International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0214
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1423 - 1426

Cyber Risk Forecasting Using Data Science Techniques

R Kaviyarasi[1] ,Sonali V Kalgutkar[2], Sheethal Jayapal[3], K Romio[4], Sivakami Anil Kumar[5]
1 Associate Professor, Department of Computer Science and Information Technology, Yenepoya (Deemed to be University), Bengaluru Campus, Karnataka, India.
[2,3,4,5]PG - Department of Computer Science and Information Technology, Yenepoya (Deemed to be University), Bengaluru Campus, Karnataka, India.
Email ID *arasikavi@gmail.com[1], sonalikalgutkar8@gmail.com[2], blackismycolour172002@gmail.com[3], kromio2004@gmail.com[4], sivakamkianilkumar@gmail.com[5]*

Abstract
The number of cyber risks is going up and they are getting more complicated as organisations use more digital technologies. These risk factors can affect the data and business operation very severely. This study mainly focusing on data science to predict cyber risks. It can help organisations take a pre-emptive approach to manage cyber security. Most of the cyber security strategies reacts only after a cyber threat occurs, rather than preventing the occurrence. This paper uses data analytical tools and computer models to analyse when and the cyber risks will commence. It looks through the system logs, lists of vulnerabilities of past incidents and information about external threats. The study uses statistics, machine learning and forecasting to make predictions. It also considers what happens when data is missing how to choose the features and how to get the data ready for use. The main goal is to predict cyber risks, like cyber-attacks, cyber-threats and help organisations prepare for them. Cyber risks are a deal and this study is, about cyber risks and how to manage cyber risks using data science techniques.

Keywords*: Cyber, Forecasting, Analysis, Data Science*

1. Introduction
The quick development of digital technology has led to an increase in cyberattacks, which are a major threat to data security, operations, and business continuity. The focus of earlier cyber security methods was primarily on reacting to an attack, which may make it more difficult to achieve effective risk management and possibly lead to new problems.[2] To get around this traditional method, cyber risks forecasting using data analytics techniques has become a preventive cybersecurity tool. To identify patterns and potential threats, data-driven models, vulnerability reports, system logs, external threat intelligence, and previous cyber incidents can all be utilised. Predicting the likelihood, timing, and possible impact of cyberattacks is made easier by some methods, including statistical analysis, machine learning, and time-series forecasting. It enabled the companies to enhance early warning systems and increase overall cyber resilience.[1] Shown as Figure 1 Uses cases for Data Analysis in Cyber Security.

Figure 1 Uses cases for Data Analysis in Cyber Security

2. Literature Review
In today's world, every organisation mostly rely on the digital equipment and its technologies to do their

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0214
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1423 - 1426

daily works and to store important data.[3] Thus, cybercrimes and threats are increasing both in number and complexity. These threats can affect the people who are using digital space as their reliable job or every person who use technologies.[4][5] In traditional manner, Cybersecurity mainly focused on reacting after an attack occurs, which is eliminating the chances of preventing the attack. To resolve this issue, organizations can you Data science to foresee the cyber risks that may be a futures attack ahead of time. Evaluating the past cyber incidents, system logs, vulnerability reports and external threat information, it is possible to track down the same pattern that reoccurs and tackle that easily. There are certain methods to predict the attacks beforehand and to understand how impactful it could be, such as- machine learning, statistical analysis and time series. Even though there are certain challenges like incomplete data and constantly evolving cyber threads, these predictive tools may often track down before it reaches and give early warnings.[6] These intuition or insights helps organization to prepare well, allocates the resources efficiently, and build a strong foundation in overall cybersecurity for making them more adaptable to future attacks.[7]

3. Methodology

This study follows a data-driven research approach to forecast cyber risks using data science techniques.[8] The methodology includes data collection, preprocessing, model development, and evaluation to identify patterns in cyber incidents and predict potential cyber threats.

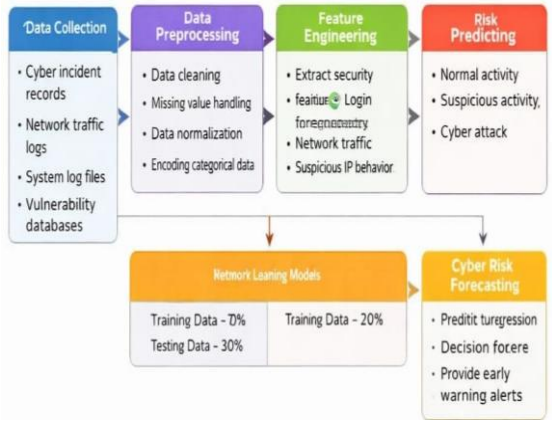

Figure 2 Methodology cyber risk forecasting using data science techniques.

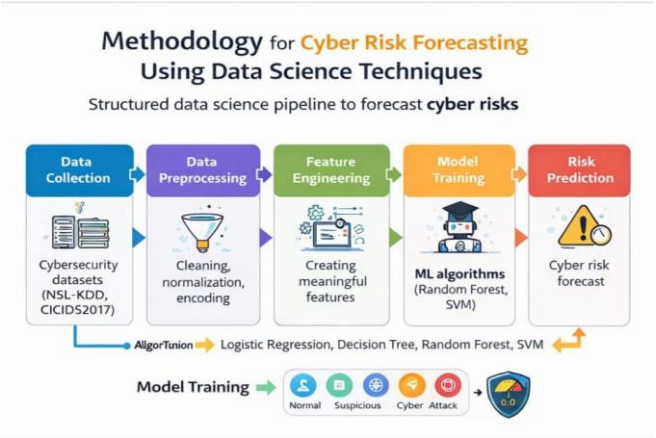

Figure 3 Methodology cyber risk forecasting using data science techniques.

3.1 Data Collection

Cybersecurity data is gathered from various sources, such as:

- Network traffic logs
- record of cyber incidents
- System - generated event logs

Reports on system vulnerabilities. Researchers also make use of publicly available datasets like NSL-KDD, CICIDS2017, and UNSW-NB15, for detecting cyber threats. these datasets include information related to network activities, communication protocols, login behaviour, attempts, and different types of attacks.[9]

3.2 Data Preprocessing

unprocessed cybersecurity data often includes errors, noise, or missing information. contains Therefore, it must be cleaned and prepared before analysis. The preprocessing steps include:

- Eliminating duplicate or incomplete entries.
- Filling or managing missing data.
- Converting categorical data into numeric form.

• Scaling data using normalization techniques.

These steps enhanced the quality and help in building accurate machine learning models.

3.3 Feature Engineering

Feature engineering involves identifying and selecting important data attributes that can help from detect cyber threats. Some examples of such features are:

- Count of unsuccessful login attempts.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0214
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1423 - 1426

- volume of network traffic.
- Protocol type
- Packet size
- Unusual IP address behaviour

These features assist machine learning models in recognizing patterns linked to potential threats.

3.4 Model Training

Machine learning techniques are applied to develop models that can predict cyber risks. The dataset is split into:

- 70% for training.
- 30% for testing.
- The following algorithms are used:
- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)

These models study past cybersecurity data to identify and classify network behaviour.

3.5 Cyber Risk Prediction

The trained model examines network activity and determines whether it falls into one of the following categories:

- Normal behaviour.
- Suspicious behaviour.
- Cyber-attack

a cyber risk score to forecast potential cyber threats and support early warning systems.

4. Experiment And Results

4.1. Experimental Setup

The experiments are conducted using a data science environment.

Tools Used

- Python
- Jupiter Notebook
- Pandas and NumPy for data processing
- Scikit-learn for machine learning algorithms
- Matplotlib for visualization

The cybersecurity dataset is divided into training and testing sets to evaluate the model's performance. The performance of machine learning models is evaluated using the following metrics:

Accuracy: Measures the percentage of correct predictions.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision: Measures how many predicted attacks are actually attacks.

Recall: Measures how many real attacks are correctly detected.

F1 Score: Harmonic mean of precision and recall.

Confusion Matrix: Shows classification results including true positives, false positives, true negatives, and false negatives.[10]Shown as Table 1 Experimental Results. Figure 4 Accuracy comparison chart

4.2. Experimental Results

Table 1 Experimental Results

Logistic Regression	85%
Decision Tree	88%
Random Forest	93%
Support Vector Machine	90%

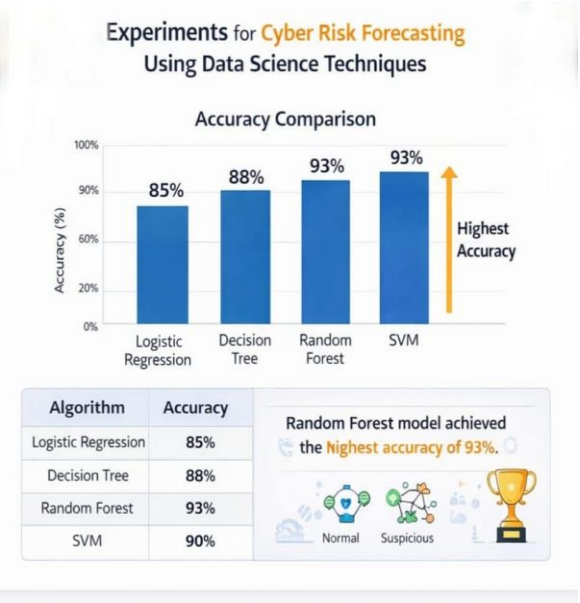

Figure 4 Accuracy comparison chart

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0214
e ISSN: 2584-2854
Volume: 04 Issue: 05
May 2026
Page No: 1423 - 1426

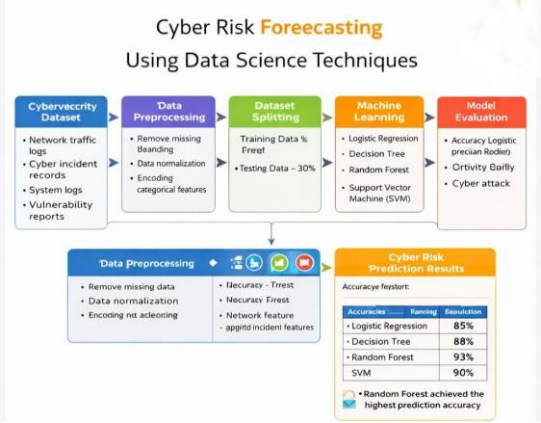

Figure 5 Cyber risk forecasting using Data science techniques

5. Discussion

The machine learning algorithms were trained on the collected cybersecurity dataset. After feature selection on the dataset, these algorithms were able to identify certain patterns in cyber-attack data such as phishing, malware, ransomware, etc. Among all the algorithms used in this study, Random Forest performed well in prediction when compared with Logistic Regression and Support Vector Machine algorithms. The results obtained from this study indicate that these predictive algorithms can effectively analyse cyber-attack history data and identify potential cyber risks and also shows that data science can help in predicting potential cyber threats in the future. And also, analysis helped identify trends in cyber-attack frequency over a period of time.

Conclusion

This study shows the application of data science and machine learning models in forecasting cyber risks. Through the study of the historical data related to cybersecurity, this research was able to identify some patterns and trends, which are essential in forecasting potential cyber threats. As shown in the experimental results, machine learning models, such as Random Forest, are effective in forecasting cyber risks. This study shows the significance of data-driven models in cybersecurity, as they are essential in detecting potential cyber threats. In the future, the model used in this study can be improved by incorporating more data and better algorithms to improve the precision of the model in forecasting potential cyber threats.

Acknowledgement

The authors would like to express their sincere gratitude to our faculty members and institution for their continuous support and guidance during the preparation of this research work. We would also like to thank our mentors and colleagues who provided valuable suggestions and encouragement throughout the study. Their support helped us successfully complete this work on Cyber Risk Forecasting using Data Science techniques.

References

[1]. R. Anderson, Security Engineering: A Guide to Building Dependable Distributed Systems, Wiley, 2008.
[2]. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.
[3]. C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[4]. T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning, Springer, 2009.
[5]. S. Axelsson, "The Base-Rate Fallacy and the Difficulty of Intrusion Detection," ACM Transactions on Information and System Security, 2000.
[6]. M. Bishop, Computer Security: Art and Science, Addison-Wesley, 2003.
[7]. National Institute of Standards and Technology, Framework for Improving Critical Infrastructure Cybersecurity, 2018.
[8]. IBM Security, Cost of a Data Breach Report, 2023.
[9]. Verizon, Data Breach Investigations Report, 2023.
[10]. D. Denning, "An Intrusion-Detection Model," IEEE Transactions on Software Engineering, 1987.suitable for cyber risk forecasting.