Applying Machine Learning to Optimize the Correlation of SecurityScorecard Scores with Relative Likelihood of Breach

BOB SOHVAL, PHD
VICE PRESIDENT, DATA SCIENCE

SecurityScorecard

SecurityScorecard.com
info@securityscorecard.com
©2022 SecurityScorecard Inc.

Tower 49
12 E 49th St
Suite 15-001
New York, NY 10017
1.800.682.1707

Introduction

SecurityScorecard ratings provide a means for objectively monitoring the cybersecurity hygiene of organizations (including their vendors) and gauging whether their security posture is improving or deteriorating over time. The ratings are valuable for vendor risk management programs, determining risk premiums for cyber insurance, executive-level and board reporting, enterprise cyber risk management (self-monitoring), and for assessing compliance with cybersecurity risk frameworks.

Cybersecurity ratings can be compared to financial credit ratings. Just as a poor credit rating is associated with a greater probability of default, a poor cybersecurity rating is associated with a higher probability of sustaining a data breach or other adverse cyber event(s).

SecurityScorecard provides security ratings (scores) on more than 1.5 million organizations worldwide. The published score is a weighted average of ten underlying factor scores, which capture an organization's security posture across multiple dimensions. The 10 factors comprising the total score are listed in the table at left.

We recently conducted a study, investigating the use of Machine Learning (ML) to tune the weighting of each of the factors so that the total score is optimally correlated with the relative likelihood of incurring a data breach.

Factors
Application Security
Cubit Score
DNS Health
Endpoint Security
Hacker Chatter
Information Leak
IP Reputation
Network Security
Patching Cadence
Social Engineering

Materials and Methods

The analysis was carried out by backtesting over a 3-year period encompassing 2017, 2018, and 2019. During this period, the number of organizations monitored and scored on the SecurityScorecard platform increased from just under 100,000 to more than 1,300,000. To ensure consistency, the current investigation evaluated the 99,076 organizations that were on the platform and scored throughout the entire 3-year period. These organizations are geographically diverse and span 18 different industrial sectors.

A total of 10,122 data breach reports were collected during this period from public and commercial sources, including the Vocabulary for Event Recording and Incident Sharing (VERIS) Community Database, Vigilante, and US HHS, as well as inquiries to states attorneys general under the Freedom of Information Act. A subset of these breaches, as described below, was used in this analysis.

In general, the exact date on which a breach occurred is not known with precision. Publicly disclosed breaches are reported by the affected organization with estimated begin- and end-dates. It is widely acknowledged that the elapsed time between the occurrence of a breach and its detection is typically several months. Based on data from the VERIS database, the median time to discovery of a breach was found to be 90 days.[1]

The reference date and reference factor scores for each breached organization were calculated as follows: the reference date is defined as the date 90 days prior to the halfway point between the reported begin- and end-dates. The 90-day offset accounts for the typical elapsed time between breach

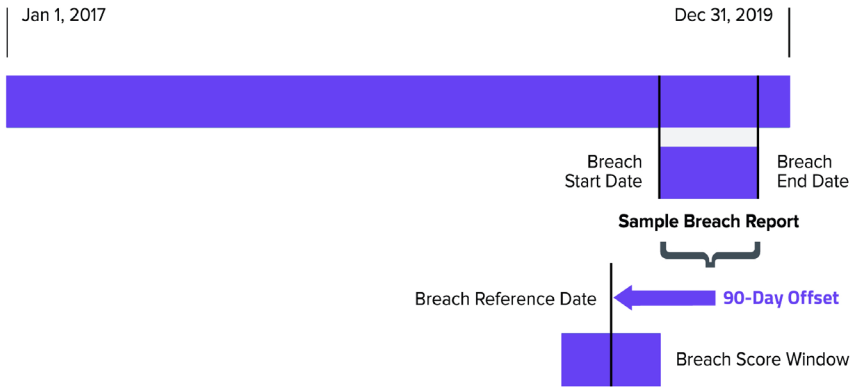
Study Parameters

Evaluation Period	3 Years
Period Start	Jan 1, 2017
Period End	Dec 31, 2019
No. Data Breaches	2,228
No. Orgranizations	99,076

1 http://veriscommunity.net/index.html

occurrence and detection. Since this is an estimated breach date, the breached organization's reference factor scores were averaged over a window extending from 4 weeks prior to 4 weeks after the estimated breach date. If the breached organization was not scored during this window (for example, the organization was added to the platform and was first scored after the breach occurred), it was excluded from the analysis.

Timing diagram for estimating breach date & score



Following this procedure, 2,228 eligible breaches were used in this analysis.

In addition, each organization was assigned a size tag corresponding to its digital footprint, as measured by the number of IP addresses owned or controlled by the organization.

The factor scores for breached organizations were determined as described above. The factor scores for non-breached organizations were based on their average scores throughout the 3-year period.

No. IPs	Size
≤	S
30 to 1000	M
> 10000	L

Calculating Total Score from Factor Scores

For each organization, the Total Score is calculated as the weighted average of the individual factor scores:

$$TS_d = \frac{\sum_f \theta_f \text{ x } g(FS_{df}) \text{ x } FS_{df}}{\sum_f \theta_f \text{ x } g(FS_{df})}$$

where TSs_d is the total score for domain d, , θ_f is the severity-based weight for factor f, FS_{df} is the factor score for domain d and factor f, and $g(\cdot)$ is a non-linear weighting function which gives greater emphasis to low factor scores. The rationale is that in a security context, "a chain is only as strong as its weakest link." Giving greater weights to low factor scores helps pull down the total score when the entity has low factor scores, reflecting a degraded overall security posture.

Score	Grade
≥ 90	A
80 to 90	B
70 to 80	C
60 to 70	D
< 60	F

Finally, the numerical scores for each organization were mapped to letter grades in accordance with the table shown at left.

SecurityScorecard has published a comprehensive description of its scoring methodology[2].

The factor weights θ_f were initially established prior to 2017 by subject matter experts (SME). The goal of the current analysis was to apply machine learning to tune the factor weights so that the total scores would be optimally aligned with breach likelihood.

2 https://securityscorecard.com/resources/ deep-dive-scoring-methodology

Machine Learning

Since this is a binary classification problem — each record was labeled as either a breach or a non-breach — the analysis was carried out using logistic regression, with the ten factor scores for each record serving as features and the breach status serving as the label.

A regularization term was added to the logistic cost function. The regularization term served as a Bayesian prior, adding a bias to penalize and thus prevent large swings in total score.

The cost function J(θ) for logistic regression used in this analysis is given by the following expression, where the second term captures the regularization constraint:

$$J(\vec{\theta}) = -\frac{1}{m}\sum_{i=1}^{m} y^{(i)} log\,(h_\theta(x^{(i)})) + (1-y^{(i)})\,(1- log(h_\theta(x^{(i)}))) + \frac{\lambda}{m}\sum_{i=1}^{m}((x^T\theta)^{(i)} - (x^T\theta_0)^{(i)})^2$$

with parameters defined as follows:

J(θ) cost function

(θ) vector of unknown factor weights

$(\theta)_0$ vector of initial factor weights (determined by SME)

$x^{(i)}$ set of factor scores for the ith record

$y^{(i)}$ breach label for the ith record (either 0 or 1)

m number of records in the data set

λ regularization strength

X matrix of factor scores

h_θ is the logistic function, given by:

$$h_\theta(x) = \frac{1}{1+e^{-x\theta}}$$

The essence of the ML problem is to find the set of factor weights θ that minimizes the cost function J(θ).

The cost function is minimized when the factor weights minimize the difference between predicted breaches and actual breaches. The solution was found with an iterative gradient descent algorithm using the following expression for the gradient of the cost function (expressed in matrix notation):

$$\nabla_\theta J(\vec{\theta}) = \frac{1}{m} X^T(h_\theta(x) - Y) + \frac{2\lambda}{m} X^T X(\vec{\theta} - \vec{\theta}_0)$$

Principal Component Analysis was used to reduce the dimensionality of the data set and mitigate residual collinearity. The retained principal components accounted for 98% of the observed variance in the data.

Validation

The factor weights derived from the gradient descent solution were used to recalculate total scores for all entities in the breach and non- breach cohorts.

To evaluate and validate the machine learning results, a sampling method was applied to measure the breach likelihood ratios. The cohort of non-breached organizations was sampled with replacement. In each trial, a randomized sample of 30,000 organizations was created by drawing 10,000 non-breached organizations from each size cohort ('S', 'M', 'L'). All of the eligible breaches were added to the trial sample. Breach likelihood ratios were calculated for the sample using a weighted linear regression model, where greater statistical weights were assigned to data points with a greater number of breaches. This process was repeated 100 times to generate 100 trials. The average breach likelihood ratios and the standard deviation were then calculated.

The relative breach likelihood ratio R(g) for grade g, where g = {A, B, C, D, F}, was calculated as follows:

$$R(g) = \frac{r(g)}{r(\mathrm{A})}$$

Where r(g) is the ratio of the number of breaches of organizations with grade g compared to the number of organizations with grade g:

$$r(g) = \frac{n_{breaches}(g)}{n_{organizations}(g)}$$

By definition, the breach likelihood ratio for a grade of A is 1.0. If poor grades are correlated with greater breach likelihood, then the breach likelihood ratio R(g) should be greater than 1.0 for worse grades.

Results

The new ML-tuned factor weights returned from the logistic regression analysis described above are compared categorically (low, medium, high) to the original factor weights (as determined by Subject Matter Experts) in the table below. The table also indicates directional change in the magnitude of the individual factor weights.

Factor Weights

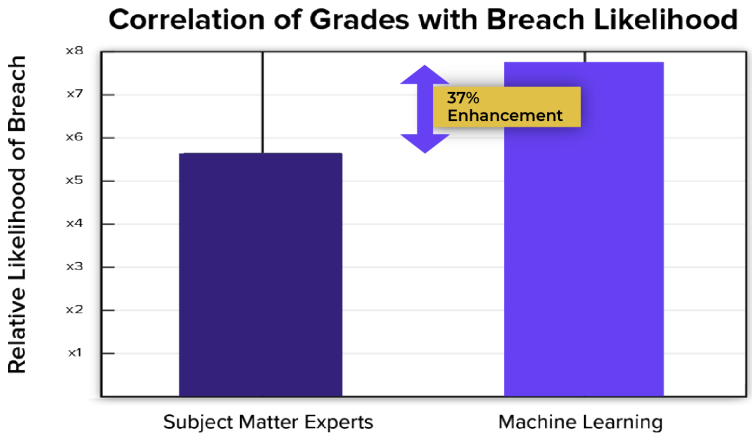
Factor	Old (SME)	New (ML)	Directional Change
Application Security	med	med	↗
Cubit Score	low	med	↗
DNS Health	med	med	↔
Endpoint Security	med	high	↗
Hacker Chatter	low	low	↘
IP Reputation	high	high	↔
Information Leak	med	low	↘
Network Security	med	med	↔
Patching Cadence	med	low	↘
Social Engineering	low	low	↘

*Within each factor weight (High, Medium, Low), there are slight variations. This explains why a factor may have an increase or decrease in directional weight but remain in the same category of severity.

Total scores for the entire breached and non-breach cohorts were recalculated using the ML-tuned factor weights. Relative breach likelihood ratios were then calculated as described above using the new total scores and a weighted linear regression was applied. The results are shown in the chart below.

The analysis indicates that, using the ML-tuned factor weights, organizations with a cybersecurity grade of F were approximately 7.7 ± 0.9 times more likely to sustain a publicly disclosed breach compared to organizations with a cybersecurity grade of A. The risk of breach increases monotonically as the grade worsens from A to F.

As illustrated in the comparison chart below, ML-tuned factor weights achieve a 37% enhancement in the correlation between SecurityScorecard grades and the relative likelihood of breach compared to results with the original SME-determined factor weights, increasing from 5.6x to 7.7x.



Discussion

SecurityScorecard scores are calculated as a weighted average of 10 factor scores, each of which assesses a different aspect of an organization's cybersecurity posture. The numeric values of the factor weights were originally determined by subject matter experts in cybersecurity.

Using the SME-determined factor weights, a regression analysis evaluating the factor scores of 2,228 breached organizations and 99,076 non-breached organizations over a 3-year period spanning 2017-2019 found that organizations with grade **F** were 5.6x more likely to incur a publicly disclosed breach than organizations with an **A**.

A machine learning analysis based on logistic regression and utilizing regularization of the total scores — which effectively served as a Bayesian prior — was applied to the same set of factor scores to generate a revised set of ML-tuned factor weights optimally aligned with breach likelihood.

In the present study, cohorts were randomly sampled, drawing equally from large, medium, and small sized in order to mitigate potential numerical biases associated with size. Using sampling with replacement across 100 trials to reduce statistical error and combining data for different sized and geographically diverse

organizations, the likelihood of sustaining a publicly disclosed breach was found to increase monotonically as the grade worsens. Using the ML-tuned factor weights, organizations with a grade of **F** were found to have 7.7± 0.9 times higher likelihood of breach compared to organizations with a grade of **A**.

The transition from factor weights determined by subject matter experts to factor weights determined by machine learning enhanced the correlation of Securityscorecard grades with likelihood of breach from 5.6x to 7.7x, a gain of 37%.

Conclusion

Companies managing the cyber risk of a portfolio of organizations — for example as part of a vendor risk management program — may use these results to make more informed risk assessments. While actual risk values will likely vary depending on the precise composition of a given portfolio, the results from the present analysis are believed to be representative, and can assist cybersecurity practitioners and risk managers to more accurately assess breach risk.

The application of machine learning to further improve the correlation of SecurityScorecard grades with breach likelihood by 37%, backtested and validated over a 3-year period, constitutes a significant milestone and important advance in the maturity and accuracy of cybersecurity ratings.

ABOUT SECURITYSCORECARD

Funded by world-class investors including Evolution Equity Partners, Silver Lake Waterman, Sequoia Capital, GV, Riverwood Capital, and others, SecurityScorecard is the global leader in cybersecurity ratings with more than 12 million companies continuously rated. Founded in 2013 by security and risk experts Dr. Aleksandr Yampolskiy and Sam Kassoumeh, SecurityScorecard's patented rating technology is used by over 30,000 organizations for enterprise risk management, third-party risk management, board reporting, due diligence, cyber insurance underwriting, and regulatory oversight. SecurityScorecard is the first cybersecurity ratings company to offer digital forensics and incident response services, providing a 360-degree approach to security prevention and response for its worldwide customer and partner base. SecurityScorecard continues to make the world a safer place by transforming the way organizations understand, improve and communicate cybersecurity risk to their boards, employees, and vendors. Every organization has the universal right to its trusted and transparent Instant SecurityScorecard rating.

FOR MORE INFORMATION, VISIT OUR TRUST PORTAL FOR A DEEPER DIVE INTO OUR SCORING METHODOLOGY TRUST.SECURITYSCORECARD.COM OR CONNECT WITH US ON LINKEDIN.

SecurityScorecard.com

info@securityscorecard.com
©2022 SecurityScorecard Inc.

Tower 49 12 E 49th St
Suite 15-001
New York, NY 10017
1.800.682.1707

SecurityScorecard