**A Data Mining Approach to Analyzing ICT Sub-Sector Performance in the National Capital Region, Philippines**

Patis, Mick Phron D.

Bachelor of Science in Information Technology  
Marikina City  
09667011362

qmpdpatis@tip.edu.ph

Malbas, Lance Spencer B.

Bachelor of Science in Information Technology  
Antipolo City  
09272537801

qlsbmalbas@tip.edu.ph

Santiago, Leanne Raine T.

Bachelor of Science in Information Technology  
Antipolo City  
09953541459

qlrtsantiago@tip.edu.ph

Gabriel, Charles G.

Bachelor of Science in Information

Technology  
Montalban, Rizal  
09777067220

qcgabriel@tip.edu.phMuñoz, Terence C.

Bachelor of Science in Information

Technology  
Pasig City  
09281977267

qtlcmunoz@tip.edu.ph

**ABSTRACT**

This study analyzes the performance of the Information and Communication Technology (ICT) sector in the National Capital Region (NCR), Philippines, focusing on key sub-sectors such as ICT Manufacturing, Trade, Services, Software Publishing, Telecommunication Services, Programming and Consultancy, Data Processing and Web Portals, and Computer Repair. Drawing on datasets from 2015, 2017, and 2019, the research examines indicators including the number of establishments, access to computing and communication equipment, internet connectivity, employment levels, and the proportion of employees using computers and the internet.

By applying data mining techniques—preprocessing, regression analysis, classification, clustering, and association rule mining—the study uncovers structural characteristics and relationships within the ICT sector. The analysis is descriptive rather than predictive, aiming to support informed decision-making among policymakers, industry stakeholders, and educators in enhancing digital infrastructure, workforce capabilities, and strategic planning for the NCR’s digital economy.

**CCS Concepts**

• **Information systems → Data mining;**

• **Computing methodologies → Machine learning;**

**• Applied computing → Business intelligence;**

**Keywords**

ICT sector; economic productivity; data mining; regression analysis; classification; clustering; association rules; Indian cities

# INTRODUCTION

The Information and Communication Technology (ICT) sector is a major driver of economic progress in the Philippines, particularly in the National Capital Region (NCR), which functions as the country's technological and economic core. As digital transformation continues to reshape industries, gaining a deeper understanding of the sector’s performance has become increasingly important for policymakers, urban planners, and industry stakeholders. NCR is home to a wide range of ICT activities that contribute uniquely to employment, infrastructure, and digital adoption. However, the economic and technological factors influencing these activities are complex and require data-driven analysis for effective planning and policy-making.

This study investigates ICT performance in NCR using a dataset spanning 2015, 2017, and 2019. It focuses on key indicators such as the number of establishments, access to computing and communication tools, internet availability, employment levels, and the proportion of employees using digital tools. Using data mining techniques such as preprocessing, regression, classification, clustering, and association rule mining; the study aims to uncover performance patterns and sector relationships. Rather than making forecasts or rankings, the analysis seeks to describe existing structures to support three key objectives: (1) examine how infrastructure and economic indicators influence performance; (2) identify linkages between technology usage and employment; and (3) group sub-sectors with similar characteristics. These insights aim to inform digital infrastructure investment, workforce development, and policy strategies for sustainable ICT sector growth in the region.

# BACKGROUND OF THE STUDY

The United Nations Sustainable Development Goals (UNSDGs) emphasize inclusive economic growth, sustainable industrialization, and innovation—goals that closely align with the advancement of the ICT sector. The Philippines, particularly NCR, is well-positioned to capitalize on these opportunities through its fast-growing digital economy. With its diverse range of ICT activities, NCR plays a pivotal role in the country’s innovation landscape.

This study conducts a data-driven analysis of ICT sub-sector performance in the NCR using a dataset from 2015, 2017, and 2019, which includes metrics on the number of establishments, access to computing and communication equipment, internet access, number of employees, and the proportion of employees using computers and the internet. By applying data mining techniques, the research aims to uncover patterns and characteristics that will enable policymakers, businesses, and educational institutions to develop evidence-based strategies to strengthen digital economies, enhance employment opportunities, and advance sustainable development goals in the Philippines.

# PROBLEM STATEMENT AND RESEARCH QUESTIONS

The Information and Communication Technology sector is a key driver of economic growth in the National Capital Region, Philippines. However, little research uses data to explore how factors like the number of establishments, technology access, employment, and computer or internet use affect ICT sub-sectors such as ICT Manufacturing, ICT Trade, ICT Services, and others. Despite investments in digital infrastructure, differences in these sub-sectors’ performance remain, and these need to be studied using detailed data.

This study addresses the need to understand what influences ICT sub-sector performance and how sub-sectors can be grouped by their characteristics. Without this knowledge, policymakers and businesses may struggle to make decisions that support digital growth and sustainability. Using a dataset from 2015, 2017, and 2019, the research applies data mining techniques—preprocessing, regression, classification, clustering, and association rule mining—to analyze patterns and relationships. The focus is on understanding the sector, not predicting trends or ranking sub-sectors.

This study aims to address the following research questions:

* Which indicators, like technology access or employment, most affect ICT sub-sector performance in the NCR?
* How can clustering group ICT sub-sectors by their performance profiles?
* What patterns exist between indicators like internet access and computer use in ICT sub-sectors?
* How can regression and classification analyze relationships and categorize ICT sub-sectors?
* How can the findings help stakeholders support sustainable ICT growth in the NCR?

# OBJECTIVES OF THE STUDY

This study aims to evaluate the performance of ICT sub-sectors in the National Capital Region (NCR), Philippines, from 2015 to 2019, focusing on how indicators such as the number of establishments, technology access, employment, and computer or internet use shape the sector.

* To identify key indicators, such as technology access and employment, that influence ICT sub-sector performance in the NCR from 2015 to 2019.
* To group ICT sub-sectors based on their performance profiles and uncover patterns in the sector.
* To analyze relationships between indicators like technology access and computer or internet use.
* To explore associations between factors like internet access and technology use to guide workforce and infrastructure strategies.
* To detect actionable patterns in ICT sub-sector performance to inform evidence-based policy and planning for digital development in the NCR.

## General Objective (SMART)

The general objective of this study is to analyze the performance of ICT sub-sectors in the National Capital Region (NCR), Philippines, from 2015 to 2019, by examining how indicators such as the number of establishments, technology access, employment, and computer or internet use influence the sector. The study will produce measurable insights into patterns and relationships within six months, enabling government, businesses, and educational institutions to align strategies with the digital economy’s needs in a practical and sustainable manner.

# RELATED LITERATURE

The ICT sector in the Philippines, particularly in the National Capital Region (NCR), has emerged as a critical driver of economic growth, supported by government-led digital transformation strategies since 1992, such as the e-Government Master Plan (2012), the Digital Transformation Strategy (2022), and the Philippine Development Plan 2023–2028, which emphasize infrastructure development and digital inclusion (Treceñe, 2021; U.S. Department of Commerce, 2024). These initiatives have fostered an enabling environment for ICT expansion, reflected in the digital economy’s $36.5 billion contribution in 2022—9.4% of the national GDP—with telecommunications services making up 77.2% of this figure (PSA, 2022). NCR, in particular, has benefited from concentrated infrastructure investments including 5G rollouts and data centers. Subsector developments such as the liberalization of telecom ownership, increased cybersecurity spending, BPO-driven enterprise software demand, and rapid growth in hyperscale data centers underscore the sector’s diversification (U.S. Department of Commerce, 2024; ITU, 2020; S&P Global, 2023).

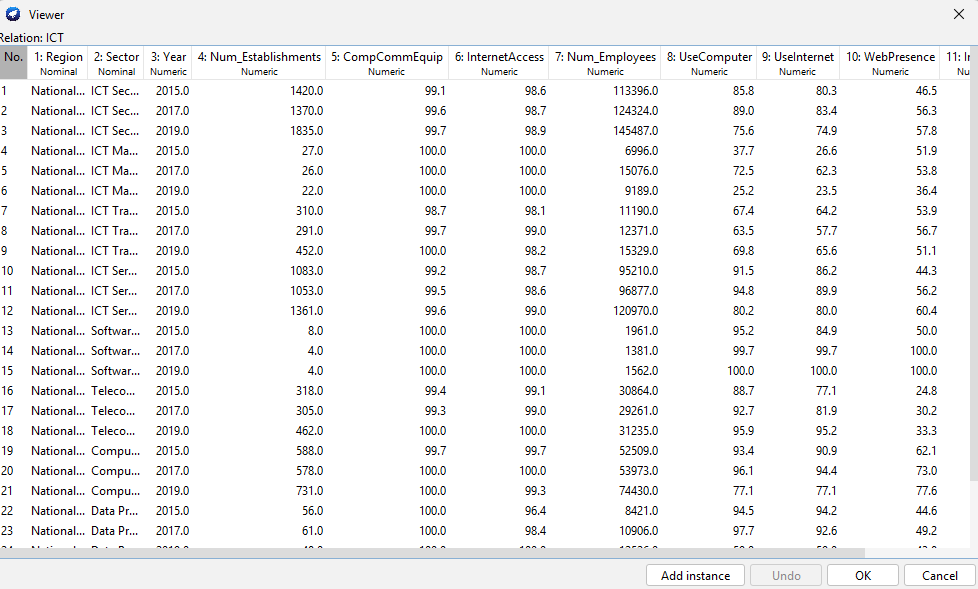
Serafica and Oren (2022), in their discussion paper titled “The Philippine Digital Sector and Internet Connectivity: An Overview of the Value Chain and Barriers to Competition”, examine the structure and challenges of the digital sector in the Philippines. They emphasize that the digital value chain is composed of interdependent markets where barriers to entry—particularly in internet connectivity—can significantly impact the entire ecosystem. The study highlights how regulatory, strategic, and natural barriers limit competition, especially in the telecommunications industry dominated by two major vertically integrated firms. Their findings stress the importance of fostering competition, strengthening access regulations, and promoting public-private partnerships to enhance digital inclusion and economic productivity. This aligns with the goals of analyzing ICT sector contributions to economic growth, especially in Philippine urban centers.

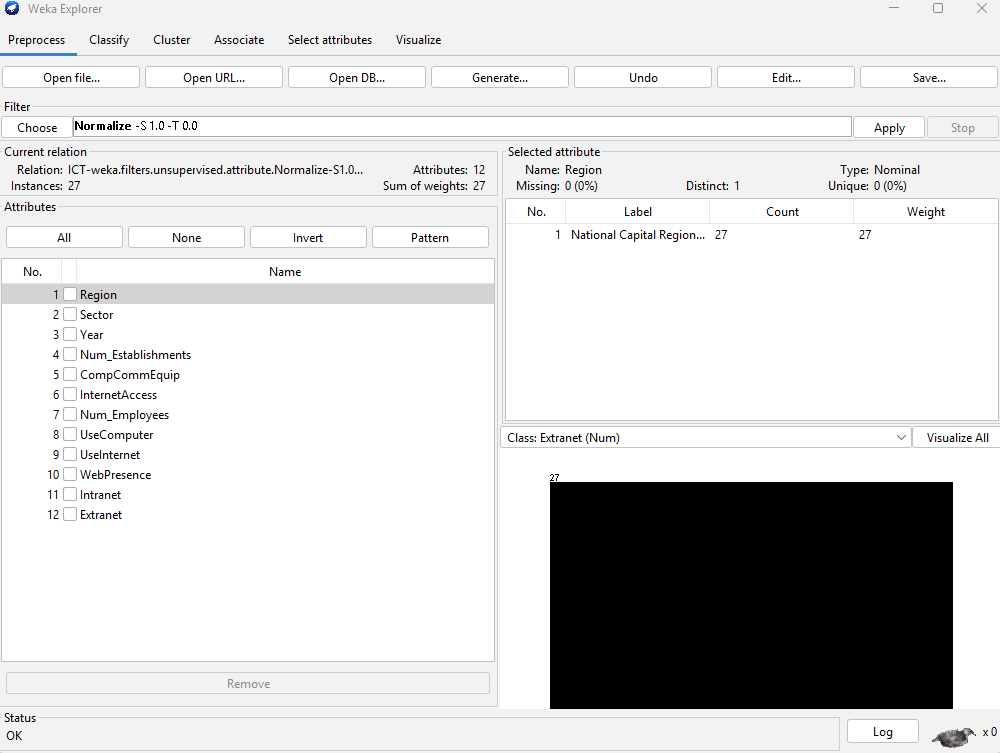
# METHODOLOGY

asdasdasdasdasdasdasdasdasdasdasdasdasdasdasdasdasdasdasdasd

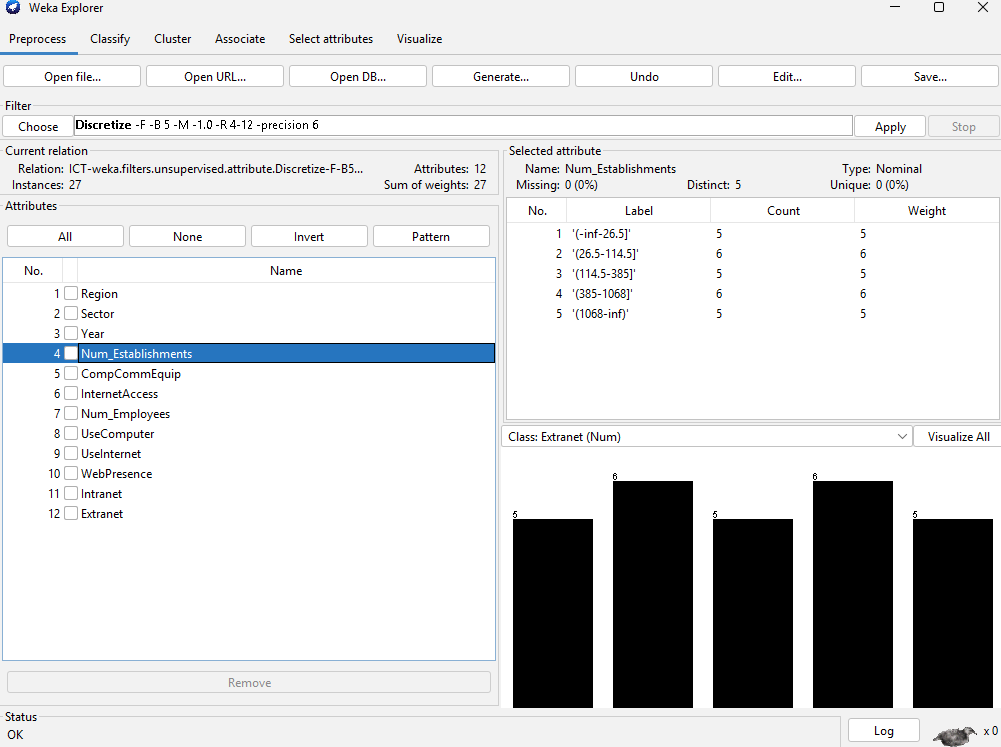
## Data Preparation

In preparing data for analysis using data mining techniques, it is essential to perform several preprocessing steps to ensure accuracy and consistency. **Replacing missing values** is a critical first step, where any null or blank data entries are filled in with suitable default values typically the mean for numerical attributes or the mode for categorical ones. This helps avoid distortions or errors in the modeling process. After that, **normalization** is often applied to numeric attributes to rescale the values to a standard range, usually between 0 and 1. This step ensures that all features contribute equally to distance-based algorithms like k-means clustering or k-nearest neighbors, which can otherwise be biased by attributes with larger numeric ranges. Lastly, **discretization** involves converting continuous numeric data into categorical bins or intervals. This process simplifies complex data and is particularly useful for algorithms like decision trees and Naive Bayes, which handle categorical inputs more effectively. Together, these preprocessing techniques help prepare the dataset for accurate and meaningful models.

  
**Figure 1:** Data Cleaning



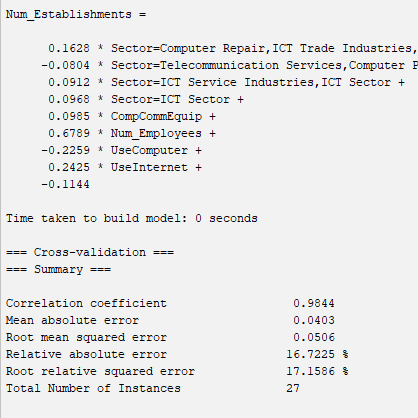
**Figure 2:**Preprocessing

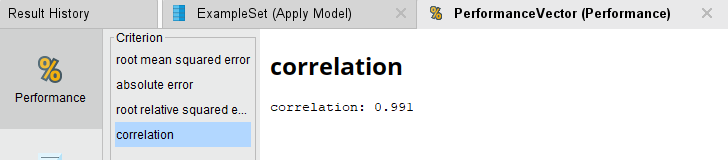


**Figure 3:**Discretize

* 1. **Model Generation and Prediction**
     1. Regression

The linear regression model was developed to examine how various factors influence the number of ICT establishments in the National Capital Region (NCR). With a high correlation coefficient of **0.9844**, the model shows a strong relationship between the predictors and the outcome variable. The **number of employees** and **internet use** had the strongest positive effects, while **computer use** showed a slight negative impact. Sector classifications and access to communication equipment also influenced the results. Overall, the model effectively explains the key factors driving ICT sub-sector growth in NCR.

  
**Figure 4: Linear regression in Weka**

****

**Figure 5: Linear regression in RapidMiner**

* + 1. Classification
    2. Association
    3. Clustering

## Model Evaluation

* 1. Regression
  2. Classification
  3. Association
  4. Clustering

# RESULTS AND FINDINGS

## Data Preparation

## Model Generation and Prediction

* 1. Regression

The linear regression model was developed to analyze how various factors influence the number of establishments in different ICT sub-sectors in the National Capital Region (NCR). The model demonstrates a strong relationship between the independent variables and the number of establishments, as evidenced by a correlation coefficient of **0.9844**, indicating a high level of accuracy in explaining variability across sectors.

Among the key predictors, the **number of employees** showed the most significant positive influence (**coefficient = 0.6789**), suggesting that sub-sectors with a larger workforce tend to have more business establishments. **Internet use** also contributed positively (**coefficient = 0.2425**), highlighting the importance of digital connectivity in the growth of ICT-related businesses.

Access to **computing and communication equipment** had a modest but positive effect (**coefficient = 0.0985**), indicating that infrastructure availability supports business presence. In contrast, the **use of computers** was associated with a slight negative effect (**coefficient = -0.2259**), which may suggest that higher internal technology use could reduce the need for multiple physical establishments, possibly due to efficiency or centralization.

The influence of sector classifications showed mixed effects. Broader ICT sectors such as Computer Repair, ICT Trade, and Programming contributed positively (**coefficient = 0.1628**), while narrower or more specialized sectors like Telecommunication Services had a minor negative impact (**coefficient = -0.0804**). ICT Service Industries and the general ICT Sector both showed small positive effects (**coefficients = 0.0912 and 0.0968**, respectively).

The model’s error rates, including a **mean absolute error of 0.0403** and a **root mean squared error of 0.0506**, confirm that the predictions closely align with actual data. These findings reinforce the role of employment, internet access, and sector classification as important drivers of ICT sub-sector growth in NCR. The regression model provides valuable insights for policymakers and stakeholders aiming to support and expand the digital economy through data-informed planning.

* 1. Classification

# CONCLUSION

Conclude based on the research questions and objectives

# RECOMMENDATIONS Recommend based on the results and findings

# FUTURE DIRECTION

Highlight the parts that are not included in the result that you intend to do in the next research

# ACKNOWLEDGMENTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.

# REFERENCES

1. 5. 2024. Philippines - Information and Communications Technology. International Trade Administration | Trade.gov. Retrieved May 7, 2025 from https://www.trade.gov/country-commercial-guides/philippines-information-and-communications-technology
2. Ramonette B Serafica and Oren, Queen. 2022. The Philippine Digital Sector and Internet Connectivity: An Overview of the Value Chain and Barriers to Competition. Econstor.eu (2022). DOI:http://hdl.handle.net/10419/284584

**Columns on Last Page Should Be Made As Close As Possible to Equal Length**