International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Cybersecurity Awareness Prediction Among College Students Using Data Analytics

P Jithesh

Assistant Professor, ITM College of Arts and Science, mayyil

Abstract

Cyber security incidents targeting individuals have increased significantly with the rapid adoption of digital devices. College students remain one of the most vulnerable groups due to heavy dependency on online platforms. This study aims to predict cyber security awareness levels among college students using data analytics techniques. A structured questionnaire was used to collect data on knowledge, behavior, and digital habits. Machine learning models such as Logistic Regression, Decision Trees, and Random Forest were applied to classify students into awareness categories. The results show that digital behavior, password practices, and social media usage are the strongest predictors of cyber awareness. The study provides insights for designing targeted training programs for higher education institutions.

1. Introduction

The growing digital environment has exposed students to cyber threats such as phishing, identity theft, malware attacks, and data breaches. Although several awareness programs exist, the actual level of cyber security literacy among students remains uncertain. Predicting these awareness levels can help institutions deliver customized training.

Cyber security awareness prediction involves collecting student behavioral and knowledge data and applying analytical techniques to classify or forecast their awareness levels. This paper explores how data analytics can be applied to understand predictors of cyber awareness among college students.

2. Problem Statement

Many college students lack adequate knowledge of cyber security practices, making them easy targets for cybercriminals. Institutions often conduct general training sessions, but these programs fail to address specific vulnerabilities.

There is a need for a **data-driven approach** to identify awareness gaps and predict which student groups are at higher risk.

3. Objectives of the Study

1. To assess the cyber security awareness level of college students.
2. To identify behavioral and demographic factors influencing awareness.
3. To build predictive models for classifying awareness levels.
4. To propose recommendations for targeted cyber security education.

4. Literature Review

Studies in cyber security behavior indicate that young adults often underestimate personal cyber risks.

Prior research highlights:

- Social media habits influence vulnerability.
- Strong passwords and authentication techniques correlate with better cyber hygiene.
- Machine learning methods such as Decision Trees and Logistic Regression have been successfully used to predict digital behavior patterns.

However, there is limited research specifically targeting awareness prediction among college students using data analytics, especially in the Indian/Higher Education context. This study addresses this gap.

5. Research Methodology

5.1 Data Collection

A structured questionnaire was shared with undergraduate students. The survey included:

- Password habits
- Email and phishing awareness
- Social media behavior
- Device security practices
- Cybercrime experience

Sample size: 100 to 125 students

5.2 Data Preparation

- Missing values handled
- Responses encoded numerically
- Awareness score calculated using weighted indicators

Awareness levels categorized as:

- **High Awareness**
- **Moderate Awareness**
- **Low Awareness**

5.3 Analytical Techniques

The following algorithms were used for prediction:

- Logistic Regression
- Decision Tree Classifier
- Random Forest Classifier
- Naïve Bayes

Evaluation metrics:

- Accuracy
- Precision
- Recall
- Confusion matrix

5.4 Tools Used

- Python
- Pandas, Scikit-learn, Matplotlib
- MS Excel for initial data cleaning

6. Results and Discussion

6.1 Descriptive Analysis

- Majority of students used social media frequently.
- Only a small percentage used strong, unique passwords.
- Many were unaware of phishing indicators.
- Most students lacked knowledge of data privacy settings.

6.2 Predictive Model Results

Random Forest produced the highest accuracy.

Important predictors:

- Password strength
- Frequency of updating passwords
- Awareness of phishing emails
- Social media privacy settings
- Experience with cyber incidents

6.3 Interpretation

Students with risky digital habits showed lower awareness scores. Behavioral indicators were stronger predictors than age or academic stream. Targeted training focusing on password management and phishing awareness can significantly improve security behavior.

7. Conclusion

This study demonstrates that data analytics can effectively predict cyber security awareness levels among college students. Machine learning models identified key behavioral attributes influencing awareness. The findings help colleges design personalized training interventions to reduce cyber security risks. Future research may include larger sample sizes, multi-campus data, and advanced deep learning models.