Analysis of Cybersecurity Risks and Teenage Digital Behavior Patterns

Eric McCloy, Samuel Nimako-Mensah, Albert Samigullin

29th Colloquium for Information Systems Security Education

12 November 2025



CityU of Seattle

Agenda

- Problem Statement
- Introduction
- Related Work
- Background
- Methodology
- Key Findings
- Recommendations
- Key Reference
- Q&A

CityU
of Seattle

Problem Statement

- Teenagers face significant cybersecurity risks online
- Lack of awareness makes them vulnerable
- **Research question**: What insights can be gained from analyzing the Teenage dataset to help understand behavior patterns and predict?

CityU of Seattle

Introduction

Rising Risk Exposure

Teenagers are increasingly exposed to cybersecurity threats due to high digital engagement and limited awareness.

Education Gaps

Traditional awareness programs often fail to influence secure behavior in meaningful ways

Stakeholder Support

Actionable insights needed for parents, educators and policymakers to develop effective interventions.

CityU of Seattle

Introduction: Related Work

Digital Literacy Paradox

More digitally literate teens face more risks due to greater online time (Vissenberg et al., 2022).

Parental Control

Strict controls may backfire, increasing risk-taking behavior (Kang et al., 2022).

Well-being impacts

Cyber risks correlate with anxiety, depression, and poor academic performance.

CityU of Seattle

Methodology: Background on Dataset

Teenage Online Cybersecurity Risks Dataset

Network activity logs and e-safety monitoring systems across various educational institutions and households in Texas and California over a 7-year period (2017-2024) .

Records and Meta Data

The dataset includes 67,921 records across 30 columns, categorized into demographics, device usage, cybersecurity threats, and online behavior.

Teenagers aged 13-16 and mobile phones dominate

Mobile phones are the dominant device type at 70.2% and teenagers between 13 and 16 made up the majority of the dataset indicating a mobile first mindset among teens accessing the internet.

CityU
of Seattle

Methodology: Data Acquisition

Dataset Details

The dataset was downloaded from the Kaggle in CSV format. Structured logs on device activity, behavior and cybersecurity incidents were included

Initial Inspection

Using pandas in Python, we analyzed dataset dimensions and identified missing fields and data types that needed transformation.

Dataset Composition

The full dataset included 30 columns representing numeric and categorical variables across digital behavior and safety dimensions.

CityU of Seattle

Methodology: Data Cleaning

Handling Missing Data

Columns with excessive null values, like Education_Content_Usage (with over 47,000 missing values), were removed entirely

Date Standardization

Timestamps were parsed to ensure consistent formatting for trend analysis and time-based risk mapping.

Derived Features

We created new fields such as Completeness Score and Total Threat Exposure to capture data quality and individual risk levels.

Final Cleaned Data

After cleaning, 91.68% of records were 100% complete, allowing for reliable modeling and interpretation.

CityU of Seattle

Methodology: Feature Engineering

Behavioral and Security Features
We engineered over 17 new features including Password Security Score, Risky Browsing Ratio, and Defensive Posture Score.

Composite Indices
Some features, like Overall Security Posture and Connection Safety Score, were computed using weighted formulas based on behavioral inputs.

Risk Classification
Categorical features such as Protection Match Category and Security Posture Category enabled segmentation of user profiles by safety level.

Correlational Metrics
We used these derived features for further statistical modeling including Spearman correlation and clustering.

CityU of Seattle

Data Analysis: Correlation Analysis

Statistical Method
Spearman correlation was selected for its ability to handle ordinal and non-linear relationships typical in behavioral data.

Key Positive Correlation
Strong correlations were found between safety measures such as VPN usage and overall security posture (e.g., $r = 0.75$).

Key Negative Correlation
We also found that poor supervision was highly correlated with increased threat severity and exposure ($r = -0.97$).

Significance Testing
All identified relationships were validated using hypothesis testing to ensure statistical significance ($p < 0.05$).

CityU of Seattle

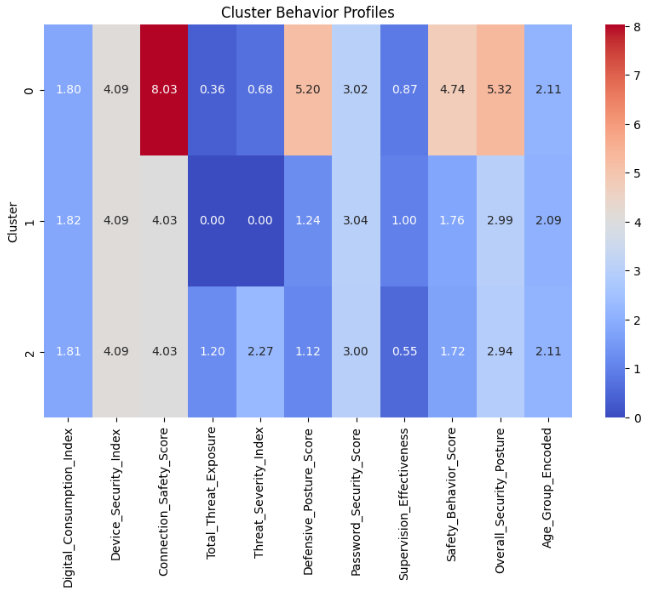
Data Analysis: Clustering Analysis

Clustering Methodology
We applied K-Means clustering using 11 behavioral and demographic features to group users into distinct profiles.

Cluster 0 – Risk-Aware Users
These teens practiced strong security habits, had high awareness scores, and showed the lowest exposure to threats.

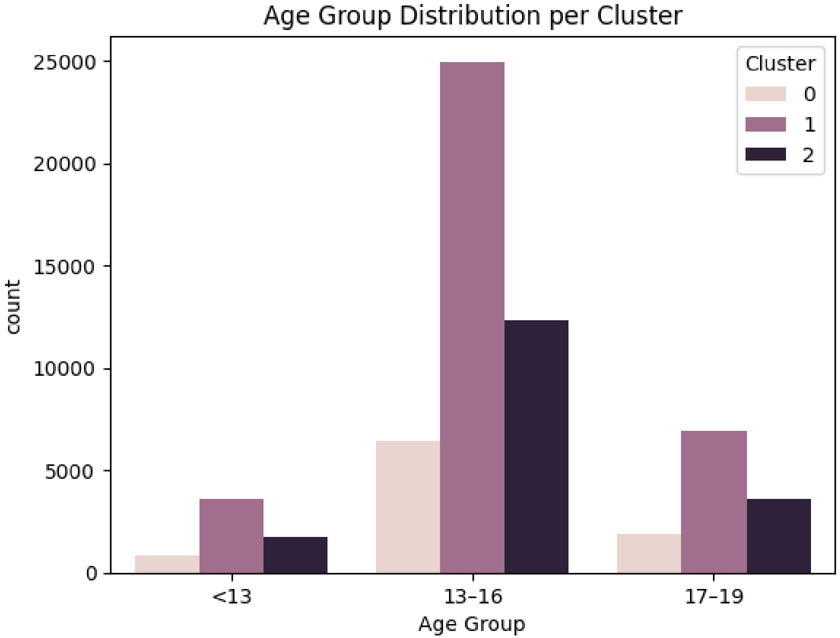
Cluster 1 – Low-Engagement Users
Users in this group rarely engaged online, resulting in low exposure but also low awareness and low parental supervision.

Cluster 2 – Moderate-Risk Users
This group had moderate security habits but high threat exposure and the weakest parental oversight, placing them at risk.



CityU of Seattle

Age distribution



Data Analysis: Behavior Insights

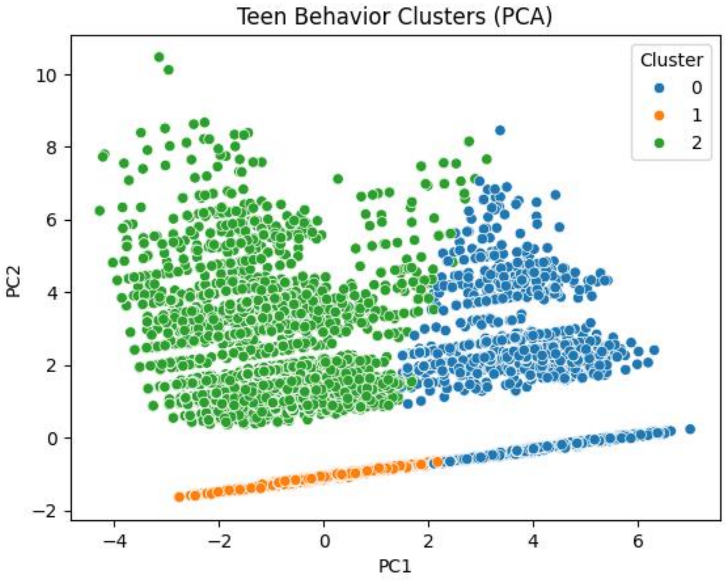
PC1 (horizontal) = *riskiness / unsafe behavior orientation*
PC2 (vertical) = *threat exposure (high up, low down)*

User Profile Insights
Cluster 0 teens were digitally literate and cautious, showing strong alignment between awareness and action.

Minimal Engagement Patterns
Cluster 1 users may have been limited by parental controls or simply lacked access, contributing to their low-risk profile.

Vulnerable Demographics
Cluster 2 demonstrated that without proper supervision, even moderately aware teens could still face frequent cybersecurity threats.

Age Group Distribution
All age groups were represented across clusters, showing that risk is more influenced by behavior and oversight than age alone.



CityU of Seattle

Findings

Digital Literacy ≠ Safety
While digital literacy improves awareness, it also increases online exposure and does not ensure safe behavior on its own.

Parental Mediation Is Critical
Teens with active, collaborative parental involvement had better online safety habits than those under strict rules.

Clustering Enables Personalization
Identifying user clusters allows us to move beyond generic awareness programs toward customized interventions.

CityU of Seattle

Recommendations

Parental Involvement
Encourage parents to shift from setting rules to having open conversations with teens about online safety and choices.

Curriculum Integration
Schools should teach cybersecurity fundamentals as part of their digital literacy programs, starting in middle school.

Policy-Level Actions
Policy frameworks should support adaptive tools and campaigns that adjust to different behavioral risk levels among youth.

Platform Design Principles
Technology platforms should embed built-in nudges, prompts, and controls to guide safe behavior, especially for teen users.

CityU
of Seattle

Key References

- Vissenberg, J., d'Haenens, L., & Livingstone, S. (2022). Digital literacy and online resilience as facilitators of young people's well-being?: A systematic review. *European Psychologist*, *27*(2), 76–85. https://doi.org/10.1027/1016-9040/a000478
- Adorjan, M., & Ricciardelli, R. (2021). Smartphone and social media addiction: Exploring the perceptions and experiences of Canadian teenagers. *Revue Canadienne de Sociologie [Canadian Review of Sociology]*, *58*(1), 45–64. https://doi.org/10.1111/cars.12319
- Kang, H., Shin, W., & Huang, J. (2022). Teens' privacy management on video-sharing social media: the roles of perceived privacy risk and parental mediation. *Internet Research*, *32*(1), 312–334. https://doi.org/10.1108/intr-01-2021-0005
- Mwagwabi, F., & Jiow, J. H. (2021). Compliance with security guidelines in teenagers: The conflicting role of peer influence and personal norms. Australasian Journal of Information Systems, 25, Article 1. https://doi.org/10.3127/ajis.v25i0.2676

CityU of Seattle

Q&A