International Journal of Advance and Applied Research (IJAAR)
Peer Reviewed Bi-Monthly

ISSN – 2347-7075
Impact Factor – 7.328
Vol.8 No.4 Mar–Apr 2021

THE IMPORTANCE OF UNDERSTANDING THE CONFIDENCE IN INTELLIGENCE ORGANIZATIONS

Dr. Anurag Upadhayay
Assistant Professor,
P.G Department of Psychology, K.B.P. G College Mirzapur.
Mahatma Gandhi Kashi Vidyapith, Varanasi.

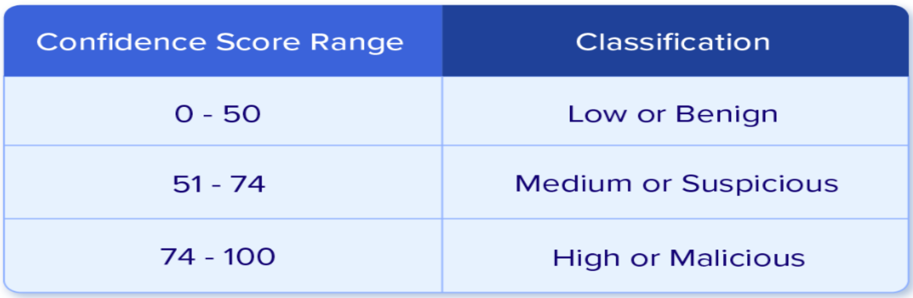
ABSTRACT:

Any person tested multiple times would produce a range of values due to test conditions, examiner, the child's state (e.g., fatigue, boredom, lack of motivation, illness), etc. These scores may not be reflective of the child's "true score." For this reason, it is important to include a confidence interval whenever a score is reported. The confidence interval is a range of values surrounding the score obtained from the administration of a standardized test. Since the standard score may not be the child's true score, the confidence interval increases the likelihood that the test has produced accurate information by including a range of probable scores. New York State has recognized the importance of treating a standard score as within a confidence interval. "Assessment professionals should be careful to treat each score from standardized tests as falling within a confidence interval whose size is determined by the reliability of the test. This represents a more appropriate description of the student's ability. It also makes a clear statement of our recognition of the inherent limitation in the technology of standardized tests. .A confidence interval changes depending on the level of confidence. Most tests provide confidence intervals at 90% confidence level and at 95% confidence level. This percentage indicates the likelihood that the "true score" falls within the confidence interval. The confidence interval gets bigger as the confidence level increases because a wider range of scores must be included to ensure that the "true score" falls within it. Confidence intervals, however, also demonstrate the lack of information gained from using and scoring standardized tests in order to identify language impairment or cognitive delay. Consider the scores included within confidence intervals in the following standardized tests.

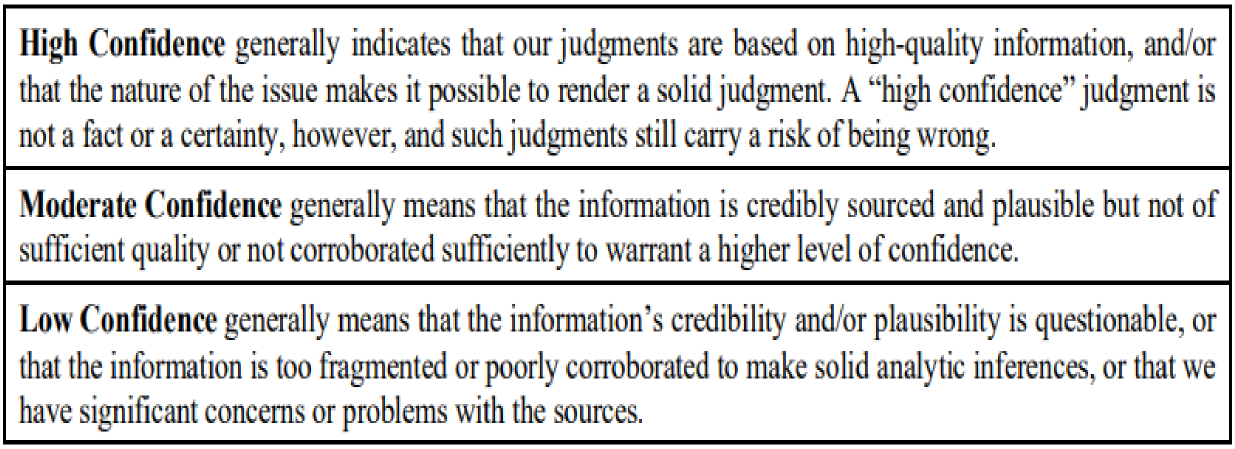
INTRODUCTION:

A major challenge that security teams face is to operationalize the massive volume of threat indicators of compromise (IOCs) ingested from multiple threat Intel sources. These IOCs usually include many false positives and noisy data points. Deriving contextual and actionable threat intelligence from this raw threat information can be a laborious process if these ingested IOCs are not automatically correlated. Security teams need to be quick and meticulous in decision-making. To correlate IOCs for threat intelligence contextualization, confidence scoring plays a significant role. Confidence level helps eliminate false positives and prioritize activities related to rising threats. With the aggregation of massive threat intelligence, confidence scoring has become essential for security teams who must leverage robust threat intelligence platforms that can help them improve their threat detection and operationalize the intelligence in a useful way. Based on their maliciousness, IOCs are classified and assigned a rating, which is known as a confidence score. A confidence score is a value ranging between 0 to 100; while 0 confidence suggests that an IOC is non-malicious, a score of 100 suggests the indicator is highly malicious.

DISCUSSION:



Confidence Score Range	Classification
0 - 50	Low or Benign
51 - 74	Medium or Suspicious
74 - 100	High or Malicious



High Confidence generally indicates that our judgments are based on high-quality information, and/or that the nature of the issue makes it possible to render a solid judgment. A "high confidence" judgment is not a fact or a certainty, however, and such judgments still carry a risk of being wrong.

Moderate Confidence generally means that the information is credibly sourced and plausible but not of sufficient quality or not corroborated sufficiently to warrant a higher level of confidence.

Low Confidence generally means that the information's credibility and/or plausibility is questionable, or that the information is too fragmented or poorly corroborated to make solid analytic inferences, or that we have significant concerns or problems with the sources.

BENEFITS OF AUTOMATED CONFIDENCE SCORING:

Automated Threat Actioning:

Advanced threat intelligence platforms enable security teams to automate actioning based on confidence scores. Security teams can build rules to automate proactive threat mitigation tasks such as blocking of IP in firewalls based on confidence scores.

Faster Threat Investigations:

Confidence scores allow security analysts to generate finished Intel reports by including tags TLP, MITRE ATT&CK mapping, and investigations. These reports can be employed to create contextualized and rich Intel, helping analysts to expedite their threat investigations.

Contextual Threat Information Sharing:

With confidence scores in hand, security analysts can create and share threat bulletins with their subscribers, members, or other organizations, equipping them with the right threat data for investigations. Threat Bulletins enable security teams and stakeholders to make smarter business decisions while helping them keep pace with the evolving threat landscape.

High Confidence:

Some aspects that support this selection:

- Intelligence is correlated from more than one source
- Intelligence is correlated from more than one system/tool
- Source of information is trustworthy (e.g., based on previous intelligence)
- Minimum assumptions and strong logical reasoning/inference

Medium Confidence:

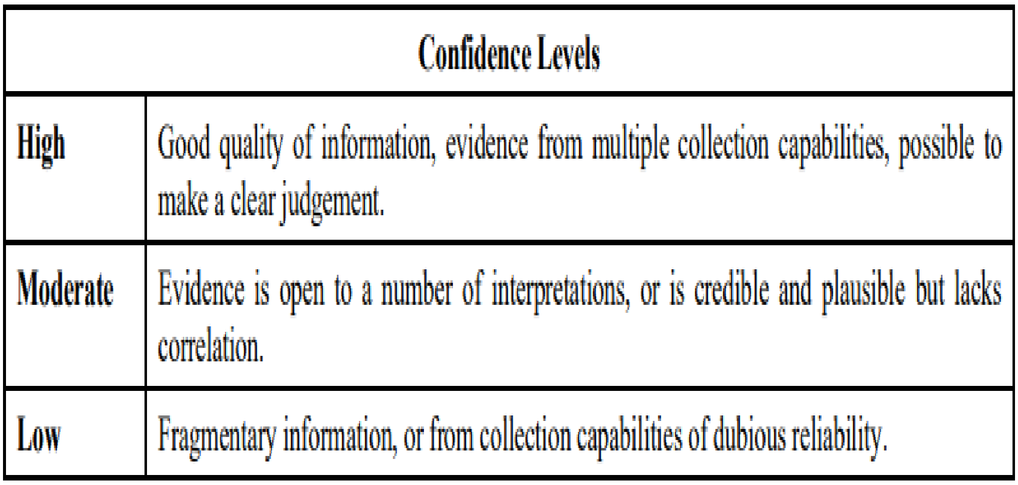
Some aspects that support this selection:

- Intelligence is partially collaborated from more than one source.
- Intelligence source or system is partially tested or has previous accepted levels of confidence.
- Low contradictions, assumptions, etc.

Low Confidence:

Some aspects that support this selection:

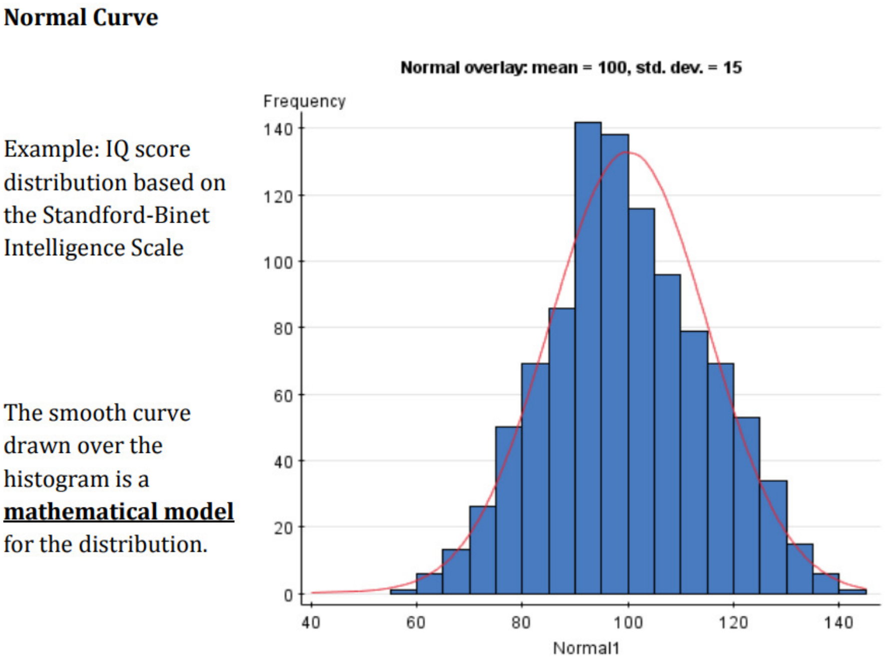
- Intelligence system or source is new, unverified, etc.
- Several assumptions and/or contradictions exist.
- Intelligence comes from difference sources with conflicting information.



Confidence Levels	
High	Good quality of information, evidence from multiple collection capabilities, possible to make a clear judgement.
Moderate	Evidence is open to a number of interpretations, or is credible and plausible but lacks correlation.
Low	Fragmentary information, or from collection capabilities of dubious reliability.

Method:

Intelligence Procedures outlines confidence communication procedures int ended for use by NAT members as well as external partners. Analytic confidence is assessed on a qualitative, level scale and is associated with information credibility, source reliability, correlation, and number of collection capabilities utilized. NAT O doctrine emphasizes that “throughout interpretation and all-source fusion, the analyst should attempt to find confirming information for any all types of problems.



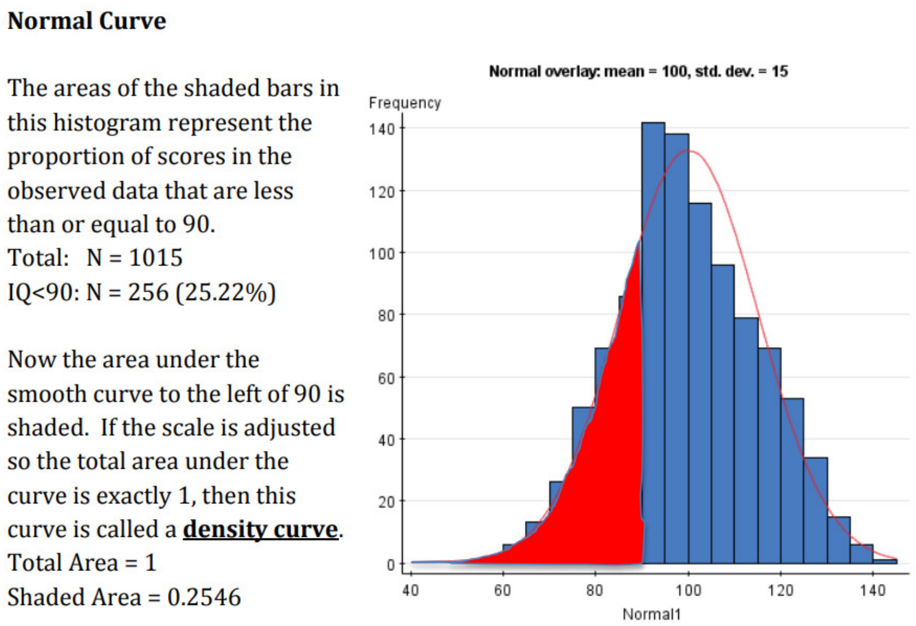
Normal Curve

Example: IQ score distribution based on the Standford-Binet Intelligence Scale

The smooth curve drawn over the histogram is a **mathematical model** for the distribution.

Dr. Anurag Upadhayay

The histogram in this image represents a distribution of real IQ scores as measured by the Standford-Binet Intelligence Scale. ♣ The blue bars represent the number of individuals who recorded IQ scores within a certain 5-point range. ♣ The main purpose of a histogram is to illustrate the general distribution of a set of data. ♣ This variable has a mean of 100 and a standard deviation of 15. ♣ The curve that is drawn over the histogram is the Normal curve, and it summarized the distribution of the recorded scores.



Normal Curve

The areas of the shaded bars in this histogram represent the proportion of scores in the observed data that are less than or equal to 90.
Total: N = 1015
IQ<90: N = 256 (25.22%)

Now the area under the smooth curve to the left of 90 is shaded. If the scale is adjusted so the total area under the curve is exactly 1, then this curve is called a **density curve**.
Total Area = 1
Shaded Area = 0.2546

The entire area under the curve represents all the individuals in the sample.

If only part of the area is shaded, this represents the proportion of individuals who scored below a certain point.

In this above example, the area under the curve represents all the individuals in the sample. In this case, they add up to 1,015. This number represents 100% of the sample. The shaded area in the above example represents the individuals who had an IQ score below 90. This group consists of 256 individuals.

To find the percentage, divide the number in the group by the total number, and then multiply by 100. In this case, 256 divided by 1015 times 100 results in a percentage of 25.22. This means that 25.22% of the individuals in this sample had an IQ score below 90.

The Normal curve is used to find proportions from the entire population, rather than just from the sample. The values for the entire population are often unknown, but if the variable has a Normal distribution, the proportion can be found using only the population mean and standard deviation for that variable.

Rather than using percentages, statisticians use decimals. Therefore, the entire area under the curve is 1. Using the properties of the Normal curve, the shaded are in the above example is 0.2546. This will be explained in greater detail later.

FUTURE PROSPECTIVE AND CONCLUSION:

- Intelligence collection and analysis activities produce results tied with accuracy or confidence levels. Intelligence audience should not look at the intelligence information while ignore its accuracy. Rumors may convey important, serious, or valuable information, but with very low accuracy and hence significance. Intelligence can vary from "facts: with very high confidence" to "rumors, or assumptions: with very low confidence."
- Data analysis algorithms (e.g., clustering, classification, prediction) produce different types of performance measures or metrics (e.g., confusion matrix, area under curve, AUC, F-measure, root mean square error, RMSE). For data analytics, it is very important to study all relevant performance metrics to an analytic activity, their implications, meanings, interpretations, etc.
- For intelligence users, trying to study and understand such metrics can be very time-consuming and confusing. Hence, it is the job of intelligence collection and analysis teams to create accuracy or confidence levels that are more readable or easier to interpret and understand by users who are typically with no technical background/skills. For example, accuracy of intelligence collection or analysis results can be summarized into three confidence levels:

REFERENCES:

1. D., and Mandel, D.R. (2018). Methods for Communicating Analytic Confidence in Intelligence to Decision Makers.
2. Intelligence consumers, intelligence producers. Washington Quarterly, 15(1):157-168. [3] Schneider, M. (2014).
3. The North Korean nuclear threat to the U.S. Comparative Strategy, 33(2):107-121. [4] Defense Intelligence Agency