiye.

CHAPTER 2
POPULATION, SAMPLING DESIGN, AND CONSTRUCTION OF THE ANALYTICAL SAMPLE

Target Population and Scope of the Study

The empirical foundation of this book is based on microdata derived from the Household Information and Communication Technology (ICT) Usage Survey conducted by the Turkish Statistical Institute (TÜİK) in 2021. This survey represents the most comprehensive and authoritative official data source for monitoring the diffusion of information and communication technologies in Türkiye. It is designed in accordance with the methodological framework developed by Eurostat and implemented across European Union member and candidate countries to ensure harmonisation and cross-national comparability.

The survey systematically collects data on a wide range of digital indicators, including household-level ICT access, individual internet usage patterns, engagement with e-government services, participation in e-commerce activities, and the possession of digital skills. By integrating these dimensions within a single survey framework, the dataset allows for a holistic analysis of digital participation and digital inequality at both individual and household levels.

The target population of the survey consists of individuals aged 16–74 years residing in private households within the territorial boundaries of Türkiye. Individuals living in institutional or collective living arrangements—such as dormitories, nursing homes, hospitals, military barracks, prisons, hotels, and similar establishments—are explicitly excluded from the survey universe. This exclusion is consistent with international ICT survey standards and is intended to ensure conceptual and behavioural homogeneity in the measurement of household-based digital practices (TÜİK, 2021).

By adopting this population definition, the study focuses on individuals who are most likely to engage with digital technologies in daily life, including internet use for communication, access to public services, and participation in online markets. This scope is particularly appropriate for analysing e-commerce behaviour and digital skills, as these activities are predominantly embedded in household contexts and individual decision-making processes.

Sampling Design and Sample Selection Procedure

The Household ICT Usage Survey employs a two-stage stratified cluster sampling design, which enables the production of statistically reliable and representative estimates at both the national level and the NUTS Level-1 (İBBS Düzey-1) regional level.

In the first stage of sampling, clusters consisting of approximately 100 households are selected as primary sampling units. These clusters are drawn from address-based population registers using probability proportional to size (PPS) sampling. This method ensures that larger population units have a higher probability of selection, thereby improving representativeness and reducing sampling bias.

In the second stage, households within the selected clusters are chosen using systematic random sampling. All eligible individuals within selected households are included in the survey, subject to the age criteria. Stratification is applied at the NUTS Level-1 regional classification to capture regional heterogeneity in digital access and usage patterns and to reduce sampling variance.

Sample substitution is not permitted in the survey design. Instead, anticipated non-response is incorporated into the initial sample size determination. This approach preserves the probabilistic integrity of the sample and aligns with best practices in official statistics production (TÜİK, 2021). As a result, survey estimates can be generalised to the national population of individuals aged 16–74 residing in private households.

Weighting and Calibration Procedures

Due to the complex sampling structure, survey data are weighted to restore population representativeness and to correct for unequal selection probabilities. Initial base weights are calculated as the inverse of the probability of selection for each observation.

These base weights are subsequently adjusted for unit non-response and coverage errors, accounting for households or individuals that could not be reached or declined participation. Final weights are derived using an integrated calibration approach, whereby survey estimates are aligned with projected population totals by age group, sex, and regional distribution.

The calibration process significantly improves estimation accuracy and reduces bias associated with differential response rates across demographic groups. All empirical analyses presented in this book are conducted using weighted microdata, ensuring that reported results reflect the population-level distribution of digital behaviours in Türkiye (TÜİK, 2021).

Data Collection Method and Questionnaire Structure

Since its inception in 2004, data collection for the Household ICT Usage Survey has primarily been conducted using Computer-Assisted Personal Interviewing (CAPI). However, due to public health restrictions introduced during the COVID-19 pandemic, the 2020 and 2021 survey waves were implemented using Computer-Assisted Telephone Interviewing (CATI) methods.

The questionnaire is structured in accordance with the Eurostat model survey and is periodically revised to reflect emerging digital trends and national policy priorities while maintaining conceptual continuity over time. In 2021, the questionnaire consisted of eight thematic modules:

1. Household characteristics
2. Individual demographic characteristics
3. Mobile phone usage
4. Internet usage
5. E-government services
6. E-commerce activities
7. Digital skills (e-skills)
8. Privacy, data protection, and online security

Each module contains standardised questions designed to capture both behavioural frequency and functional digital competence, allowing for detailed analysis of digital engagement.

Fieldwork Outcomes and Response Profile (2021)

In the 2021 survey wave, data were collected from 16,270 households, of which 14,264 households provided valid responses. At the individual level, interviews were conducted with 46,362 individuals, among whom 33,006 individuals fell within the target age range of 16–74 years. After accounting for non-response and incomplete interviews, the final number of responding individuals in this age group was 30,530.

These figures indicate a high response rate and confirm the robustness and reliability of the dataset for advanced empirical analyses, including machine learning and data mining applications.

Microdata Structure and Units of Analysis

The microdata set comprises 30,531 individual-level observations and 153 variables, where rows represent individual respondents and columns correspond to survey questions and derived indicators. Variables are organised at both the household and individual (fert) levels.

Household-level variables include indicators such as internet access availability, type of internet connection, household income, and ownership of digital devices. Individual-level variables encompass demographic characteristics, internet usage patterns, engagement with e-government services, e-commerce behaviour, digital skills, and attitudes towards online privacy and security.

This dual-level structure enables the integration of socioeconomic context with individual digital behaviour, providing a significant analytical advantage for modelling digital participation.

Construction of the Analytical Sample for E-Commerce Analysis

For the purposes of this book, the analytical focus is placed on individual-level determinants of e-commerce usage frequency. Accordingly, several preprocessing steps were applied to construct the final analytical sample.

First, individuals who reported never using the internet or not using the internet within the last 12 months were excluded from the analysis, as e-commerce-related questions are only administered to recent internet users. This step resulted in the exclusion of 6,202 observations.

Second, for variables related to purchases made within the last three months, individuals who had not engaged in e-commerce during that period were recoded as "no purchase" rather than treated as missing. This ensured logical consistency and prevented structurally defined non-use from being misinterpreted as item non-response.

After the initial data cleaning procedures, a targeted variable selection process was applied in order to retain only analytically relevant indicators related to e-government usage, e-commerce behaviour, digital skills, and key sociodemographic characteristics. As a result, several non-informative or out-of-scope variables were excluded from the dataset. Following these preprocessing and variable reduction steps, the final analytical dataset consisted of 24,329 valid observations, which were subsequently used in all data mining and machine learning analyses presented in the following chapters.

Relevance of the Dataset for Data Mining Applications

The large sample size, rich variable structure, and nationally representative design of the TÜİK Household ICT Usage Survey make it particularly suitable for data mining applications. Unlike small-scale survey studies, this dataset allows for robust training and testing of classification algorithms—such as decision trees and support vector machines—while maintaining strong external validity.

Consequently, the dataset provides a solid empirical foundation for analysing the determinants of e-commerce usage frequency and e-government engagement in Türkiye, as well as for developing predictive classification models with high generalisation capacity.

Variables, Definitıons, and Analytical Framework

This chapter presents the analytical framework of the study, including the selection of dependent and independent variables, the operationalisation of digital skills as a composite index, and the formulation of the empirical models. The chapter establishes the conceptual and methodological basis for the data mining analyses conducted in subsequent chapters.

The analytical strategy adopted in this book treats e-government usage, e-commerce participation, and digital skills as interrelated yet analytically distinct dimensions of digital participation. Rather than relying on a single outcome variable, the study employs a multi-model approach in which different aspects of digital engagement are examined through separate but complementary empirical models.

Dependent Variables

Two main dependent variables are defined in this study.

The first dependent variable represents e-commerce usage. This variable is derived from the survey question asking respondents when they last purchased or ordered goods or services over the internet for private use. Based on this information, e-commerce behaviour is operationalised as a categorical variable capturing different levels of engagement, ranging from recent and frequent use to non-use. This approach allows the analysis to move beyond a binary distinction between users and non-users and to capture the intensity and continuity of participation in online markets, which is widely recommended in the digital economy literature (OECD, 2020; Eurostat, 2025).

The second dependent variable represents e-government usage. This variable captures whether individuals have used digital public services through official websites or mobile platforms for purposes such as obtaining information, downloading official forms, or submitting completed forms online. E-government usage is treated as a separate analytical outcome in order to examine whether engagement with digital public services follows patterns similar to, or distinct from, those observed in private digital market activities. This distinction is particularly important, as prior research indicates that public and private digital services may be shaped by different institutional, trust-related, and skill-based factors (UN, 2022; van Dijk, 2020).

By modelling e-commerce usage and e-government usage separately, the study avoids conceptual conflation and enables a clearer assessment of the determinants of each form of digital participation.

Independent Variables and Control Factors

The set of independent variables used in both models includes demographic, socioeconomic, and behavioural indicators drawn from the TÜİK Household ICT Usage Survey microdata.

Demographic variables include age, gender, and educational attainment. These variables represent structural characteristics that have been consistently associated with digital inequalities in previous studies (Helsper, 2021; van Dijk, 2020). Socioeconomic variables include employment status and household income, which reflect individuals' material resources and economic capacity to engage with digital technologies.

In addition, variables capturing general internet usage behaviour—such as frequency of internet use, types of devices used, and common online activities—are included to control for baseline digital exposure. These variables ensure that the estimated effects of digital skills and other factors are not merely proxies for overall internet familiarity.

Construction of the Digital Skills Composite Index

A central contribution of this book is the operationalisation of digital skills as a composite index, which is used as a key explanatory variable in the empirical models. Rather than relying on single-item indicators, digital skills are conceptualised as a multidimensional construct encompassing functional, informational, and security-related competencies.

The digital skills index is constructed using survey items that capture individuals' ability to perform a range of digital tasks, including installing and updating software, managing device settings, using office and productivity applications, creating or editing digital content, and applying basic online security and privacy practices. Each skill item is coded to reflect whether the respondent is able to perform the task independently.

These items are then aggregated to form a composite digital skills index, where higher values indicate higher levels of digital competence. The index construction follows established practices in digital skills measurement and aligns with Eurostat and OECD recommendations for assessing functional digital literacy (Eurostat, 2020; OECD, 2019).

Using a composite index provides several analytical advantages. First, it captures the cumulative nature of digital competencies rather than isolated abilities. Second, it reduces measurement error associated with single-item indicators. Third, it enables the examination of how overall digital competence influences engagement with both public and private digital services.

Analytical Models and Conceptual Structure

Based on the variable structure described above, the empirical analysis is organised around two main analytical models.

The first model examines the determinants of e-commerce usage, with e-commerce participation as the dependent variable and demographic characteristics, socioeconomic factors, internet usage patterns, e-government usage, and the digital skills index as explanatory variables.

The second model focuses on e-government usage as the dependent variable. In this model, the digital skills index plays a central explanatory role, alongside demographic and socioeconomic controls and general internet usage indicators. This model allows for the assessment of whether digital skills are more strongly associated with public digital service use than with private market activities.

By estimating these two models separately, the study is able to compare the relative importance of digital skills across different domains of digital participation and to identify domain-specific patterns of digital inclusion and exclusion.

Methodological Implications

The analytical framework presented in this chapter enables a nuanced examination of digital participation in Türkiye. By distinguishing between e-commerce and e-government usage and by incorporating a composite measure of digital skills, the study moves beyond simplified adoption metrics and provides a multidimensional perspective on digital inclusion.

The variables and models defined in this chapter form the empirical foundation for the data mining and machine learning analyses presented in the next chapter, where classification algorithms are applied to identify key determinants and predictive patterns of digital behaviour.

CHAPTER 3
DATA MINING METHODS AND MODEL SPECIFICATION

This chapter presents the methodological framework employed to analyse e-commerce usage and e-government participation in Türkiye. The chapter introduces the data mining approach adopted in the study, explains the rationale for using classification-based methods, and describes the model specification, training–testing strategy, and evaluation criteria applied to the empirical analyses.

The analytical strategy is guided by the objective of identifying complex and potentially non-linear relationships between digital skills, socioeconomic characteristics, and patterns of digital participation. Given the large sample size, categorical variable structure, and multidimensional nature of the data, data mining techniques provide a suitable and robust methodological framework for this purpose.

Rationale for a Data Mining Approach

Traditional regression-based methods are widely used in studies of digital inclusion and technology adoption. However, such methods typically rely on strong assumptions regarding linearity, normality, and functional form. In contrast, data mining techniques are specifically designed to handle large-scale datasets with complex interaction structures, mixed variable types, and non-linear relationships (Han, Kamber, & Pei, 2012).

The TÜİK Household ICT Usage Survey microdata exhibit precisely these characteristics. The dataset includes a large number of categorical and ordinal variables, behavioural indicators with asymmetric distributions, and potentially interacting predictors related to demographic characteristics, internet use patterns, digital skills, and public service engagement. Data mining methods are therefore particularly appropriate for uncovering hidden structures and classification patterns within the data.

In this study, data mining is not employed as a purely predictive exercise. Instead, it is used as an exploratory–analytical tool to identify key determinants of digital participation and to assess the relative importance of digital skills in shaping both e-commerce usage and e-government engagement. This approach aligns with recent methodological developments in social science research, where machine learning techniques are increasingly integrated with theory-driven analysis (Athey & Imbens, 2019).

Model Structure and Dependent Variables

Two separate classification models are specified in accordance with the analytical framework established in Chapter 3.

The first model focuses on e-commerce usage, where the dependent variable captures individuals' engagement in online purchasing activities. This model aims to identify the factors associated with different levels of e-commerce participation, with particular attention to the role of digital skills, demographic characteristics, and socioeconomic conditions.

The second model examines e-government usage as a distinct outcome variable. In this model, the dependent variable reflects whether individuals have used digital public services for information retrieval, form downloading, or online submission of official documents. Treating e-government usage as a separate dependent variable allows for a comparative assessment of how digital skills operate across public and private domains of digital participation.

In both models, the digital skills composite index constitutes a key explanatory variable, while demographic, socioeconomic, and general internet usage indicators serve as control variables. This parallel model structure enables systematic comparison of the determinants of e-commerce and e-government engagement.

Classification Algorithms and Training Procedure

The empirical analysis employs supervised classification algorithms, which are well suited for modelling categorical outcome variables. Among the available data mining techniques, decision tree–based methods and margin-based classifiers are particularly appropriate for survey data due to their interpretability and robustness.

Decision tree algorithms are used to identify hierarchical decision rules that segment the population into groups with distinct digital participation patterns. These methods allow for intuitive interpretation of results by highlighting threshold effects and interaction structures among predictors. In addition, support vector machines are employed to assess classification performance in high-dimensional feature spaces and to validate the robustness of the findings.

The analytical dataset is randomly divided into training and testing subsets, following standard practices in machine learning research. Models are trained on the training set and evaluated on the testing set to assess out-of-sample performance. This procedure reduces the risk of overfitting and ensures that the reported results reflect generalisable patterns rather than sample-specific noise (James et al., 2013).

All models are estimated using weighted data, ensuring that classification results reflect population-level distributions rather than unweighted sample characteristics.

Model Evaluation and Performance Metrics

Model performance is evaluated using multiple criteria to ensure a comprehensive assessment of classification quality. Accuracy rates are used as a basic measure of overall classification success. However, given the potential imbalance between user and non-user categories, additional metrics such as precision, recall, and the F1-score are also considered.

For comparative purposes, performance metrics are reported separately for the e-commerce and e-government models. This enables the evaluation of whether digital skills contribute differently to classification accuracy across public and private digital domains.

Beyond predictive performance, particular emphasis is placed on variable importance measures derived from the classification models. These measures provide substantive insights into which factors play the most influential role in shaping digital participation. In line with the analytical objectives of the book, the interpretation of results prioritises explanatory relevance over purely predictive accuracy.

Methodological Contribution

The methodological approach adopted in this study contributes to the digital inclusion literature in two key ways. First, by integrating data mining techniques with official survey microdata, the study demonstrates the analytical potential of large-scale public datasets beyond traditional descriptive analysis. Second, by modelling e-commerce usage and e-government participation separately while incorporating a composite digital skills index, the study provides a nuanced empirical framework for understanding multidimensional digital participation.

The methods described in this chapter form the basis for the empirical findings presented in the next chapter, where classification results are reported and interpreted in detail.

Methodological Box 3.1

Construction of the Digital Skills Composite Index

This study operationalises digital skills as a multidimensional construct by developing a composite digital skills index based on individual-level microdata from the TÜİK Household ICT Usage Survey. Rather than relying on a single proxy indicator, the index captures functional, operational, and cognitive dimensions of digital competence, in line with contemporary digital inclusion frameworks (van Dijk, 2020; OECD, 2021).

Conceptual Framework

Digital skills are conceptualised as an enabling capability that allows individuals to effectively access, use, and benefit from digital technologies across both public and private domains. Consistent with the Digital Competence Framework (DigComp) proposed by the European Commission, the index reflects four core dimensions of digital skills:

1. **Operational and technical skills**
2. **Information and communication skills**
3. **Content creation and software use skills**
4. **Digital safety and privacy awareness**

This multidimensional structure recognises that meaningful digital participation depends not only on basic access or usage frequency but also on the ability to perform diverse digital tasks and manage online risks.

Selection of Skill Indicators

The composite index is constructed using binary indicators derived from the e-skills and privacy modules of the TÜİK survey. Selected indicators include individuals' ability to:

- Install software or mobile applications
- Change device or application settings
- Use word processing software
- Use spreadsheet software
- Create digital presentations
- Edit photos, videos, or audio files
- Transfer files between devices
- Write code in a programming language
- Adjust privacy settings and manage cookies
- Verify the reliability of online information

Each indicator is coded as:

$$x_{ij} = \begin{cases} 1, & \text{if individual } i \text{ reports ability to perform skill } j \\ 0, & \text{otherwise} \end{cases}$$

where $i = 1,2, \dots, N$ individuals and $j = 1,2, \dots, K$ skill indicators.

Index Construction and Normalisation

The digital skills composite index is calculated as the normalised sum of skill indicators for each individual:

$$DSI_i = \frac{1}{K}\sum_{j=1}^{K} x_{ij}$$

where:

- DSI_i denotes the digital skills index score of individual i,
- K represents the total number of skill indicators included in the index.

This formulation yields a continuous index bounded between 0 and 1, where higher values indicate greater levels of digital competence. Normalisation ensures comparability across individuals and facilitates integration of the index into classification and machine learning models.

Categorisation for Classification Models

For the purposes of classification analysis, the continuous index is further categorised into ordinal skill levels using distribution-based thresholds:

- **Low digital skills**: $DSI_i < P_{33}$
- **Medium digital skills**: $P_{33} \leq DSI_i < P_{66}$
- **High digital skills**: $DSI_i \geq P_{66}$

where P_{33} and P_{66} denote the 33rd and 66th percentiles of the index distribution. This categorisation balances interpretability with statistical robustness and avoids arbitrary cut-off points.

Reliability and Analytical Validity

The internal consistency of the digital skills indicators was assessed prior to index construction. The composite structure ensures that the index captures variation across multiple functional domains rather than reflecting a single technical ability. Previous studies employing similar additive or normalised indices have demonstrated strong analytical validity in modelling digital participation outcomes (Helsper, 2021; van Deursen & van Dijk, 2014).

By incorporating the digital skills composite index as a central explanatory variable, the study explicitly models skills as a mediating factor between access conditions and digital participation outcomes, including both e-commerce usage and e-government engagement.

Analytical Contribution

The construction of a composite digital skills index constitutes a methodological contribution of this book. Unlike studies that treat digital skills as isolated indicators, this approach provides a holistic and empirically grounded measure of individual digital competence. The index enables direct comparison of skill levels across demographic and socioeconomic groups and enhances the explanatory power of classification models applied in later chapters.

Methodological Box 3.2

Operational Definition of the E-Government Usage Variable

In this book, e-government usage is treated as a distinct dimension of digital participation and modelled as a separate dependent variable. This approach reflects the conceptual distinction between engagement with digital public services and participation in private digital markets, such as e-commerce. While both domains rely on digital infrastructure and skills, they differ substantially in terms of institutional context, motivation, trust mechanisms, and usage objectives (OECD, 2020; UN, 2022).

Conceptual Rationale

E-government usage refers to individuals' interaction with public authorities through digital channels for administrative, informational, or transactional purposes. In the digital governance literature, e-government is widely recognised as a core component of digital citizenship, enabling citizens to access public services more efficiently and to interact with the state beyond traditional bureaucratic channels (UN, 2022; Bannister & Connolly, 2014).

Importantly, the use of e-government services requires not only internet access but also sufficient levels of digital skills, procedural understanding, and trust in digital public institutions. For this reason, modelling e-government usage as a separate dependent variable allows for a more precise assessment of how digital skills and socioeconomic factors shape participation in digital governance.

Derivation from Survey Data

The e-government usage variable is derived from the e-government module of the TÜİK Household ICT Usage Survey. Respondents are asked whether they have used digital public services via websites or mobile applications within a specified reference period. These services include activities such as:

- Obtaining information from public institution websites,
- Downloading official forms,
- Submitting completed forms electronically,
- Conducting administrative procedures through e-government portals.

Based on responses to these items, an individual-level indicator of e-government usage is constructed.

Variable Coding and Classification

The dependent variable is operationalised as a binary indicator defined as:

$$EGOV_i = \begin{cases} 1, & \text{if individual } i \text{ has used any e-government service} \\ 0, & \text{if individual } i \text{ has not used e-government services} \end{cases}$$

This binary specification is consistent with international ICT survey practices and ensures compatibility with supervised classification algorithms used in the empirical analysis (Eurostat, 2020).

Individuals who reported never using the internet or not using the internet within the last twelve months were excluded prior to variable construction, as e-government questions are administered only to recent internet users. This preprocessing step prevents structural non-users from being misclassified as non-participants.

Analytical Role in the Modelling Framework

In the empirical analyses presented in later chapters, the e-government usage variable serves as the dependent variable in the second classification model. Digital skills, measured through the composite index described in Methodological Box 3.1, constitute the primary explanatory variable, while demographic characteristics, socioeconomic indicators, and general internet usage behaviours are included as control variables.

This modelling strategy enables a direct comparison between the determinants of e-commerce participation and e-government engagement, highlighting whether digital skills exert similar or differentiated effects across public and private domains of digital participation.

Methodological Contribution

By explicitly defining e-government usage as a standalone dependent variable, this study advances the empirical literature on digital inclusion. Rather than subsuming public service use under general internet activity, the approach adopted here acknowledges the institutional specificity of e-government and allows for a more nuanced understanding of digital citizenship in Türkiye.

Conceptual Framework of the Analytical Models

This book adopts a dual-model analytical framework to examine digital participation in Türkiye. Rather than modelling digital behaviour as a single outcome, the framework distinguishes between private digital market participation and public digital service engagement. Accordingly, two separate but structurally comparable models are specified: one for e-commerce usage and one for e-government usage.

At the core of both models lies the assumption that digital skills function as a central enabling mechanism that mediates the relationship between individual characteristics and digital participation outcomes. Digital skills are therefore conceptualised not merely as a background characteristic, but as an active determinant shaping individuals' capacity to engage meaningfully with digital platforms.

Model Structure

Both analytical models share a common set of explanatory dimensions:

- **Demographic characteristics**, including age, gender, and educational attainment, represent structural factors associated with digital inequality.
- **Socioeconomic conditions**, such as employment status and household income, reflect material resources and opportunity structures.
- **General internet usage behaviours**, including frequency of use and common online activities, capture familiarity with digital environments.
- **Digital skills**, operationalised through the composite index described in Methodological Box 3.1, represent functional digital competence.

These explanatory dimensions jointly influence two distinct dependent variables:

- **E-commerce usage**, reflecting participation in private online market activities.
- **E-government usage**, reflecting engagement with digital public services and administrative platforms.

While both outcomes rely on similar infrastructural and skill-related prerequisites, they differ in institutional context, motivation, and trust requirements. This distinction allows the framework to assess whether digital skills and socioeconomic factors exert symmetric or asymmetric effects across public and private domains of digital participation.

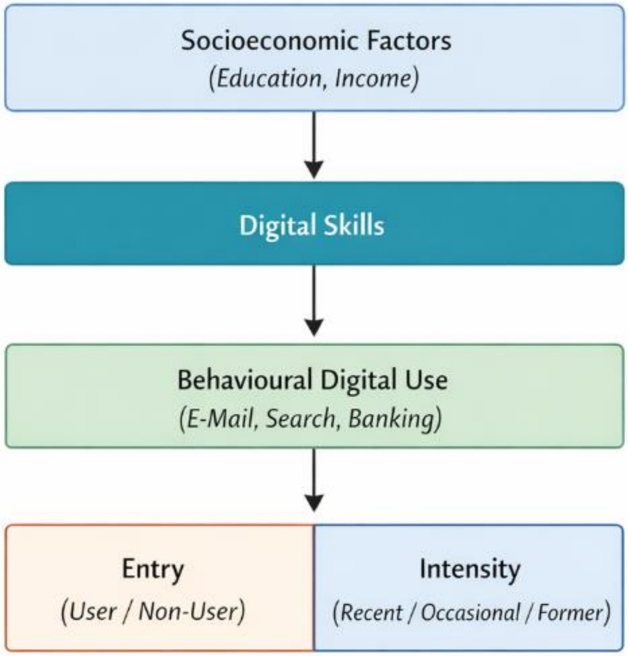

Figure 3.1. Conceptual Framework of Digital Participation Models

Figure 3.1 illustrates the dual-model structure employed in the study. Demographic and socioeconomic characteristics, general internet usage, and the digital skills composite index jointly influence e-commerce usage and e-government usage as two separate but interrelated dimensions of digital participation.

Analytical Implications

The comparative structure of the framework enables several analytical advantages. First, it allows for the identification of shared determinants of digital participation across domains. Second, it highlights domain-specific effects, particularly in relation to digital skills and trust-sensitive behaviours such as e-government usage. Third, it provides a coherent basis for interpreting classification results in a parallel and systematic manner.

By explicitly modelling e-commerce and e-government as separate dependent variables within a unified conceptual framework, the study advances a more nuanced understanding of digital inclusion. Digital participation is thus framed as a multidimensional process shaped by individual capabilities, socioeconomic context, and institutional environments.

This conceptual framework guides the empirical analyses presented in subsequent chapters and serves as an interpretive lens for discussing the findings and their policy implications.

CHAPTER 4
DATA MINING METHODS AND MODEL IMPLEMENTATION

This chapter presents the data mining methodology employed to analyse patterns of digital participation in Türkiye. Building on the variable definitions and data preparation procedures described in Chapter 3, it outlines the analytical strategy, modelling logic, and evaluation framework used to examine e-commerce behaviour. Particular emphasis is placed on the role of digital skills as an enabling factor shaping individuals' engagement with digital markets.

Given the large-scale structure of the TÜİK microdata and the predominantly categorical nature of the explanatory variables, classification-based data mining methods were selected as the primary analytical approach. These methods are especially well suited to capturing non-linear relationships and interaction effects that are difficult to model using traditional parametric techniques.

Analytical Strategy and Modelling Framework

The study adopts a supervised learning framework in which the dependent variable is observed and the primary objective is to identify the factors that best explain variation across outcome categories. In the context of digital participation, supervised classification facilitates the identification of both dominant predictors and behavioural thresholds that distinguish different user profiles.

Preliminary exploratory analyses revealed substantial class imbalance in the e-commerce usage variable, particularly among occasional and former users. To address this challenge and to reflect the sequential nature of digital engagement, a two-stage classification strategy was adopted.

In the first stage, individuals were classified according to whether they had participated in e-commerce activities (users versus non-users). This stage captures the fundamental distinction between digital inclusion and exclusion in online markets. In the second stage, the analysis was restricted to individuals with prior e-commerce experience, allowing a refined classification of users into recent, occasional, and former users. This hierarchical approach enables a more nuanced modelling of behavioural heterogeneity while mitigating the effects of class imbalance.

Classification Algorithms and Estimation Procedure

Decision tree–based classification models were employed as the primary analytical tool. Decision trees are particularly advantageous in applied social science research due to their interpretability and their ability to generate rule-based structures that clearly illustrate how combinations of explanatory variables lead to specific outcomes. These characteristics are especially valuable when analysing survey microdata composed of categorical and behavioural indicators (Breiman et al., 1998; Han et al., 2012).

All models were estimated using the CART (Classification and Regression Tree) algorithm implemented in the rpart package in R. To prevent overfitting and to improve generalisability, tree complexity was controlled through cost–complexity pruning, minimum node size constraints, and internal cross-validation. Optimal tree size was determined by selecting the complexity parameter that minimised cross-validated classification error.

The dataset was randomly divided into training and testing subsets using a 70/30 split. Models were trained on the training data and evaluated on the test data to assess out-of-sample performance, ensuring that reported results reflect generalisable patterns rather than sample-specific artefacts.

Model Evaluation Criteria

Model performance was evaluated using multiple classification metrics. Overall accuracy was reported as a general indicator of predictive success; however, given the uneven distribution of outcome categories, accuracy alone was not considered sufficient.

Class-specific sensitivity and specificity measures were used to assess the model's ability to correctly identify both dominant and minority user groups. In addition, balanced accuracy and Cohen's kappa statistics were reported to account for chance agreement and class imbalance. This multi-metric evaluation framework provides a comprehensive assessment of model quality and supports meaningful comparison across modelling stages.

Beyond predictive performance, particular emphasis was placed on variable importance measures derived from the decision tree models. These measures indicate the relative contribution of each explanatory variable to the classification process and provide substantive insights into the mechanisms underlying digital participation.

Two-Stage Modelling of E-Commerce Participation

In the first stage, a binary classification model was estimated to distinguish between individuals who had engaged in e-commerce and those who had never participated. This model achieved substantially higher predictive performance than a baseline majority-class classifier, demonstrating that digital behaviour and skill-related variables provide strong explanatory power beyond basic demographic characteristics.

In the second stage, the analysis focused exclusively on individuals with prior e-commerce experience. By excluding non-users, the model more effectively captured behavioural differences between recent, occasional, and former users. The second-stage model yielded a marked improvement in classification performance, with high overall accuracy and substantial agreement beyond chance.

Notably, recent users were identified with near-perfect sensitivity and specificity, indicating a highly distinctive behavioural and skill profile. Occasional and former users—representing transitional and heterogeneous groups—were also identified with substantially improved accuracy compared to single-stage models. These findings underscore the analytical value of a hierarchical modelling strategy when examining complex digital behaviours.

Role of Digital Skills in the Modelling Framework

Across both modelling stages, the digital skills composite index emerged as a central explanatory variable. Rather than functioning as a secondary control factor, digital skills played a structurally significant role in distinguishing both participation in e-commerce and intensity of use.

The prominence of the digital skills index relative to traditional demographic variables such as age and region highlights the importance of functional digital competence in shaping digital market participation. This finding supports conceptualisations of digital inclusion that emphasise skills and effective use—rather than access alone—as the primary drivers of meaningful digital engagement.

Summary

This chapter presented a structured data mining framework for analysing e-commerce participation in Türkiye using nationally representative microdata. By combining decision tree–based classification with a two-stage modelling strategy, the analysis addressed class imbalance and behavioural heterogeneity while preserving interpretability.

The methodological approach outlined in this chapter provides a robust foundation for the empirical findings presented in the following chapter, where results from both modelling stages are examined in detail and discussed in relation to existing literature on digital inclusion, e-commerce, and digital skills.

CHAPTER 5
RESULTS

This chapter presents the empirical findings derived from the data mining analyses of e-commerce participation in Türkiye. Building on the methodological framework outlined in Chapter 4, results are reported for both stages of the hierarchical classification strategy. The analysis focuses on descriptive patterns of e-commerce participation, model performance, classification accuracy across user groups, and the relative importance of explanatory variables, with particular attention to the role of digital skills.

Descriptive Overview of E-Commerce Participation

This section presents a descriptive overview of e-commerce participation patterns in Türkiye based on the final analytical sample. Table 5.1 summarises the distribution of individuals across e-commerce usage categories, providing an initial empirical context for the subsequent modelling results.

Table 5.1. Distribution of E-Commerce Usage Categories

E-Commerce Usage Category	N	Percentage (%)
Recent users	9,438	38.8
Occasional users	2,214	9.1
Former users	1,291	5.3
Never users	11,385	46.8
Total	24,328	100.0

As shown in Table 5.1, nearly half of the population (46.8%) has never engaged in e-commerce activities, indicating a substantial level of non-participation in digital markets. In contrast, recent users constitute approximately 39% of the sample, representing individuals who are actively integrated into online purchasing environments.

Occasional users (9.1%) and former users (5.3%) together form a smaller but analytically important segment. These groups capture transitional patterns of digital behaviour, encompassing individuals who engage in e-commerce sporadically or who have discontinued online purchasing despite prior experience. The presence of these intermediate categories highlights the heterogeneity of digital participation and suggests that e-commerce engagement should be conceptualised as a continuum rather than a binary outcome.

The observed imbalance across usage categories has important methodological implications for classification modelling and directly motivates the adoption of a two-stage analytical strategy. By first distinguishing users from non-users and subsequently differentiating between types of users, the analysis captures both the extensive margin of e-commerce adoption and the intensive margin of usage among experienced participants.

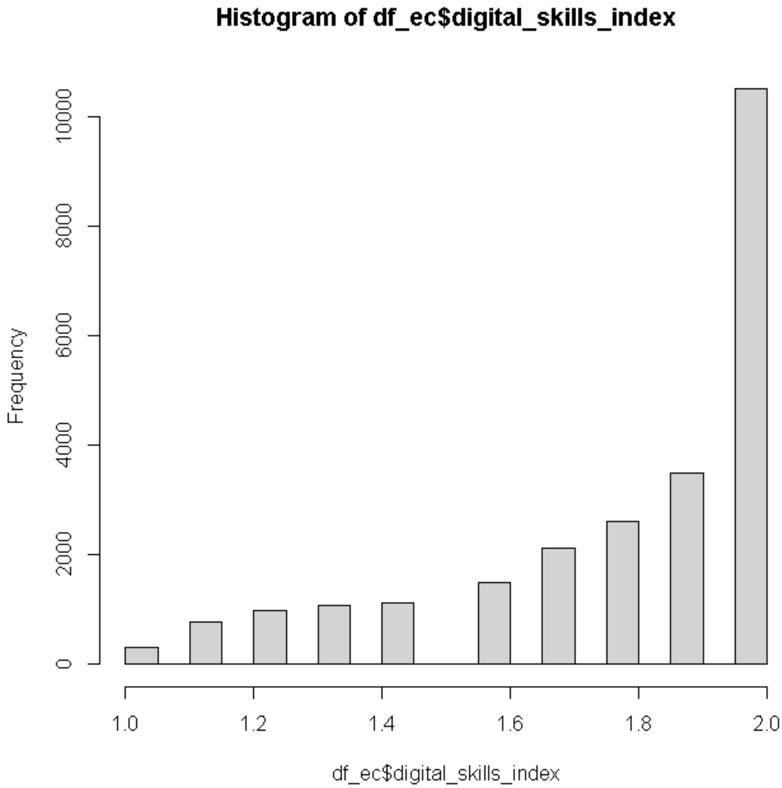

Figure 5.1. Distribution of the digital skills composite index.

The distribution is right-skewed, with a substantial concentration of individuals at higher skill levels, indicating pronounced heterogeneity in functional digital competence across the population. This heterogeneity provides an empirical basis for the strong predictive role of digital skills observed in subsequent classification models.

Stage 1 Results: User vs. Non-User Classification

The first stage of the analysis focuses on distinguishing individuals who actively participate in e-commerce from those who do not. At this stage, e-commerce participation is modelled as a binary outcome, capturing the fundamental distinction between digital inclusion and exclusion in online markets.

Recent users were classified as users, while occasional users, former users, and never users were grouped as non-users. This formulation emphasises active engagement in e-commerce rather than sporadic or discontinued use.

Model Performance

The binary classification model was estimated using a decision tree algorithm and evaluated on an independent test dataset. Table 5.2 reports the key performance metrics.

Table 5.2. Performance of the Binary E-Commerce Participation Model

Metric	Value
Accuracy	78.7%
Sensitivity (User)	67.6%
Specificity (Non-user)	85.7%
Positive Predictive Value	74.9%
Negative Predictive Value	80.7%
Balanced Accuracy	76.7%
Cohen's Kappa	0.543

The model substantially outperformed a baseline majority-class classifier, which achieved an accuracy of approximately 61%. Sensitivity results indicate that roughly two-thirds of active e-commerce users were correctly identified, while specificity exceeded 85%, reflecting strong performance in classifying non-users. Balanced accuracy and Cohen's kappa confirm substantial agreement beyond chance.

Variable Importance and Determinants of Participation

Variable importance measures reveal a clear hierarchy of determinants shaping e-commerce participation.

Table 5.3. Variable Importance in the Binary E-Commerce Participation Model (Stage 1)

Rank	Variable	Importance Score
1	Internet e-mail usage	1805.1
2	Online information search for goods and services	872.4
3	Educational attainment	821.0
4	Digital skills composite index	751.2
5	Online banking activities	641.1
6	Household monthly income	342.9
7	Online news consumption	57.6
8	Age	7.4
9	NUTS-1 region	2.1

Behavioural internet use variables and digital skills exert a substantially stronger influence on e-commerce participation than traditional demographic characteristics. Notably, the digital skills composite index ranks above age and region, underscoring the pivotal role of functional digital competence in enabling participation in online markets. Regional differences exhibit minimal explanatory power once behavioural and skill-related factors are taken into account, indicating that spatial disparities operate largely through individual digital behaviour rather than geographic location per se.

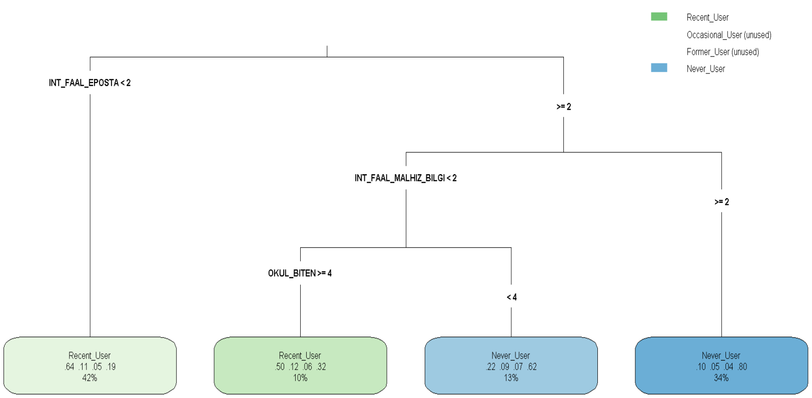

Figure 5.2. Decision tree illustrating the primary determinants of e-commerce participation (Stage 1).

Figure 5.2 provides an interpretable representation of the classification logic underlying the first-stage model. The initial splits are driven by behavioural indicators of internet use, most notably e-mail activity and online information search for goods and services. Educational attainment emerges at a later stage, reinforcing the finding that functional digital engagement precedes sociodemographic characteristics in shaping e-commerce participation.

Stage 2 Results: Differentiation Among E-Commerce Users

The second stage of the analysis focuses exclusively on individuals with prior e-commerce experience. Within this subpopulation, a multi-class classification model was estimated to distinguish between recent, occasional, and former users.

Restricting the analysis to users resulted in a marked improvement in classification performance. The second-stage model achieved an overall accuracy of approximately 88%, with a Cohen's kappa value of 0.716, indicating substantial agreement beyond chance.

Table 5.4. Confusion Matrix for Stage 2 E-Commerce User Classification

Predicted\ Actual	Recent User	Occasional User	Former User	Total
Recent User	2,831	0	0	2,831
Occasional User	0	366	177	543
Former User	0	298	210	508
Total	2831	664	387	3882

Recent users were identified with near-perfect sensitivity and specificity, indicating a highly distinctive behavioural profile. Occasional and former users were also classified with substantially improved accuracy compared to single-stage models, although some overlap between these groups remained, reflecting the gradual nature of disengagement from e-commerce.

Table 5.5. Performance Metrics for Stage 2 Classification Model Metric

Metric	Value
Overall accuracy	87.8%
Cohen's kappa	0.716
Sensitivity – Recent users	1.000
Sensitivity – Occasional users	0.551
Sensitivity – Former users	0.543
Balanced accuracy – Occasional users	0.748
Balanced accuracy – Former users	0.729

Table 5.6. Variable Importance in the Stage 2 Classification Model

Rank	Variable	Importance Score
1	Digital skills composite index	1,204.6
2	Online information search for goods and services	912.3
3	Internet e-mail usage	865.7
4	Online banking activities	733.4
5	Educational attainment	654.2
6	Frequency of internet use	421.8
7	Daily internet use	298.5
8	Household monthly income	215.6
9	Online social participation	146.9
10	Age	54.2
11	NUTS-1 region	18.7

Digital skills emerged as the most influential determinant of usage intensity among e-commerce users. Behavioural internet use variables also exhibited high explanatory power, while demographic and regional variables contributed relatively little once skills and usage patterns were accounted for.

Summary of Key Findings: E-Commerce

The results demonstrate that e-commerce participation in Türkiye is characterised by substantial heterogeneity, both between users and non-users and among user subgroups. The two-stage classification strategy substantially outperformed single-stage models, particularly in identifying transitional user categories.

Most importantly, digital skills emerged as a central determinant of both entry into e-commerce and sustained participation. The prominence of the digital skills composite index across all models highlights the importance of functional competence in navigating digital markets and provides a strong empirical foundation for the discussion presented in the following chapter.

Descriptive Overview of e-Government Participation

Table 5.7 presents the distribution of e-government usage among individuals included in the analytical sample. A substantial majority of the population (70.6%) reported using at least one e-government service during the reference period, indicating a relatively high level of engagement with digital public services. In contrast, approximately 29.4% of individuals had not interacted with e-government platforms, despite reporting internet use.

Table 5.7. Distribution of e-Government Usage

E-Government usage	N	Percentage (%)
User	17,175	70.6
Non-user	7,153	29.4
Total	24,328	100.0

This distribution suggests that e-government adoption in Türkiye is considerably more widespread than e-commerce participation. Nevertheless, the persistence of a sizable non-user group highlights the continued relevance of digital skills, usage patterns, and behavioural barriers in accessing public digital services.

Stage 1 Results: Binary Classification of e-Government Use

Table 5.8. Performance of the Binary e-Government Participation Model

Metric	Value
Accuracy	79.4%
Sensitivity (User)	90.3%
Specificity (Non-user)	49.6%
Positive Predictive Value	83.1%
Negative Predictive Value	64.9%
Balanced Accuracy	69.9%
Cohen's Kappa	0.431

The binary classification model achieved an overall accuracy of 79.4%, exceeding the no-information rate of 73.3%, confirming meaningful predictive improvement beyond class prevalence. Sensitivity results indicate that the model is highly effective in identifying e-government users, correctly classifying more than 90% of individuals who engage with digital public services. By contrast, specificity remained moderate (49.6%), reflecting a lower ability to correctly identify non-users. This asymmetry is partly attributable to class imbalance in favour of users and suggests that non-use may be shaped by more heterogeneous and potentially unobserved barriers.

Table 5.9. Variable Importance in the Binary e-Government Participation Model

Rank	Variable	Importance Score
1	Online banking activities	947.6
2	Digital skills composite index	645.5
3	Internet e-mail usage	408.3
4	Online information search for goods and services	392.3
5	Educational attainment	380.3
6	Employment status	272.4

Rank	Variable	Importance Score
7	NUTS-1 region	126.2
8	Age	106.4
9	Online news consumption	85.1
10	Online social participation	10.0
11	Daily internet use	1.8
12	Gender	0.8

Table 5.9 shows that transactional digital behaviour is central in predicting e-government participation. Online banking activities emerged as the most influential predictor, followed by the digital skills composite index and e-mail usage. Educational attainment and employment status also contributed meaningfully, reflecting the role of human capital and labour market attachment. By contrast, gender and daily internet use exhibited minimal explanatory power. Regional differences showed limited importance once individual digital behaviours and skills were accounted for, suggesting that spatial disparities in e-government participation are largely mediated through functional digital engagement rather than geographic location per se.

Summary of Key Findings: E-Government

Overall, the e-government model achieved strong performance in identifying users but demonstrated weaker discrimination for non-users, consistent with the higher prevalence and broader normalisation of digital public service use. Variable importance results indicate that engagement with transactional online services and digital skills are decisive in explaining adoption, while demographic and regional factors play a comparatively limited role once behavioural engagement is taken into account.

CHAPTER 6
DISCUSSION

This chapter discusses the empirical findings presented in Chapter 5 in relation to the broader literature on digital inclusion, e-commerce adoption, and e-government use. The discussion focuses on three interrelated themes: (i) heterogeneity in digital participation, (ii) the comparative determinants of e-commerce and e-government engagement, and (iii) the central role of digital skills in shaping both market-oriented and public digital service use.

Heterogeneity in Digital Participation

The results demonstrate that digital participation in Türkiye cannot be adequately described using binary distinctions alone. In the context of e-commerce, nearly half of the population has never engaged in online purchasing, while a substantial share exhibits intermediate usage patterns as occasional or former users. This heterogeneity supports conceptualisations of digital engagement as a continuum rather than a dichotomous outcome, consistent with prior research emphasising gradations of digital inclusion.

The two-stage classification strategy adopted in this study directly addresses this complexity. By first distinguishing users from non-users and subsequently differentiating among user types, the analysis captures both the extensive margin of adoption and the intensive margin of sustained engagement. The marked improvement in classification performance observed in the second stage confirms that behavioural heterogeneity among users is substantial and analytically meaningful.

In contrast, e-government participation exhibits a more concentrated distribution, with a clear majority of individuals reporting use of digital public services. This pattern reflects the institutionalisation and growing normalisation of e-government platforms in Türkiye. However, the persistence of a sizeable non-user group indicates that widespread availability does not translate automatically into universal adoption.

Comparing E-Commerce and E-Government Participation

A central contribution of this study lies in the comparative analysis of e-commerce and e-government participation using a consistent data mining framework. While both domains rely on digital infrastructure and individual internet access, the determinants of participation differ in important ways.

E-commerce participation is strongly driven by behavioural indicators of active and voluntary digital engagement, such as online information search, e-mail use, and online banking activities. These behaviours reflect familiarity with transactional environments and a willingness to engage in market-mediated exchanges. The relatively lower prevalence of e-commerce users suggests that participation remains selective and closely tied to individual digital competencies and consumption-oriented practices.

By contrast, e-government participation is characterised by substantially higher prevalence and stronger sensitivity in classification models. The prominence of online banking as the most influential predictor indicates that transactional trust and experience with formal digital systems are particularly important for engagement with public digital services. This finding aligns with the notion that e-government use is less discretionary than e-commerce and often linked to administrative necessity rather than consumer choice.

Despite these differences, both domains exhibit limited explanatory power for purely demographic and regional variables once digital behaviours and skills are accounted for. This convergence suggests that structural access alone is insufficient to explain observed participation patterns and that effective use capabilities play a decisive role across digital contexts.

The Central Role of Digital Skills

Across all models estimated in this study, the digital skills composite index emerged as a consistently strong predictor of digital participation. In the e-commerce models, digital skills influenced both the likelihood of entry into online markets and the intensity of continued use. In the e-government model, digital skills ranked among the top predictors, second only to online banking activities.

These findings provide robust empirical support for theoretical frameworks that conceptualise digital inclusion in terms of skills and effective use rather than mere access. The right-skewed distribution of the digital skills index further underscores the unequal distribution of functional digital competence within the population. Individuals with higher skill levels are not only more likely to adopt digital services but also more likely to sustain and diversify their engagement.

Importantly, the relatively weak influence of age, gender, and region once skills are controlled for suggests that observed sociodemographic disparities operate largely through skill acquisition and behavioural pathways. This has significant policy implications, as it implies that interventions aimed at improving digital skills may yield broader returns than those focusing solely on infrastructure expansion.

Methodological Implications

From a methodological perspective, the findings highlight the value of decision tree–based classification models in analysing complex digital behaviours. The interpretability of decision trees allows for transparent identification of behavioural thresholds and interaction effects that are difficult to capture using traditional parametric approaches.

The two-stage modelling strategy proved particularly effective in addressing class imbalance and uncovering latent heterogeneity among user groups. The near-perfect classification of recent e-commerce users in the second stage indicates that sustained digital engagement is associated with a distinct and stable behavioural profile. At the same time, the overlap observed between occasional and former users reflects the fluid and transitional nature of disengagement from digital markets.

Implications for Digital Inclusion Policy

The comparative evidence from e-commerce and e-government participation suggests that digital inclusion policies should move beyond access-oriented frameworks and prioritise skill development and functional digital engagement. While e-government platforms have achieved relatively broad uptake, non-use persists among individuals lacking transactional experience and digital confidence.

For e-commerce, the findings indicate that market participation remains uneven and strongly conditioned by digital skills and behavioural familiarity. Policies aimed at enhancing consumer digital literacy, trust in online transactions, and exposure to practical digital applications may therefore contribute to more inclusive digital market participation.

Concluding Remarks

Overall, this study demonstrates that digital participation in Türkiye is shaped by a complex interplay of skills, behaviours, and institutional contexts. While e-government use appears more normalised and widespread, both e-commerce and e-government engagement are fundamentally anchored in functional digital competence. By integrating data mining techniques with a two-stage analytical framework, the study provides nuanced insights into the mechanisms underlying digital inclusion and offers a robust empirical foundation for future research and policy development.

CHAPTER 7
CONCLUSION

This study examined patterns of digital participation in Türkiye through a comparative analysis of e-commerce and e-government engagement, employing a data mining–based classification framework. Using nationally representative microdata and decision tree algorithms, the analysis aimed to identify the determinants of digital participation and to assess the role of digital skills in shaping both market-oriented and public digital service use.

Summary of Main Findings

The empirical results demonstrate that digital participation in Türkiye is characterised by substantial heterogeneity. In the domain of e-commerce, participation remains uneven, with nearly half of the population never engaging in online purchasing and a non-negligible share occupying intermediate positions as occasional or former users. These findings confirm that e-commerce engagement cannot be adequately captured through binary classifications alone.

By contrast, e-government participation is considerably more widespread, reflecting the institutionalisation and increasing normalisation of digital public services. Nevertheless, the persistence of a sizable non-user group indicates that access to infrastructure and services does not guarantee universal adoption.

Across both domains, behavioural indicators of digital engagement and digital skills consistently emerged as the most influential determinants of participation. Traditional sociodemographic variables such as age, gender, and region exhibited limited explanatory power once skills and usage patterns were taken into account. This finding underscores the central role of functional digital competence in enabling meaningful engagement with digital platforms.

Methodological Contributions

A key contribution of this study lies in its methodological approach. The adoption of a two-stage classification strategy for e-commerce participation proved particularly effective in addressing class imbalance and uncovering latent heterogeneity among user groups. The marked improvement in classification performance in the second stage highlights the analytical value of distinguishing between entry into digital markets and sustained intensity of use.

Decision tree–based models offered an interpretable and flexible framework for analysing complex behavioural data, allowing for transparent identification of dominant predictors and interaction structures. The consistency of variable importance rankings across models further strengthens the robustness of the findings.

Implications for Digital Inclusion Policy

The results carry important implications for digital inclusion strategies in Türkiye. The strong and consistent influence of digital skills suggests that policies focusing solely on infrastructure expansion are insufficient to ensure inclusive digital participation. Instead, effective digital inclusion requires targeted interventions aimed at enhancing functional digital competence, transactional confidence, and practical engagement with digital services.

In the context of e-commerce, improving consumer digital literacy and familiarity with online transactions may help reduce persistent non-participation and support sustained engagement. For e-government, the findings indicate that strengthening users' transactional experience and trust in digital systems—particularly through integration with commonly used services such as online banking—may further broaden adoption.

Limitations and Directions for Future Research

Despite its contributions, this study is subject to several limitations. First, the analysis relies on cross-sectional survey data, which limits the ability to draw causal inferences or to capture dynamic transitions in digital participation over time. Second, self-reported measures of digital behaviour and skills may be subject to reporting bias.

Future research could address these limitations by employing longitudinal data to examine trajectories of digital engagement and by integrating qualitative approaches to better understand the motivations and barriers underlying non-use and disengagement. Additionally, extending the two-stage modelling framework to other domains of digital participation—such as online education or digital health services—would further test its generalisability.

Concluding Remarks

In conclusion, this study provides a comprehensive and comparative analysis of e-commerce and e-government participation in Türkiye, highlighting the central role of digital skills in shaping digital inclusion. By combining interpretable data mining methods with a hierarchical analytical strategy, the research offers nuanced insights into the mechanisms underlying digital engagement and contributes to both the empirical literature and policy-oriented discussions on digital transformation.

REFERENCES

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics, 11*, 685–725.

Bannister, F., & Connolly, R. (2014). ICT, public values and transformative government. *Government Information Quarterly, 31*(1), 119–128.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

Castells, M. (2010). *The rise of the network society* (2nd ed.). Wiley-Blackwell.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Eurostat. (2020). *European compiler's manual for statistics on the use of ICT in households and by individuals* (Methodological guidelines). Publications Office of the European Union. Retrieved January 8, 2025, from https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/w/ks-01-25-011

Eurostat. (2025). E-commerce statistics for individuals (Data extracted: February 2025). Statistics Explained, European Commission. Retrieved January 8, 2025, from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-commerce_statistics_for_individuals

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Helsper, E. J. (2021). *The digital disconnect: The social causes and consequences of digital inequalities*. SAGE.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

OECD. (2019). *Measuring the digital transformation: A roadmap for the future*. OECD Publishing.

OECD. (2020). *Digital government index: 2019 results*. OECD Publishing.

OECD. (2021). *OECD framework for digital talent and skills*. OECD Publishing. https://doi.org/10.1787/4e7c3f58-en

Tanır, D., & Ramazanov, S. (2023). Veri Madenciliği Yöntemleri ile Türkiye'de Fertlerin E-Ticaret Kullanımını Etkileyen Faktörlerin Analizi. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 25(44), 46–65.

TÜİK. (2021). *Hanehalkı bilişim teknolojileri kullanım araştırması (2021): Metodoloji ve mikro veri dokümantasyonu*. Türkiye İstatistik Kurumu.

United Nations. (2022). *United Nations e-government survey 2022: The future of digital government*. United Nations (UN DESA).

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2014). The digital divide shifts to differences in usage. *New Media & Society, 16*(3), 507–526.

van Dijk, J. A. G. M. (2020). *The digital divide*. Polity Press.